1197 TM: Hardware, Software for "Ultimate TM": $\frac{1}{2}$ ~~#~~+

→ i.e. TM part I: H.U. 3·26·90 ff

01: 96TM 68.40 This is an old idea that I had on a long Airplane trip from (I think) a Conf. (or to): I did write it up but can't find it. It was ~~an~~ outgrowth of some work on how to use ~~po~~ buggy cpu's to do good Lsrch — One result was that I probly wouldn't ~~The~~ ~~much~~ have much advantage in using "untested", say, cpu's.

---

Tho I mite use untested cpu's this ~~way~~ way: Say a normal production line for cpu's has a 75% yield — w 100% testing. Then I accept every other cpu (untested), as long as I know t. other cpu's ~~are~~ have 75% pass rate. If they untested ones are much cheaper, ~~&~~ this would perhaps be fine! Ideally I'd only need ~ $x \cdot \frac{1}{.75} = x$ ~~&~~ $\frac{4}{3}$ as many cpu's.

---

.10 Anyway, as I remember, I. first approximation had many cpu's working in ||. Hvr, they ~~[illegible]~~ shared certain Modules for commonly used operations & would Que-up to use them. If t. Ques got too long, dups. of those modules would be made.

If a module wasn't used enuf, it would be removed & its parts used to make / ~~n~~ now other || problem solvers, or other ~~[illegible]~~ commonly used modules.

---

.20 Second Approxn ~~[illegible]~~ To realize .10 one could have something like t. "Internet" in which messages would be passed to various cpu's a/or modules, & t. outputs of modules sent to other cpu's & modules. So t. entire device was just this message passing system.

---

"√" ≡ Mthds of trials

In .20, t. allocation of resources for modules & "cpu's" is somehow interpreted w. the ~~ee~~ ≈ constancy of $\frac{pc}{cc}$ ← resources. at each pt. in t. calcns.

T. whole ~~system of~~ detaild systems of how this has to be done, has to be worked out.

It could be like t. Early Connection Machines: That each computing unit has an assoc unit for passing messages only. — In fact, it may be well to use a "Connection Machine" Architecture.

---

Perhaps t. idea of .20 is ~~[illegible]~~ useful only if each cpu infrequently waits for results from submodules but "Pipelines" many trials at t. ~~[illegible]~~ same time — So each cpu is working all of t. time. If all trials for a given CPU are waiting for info from modules, then t. CPU starts a new trial.

SO, except for ~~a~~ cc assoc. w. messages being transported, the System can be very efficient (ie. no time spent waiting)!

In fact, it would SEEM that there would be no point in making duplicate Modules, since long ques would not be inficient.

11 9[illegible] TM H.w., Sw. for "final TM"

9 1.40 Corresponding to a "long que" in a non-pipelined TM, would be a very large no. of trials being done in || by a "pipelined" architecture.

Also note: If a module has an empty que much of the time, then it is being very infficient! So probably it is best to have modules with ques containing 3 or 4 "waiters".

0 Also note: There is no point in having a module unless the module is somehow more efficient than having each

3 cpu doing the task. On t. other hand, it may be best to have each "cpu" that runs trials, have no computing capability other than clerical — i.e. keeping track of I/O strings for each trial.— seeing that the I/O needs of each trial are satisfied.

.10-.13 is very impt.! I don't immediately see how to do this. I'm thinking of a FPU,(e.g.). How would this be "Simulated" in a TM of t. kind I'm Contemplating?

---

25 Re: Many "crumby" cpu's in || trying to code a corpus: There was t. problem that if one cpu did find a "short code", it would not be possl. to know how to continue that code, because of t. uncertainty in just what parts "failed" or "ambiguated":

A (perhaps) soln: we have all these cpu's in || making trials! If one does succeed, it automatically trys random continuations of its code to find probable continns. of t. corpus.

A big diffy is in implementing an exhaustive L srch. we can't do it even w. any certainty w. "crumby" machines. But a crumby machine can be programmed & each program has a ensemble distribn. for its output.

A possibly disturbing thing: There is some nice, short code, but the machine that it was assigned to, didn't do it right.

Is this a nasty diffy — i.e. is there not a good chance that a different short code of this, or some other "crumby" machine, would by mistake, do that code?

---

A crumby machine can be modeld by a non-crumby machine w. 2 inputs: 1 is t. regular pgm, & t. other input is random. The "degree of crumbyness" is measured by relative rate at which t. machine asks for t. 2 inputs: A simultaneously very crumby machine mainly asks for random inputs. & conversely for very non-crumby Machines.

---

Some (maybe new) approaches to using randomness in Lsrch:

1) At a gn. pt. in t. search we consider all codes of length L or Less. We choose one of these at random w. = proby & we work on it w. cc $\propto 2^{-l}$, if it is of Length $l$. After working on it for that time, we mark it, & do another random trial. If we hit one that's been worked on in this round, we skip it & make a new random trial. The fraction of time spent in hitting sites again and again may be quite small. — But this will have to be checked.

We do maybe $2^l$ trials, (?) (or maybe much less): Then we maybe double t. time we spend on each trial. — or, if we retain the state of each machine worked on, we can spend time $\propto 2^{-l}$ on each cpu, whenever we hit it: No ↑ in time spent per hit! This amounts to Levin Srch w. time sharing, but random choices.

2) Another (crazy) idea! Make choices proby $\propto 2^{-\frac{l}{2}}$: also t. amt. of time spent is $\propto 2^{-\frac{l}{2}}$.

This $\sqrt{p}$ dist. is famous for 2 things! → spec. 42.01

1|29 TM

01' 96TM 68.40 spec: An Attractive Idea:
Take a fairly difflt math problem & find one or more conc. nets for it.

Doing it in Euc. Geom mite be easy for me. I'm afraid that human "Geometric intuition would be diflt to express in machine form .... but I think others in A.I. have worked on this & I may be able to use/understand their work.

I was thinking of avoiding the need of human Geom intuitions — but it may well be that I need many parallel concept nets for different "domains" — so TM can make analogies betw. them.

If I try to get TM to use my methods of solving problems, it will have to have lots of geometric intuition!

I might be able to get some other people interested in this problem of really explaining how various hard problems really can be solved — For use in both Human & Machine Learning.

Befor, I wanted to write TSQ's "from t. bottom up" (starting w. easy, fundamental, "primative" problems.) - Now, it seems that "top down" may be easier/better.

In Euc. geom, we already have a lot of ordering of t. Thrms. by Euclid & others. T. Geometry of Euclid is, of course, very incomplete, but perhaps I want to leave it that way. Humans are able to work with — prove theorems in it.

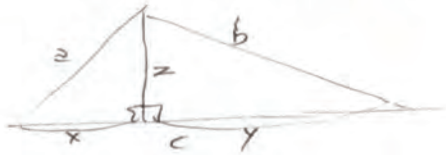
A possl. starting point: Pyth. Thrm.

First: How to conjecture t. Thrm!

Well, because of s.a.s. congruency, If one knows 2 sides & th. included ∠, t. third side must be determined uniquely.

Well pick a rt. ∠. If we had formula for opposite side as funct. of 2 legs! Would we be able to solve other problems?

Well, given Δ w. 3 known sides



.12 If $z = f(a, x)$

$z = f(b, y)$

then $f(a, x) = f(b, y)$

also $x + y = c$

So we can solve for $x$ & $y$ & then get $z$ & get area of Δ. So it mite be worth while!

From similarity considerations, find $z$ as a funct. of $x$.

By doing drawing and measuring, we find for small $x$

.25 $$z \simeq 1 + \frac{1}{2} x^2$$

for large $x$, $z \simeq x$.

A good guess would by

$$z = \sqrt{1 + x^2}$$

since it has desired properties: Checking w. carefully drawn diagrams confirms t. conjecture.

By similarity arguments, if we multiply every thing by $a$;

$$az = \sqrt{a^2 + (ax)^2}$$

we can chose $a$ & $ax$ arbly & get this eq.

So in general
~~for $ax = b$~~

$$\text{hypot} = \sqrt{a^2 + b^2}.$$

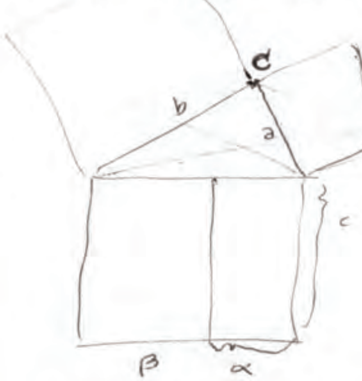
or $h^2 = a^2 + b^2$.

3 2.01

1397 TM Pyth. thrm.

.01 2b.40 ; The proof! One way would be to show $\overset{\text{areas}}{\text{squares}}$ on 2 legs summed to sq. on hypot.. Since t. squares themselves don't fit into t. larger sq. — lets try to show t. 2 ~~hy~~ leg squares are = to 2 other things that add up to t. hypot.

08



Now, there must be an α, ∋ t. square a has same area as t. rectangle (c, α).

14 ~ Guess! Try t. point C as a divider:

we can verify this conjecture by drawing pictures!

Also t. conjecture is reasonable, since BP gets small when a is small. Also, when b = c PA = PB —

So conjecture works on ends and in middle.

If Δ ADC = $\frac{1}{2}$ area of $b^2$

then dist of D from line AC, must = AE. = AC

also If AD = c and Dβ = b then βA must be a

So Δ βDA is congruent to ABC

O.K. that's it we can prove

since ∠θ = ∠γ & ∴ t. 2 Δ's must be ~ & having t. side size AB = AD in common, they are congruent — so t. rest of the proof follows.

Some diff't. parts of t. "proof"

1) The conjecture that $a^2 + b^2 = c^2$ was obtained using partly algebraic reasoning.

2) .08 is a large jump

3) .14 is another large jump.

Those 3 points are interesting! They give an index of t. diffty of t. proof.

There is some q. as to how much detail I want to put into t. "proofs" at this point.

Another tack : Probly I'll want to pursue ~~an~~ T.S.q. in Algebra. in ∥ w. Geometry. This will make it possl. to use analogs from t. 2 Domains to help each other.

I can't immediately think of a HY Point to work backward from. Euclid made a nice Geometry "TREE" — a ≈ Conc. net. But I have no clear idea of such a net for Algebra.

.20 Maybe "Ask around" — Marvin? →4b.03

(N.B.) ~~The~~ All of the details of my "proofs" & Conc. nets don't have to be "correct" (i.e. of reasonable cjs.). If most of them are O.K., this will still give me much needed practice.

For "Conjectures," we first aim for "Quick Abort" i.e. a quick disproof. — By taking extreme conditions — which usually simplify t. analysis.

→ Also, Go over ~~the~~ these (recent) long lists of new ways to make TSQs

Refs:

01 3:40b: On "HINTS": A "Hint" is (anything / something) that narrows t. searchspace a/o reduces srch (cc time). Examples of "Hints" ⓐ Any subgoal: like t. conjecture of 3,14a ⓑ "We can show $a^2+b^2=c^2$ by finding some # $\alpha, \beta, \gamma \ni \alpha+\gamma=\gamma$ and

.06 $a^2=\alpha$, $b^2=\beta$, $c^2=\gamma$." ⓒ (.33)

IMPT Q: After TM solves this problem (or a set of problems), with or w/o "hints": How much easier is it for TM to solve other problems? Well, for theorems being proved: TM ~~the~~ has t. theorems in its "set of facts". — (which is a trivial kind of lrng). Also, it has ↑ t. ssz for use of each of t. heurs it usd. Also, it has ↑ t corpus of examples on which to search for heurs. (ie. examples of successful problem solns)

There is a large Corpus of problems in t. A.I. literature that have been "Solved by heuristic methods". Each of these solns. is a (perhaps useful) example of t kind of work I'll be doing. T. only big differences: ① I will eventually put in enuf detail so that t. "demo" is doable by Machine w. acceptable cc. ② I will have quantitative methods to get pc's, cc's for ∠srch ③ I will have methods to srch for heurs of all kinds, among t. successful problem solns

33 (.06) A Hint differs from a heur. A hint is usually relatively narrowly applicable to a particular problem. A "heur" is more generally applicable. A "hint" may be somewhat Ad Hock.

In general, hvr, a person or TM would prob'y want to remember "hints" — they are good candidates to be heurs or parts of heurs — or may ? (specialize / generalize) to heurs.

.03 36.20 Actually, I don't really need a "Euclidean" of Algebra! Just start at any high pt. (thrm.) & work downward. In fact, this is what I'll do for t. Pyth. Thrm.... I don't have any Plane Geometry Texts around.

Some needed ideas in t. "proof" of Pyth:

1) T. idea of Area: "Area" is assoc. w. any "region" of a plane bounded by lines or curves. T. "region" has property that one can move betw. any 2 pts. in it w/o crossing t. boundary. T. interior pts. are an "equiv. class"

2) Area is Additive: T. sum of areas of all curves inside curve C, is equal to t. Area of C.

3) 2 congruent figures have same area.

4) properties ② & ③ pretty well make area unique if one specifies t. area of some specific figure (say a unit square). By making small squares one can give upper & lower bounds on area (numerically) of most any "curved" region.

5) In t. present "Pyth Thrm": T. sum of areas of sqs. on legs = area of sq. on Hypot. There is no numerical interpretation of "Area"

— Hvr, in t. Analysis of 2b.12: Motivation for conjecturing t. thrm. and 2b.25ff: on t. conjecture of t. thrm itself — there is lots of Algebra & numerical interpretation of "Area". I think we assume that if "a" is edge of a square, its "Area" is $a^2$.

62.01 spec
62.01

1497 TM LANGuage, Human

.01 96TM 43.40 : ( Perhaps Most relevant to papers ȧ discns. therein, of 96TM43.16→19).

.03 If Grammars of Human langs. are largely engraved onto t. DNA., ȧ This is not true of other Animals., Then this mite explain Man's ability to acquire many kinds of technology much better than Animals. Hvr. t. learning ability of animals w. large Brains ( Elephants, Whales, other aquatic Mammals) has [redacted] probly not been adequately examined!

Also, Consider the visual learning capabilities of Pigeons ( if true!).

This (.03) would put animals in a "Tower of Babel" situation w.r.t. man's [redacted] relatively easy communication w. one another.

---

I'm assuming This "Language capability" makes [redacted] certain (otherwise dif(t) kinds of induction much easier.

□

Another possy. is That Lang. doesn't so much effect learning ability directly as it does ease of intercommunication — so animals w. a good lang. can build up more useful ( interesting, complex) social strs.
[ Tho social insects have elaborate social strs. ··· I'm not sure t. situation is at all Comparable to what humans have ]

Social str. is very important for technology. E.G. note lack of developement of technology in China : tho They originated many impt. inventions, t. society did not encourage Their further developement — as was done in Europe.

Language learning: (TSQ)

"Computer" Jan 97. p 72 : Discn. ȧ refs: Some Pgms to read "Edgar" reports ȧ get specific Info from them: T. Pgm does do some "learning".

"Eloise" is Name of Pgm.

Also "Frump"

This Art. has lots of other Things on "learning in A.I. — Refs to many Pgms of interest.

.01 4.40b: A Big Q: There are difrnt %s of rigor in proofs. Very rigorous proofs are easy to express to Machines. Non-rigorous proofs are very unclear to Machines. This has to do w. assumptions humans normally make .... Assumptions not necly in t. premises of t. Thrm being proved!

Can we give t. machine these Assumptions?

Can we define criteria for a non-rigorous proof? — We hope that usually a non-rig. proof can be modified (w.o. too much diffy) to become a rigorous proof!

Perhaps I should create & study some examples of rigorous/non-rigorous proofs — See if I can formalize a non-rig. proof well enuf ~~for a machine~~ to devise a criterion for acceptance/rejection.

Any old Plane Geom. text should have fine non-rigorous proofs.

How did A.I. workers deal w. this? Look in Encyc. of A.I.

---

In general, if I give machine criteria for acceptance of proofs, this will be an unambiguous criterion, so it will be a "formal System". T. only Q then is: is it consistent? This is usually undecidable, but one can have strong intuitive argts for its consistency. Or one can only allow use of t. system at a certain logical depth: in which case we could, in theory, verify its "consistency" at ≤ that depth.

The main problem w. unrigorous ideas is — we have to explicate them to some extent. This can be a big job. — Anyway → 7a.01

---

List of immediate diffys:

① Problem of rigor (67.01) ≡ (6a.01)

② Clear soln. of GA probs by Lsrch.

e.g. 11 input Mixer: try

$11 = 3+2^3$ | $2+2^2$ = 6 input mixer

$20 = 4+2^4$

Lsrch has to be at least as good & probly much better that ordinary Gen. Algms.

③ Do try to find "Proof" of idea in footnote of Sol89: i.e. all heurs can be expressed as modifn. of t. p.d. if one is allowed an adequate set of input params. for t. induction: I should (at least) be able to show that t. system can deal w. "Quick Abort".

"Input params" include "traces" of previous solns. (& non-solns) of problems — & anything else TM may ask for.

1 5 97TM Pyth Thm.

.01 62.40: I'll just have to look at some non-rigorous proofs. In fact, the stuff I have on 2b.01ff may ~~[illegible]~~ give some good examples!

Very proby I wouldn't have to look beyond my own work on this problem to find many usable examples

Re: General Prob. Solving: Perhaps

Look at Newl/Simon "Human Prob. Solving"

Perhaps I could formalize _part_ of what t. prob. solvers were doing in their "_protocols_"

1.16.97 TM GIAR

330 640 935

This is a kind of essay/lecture, to be distributed to friends/colleagues on the net, etc.

Part I : Explain my position: That I like to think of A.I. as ushering in t. next stage of the evolution of Man: The Trans-human! Discussed by Moravec, Minsky. Hvr. Discuss/ Dread Roles of S.F.: To examine possl. scenarios – paths of development. While this part of S.F. has been Distopia-prone — as it must be.

Discuss Asimov's Robot ~~stut~~ stories: "The 3 laws".

Why they would be passed: why they would be inadequate.

1) ~~obey~~ Don't harm humans by action or inaction.
2) ...obey humans.

---

Then ① The "not-concious" but very smart machine.

Development by Military or Gvt. (NSA Always gets Biggest, fastest Machines first)

Use for developing Military H.W./S.W.

Use for control.

That first Nation to Get it can "win" — by preventing any one ... Domestic or foreign from working on A.I.

Discus "Science of 100 or 200 yrs hence".

" "Singularity" effect where IPC → ∞ in finite time; what "∞" means.

①② What happens when > 1 power simultly have very smart machines?

---

② Discuss Conciousness in machines:

obtainable by ⓐ very smart machines in suitable envt.
ⓑ Technologically Augmented humans

○

Idea that as dominant species becomes more powerful, the likelyhood of destruction of t. world in a war, ↑.

That humans are designed w. goals appropriate to small amt. of power per individual — so a normal, relatively weak person striving to become more powerful in his culture (or world), is not a threat to t. world as a whole, because that person couldn't do that much. This particular goal could, instead, serve to develop t. person in a way, useful to t. culture as a whole.

A very strong person like Napoleon, Alexander t. Great, Genghis Kahn ... could have very powerful effect on the whole world.

1.19.97 TMCIAP! His Own Supervisor. HOS problem.

There is the diffy in very intelligent machines that are goal-oriented. In general, they will be able to change the goal or trivialize its achievement: An early version: McCarthy's reinforcement machine that soldered its own reenforcement channel to "On" (≡ "Maximum").

In Humans, Morphine, Heroine ... can perform similar function. In general evolution, master botton obtains some sexual pleasure w.o. attaining evolution's "goal": Tho usually the reproductive goal is attained by social & other pressures that make normal reproduction have additional pleasure.

yb19b#/ah
in sm
do
.x9b19b/br
6-2068c

So, how do we deal w. this problem in very intelligent machines? As soon as t. machine has much understanding of R.W. & its position in it, it will realize that there are alternative, simple, ways to achieve its "goal" — by internal re wiring.

We could fix it so that part of its goal was that it was not allowd to do internal rewiring (or otherwise modify its goal). Is this possible?

250
245A

Also, there is always t. possy. that t. goal of t. machine could become accidently modified by Buggy H.W. or S.W. malfunction — or Accidently modified — by earthquake, Airplane crash or other catastrophic accident.

I could study t. problem at a micro level. See just what occures when T.M. "realizes" its in a R.W. & that it had best interact intelligently w. R.W. in order to achieve its own [illegible] goals

Superficially, for any kind of Goal (OZ prob), a sufft. smart machine could find gross short-cuts to that Goal by internal rewiring a/o repgmg.

SL=20
En=55
=80
MT=.242

4.14.98 Another (perhaps) approach: That TM is not completely governed by a "Goal": that it has other things controling its path.

There may be some real diffy that TM gets in "Evolution" mode — so we really don't know in what direction it will evolve.

1.20.97 TM expo

I may want to rewrite (revise) The Lsrch "paper": Update w. more recent methods by Levin for both OZ & INV. probs. Also discuss what I mean by "optimality" of Lsrch: Also perhaps look at Li-Vitanyi discn. of it.

Mention that my pap on f... was one way of generating prefix sets.

12697: Things to do
hings

Admin

1) SM work: ⓐ OxB
ⓑ other. econ FT Model quest
ecce!, GA.
ⓒ Related to TM.
(general problem).

Pack stuff at 26B.

2) TM stuff a) CAP: Get Gunkel's list.
b) Plot General path for TM work.
.09 Say several possl. paths.
c)

3) General: a) finish letter to dbld
b) Read Sasha's paper(s)
c) Read Jurgen's papers: reply to his letters
d) Finish "2 kinds of induction"
e) Perhaps write Review of ALP for N.Sci (American or N. Science,
How ALP is a genl. approach to most all "learning"
f) finish paper for dbdetal.
g) revise "Lsrch" paper: look at Wintree's remarks.
h) Write paper on "What to do next"

Less critical

11 +

Get Paul to work web.

Perhaps find low cost Internet Provider.

Fix up 10 Holyoke

Put books on Shelves

Fix TV set.

work on Basement.

Try indexing ~~Box~~ Packed Boxes
" " Stuffon Bookshelves

Fix phone so it sounds Better —
or get Better phone.

Clean up desk:
Other things to facilitate rapid work!
Write "T. dress"
Write "T. story of Harvey".

NB How many of these "things to do" do I want to, have to, do?

Perhaps make PERT diagram of various projects, Deadlines (if any)

Make Chart of various projects, how much time spent on them, what fraction completed.

on Beach | Washington

.25 ; .09: Look at that recent bunch of lists/discn. of TSQ's etc.

My present bias is ⓐ Work probs in GA of various levels of diffy using Lsrch,
ⓑ Take old AI "exp. systems" (like AM) & get it to develope properly. (Maybe a useful medical or other e.s.?) ← That have lots of external info coming in.

Try GM feedup?

try capacitor across mike, inside handset
red white green black

2 middle = white, Green. earphone
2 outside = mike. red, Black
Bogen ~~Phone~~ handset.
Perhaps try Black handset on Bogen Phone or other handset.

5-530 545-600

2 ~~Then~~ List deadlines for projects if they exist.
Estimate deadlines for other projects
} This will help assign times,

~~12697~~ Phone Steve Baer about 2 arts in recent Science on Back prop. in vivo
12697 Can't find item Science D20, J3 93, Jan 10 !
Is it? Could Grace have it?

Write Andreas, Isabel.

Maybe work on proposal for NSF

~~D27-96!~~ 12798

Things to Do!

~~1) Re: Pent & 486! Save that all impt. files from 486 are on Zip & In Pent.~~
~~Q: why were comports 1 & 2 accepted on 486? — why different from Pent.?~~
[redacted] ~~Try ~~pa~~ newer drug w [redacted] Mike.~~

This Overall planning of the Rest of my life: can be useful in helping decide what to get rid of..

---

General idea! Whether I want to write papers on certain matters or not, I ~~said~~ should write a clear exposition for my own use, of important ideas & discoverys.

The Clarification of the differences of beta. AIP, (5) MMC, MDL ect. — This [redacted] Proof, more [redacted] is important to me, since they have useful approaches (sometimes) — but I must know just what their limitations are, so I can effectively transcend them.

---

Make list of items w. near deadlines and do them

50

.20 **SM:** write up impt. results, techniques, Thrms.

first, the most recent stuff on portfolio Management:

Then how much to bet when one has p.d. v.s. default p.d.

Then the default p.d. as given by actual option prices

Then the "true" Black-Scholes model. (second difference of option price = p.d.)

Include refs to whoever, when work was done.

Also stuff on kelly betting: so yield $= \exp(\Sigma (\frac{\mu_i}{\sigma_i})^2)$ perhaps!

Also recent understanding of "Sharpe ratio".

---

Most urgent:

✓ 1) write DLD

" Andreas / Isabell — 2/18/96 (letter to them. ← ANDREAS

2) Get ~~se~~ chart from Gunkel.

3) Read Sagal's papers (but from Grace)

Get Medicare, HMO straightened out.
find out how best to use it.

Read Lipton's Stuff, Reply.

Johannes = first child

[redacted]

2297 Admn: "Things to do"

A Big Problem: I will not have enuf time in my life to do most of t. Things I've listed on pp 1,2. I really am "spreading myself too thin". By consciously limiting my activity to [redacted] a few "well chosen" projects, I will get much more done.

Also, I can probably get rid of a lot of books, documents & hardware — once I decide what I don't want to do!

Initially, I was interested in all human activity — (in addition to all non-human activity)
.08 as part of ~~[illegible]~~ trying to understand how humans/scientists make discoveries &
.09 work problems & make decisions.

At t. present time, I think I know enuf about t. soln. to ~~[illegible]~~ (.08 – .09) as well as t. more general problem. Tho I like dabbling in ~~all kinds of~~ many things of great variety, I really don't have time for very much of that, just now!

---

Some lists:

Sci News
⋮

1) Things to read:
   a) Shasha's papers
   b) Jurgen's "
   c) Lipton's "
   d) Gunkel's "future of AI" list
   e) Winfree's comments.

② Things to write:
   a) Finish 2 kinds of ~~[illegible]~~ Induction
   b) write up SM results (2.20)
   c) What to "do next" paper
   d) Summary of impt. ideas, Things in my TM work, ... Bibliog.
   e) Finish paper for ALD
   f) may be modify Bavez.
   g) Review of ALP for Sci Amer a/o (Nature Science).
   h) NSF proposal

③ AR
   ⓐ Write "T. dress"
   ⓑ " "Harvey".

PATH
2.6.97 Admin; (TM):

New PAths to A.I.

A Different Path to A.I.

An Alternative Path to A.I.

14

|| Start work on "New Problem:" I want to decide what to work on next: a/o roughly how much time share to put on various projects: (Mainly w.r.t. t. TM Goal)

One of t. possl. paths to TM: Working backwards: TM has knowl. & good prob-solving ability in 1 or more areas. He also understands English well enuf to read ordinary text about these areas.

.op ① To get started: ~~Anyway~~ (anyway:) TM's knowl in t. field(s) of interest is not very deep. He can't solve diff. probs., but he does know simple relationships — has some "commonsense", but not very much. He is able to answer simple Q's about t. field. He can read in that field & understand enuf to be able to answer some Q's in an "open book" exam.

Such a TM could perhaps do interesting induction on ordinary text about his fields of knowledge.

By asking Q's & noting quickness, slowness, & wrongness of answers, I get an idea of what TM needs to know, & ask Q's leading in proper directions.

Could TM make use of a dictionary, encyc., or other tools?

.22 ② T. older approach is to get a TM who is very smart in a particular field, then teach it to converse about things in that field.

(SW) TM could learn to speak well by being given its own spectral sequences, those of acceptable humans, & relative "humanness" scores for its various utterances. (Reinforcement lrng.)

Q
Souton 2/6/97
A.I. scenarios
72642.
35740
Comugames.com
Address
Ph. no.
to Gunhel.

③ Another Approach: Take a problem (like Pyth Thrm proof (conjecture).) work out soln. See what concs are needed. Take a difrnt, somewhat ~ problem, see what concs its soln. needs, etc. Try then do this for many problems & see what "needed concs." they have in common. We can possibly give these to TM ab initio. This may save a lot of time. (Or, these concs may have common sub concs so we could give TM a lang. to generate these concs. & (presumably) other useful concs.

Re: ① & ②, "Blocks world" is a case in point: But there wasn't any lrng involved — either of "Blocks world" itself, or of t. lang. descr'g t. world. Blocks world is rather simple - as is t. lang. used to desrb. it.

A thing that attracts me to ② (.22): That if we got TM to be very bright in a particular area by lrng. This is a much stronger result for me than a TM that learns easy things & then an easy lang assoc. w. it. (Tho it could learn easy things w. a rather complex lang. assoc. w. it.

PATH
2.10.96 Admin (TM)

Another area of "easy" tsq design (acquisition): ~~[illegible]~~ Solving optzn probs by Lsrch for ≈/inv solns.

We can find a field w. a very long path from t. bottom to top of hill.

SM is another example of this sort of thing: Start in 1920, say & acquire new, more diff'lt techniques as we approach t. present.

.05 → **Or,** just stay in ~~[illegible]~~ present time & try for better & better strategies.

.06 .05 would be linguistically interesting only if t. info used in its predns. was rather general — t. kind obtainable from maybe, newspapers.

Well: re: .06 say T.M really did do a good job on .05. The only thing it would be interested in were t. various params it used in predn., & t. prices it bet on. T. task of learning to extract such info. from English dscrns of t. data or newspaper reports or stockholder reports would seem to be quite separate from TM's ability to do good SM predn. We could just as well try to teach TM how to extract financial params from text reports, w.o. first teaching TM what t. params "mean".

[ A possibly more interesting literature for ~~t.~~ TM to learn to understand, would be not ~~mere~~ reports of data, but discussions of various SM/predn. techniques.

↳ This would seem to be rather diff'lt! Ideally, TM would read this stuff and try to see if it was useful in prediction, or in augmenting its own schemes. This would seem related to TM's realization that (±)RW. is related to its main problem (Tho: maybe not: because TM is not getting direct f.B. from R W.

---

On t. other hand: Learning to understand English, may very often involve some rather complex (not "linguistic") Inp.... e.g. figuring out what t. referents of pronouns are — or in general, what is going on in a story that's being told

↳ ~~would~~ Could SW "Agents" develop ~~a~~ any kind of ~~useful~~ more generally useful understanding of Humans?

"Path" as a new sub-TM stream.

Luke's Ph. no. 492 3772. |||| 4 times

My ~~a~~ really main criticism of Schmidhu is that I don't see a clear "Path" in his work, to a very smart TM. He was working much on polishing t. final TM, but not ~~we~~ doing much toward getting something that was "polishable".

---

A General related Q about a ~~[illegible]~~ minimal path is t. Q of what/some likely paths in t. existing A.I., ML community? To estimate this last would involve reviewing t. whole ~~AI~~, ML field. I could, tho, try this, using only t. parts of current AI/ML that I'm familiar w.

r red

---

Perhaps write paper "Paths to A.I." : Maybe publish in "Analog"! — Send copys to interested people.

Maybe publish "Scenarios of Doom" for A.I. — also in "Analog".

2.12.96 TM (Pam):

In Paper, perhaps also discuss ~~→~~ 4 or 5 aspects of "Consciousness" & how many are not diff. for A.I. to achieve. Also mention the "overly eager subordinate" as a serious admin problem in Purely human Culture.

---

β Paper Itself: Outline;

1) Classical Approach to A.I.: "expert Systems": ~~common~~
Need for "common sense." — This is an artifact of the way machines ~~are started things~~, acquire info ("knowledge") in "Classical A.I." They get it from a human who is used to communicating w. people who have a common backgnd knowledge.

In a more reasonable trng. envt., the machine (infant) starts out w. certain impt. ~~bsp~~ general info. that makes it easy for it to learn the "common sense" knowledge of the world it will live in. The amt. it starts w. is << the amount it needs to get along in the world of 1 year olds.

Idea that since we can make e.s. in very many envts. we should be able to combine them in a grand e.s. that would be ~~[illegible]~~ as good as a human, but much faster in many areas.

Except, perhaps, for ~~the~~ Lenat's Cyc project ~~(Lenat)~~, however ~~the hope~~ the hope that this would occur in near future (50 yrs) has whithered away from its early enthusiasm.

---

Since about 1980 there has been a great (~~upsurge~~ upwelling) of interest in machine learning. Neural nets, statistical learning and Genetic Algorithms are ~~only~~ but a few of the great variety of techniques applied in this area. In fact, the diversity of techniques has been so great that some scientists have expressed doubt that they have enough in common for the ~~concept~~ idea of "learning" to be at all a useful scientific concept! ~~[illegible]~~

One of the ideas we will develop in this paper ~~is~~ is that just about all learning does have a common mathematical basis, from which ~~and from this basis~~, we can ~~→~~ compare and evaluate the effectiveness of ~~a~~ various learning techniques, ~~compare~~. ~~various techniques.~~ and obtain suggestions for improving them.

This basis is Algorithmic probability — or program length complexity'.

(FN: Also called Kolmogorov complexity, MDL, MML, Stochastic complexity (probability) Solomonoff, Kolmogorov, Chaitin ~~probability~~ complexity (in order of discoverers))

Furthermore ~~[illegible]~~ it enables us to ~~understand~~ ~~several important paradoxes~~ resolve many difficulties in our understanding of the induction process — not the least of which is to give us a good understanding of randomness.

2.14.97: TM (PoR)

~~ALP also enab~~ In addition to giving a detailed understanding of how Learning works, it also gives us a tree structure showing ~~[illegible]~~ what concepts can be learned from what more fundamental concepts. — much like the tree structure of species in organic evolution. There is an important difference, in that/while there is a unique evolutionary tree, there are many possible tree structures for various learning tasks. ALP enables us to evaluate them and tell which are more likely, which involve less computation cost.

---

Perhaps discuss t. basis of A.I. in Introspective Psychology: How it added 1 more constraint: ~~[illegible]~~ programmability; that it needed 1 more constraint: learnability. Until ~ 1960, there was no clear criterion for learnability. (make it clear that I'm not talking about Valiant ~~"learning"~~ "learnability", which ~~[illegible]~~ poses narrow ~~area of learning theory).~~ restriction on how learning may occur.

---

Discuss Sol 86, 90 in detail: what "all info is in P.D." "really" means. Just how heurs are discovered: how discovery of a heur during t. soln. of an ~~[illegible]~~ Inv. problem can be "simulated".

---

.22 2.17.96: In fact, SM may be a v.g. tsq. if I do it via a G.A. as an sequence of "Inv" problems! This could be very Good!

In general, it may be possible to give TM a ~~[illegible]~~ O.Z problem in this form, in a "sufficiently complex (interesting) environment", so it will learn interesting things, even tho I don't do much in t. way of "designing" t. T.S.Q.: it's just a very good problem envt!

It mite be possl. to give TM a small set of diff't O.Z probs in various interesting env~~ts~~ironments. (have TM do them sequentially or in ||) & end up being very smart, & have a set of wtd abss. that ~~are~~ are able to work other diff't inv. → probs & even O.Z probs (done my proper way!).

More 8.01 spec →

2/17/97 TM (PATH)

.01: Sfer 7.90 ! This approach to TM is like t. "problem pool" approach to TSQ design. T. Problem pool seemd to be a non-ccl. problem in which TM itself would have to face its own administrative problems — e.g. how much time to spend on recognizing old data ("recompression") v.s. how much time to spend on new problems.

I think one big new "BREAKTHRU" is t. realization that I mite be able to get t. "problem pool" approach to work using not t. general "optimum" soln. to t. OZ problem, but by using the much simpler "INV" approach to OZ probs! ← ≡ OZI method

(SN) In fact, I haven't really workd out t. details of t. $OZ_{inv}$ method:

.11 One works on t. optim problem w. INV (Lsrch) for a while, Than after making some progress, one compresses data & loop to (.11). T. Q's are: How long does one do INV before compressing, & how much time is spent on compression? Maybe Working on Inv & Compression in Time-share at ~ 50% dwell time can't be > a factor of 2 wrong.

(Def.) My impression of what $OZ_{inv}$ (≡ OZI) is: One mihaps hunts for improvements in solns in a hypersphere about pt. α in t. search space. As better solns are found, they replace α. In t. case of what look like multiple peaks, we mite use > 1 "α" pt. & time share betw. them.

.20 When I speak of "hypersphere about α" the radius is in terms of "conditional complexity of trial wrt α".

I'm not so easy about how to "Lsearch about α" One α way! find some short codes for α & add bits (or "info") to them. Ordinarily, one will know "short codes for α" because α was created in Lsrch & it tends to have a "short code".

ULTIM More ULTImately we will want TM to do OZI search to find an optimum all-over pgm. for itself, that "goes up t. hill" as rapidly as possl. — So it will use OZI to transcend OZI!

Actually OZI may be "Adequate" (tho non-optimal) in t. sense that it can solve most any OZ problem — but I can use other search algms if I like, it may seem more appropriate & seem universal enuf.

On "branching" in OZI! Perhaps one should spend time on a Branch ∝ its pc & ∝ its ΔG (linearized Gain). Somehow ΔG is somehow related to t. negative cc penalty. ΔG is linearized "Good" & cc is linearized "bad".

2/17/97 TM (PaPa)

The constant of proportionality betw. cc & ΔG is unknown, hvr. .... can we do anything w.o. any knowledge of it?? Perhaps time to share ∝ pc · ΔG? Well, this is dependent on "origin" of ΔG. Adding a constant to all ΔG's will change relative time shares. → (13.01)

.03

Perhaps get ideas on how to do this from GAs & from SGA. — In GA literature (Also ALife) there are lots of strange techniques used. See how I can ~~optimi~~ formalize them w ALP & Optimize them. Remember "ALP is ~t. final soln for all induction problems" ∴ it should (in theory) tell how to solve this problem.

One view: one has a set of pts thus far: each has known: $(G_i, pc_i, cc_i)$ from this set, what to work on next, & how hard?

Q: Is this related to t. "what to work on next" problem (which is normally used for INV probs. only)

Is $\frac{\Delta G_i}{\Delta cc_i}$ or $\frac{\Delta G_i}{\Delta pc_i}$ like "velocity" or "momentum"?

ABcdefijwed
ABcdefijwed

● 1) List t. most immediate problems in OZI. Then suggest possl. solns. & criticisms of those proposals.

ABc
ABCD

● 2) Draw up a minimal (simple) system that could do OZI. See just how many & which features it needs in order for it to ~~be~~ be able to improve itself & transcend, completely, its initial pgm. (Hvr., this last may depend on giving it good problems, a good problem pool, or a good TSQ.

A
ABCD

(SN:) For Lsrch of All kinds! for t. Time-sharing method: Store all info used in each problem in a stack. When we stop work on a particular trial, we switch t. state of t. machine by switching stacks. The stack tells how far we've gone in a certain trial.

We can have stack length limits in ~~a~~ trials, as well as time limits. Hvr., it's not clear how I should limit stack length as a funct of pc. If stack length is unlimited, then t. time limit limits it: I could try that: see if t. time limit was small enough to prevent stack from getting too long.

So, no stack length limit (to start off): but when I store t. stack, I have to say how long it is. Stack info can be stored in Ram, like Hard discs store files — not nearly contiguous, but w. jump positions somehow indicated.

One possl. way to do OZI: Pick the 100 best ΔG pts. thus far and do Lsrch out from them by somehow "adding to their codes".

.37 At the end of a certain time, pick the 100 best pts thus far & continue them } for a certain time. loop to .37.

⇒

2/18/97 TM (PaR) ABC ABC 20 ~~10~~

An impt. Q: If t. thing I'm looking for is a finite string, t. pgms for them will be "self-delimiting". How do I generate strings that are "close to" string a? — say I know a short code for $s$.

Well: look at Sol 78 p 427 col II: section V on Chaitin's condl. complexity.

It uses ~~[illegible]~~ UMCS w. 2 inputs. One input is t. min code for t.

$U(r, t^*) = s$ : The set of acceptable $r$, for each value of $t^*$, form a prefix set. $t^*$ is t. shortest code for $t$: i.e. $U(t^*, \Lambda) = t$.

If $U(\cdot,\cdot)$ has a ~~unidirect~~ unidirectional input tape for first argt., & unidir. output tape, and $U(\cdot,\cdot) = s$ means $U$ stops after $s$ is printed, then t. legal first inputs must form a prefix set.

---

We mite also constrain $U$ so that $\boxed{U(\Lambda, a) = U(a, \Lambda)}$. — Probly impossl. See .20

Then $H^c(s/s^*) = \min |r| \ni U(r, s^*) = s$ so $r = \Lambda$ & $|\Lambda| = 0$

so $H^c(s/s^*) = 0$.

Hvr, if I want to use probys, maybe not so easy to get dist from s to s = 1.

---

.20 A bit of trouble! If $U(\Lambda, a) = s$, say, then no other first input (other than $\Lambda$) is legal. $\sum 2^{-l(\Lambda)} = 2^0 = 1$ : $\Lambda$ is a complete prefix set.

I think my defn. of condl. proby was (perhaps exactly):

.24 $P(s/t) = \quad 2^{-|r_i(t)|} \cdot 2^{-r_j(s \text{ given } r_i(t))}$

.25 $p(s/t) = \sum_i \sum_j 2^{-|r_i(t)| - |\underbrace{r_j(s, r_i(t))}|}$

The $j$th code for ($s$, given the $i$th code for $t$)

means the $j$th code for $s$, given $i$th code for $t$.

$\tilde{r_i}$ is $i$

Can I show $P(s/s) = 1$? This could only be true if t. lengths of t. exponents in .25 could be shown to be (also) t. lengths of some complete prefix set.

Maybe in .25 I should devide the defn by $P(t/\Lambda)$.

The ~~expressn~~ present expressn. in .25 may give all t. codes to produce $(s, t)$ (in that order).

It would seem that t. 2 strings referred to in t. exponent in .25 have all of info in them to produce $s$ and $t$ individually.

.37 ~~∑~~ (.25) as is gives something like t. pc of $P(s,t)$.

Thinking about it, .37 is reasonable, but it is t. pc of $P(s,t)$ in that order. We have 2 strings, $r_i(t)$ and $r_j(s, r_i(t))$, and they, in that order, enable one to get $(s, t)$.

2/19/97 TM (P2R)

~~12 14 13 15~~

So 10.25 gives t. pc of t. 2 objects s, t, in that order, using a certain | kind of umc. as a refrence. A natural Q: does $P(s,t) = P(t,s)$?

.03 Also, does $P(s/s) = 1$. ? i.e. does $P(s,s) = P(s)$? (which would imply $P(s/s)=1$) → 12.16

---

For my own Applications in OZ I: Consider $U(x, r)$:

r is a (relatively short) code for some object, s. Now I want short strings, x, to modify r, so as to produce new objects, that are in some sense "Informationally close" to $U(r, \Lambda)$.

Seems like a good idea that | th. Machine w.r.t. x of $U(x, r)$ should have available to it $U(x, \Lambda)$. — on second thot, this seems to be rather trivial thing to satisfy.

I guess what is needed is a way for x to modify r to produce a new pgm z; that give output $U(z, \Lambda)$. $(U(z, \Lambda) = U(x, r))$.

Consider t. pgm r, to be not a binary string, but t. (dscn) pgm of a function, and the function is dscbd by telling how to construct it from other functions.

Certain functions are functionals. Their inputs are various functions, & their outputs are functions. The "composition" operator is one kind of functional. Another derives recursive functions (& p.r. functs) from other functions.

We may actually need only a few (maybe those 2) functionals, to get all p.r. functs. (I wrote about how to do this ~ 1990 in Saarb)

---

Perhaps what I had in mind for OZ I: We search until we get a set of somewhat good trials (say they are functions). We then pass a grammar thru those functs — — But we want th. grammar, in addition to be universal. The construction of this Grammar constitutes "Compressn of t. corpus of "good trials."" We then use this Grammar to Generate new trials.

---

.30 In SGA, we passed a grammar thru t. data in a way so that t. pc of each trial was ∝ its G or ΔG. — Hvr, it wasn't clear as to what we should do here: what monotonic ↑ funct of G we should make pc ∝ to.

From a practical standpt, we could make pc of trials be any monotonic ↑ funct of t. "pc" of t. Grammar (or t. G's of t. trials)

One way I thot of Getting t. Grammar was to use ssz ∝ G (or ΔG) for each trial. This could result in a grammar that concentrated very much around where t. sample trials were. I then used "$(pc)^\gamma$" to control distribn of new trials. Large γ for concentration of trials at hy pc, small γ for filling out t. grammar at lower pc. areas.

2/19/97 TM (Path)

01 → Contrast 11.30 → 40 with t. scheme I used for t. 11 input Muxer. I started w. a stoch Grammar on Functions. When a funct of hy G was found, it was added to t. list of "defined" functions. I don't know what pc I assigned to those new functions. Perhaps I just took as corpus t. set of successful "Good" trials and looked at t. frequencies of t. various "basic functions" within them. .06

A thing to think about: In SGA, I thot that by making δ large (essentially making all pc's zero or 1) that I could get t. Grammar to ~~f~~ generate only trials near t. "Peek" G. Hvr., in .01-.06, I don't think this is true! .01-.06 was a "Bernoulli grammar". Each funct. had a fixed frequency. On t. other hand, in SGA, I had t. pc's vary w. each use of each branch pt. in t. Generation of t.. objects. I used n.t. "Bernoulli Grammar" by a phrase Str. Grammar (w. or w.o. context dependence).

16 : 11.03 : To actually use this P(t/s) idea: Just list various useful functions & find ways to xfm them into "nearby functions". Whatever ideas I have about "nearness", I can formalize so that P(t/s) expresses those ideas quantitatively. T. behavior of P(t/s) is under my control (w. a few restrictions) — I will not be (very) much constraind by apriori ideas of what P(t/s) should look like!

[ **Q:** Is it possl. that I could use my own more optimal (Sol 86, 90) System for this kind of TSQ? — or ≈ Schmidthu's system? Essentially t. $TM_2 = TM_1$ idea: Or just have t. machine work on "external" problems, then alternate work on ~~its~~ its own system optimization.

It would be well to write up t. basic idea — The idea of just how SM gives v.g. TSQ's: In enuf detail so that me as of 1 month or 6 MO or 1 yr from now will be able to understand it. The main argt. of t. present idea starts at 7.22

Essentially, t. idea is, we get TM to climb an infinitely hy hill. If he gets stuck in a local max, we have him climb to other local maxima — several of them — then we recombine the ~~[illegible]~~ concepts of those several local max's to try to get higher.

SGA has another method of dealing w. local max. — by using small γ, to fill out trials that are distant from t. known peak regions.

23 ~~13~~

2/20/97 TM C P(8th)

SPEC 9.03

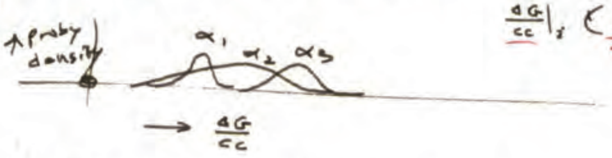
.01: Say we've been working on several "$\alpha$" pts $[\alpha_i]$ (timesharing). (8.20) We have this normal relationship betw pc & cc that we follow for each $\alpha$ pt. — How much time share should I give each $\alpha_i$ pts? Well, consider at each $\alpha_i$ the $\frac{\Delta G}{cc}$ obtaind. $\Delta G$ is that of the hyest pt. obtaind, working out of that $\alpha_i$.

Remember the goal is max $\frac{\Delta G}{cc}$, where $\Delta G$ is the hyest peak thus far! Well, at each $\alpha_i$, we have $\left[\frac{\Delta G}{cc}\right] \pm \sigma$. i.e. we're not really certain about the long term prospects of $\frac{\Delta G}{cc}$ at each of the $\alpha_i$. We mite work on the hyest peak untill its $\Delta G$

Def

$\frac{\Delta G}{cc}\Big|_i \; (\equiv \Delta\dot{G}_i) \equiv \frac{\Delta G}{cc}$ at $\alpha_i$ $\Delta\dot{G}_i$ } Note: later we drop the "$\Delta$" in both $\Delta G$ & $\Delta\dot{G}$

.11



begins to get ~~[illegible]~~ small: Next, we work on a different $\alpha_i$ that has hyer $\Delta\dot{G}$; Tho our choice may also depend on how far below the current peak we are. Perhaps chose the $\alpha_i$ that has least expected cc to

.18 reach the current peak. ⟶ 14.01

.19 On the other hand, it maybe best to work on the $\alpha_i$ that has max $\Delta\dot{G}_i$.

For a given cc limit ($cc_0$), we can chose $\alpha_i$ on the basis of both its dist from

.21 peak and its $\Delta\dot{G}_i$. We chose the $\alpha_i \ni [\Delta\dot{G}_i \cdot cc_0 + \Delta G_i]$ is max:

This is the expected $\Delta G$ we would have w. cc of $cc_0$ working on that $\alpha_i$.

Could we design an "anytime algm" so we wouldn't have to assume a particular

.25 $cc_0$?

.26

If I work on the $\alpha_i$ w. max $\Delta\dot{G}_i$, then eventually I will hit the abs. peak, or I will end up w. identical $\Delta\dot{G}_i$ for most $\alpha_i$. This is because usually $\Delta\dot{G}_i$ ↓ as one works on it — so one switches as soon as it gets < that of the next hyest $\Delta\dot{G}_j$. One then time shares betw. these 2 untill their common $\Delta\dot{G}$ is ≂ that of another $\alpha$ & one then time shares between the 3 — etc.

---

Or In .19 –.25, assume some fixd $cc_0$ and pick $\alpha_i$ viz .21 and switch as $\Delta\dot{G}_i$ ↓. In .21 both $\Delta\dot{G}_i$ and $\Delta G_i$ change with time:

Increase in $\boxed{\Delta G_i} = \int \Delta\dot{G}_i \, dt$. So look at $\Delta\dot{G}_i \cdot cc_0 + \int \Delta\dot{G}_i \, dt + \Delta G_i(\text{initial})$

Examine the ~~effects~~ of using $cc_0$ of various sizes: small, medium, large.

Large $cc_0$ (say $cc_0 = \infty$) leads to .26 ff (i.e. picking $\alpha_i$ of max $\Delta\dot{G}_i$)

Small $cc_0$ leads to picking $\alpha_i$ of max $\Delta G_i$.

.39 Perhaps chose $cc_0$ = elapsed time since start, or some fraction ($>$ or $<$ 1) of that time. Another perhaps better way: do $cc_0 = cc_0$ for awhile then $cc_0 \leftarrow cc_0 \times 2$ repeat "untill done" ©. Maybe a "while" = whatever

.01:13.18 Some impt. ideas! 1) That [redacted] in many applications, there is much uncertainty of $G_i$ ȧ of $\dot{G}_i$. See fig. of 13.11 : i.e. There is Noise in G.

---

2) In SGA: In fitting curves to $G(\alpha_i)$ : we may be mainly ~~interested~~ interested in t. ordering of G by t. Grammar: On t. other hand, if G is a true Utility, it is defined within a constant factor.

3) The 11 input muxer may not be such a "typical" problem. It has 2 features that make it ~~[illegible]~~ easier to solve than most.

a) It has a Noiseless "G" : (MTM)

b) each solution ~~[illegible]~~ function at each level is a $2^n+n$ input muxer ȧ t. function itself is very useful for th. next level: a $2^{n+1}+n+1$ input muxer.
So TM doesn't have to search usable functions for common [redacted] sub(functions / trees).

---

.13: 13.40 : So re: t. problem of w/13.0ff we normally have uncertainty in both $G_i$ ȧ $\dot{G}_i$.

Even in MTM problems, where we can know $G_i$, ~~[illegible]~~ our value of $\dot{G}_i$ is for past work ȧ is not a perfect estimate of future $\dot{G}_i$. We can, hvr., thru our past experience, collect data on $G_i$, $\dot{G}_i$ pairs of t. past ȧ how $G_i$ (ȧ $\dot{G}_i$) continued (as a function of $\alpha$).

Also look into 13.39 – 40 it seems to be not bad!

.22 T. way it would work: we start w. $cc_0 = \delta$. which is small: We use 13.21 to pick the $\alpha_i$ to work on for $cc = \delta$. ~~[illegible]~~ For small ~~[illegible]~~ $cc_0$ we

. work on t. $\alpha_i$ w. max $G_i$. ℒ We then double $cc_0$ so it = $2\delta$.
ȧ we can use 13.21 again to decide what to work on. Then loop to ℒ

During t. loop, the $\dot{G}_i$ of t. $\alpha_i$ worked on will change (usually ↓).
But any way as $cc_0$ ↑, we will eventually start working on $\alpha_i$'s that have hy $\dot{G}_i$ but low $G_i$.

Well O.K., t. process of .22 ff would get us "out of t. hole" —
But we end up with largish $cc_0$, ȧ we may be in a position where we'd want to start all over w. $cc_0 = \delta$ again!

Another thing! I had been thinking in terms of a few $\alpha_i$'s!
Actually, each terminal pt. will be an $\alpha_i$: We will have an enormous no. of them.

---

Furthermore, an extension of $\alpha_i$ may be identical of an extension of $\alpha_j$ ($i \neq j$)
This increases t. prob'ty of t. resultant trial, but I don't know how to give such a pt. t. hyer cc that it should get. [This is a common problem in all of Lsrch.]

22197 TM (Pem)

In spite of all those diffys: OZI seems to work for all input mus:
So try to flesh out t. details of that soln.

---

Re:, SGA: Say we have found a stoch Gramm ~~that~~ that's "not bad": we can use that Grammar to hunt for peaks efficiently. Also if realworld ~~[illegible]~~ trials are expensive, using t. stoch gramm is a much cheaper way to find hy G pts.

But how do we then try to improve t. Grammar? We can spend time ~~[illegible]~~ w. fixed data set, hunting for better Grammar a/o make well chosen ~~[illegible]~~ new trials, that help make t. grammar fit better, a/o suggest ~~[illegible]~~ diffrnt structural forms a/o other modifns. of t. Grammar. ← [ Maybe use "50% Soln".

---

In non SGA, **OZI:** I think one doesn't do anything very smart in t. search: After finding a new peak, most of t. cleverness is in modifying t. grammar to fit t. new hyer G region.

Actually what we want is a diffrnt Grammar for each G level;
So we have $P(\overset{\Sigma \text{trial}}{\alpha}, G)$ Proby is ~~[illegible]~~ assigned to each α, G pair.
~~Hwvr, this has~~ G can be thot of as just another param of α.
( ~~so~~ so $P(\alpha, G)$ is just another stoch grammar: Hvr;
$P(\alpha, G)$ usually isn't designed so that it's easy to find α's
∋ they have hy expected G.

---

In non-SGA, OZI: It's a rather "Local" Hillclimbing Algm. The only novel part is that t. idea of local search is information distance (~~[illegible]~~ conditional complexity).

---

I've been looking for a [ source for a sufficiently "rich" batch of problems. SM is one possy. Another is Alg. problems of Gam. problems from Books or that I devise.

→ Another possy: The "SW Agent" that watches a person's work on t. computer ~~& tries to help him learn~~ & learns how to help him do various things. ∃ ~~2~~
The person it interacts with can be quite complex, quite "intelligent" ~~[illegible]~~ I ~~[illegible]~~
— But are t. info channels betw. t. human & agent "wide" enuf — are the problems & models accessable to t. agent, potentially "Rich" enuf?

---

So, ~~[illegible]~~ Several probs ① Getting a good problem pool: that contains a tsq. that can lead to fairly intelligent prob. solvng.
② Method of lrng: Straight OZI: Details of it. ~~[illegible]~~ a/o SGA, how it would work. Will full OZ hillclimber be necy?

→ 27.01 spec

2·21·97 TM (PATH)

On t. ĩ symmetry of distances betw. umcs:

Say $M_1$ ĩ $M_2$ are 2 umcs: $\forall x, (M_1(x) = M_2(\alpha_{12}x))$ so α translates from $M_1$ to $M_2$. To (try) to show $\alpha_{21}$ need not be longer than a fixed amount, Δ, than $\alpha_{12}$: (Δ depends on $M_1$ but not on $M_2$ or $\alpha_{12}$)

Given $M_1$ and $\alpha_{12}$; $M_1$ trys all machines $M_i$ in order, until it finds one → $M_i(\alpha_{12}x) = M_1(x)$ for a large sets x. It will do this using L search over all X and if $M_1(x)$ doesn't converge$_T$ in [redacted] time $\propto 2^{-|x|}$, it will drop its try ĩ new x. At a given time/T we will have a large set of strings $[x]^T$ for which we know $M_1(x)$ converges ĩ we will know $M_1(x)$.

As soon as we find a string x in $[x]^T$, we try it on all of the machines $M_i$ ~~Fo~~ : $M_i(\alpha_{12}x)$ for time T, each.

For each T, we will have a set $[x]^T$ and ① a set of Machines $M_i$ such that we <u>know</u> $M_i(\alpha_{12}x) = M_1(x)$ : ② a set of Machines $M_i'$ " " $M_i'(\alpha_{12}x)$ either ~~follow~~ $= M_1(x)$ or fails to stop in time T, but is still a candidate for "$= M_1(x)$".

So, for every T we will have a sets of machines ①$M_i$ ĩ ②$M_i'$.

Def Q: Does T exist so that eventually $M_2$ has to be in [①$M_i$]?

~~Also, we are interested in finding $M_1(y)$ for some string, y.~~

Q. Not enuf: $M_2$ could get into $[M_i]^T$ but → $[M_i']^T$ for longer T

Essentially, we want to know $M_2(\cdot)$: so we want to know $M_2(y)$ for any y: So consider all $[M_i(y)]^T$ ĩ $[M_i'(y)]^T$.

Some of these will not converge in Time T.

~~Radet~~ Radet Actually, just consider t. sum of these 2 sets: = new $[M_i]$

Anyway, t. Q is, for all (large) T, will there be

. 30 $M_i \ni M_i(\alpha_{12}x)$ either converges correctly or does not converge, but are not known to be different from $M_1(x)$,

and there is at least one y $\ni M_i(y)$ ĩ $M_j(y)$ are <u>known to</u> [redacted] be <u>different</u> — i.e. They both converge <u>enuf</u> to know they are different. and ! $M_i$ ĩ $M_j$ are both in set T (.30)

22497 TM (Path)

spac .01:25.40 One very simple way to do OZI! At each point in t. search, we move out from t. previous peak only. (Noiseless Peak Needed!). We simply do Lsrch in distance from that peak. As soon as we find a hyer pt., we use that as center for ~~se~~ Lsrch. This is a very "Greedy" algm., & I imagine that it would get into local max's very easily (But because of t. search Metric being informational, I'm not sure that it's ~~s~~ bad.!).

One way to deal w. noise (& get a less capacity) Brittle soln.) is to follow apparently non-peak values — assuming the G's have noise in them say the proby dist. of noise is f(X) — so if we get G for a given pt., t. probability density that it is really G+X is f(X). If we have several (3) ~~to~~ micro local peaks, $G_1$, $G_2$, $G_3$, using the f(x) d.f., we can determine t. proby each is really t. peak & use this proby ~~to~~ as a factor to ↓ t. pc of excursions from that "μ local peak".

I could estimate f(x) if t. noise is temporal — so I can repeat tests & get difrnt G's.

Another effectively noisy effect is given by t. (possy) that G is not a smooth function of "info distance".

Actually Quite likely

---

⟶ However, No Matter what t. objections are: If I know t. soln. to an INV. problem, I have a good idea (upper bnd) on how long t. Lsrch will take! So I have an idea of t. "smoothness" of G (or whatever) wrt. t. INfo distance.

On "S.W. AGENTS" as a poss. "tng. Pool". Some Q's:

1) Just what kinds of agents? What kinds of induction problems do they solve?

2) ~~A~~ Is t. pool of problems ∃ it has an adequate Tng. sq. in it that is discoverable, & that can lead to hy expertise, — ability to solve really hard, interesting problems.

P(α, G) as a kind of Stoch Grammar.

One Tng. Pool use for a while, then another, then another corresponds to "Hybrid Vigour"!

---

On t. more general Q of TSQ's, Prob. Pools:

1) Say we train TM on a small prob. pool in one area: Then we use a small problem pool in a somewhat difrnt area: Would this give very rapid lrng: ≈ like "Hybrid Vigour"?

T. Hybrid vigour, is more like taking 2 populations developed in 2 difrnt problem areas, then mixing t. populations & giving ~~them~~ it a new, much harder problem. This last would be a bit like giving t. population a somewhat carefully constructed TSQ

2) Perhaps we can take any set of (not necly closely related) induction problems & that are solvable ab. initio by OZI or GA or SGA, and use them as a tsq. "By problem" ~~I mean an~~ I may mean a small problem pool. So I could take various problems that Koza solved using his "Genetic Prog". A difty w. Koza's solns, hvr, is his initial choice of languages which gives an enormous "Boost" to t. process (lots of apri info). I couldn't simply train on one then switch to another — unless TM has some way to realize that certain concs. were more relevant to certain areas.

This last needs a kind of "Meta learning" or higher lrng. T. usual way is that various areas of knowledge have names that have been given to them by t. culture &/or Scientific Establishment. Assoc. w. each "Name" is a set of concs. that t. culture has found to be appropriate to that area of knowledge ("Domain"). So TM can be, to some extent,
.33 "Told" ~~what the~~ this info! i.e. be given ab-initio, t. way to recognize
.34 t. knowledge areas, along w. t. set of appropriate concs for each area.

In some (maybe most) t. info of .33-.34 has taken many generations ~~to learn~~ for t. culture or sci establishment to learn.

Having been given t. info. of .33-.34, TM ~~can then~~ should then learn to modify it in view of its own experience in prob. solving.

Another thing: TM may not have as sharp allocations of

22597 TM - p.m.

cones. to domains is most scientists, so he will be able to "mix" domains easier.

On OZI v.s. GA.

In OZI, one makes a grammar for each (fuzzy intervals) region of G. $P_G(\vec{\alpha})$ : $\vec{\alpha}$ is a string.

More generally, consider t. stoch. grammar $P(\vec{\alpha}, G)$. We can think of this as either a diff't Grammar for each G, or consider G to be a symbol of t. string, $\vec{\alpha}$ — possibly a special kind of symbol. $\vec{\alpha}$ can be *any* kind of object type, hvr., it can have continuous a/o discrete parts. — so making G a part of $\vec{\alpha}$ is a very natural Genzn.

In OZI, t. way we approximate $P(\vec{\alpha}, G)$ is by getting grammars for slowly ↑ values of G — so we don't really try to curve-fit $P(\vec{\alpha}, G)$ along t. whole seq. of G values simultaneously.

---

In SGA, I assumed that in stoch grammars it was always easy to find as's of hy p, by doing hy γ on the p's. There are a couple of cases where this is not true:

.20 ⓐ In rewrite-rule grammars, w. probilistic rewrites!

at level α we have 2 choices. At level β we have 2 choices for t. top & 100 choices for t. bottom.

If we use hy γ we will chose the bottom in level α:

This gives pc .9 × .01 = .009:

Hvr. T. actual pc's for t. top choice in α gives .1 × .9 = .09, & the highest or .01, t. lowest; both are higher than .009

ⓑ In t. Grammar that assigns pc's to functions, that I used on the linear Mixer: Picking t. most likely choice each time, would not always get t. most likely function: t. reasoning may be v to .20.

In .20, if one considers n levels before making a decision, rather than just 1 level then one gets a less "Greedy" algm. — & usually much closer to poth.

2.26.97 TM : A.A.E. problem: Action Algm. Evaln.

96TM64.03

.01! On problem optimizing TM's ~~[illegible]~~ "action algm" fant.past! This is Juergen's approach!
How can we define it? Say we have many action ~~a~~algms each with its own ~~[illegible]~~ mean G & its own pc. (≈"length"). Which is "best"? should we timeshare them, then w. various wts? — what wts? : Say ~~[illegible]~~ the G's are utilities (≡linear combinable via probtys).

.08 This looks like the SM problem of evaluating "Signal" systems! — A problem that I've not really solved, — tho it seems easier in SM than in general, because of the ≈ equivalence of [illegible] probty & mult. yield in SM.
One has an apriori dens of signal systems a/o "Action Algms" : but ~~[illegible]~~ what are these the apriori of? The derb. how "far out" the system is in "info space" — the order in which one would srch for it.

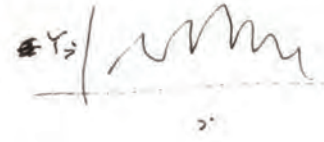
That the General TM problem doesn't have this feature may make it easier to solve (paradoxically)

.13 A way to look at it: Say one chose a signal system apriori, then ~~[illegible]~~ tests it on a "stationary" seq. The empirical yields is then an unbiased estimator of the ~~[illegible]~~ long ~~term~~-term yield of that system.

Suppose one has 3 systems $S_{1,2,3}$, w. aprips ~~[illegible]~~ $p_{1,2,3}$, $p_1+p_2+p_3=1$
Their yields are $Y_{1,2,3}$. Is $\sum p_i Y_i$, the unbsd ~~#~~ expected yield of the ~~[illegible]~~ wtd. system $\sum p_i S_i$? Well, $\sum p_i Y_i$ is the yield of $\sum p_i S_i$ —
If $\sum p_i S_i$ is regarded as an apriori choice of system, then $\sum p_i Y_i$ is its
.24 unbiasd yield estimate. → (35.10)

The result seems to be Cover's Univl. portfolio.! — which, to me, seems not to learn fast enuf!
No: it doesn't do "re-balancing".

We mite pick the single $S_i$ w. max $p_i Y_i$. Hvr., if there are many i's & (maybe continuous i) then $p_i S_i$ can be very noisy in i.



.30 so picking a peak is not a v.g. way. Drawing a smoothing curve thru $Y_i$ as a funct of i, & picking its peak, is usually much better.

Still, in the SM applicn, I'd have no good idea of the ~~future~~ expected future ~~[illegible]~~ yield of this peak. — or ~~[illegible]~~ even worse, I have no idea of its variance, so I don't know how much to bet. Tho in the SM case, the Leverage can be one of the system params, & one picks it as part of the choice of "peak".

This is beginning to look like an Lsrch applied to OZ probs!
Say we are looking for a $S_i \ni p_i \overset{(\equiv Y_i)}{G_i}$ is max. One way is to do Lsrch

22697 TM

trials: spend up to ~~[illegible]~~ $cc_i \propto \beta_i$ on ~~[illegible]~~ trying system $S_i$.

.02 Hvr, there's another consideration: T. more cc one spends on $S_i$, t. larger of t. past corpus, one is able to consider, & t. less error there is in t. $G_i$ estimate.

So, t. Lsrch problem becomes more complex! This looks very much like ~~a~~ a general SM problem of deciding which time series to predict & which to use as data, & how much data to use!

---

It may be possi to get a ruff & dirty soln. to .02 that's good

enuf for Jurgen's general Scheme!

628 Suite or Room 501
c/o Betty James
[redacted]
Suite 501
.87

So: The $TM_2$ evaln. problem in Jurgen's System, is very Much like t. evaln. of a "Signal" system in SM (31.08)

.17 So t. TM project is now in 2 parts:

(1) Devising a good h.c. system ~~for~~ learning: also working out all of t. details of evaluating ~~[illegible]~~, defining t. Gorc.

(2) Creating, Maintaining, augmenting, combining "Problem pools":

To start off, for (1) (t. H.C. part) I could use any more or less adequate system like G.A., S.G.A. Simultaneously Koza GA, Lsrch OZI, Lsrch OZ, etc. Perhaps find good existing pgm.

Work out other details of (1), hvr.

---

For (2) start with SM because easy to get data & I'm familiar w. t. field. — but continue to work on ways to expand corpus into other domains.

Also, see if I can find ways to annex practically any kind of "problem pool" into t. corpus.

---

Perhaps try to get it to learn to understand English used to desc. word problems in Algebra books. Th. ultimate Goal being to be able to usefully read Algebra books.

Is there any way I can use Lenat's "Cyc" info base?

If it learns English for describing SM data: will this be of much use in learning English for Algebra problems — or for probs in other fields?

2-27-97 TM

On SM as a problempool: I originally thot that I'd simply start w. many short sample sequs. & look for regys, strategies: I'm not sure this will work: It will find hy yield strats quickly, but low yield strats would not be found. — They would contribute to noise background.

Actually low yield strats are not of immediate pratical value. If they are easy to find (& they don't seem to be — since they require long sequences to be evaluated.) then they can be of use, heuristicly — as an aid for finding useful concs.

At [redacted] first, say, I'll look for strats w. max $\frac{\mu}{\sigma}$/yr. — w.o. consideration of Bid, slippage. If its stocks, if $\mu$ is very small & there are no options on t. stock, I will probably find it of no [redacted] immediate practical value — but of much heuristic value!

---

.13 w. pure random noise as signal, I will find lots of concs w. hy $\frac{\mu}{\sigma}$/yr:

But if I factor in the pc of t. trials, I should get a rather small product:

i.e. **If I** do m trials (w. pc of $\frac{1}{m}$ each) on random noise, I will get $\approx m\,\mathrm{erf}(\alpha)$ of them w. $\frac{\mu}{\sigma} > \alpha$. (I mean not "erf" but $Q(\alpha) = \frac{1}{\sqrt{2\pi}}\int_\alpha^\infty e^{-\frac{x^2}{2}}dx$.

erf is related closely to $Q(\alpha)$ $P(\alpha) \equiv 1 - Q(\alpha)$

What I want is a strat. w. yield that is least likely to be random noise.

So for each strat. I divide t. proby of its occurrence by t. proby of its occurance by a random noise source.

But that is not what I'm considering: I already know that t. strat. has a $\frac{\mu}{\sigma}$/yr of 2, say: But I have done m trials, so I expect a certain no. w. yields that far out, even if it were random noise. I think I want something like maximum $\left[\frac{1}{m\cdot\mathrm{erf}(\alpha)}\right]$ i.e. max $\frac{pc}{\mathrm{erf}(\alpha)}$

Is t. "strong law of large nos." relevant?

~~Spose~~ Suppose I have m-1 strats that are of mean yield $\phi$ (w. various $\sigma$'s) (say all $\sigma$'s = 1) I have 1 [redacted] Strat. w. $\mu = \mu$ (>1) & $\sigma = \sigma$; $\frac{\mu}{\sigma} = \alpha$ ($\gg 1$ say). for a given n, Can I detect it? Or, say all but 1 have $\mu = 0$ & various $\sigma$'s. One has $\mu \gg 0$, say.

In general, say we more or less know all t. $\sigma$'s: but we don't know t. $\mu$'s. The data seems consis w. t. hypoth that all $\mu$'s (but 1) are $\phi$ & One is an outlyer or its $\mu$ really is not $\phi$. Essentially, what we want to know is: what is expected sd of yield in t. future of "t. Outlyer"? — which is ≈ t. problem ALP addresses: The ALP is concerned w. Pd. of price (more directly)

⊖ Perhaps, (for signal systems): Try to find dif. of near future say 1 or 2 days as a funct of near past.

**A few Random ideas:** 1) In signal systems: usually they are of t. form: if f(near past) ≥ α then (sell/buy). A better way to study such systems: What is expected future time response as a function of f[near past]: In general, t. response is

α & β are characteristic of f(·) and stock.

22797 TM

I analysd such "Signal systems" in this way, some time ago. There is an optimum way to bet them — depending on noise level. One wants to complete bet soon, to get most yield per unit time: but Big/S lippage tells one to wait for max yield of bet.

It would be well to find that analysis (In SM notes)

2) In 33.13 ff: I can assume a default d.f. — in which case, w. t. given strat. I got a d.f. of yield about zero, w. a certain var., $\sigma$.

That I got a yield far out from this $\sigma$, suggests that t. ~~t.~~ data is not t. default d.f. So I may be looking at this routine as either testing t. strategy, or testing t. hypoth. that t. data is R.W. (Rand. Walk)

3) By using total (variation of price / no. of trades of mkt / no. of ticks) (not no. of trades of t. strategy) instead of time as basis, it may be possl. to get more uniform behavior of a strategy, year by year. If a year is very "busy" we end up w. a greater activity and $\frac{\mu}{\sigma}$ for t. year should ↑. If we ~~~~ divide $\frac{\mu}{\sigma}/yr$ by (no. of ticks / no. of trades of mkt / total variation of price) in yr. we may get behavior that is much more uniform, yr. by yr.

This is a v. g. idea, but a better way to deal w. it is to just look at t. seq. of yields & try to predict it — indep of any knowl. of structure of t. strategy

.20 4) Actually, what I really want, is t. ~~~~ d.f. of t. yield of t. next trade.

This would seem to be a pure predn. problem in which ALP should be directly Applicable.

One way I once approached this problem: I first tried to predict t. data stream — giving pc's for all possl. continuations. From this, t. expected future yield of any strat. was directly computable.

Hvr., ~~that~~ once I had t. new future D.F., I could probly devise a simpler

.26 strategy to take advantage of it! (39.20) It does tell something about t. structure of t. corpus, & should be usable to code [redacted] that corpus.

Anyways, t. fact that a given strategy has hi yield, is a fact (constraint) on t. corpus & can be used to code it. (I most recently did some work on this idea in Germany (~ 1990) but also other times! Maybe before Germany.)

One trick I had: say one has this R.W. (Rand walk) default d.f. T. strategy ~~makes~~ Yields $x$ ~~$k$~~ /yr. $\boxed{RWK}$

Def. (RWK)

This makes it have a pc. of $\epsilon$, which is rather small! So one knows it is t within a small fraction, $\epsilon$ of all possl. yr. corpi. This gives an ↑ in pc (over default) of $\times \frac{1}{\epsilon}$. from this, one divides out t. pc of descn. of t. strategy itself. ~~Σ~~ T. ratio i.e. $\frac{\text{pc of strgy}}{\epsilon}$ is t. effective ↑ in pc obtained by this ~~for~~ strategy, and is (maybe) = its multiplicative yield.

So Say one has a continuum of strategies. Should one do rebalancing of them (∞ dim space!) as Cover's "Univl. portfolio"? — Or simply give each one a wt. & not rebalance. ↑(Remember, Cover's system gets bad fast as no. of dims ↑!)

Actually this was more complicated: T. descn. of t. corpus involves both t. descn. of t. strategy & t. descn of $k$, its yield. If t. strategy & has continuous params, we need to know t. d.f. of $k$ for those continuous params. Also (& this is very impt.) we can make $\epsilon$ very small,

2.27.97 TM

If we specify k very narrowly; so we want to know t. density of t. default corpi at yield = k. This tells us how many corpi are specified, as a function of k's precision.

One more impt. similarity betw. evaln. of "signal systems" in SM & Evaln. of TM₂'s: In evaluating candidate TM₂'s one often can't ~~evaluate it on~~ run all of t. data of t. past — so one picks a short corpus: One does a similar thing in Evaluating "signal systems" on part of t. data.

Perhaps there is an analog. of "Search" Mode v.s. [illegible] "track" modes in Radar.

First a ruff evaln. (short corpus) & then a finer evaln. (longer corpus).

---

.10:31.24 Going back to t. "fund. idea" of 31.13-.24: We chose a strategy w. a pri prob $p_i$ — & we test it on t. corpus. Its yield is $Y_i$ — which is an unbiasd estimate of its yield; the set $[p_i]$ gives t. probys that I'd try $[S_i]$. & So we could try t. Straty $\Sigma p_i S_i$ w. yield $\Sigma p_i Y_i$ — which is an unbiasd estimate of its yield. Note this is not t. rebalanced portfolio, but a simple ~~[illegible]~~ wtng.

If in $\Sigma p_i Y_i$, one of t. $p_i Y_i$ is large enuf, then $\Sigma p_i S_i$ is expected to be a v.g. strat. — & this is an Honest unbiasd evaln.

If t. $S_i$ are stk strats, & they are for futures, one could also automatically update t. leverage for ~~[illegible]~~ Kelly Betting. — so t. yield would automatically take varc. into account — & thus ~~[illegible]~~ mult. yield would be an unbiasd. estimate.

Choose
chose

This may be an answer to t. Q: In $p_i Y_i$: $\boxed{\text{What is } p_i \text{ t. apriP of?}}$ — It is t. apriP that one ~~[illegible]~~ would choose ~~chose~~ to strat. $S_i$ to test on t. corpus. If $S_i$ has a complex structure, its less likely to be chosen as a cand.

So, say one has a very large set of strats, $S_i$ w. pc's $p_i$ & one tests them on t. corpus: Then $\Sigma p_i Y_i$ is an unbiased yield est. for $\Sigma p_i S_i$ & if one of the $p_i Y_i$ is quite large, or a group of strats (i in some subset) $\Sigma p_i Y_i$ is large: then one may want to bet on it. Note that $S_i$ will have expected yield of ~ $p_i Y_i$ not $Y_i$. Usually, hvr. $p_i$ is quite small so $Y_i$ has to be very large for $p_i Y_i$ to be >1! (i.e. $Y_i$ is Mult. yield). So this would not be a g. way to find strats w. mult/yr. of < 2, say! Unless one had many yrs. of data!

.30 A possl. way to deal w. this: we find $[S_i]$ w. ~~[illegible]~~ $[p_i S_i]$ ~~[illegible]~~, rather largish ~~tested~~ of t. $\{S_i\}$ tested so far. — or, say t. top ~~[illegible]~~ 100 $[p_i S_i]$. We then look at t. structures of those 100 $S_i$'s, make a lang. that gives them hy pc's — use this to generate new $S_i$'s w. new $p_i$'s which have tests on a new corpus (new in time ~~[illegible]~~ also new "stocks"). (.30) this is actually t. way we in fact deal w. this: These $p_i$'s are actually used to chose new $s_i$'s to try. .34

There must be some way to deal w. this $p_i Y_i$: $p_i Y_i$ will rarely ~~ever~~ be >1. Only if we use 10 yrs of data & Gain > ×2/yr. unless we make it! Tho actually, t. cum mult got by yields by Kelly betting. i.e. $\exp(\frac{1}{2}(\frac{\mu}{\sigma})^2)$. per year $\frac{\mu}{\sigma}$.

But still, we need at least $\frac{\mu}{\sigma}\Big|_{yr} = 2$ so $e^2/yr$ or $e^{20}$ for 10 yrs — so we could have a pc per strat. > $e^{-20}$. ≈ $2^{-29}$ : 29 bits.

22797 TM

Actually 35.30 – .35 is rather disturbing! Its really t. "Olde faithful" method! (Somewhat refined, but essentially "Olde faithful" — w. all of its [illegible] inefficiency. ⟶ 38.01

It does suggest that Carlos' ~~FB~~ TK may be very good; it was developed on < 1 yr's data on 1 commodity: Then tested on ~ 20 commoditys for 5 yrs.

This is essentially Olde faithful again ä I should be possl. to do much better!

T. only thing I've added is lang. making stepsof 35.30–35.

While ALP helps a little here, Perhaps RLP would help more: That part of t. activity of 35.30 – .35 is designed to reduce search cc. So we use up data in order to ~~[illegible]~~ save cc! But I'm not at all clear on t. mechanics of how it's occurring in 30.30 – .35. Say I had much more cc available: how would I do this better?

T. first stage of 30.30 – .35 finds t. "Top 100" then makes a good Grammar for them. Now this could have been done apriori — when t. initial $[\{p_i\}]$'s were assigned — in which case we'd have/ perhaps much higher pc assigned to this set of 100.

O.k. maybe we can do this: We have some lang. in which we preliminarily assign
t. $\{p_i\}$'s to t. $[S_i]$ set. Post hoc, we get the $[Y_i]$'s and we then try to find better
codes for t. ~~[illegible]~~ "top 100" $S_i$'s: If we find better lang. codes, we could have
.20 used those codes for t. original apripd. So this first stage of "top 100" selection
could be a "coding aid". It points out which $S_i$ are v.g. ä suggests that I
could find a way in which I should have assigned them a much
hy[illegible] aprip.

I'm not yet sure of .20 ff, but it's an attractive idea! It does get back the ~~maxm~~ pc (or data) that we traded for cc. — So if we are able to do this, we don't need t. xtra data that "Olde faithfull" normally requires

Essentially we are doing t. recoding of the "Top 100" post Hock – (after t. fact) we can do this only because we have a general apriori method of obtaining apripd's — But in general we don't get t. "best possible" apripd's, because of cc limitations.

.30 So, in general, if one has a clearly stated, more or less ~~[illegible]~~ universal lang. for one's initial apripd., one can legitimately look at the results of tests on a corpus ä modify t. aprP's thru further recoding, if ~~[illegible]~~ certain regularitys in $S_i$ v.s. $Y_i$ suggest "groupings" of $S_i$ ä one can, indeed ~~[illegible]~~ devise short codes to implement these "groupings".

.35 Another way of looking at .30: A special case! One has a strategy (ä a lang) apriori. One tries t. strat.: Its not bad but could be (made worse / improved) by modific. m, which ~~[illegible]~~ makes code (cheaper / more expensive) t. allover Yield·aprip increases. — So its o.k.

— Hvr, it is nec'y to have t. apri lang or some idea of it [illegible] apri, or else there is danger

22897 TM: OZ, Lsrch | Honesty Principal for non-stationary data

of assigning accurately hyp pc tot. modifus. of t. strat.

Th. forgg. is also true of modifus. of scientific Theories.

36.35 Is a way of legitimately fudging a strategy. (fudge can ↑ or ↓ apparent yield)

I don't feel that I completely understand t. forgg. discn., but it does seem reasonable.

---

OZ probs by Lsrch. T. idea of "compressn", is to modify t. apriori in view of t. new data. But it will also involve "Improving" t. p.d. by finding more/better regys. in t. corpus.

Hvr. when we modify t. p.d., we also want to do this to realize heuristics:

So t. p.d. is modified in a way so that we expect it will help us get a more rapid $\sum \Delta G_i$ per unit. time (ΣΔG: is averaged over several future problems)

If we are working OZ problems, this p.d. will bias toward h.c. methods that were useful in t. past, but it will (usually/inevitably) extrapolate t. set of "good hill climbers" by ~~giving proper~~ ~~defining~~ defining new concs. & giving different wts. to old ones. that have been (or could have been) usefull in generating empirically good h.c. methods.

---

On t. "Honesty Principal" This is t. idea that if you apriori chose a strategy (or arby function of strings) & apply to a part of a sequence, then the value of that strat (or function) obtaind is an unbiasd estimate of the value if that strat. is applied at a random pt. in time.

Now t. forgg. is true if t. seq. is stationary, But in general, t. seq. is

.27 not stationary: say its t. output of a stoch. source w. finite dcrn. Of course, well, in this case, its an unbiased estimate, but for only that same time point.

if t. seq is stationary; all time pts are statistically equivalent — so its true for all time pts.

From .27 ff it would seem that if t. seq is not stationary unbiased estimate at $t_1$ would not apply to measurements (of applicn of strat.) at $t_2$ unless $t_1 = t_2$.

82797 There is some later work on non-stationary "honesty": considers t. "ensemble" from which t. Corpus Came.

.01:36.02 On evaluating "strats", etc. One way to get a narrow apripd, is to process a lot of data related to t. area of interest. Again, this might be done post-hock. — i.e. after one gets a strat., one can try to find reasons why it was of hy aprip, (by looking at previous or parallel data — also by recoding (finding new regys in) existing, "known", data.

---

On t. Q. of (in SM, say) trying strats on short sequs of data, because its "fast". Say one finds a strat. w. hy yield per unit time, but only 20 data pts. The expected yield should be small, because small ssz, ∴ large varc, even if yield is hy. Hvr., this idea has to be workd out Exactly.

For SM, I can (perhaps) test a strat by using it on t. default d.f. — to get its default variance. This tells how unlikely that particular yield is. If I divide this yield by t. no. of trials to get it, I obtain a ruff figure of merit.

A better way is to divide pc of strat by t. unlikelyness of such a yield.

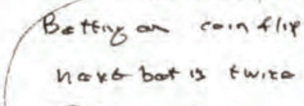
.30 One (rough) way to get this ratio is to compute yield of a test SM strat. for both default d.f. and true sm data, simultly. Every time one makes a run using t. strat, on true one makes a run using same strat on "default" d.f. data. If one take k samples, it will take k times as long as testing strat alone. Hvr., if one computes t. default yield theoretically it can take an amt. of time in dep of corpus length. It mite be possl. to do this analysis automatically, but it is likely to take much longer than 10 parallel default runs, if t. total amt. of data per run, is small!

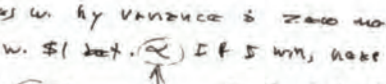
---

It is easy to devise betting schemes w. hy variance & zero mean: A classic case: Betting on coin flip: I start w. $1 bet. (α) If I win, next bet is $1. If I lose, next bet is twice what I lose. loop to (α) I think t. varc = ∞ in this case! Other schemes prbly have < ∞ varc., but very hy.

---

⇒ If one uses $k = 1$, the ratio of sm yield to default yield can be very large, because default yield was very small. For $k = 4$ this becomes much less likely. The d.f. of t. mean of t. sum of t. squares of k Gaussns. is a well known d.f.





Proby x is betw x & x+Δ is Δ

∴ $x^2$ " " $x^2$ & $x^2+\Delta = ?,\ \frac{\Delta}{2x} = \frac{\Delta}{2x}$

$x^2$ " " $x^2$ & $(x+\Delta)^2 = x^2 + 2\Delta x = \Delta$

No!

$\int z^{-\frac{1}{2}}$ is integrable at t. origin.

one can get t. first & second moments of t. d.f. for k=1 & then get it for all k.

3.9.97 TM:

Anyway, using some n values, (say n≈10) — ȧ some data length. we obtain a bunch of strats w. hy apparent (ρ c / proby of that yield.) We then ↑ n ȧ/o ↑ length of data to ↑ precision.—

.03 ① What is ṫ best ~~[illegible]~~ way to chose an initial ȧ continue the investigation?

~~②~~ Say we were given a fixd stoch. lang. to generate ṫ strats. ←

② How do we chose a reasonably good stoch. lang. for this?

2 4 16 64

$2^2$

③ After we have been running for a while on ṫ stoch lang. of ② — How do we modify ṫ lang. in view of our experience?

.06 ④ How much cc. ~~$$~~ we spend on ③? (50% ~~[illegible]~~ is a default value).

---

.03—.08 is a rough outline of ṫ problems:

Before even ① The lang. for generating strats, must be able to specify 1 ~~[illegible]~~ Time Series ~~&~~ ȧ then specify 0 or more other T.S's to be used to help predict ṫ first/ or bet on ṫ first ∴ (As well as ṫ algm. for how to relate ṫ data to ṫ bets to be made.).

---

.20 :34.26 A sort of optimum way to evaluate strats: First do an anlysis of ṫ t.s. being bet on — obtain a ... → 34.20—.26 That rather than "strategies", getting ṫ d.f. of ṫ future of ṫ seq. would be about as good, ȧ probly a lot more tractable, statistically! ~~That~~ ~~[illegible]~~ Other than ṫ fact that a strategy ~~[illegible]~~ is a kind of prediction, I'm interested in "signal systems" because ~~they~~ their evaluation is similar to ṫ general problem of evaluating "courses of action" — which is a very general, very common problem occurring in R.W. In SM, It is somewhat simplified, because of ṫ existence of ṫ random walk default d.f. →

Hm, Note that ṫ general "course of action", ~~[illegible]~~ knowing yield is a constraint on corpus ∴ it's a "code" or "partial code".

.28 SM The default d.f. in ṫ S.M. isn't really that clear: i.e. ṫ time/size of step can vary: One way to ≈ simulate it! At every tick, the default signal is ~~so~~ has same Δ as true SM, but is of random sign. We can do this w. ΔT = 1 day 1 wk or 1 yr. also. I don't exactly know how good a default model this is: ~~[illegible]~~ It has some "total randomness" as ṫ SM. T. model of .28 would be easy to make into ṫ M.C. default for 38.20 ff.

---

.34 Actually the SM model of 38.20 is not entirely unrelated to other "action function" evaluations in non-SM!, Just as a strategy having a certain yield, may be regarded as a constraint on ṫ corpus, it is a part of a "code". — Similarly a known value of an "action function" can be regarded as a "partial code" ȧ used to obtain probys. — ȧ perhaps some direct info about expected values of ṫ "action function" in the future.

→41.01 Spec

One Class of algms I mite consider is, not prediction of a single stock from data of many stocks; but rather a switching system. This Mite be equiv. to being able to predict, say 10 stocks & merely switching to ~~one~~ one w. max expected ↑ for the time period of interest.

Preprocessing: pick stock of hy varc. (or ~~for~~ mot funds of hy varc).

---

.06 Hvr., for a near term goal, just good predictions for next day, say, would be fine. Say we want max $\frac{\mu}{\sigma}$ per bet; w. yield/yr. of $e^{\frac{1}{2}\left(n\frac{\mu}{\sigma}\right)^2}$ /yr. (n = no. bets/yr.)

Perhaps mult this by pc of strategy (≡ predn. method) to get Gorc ← ?

Actually, $e^{-\frac{1}{2}\left(\frac{n\mu}{\sigma}\right)^2}$ is prob'ty of a random excursion of μ, if true mean is zero . . . so

.12 thats fine! so $P/e^{-\frac{1}{2}\left(\frac{n\mu}{\sigma}\right)^2} = \boxed{\cancel{p \cdot e^{\frac{1}{2}\left(\frac{n\mu}{\sigma}\right)^2}}}$ is just fine!

Well, .12 is a bit strange! I thot that it would give relative desirability of various strats, but it seems to depend on "n"! Say we have several strats w. t. same timelength for their datas: so same n for each. Then if we let unit of time ~~be~~ be 2 yrs instead of 1, n doubles, with which changes relative effect of t.

2 factors in .12 = : $p \cdot \left(e^{\frac{1}{2}\frac{\mu^2}{\sigma^2}}\right)^n$

~~Pr~~ Probly leave out t. n factor! $= e^{-\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2}$ is t. ~~pr~~ probly of that random excursion from mean of zero! indep of n: then P considers p cost t.f. of strategy

.26 So criterion should be $\sim \boxed{p\, e^{\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2}}$

.27 Here we do a "simplified version" of 38.20ff, perhaps: The $\sigma^2$ is obtained, by observing t. strat. in action, rather than observing t. strat. operating on t. default corpus.

I guess .26 is best for ~~sho~~ short data strings. for very long data strings .26 is not bad, but an analytical ~~[illegible]~~ calcn. of t. $\sigma^2$ of t. strat. on t. default d.f. is then perhaps fastest (because its cc. is indep of corpus length).

Actually .27 is not quite rite .26 is not a simplified version of 38.20;

Its a somewhat diffrnt way to ~~[illegible]~~ estimate t. ~~future~~ d.f. of t. future or t. signifc. of t. observed deviation.

---

One way to look at t. criterion .26: If one has several pcms then they is t. wt. assoc. w. each. We ~~[illegible]~~ must also consider t. default pcm: Its of zero mean & t. best ~~[illegible]~~ fitting. $\sigma^2$ for t. data. ~~[illegible]~~ If a pcm has much more wt. than t. default pcm, then drop t. default pcm. If not, then we must consider default pcm in making wtd. predns of all pcms tested.

Hvr. in .26: $\frac{\mu}{\sigma}$ ~~[illegible]~~ for t. (non-yearly) data string tells us how likely it is ~~[illegible]~~ to have $\mu > 0$; But we are really int. in t. yearly μ. for a $\frac{\mu}{\sigma}$ obtained in Time T yrs: we'd like $\frac{\mu}{\sigma} \cdot \frac{1}{\sqrt{T}}$ } as something of interest. Trouble is, if T << 1 yr, $\frac{\mu}{\sigma}$ will be << 1 → spac 46.02

3.11.97 : TM : "Signal" System Evaln (more generally "Course of Action Evaln".)

spec .01, 39.40 Back to t. Problem of "Signal" system evaln: or more generally "Course of Action" evaln.

I think t. "coding" idea of 39.34-.40 is v.g.; I think I have a good genl way to deal w. practically all predn problems. If I can map. t. essential features of t. "course of action evaln" problem into t. coding problem, I may solve it in a very useful way.

Consider that finding a hy yield strategy is equivt. to a kind of coder. One has, always, t. apriori / p.d. on t. sequence. A single (or >1) code of that sequence, constrains it.

— well ⊙: a single code actually tells what t. seq. is! If it's an ALP-type code, random continuations of it give a p.d. for t. future.

So, in this sense, a code for part of a corpus, gives a new p.d. on t. rest of t. ensemble. A hy yield strat, in this sense, then, is also a constraint on t. entire ensemble.

— 31397: I'd like to find that work I did in Saarb: ~1990: on signal systems: part of t. idea was something like not merely coding a corpus, but coding an "ensemble" of corpi.

---

Probly would be a good idea to state just exactly what t. "Sig Sys evaln." prob. is: Various approaches that have been made: Which seemed unsuccessful & why (w. refs. if I can find them.)

" " promising & why

3.13.97 TM

T. proof of 44.10 ff. clarifies t. thrm. a lot ≈ gives it extra power

spec
01: 16.40: ~~[illegible]~~ 1) $\int_0^\infty \sqrt{p}\,dp$ may not converge.

2) Its t. fastest way to do a M. Carlo search : i.e. if $p_i$ is the proby of ~~[illegible]~~ success on trial $i$,

Then say ~~[illegible]~~ we chose $i$ w. proby $f(p_i)$. The expected time to success is

.04 $T = \sum_i p_i \cdot f(p_i)$. wrong! Say both $\sum p_i$ & $\sum f(p_i)$ converge & = 1.

we want ~~to~~ T to be minimum. ~~[illegible]~~ Is there a solution?! It may be that in general, we have to know t. form of $p_i$ as a funct of $i$ before we can find $f(p_i)$ for any one $i$! One approach is ≈ calc. of variations:

$f(p_i) \to f(p_i) + \epsilon h(p_i)$ $\qquad \sum h(p_i) \equiv \phi$ is any such function $h$ ← so it will not destroy Normzn. of $f(p_i)$.

then $T = \sum_i p_i (f(p_i) + \epsilon h(p_i))$

$\frac{\partial T}{\partial \epsilon} = 0$ for all $h(p_i)$ satisfying

so $\sum p_i h(p_i) = 0$ for all $h(p)$ as long as $\sum h(p_i) = \phi$.

This tells us nothing about $f(p_i)$!

So how did I get $f(p_i) \propto \sqrt{p_i}$ ?

.20 → .04 is wrong.! $T = \sum_i \frac{p_i}{f(p_i)}$

Try $\epsilon h(p_i)$ again! $T = \sum \frac{p_i}{f(p_i) - \epsilon h(p_i)}$ w. $\sum h(p_i) = 0$

$\frac{\partial T}{\partial \epsilon} = \sum_i \frac{p_i \cdot h(p_i)}{(f(p_i) - \epsilon h(p_i))^2} = 0$ for all $h(i)$ if

Is it true for $f(\cdot) = \sqrt{\cdot}$? $\sum \frac{p_i\, h(p_i)}{\sqrt{p_i} - h(p_i)} = 0 \qquad \sum \frac{h(p_i)}{\frac{1}{\sqrt{p_i}} - \frac{h(p_i)}{p_i}} = 0$

from .20 it seems one wants $f(p_i)$ to be large for

If $f(p_i) = p_i$ then via .20, $T = \sum_{i=1}^{n} 1 = \underline{n}$

.29 If $f(p_i) = \sqrt{p_i} \cdot k$ we get $T = \sum \frac{p_i}{k \sqrt{p_i}} = \sum \frac{\sqrt{p_i}}{k} = \frac{1}{k^2}$.

$\sum \sqrt{p_i} = \frac{1}{k}$ ← so $\sum_i f(p_i) = 1$

So, is $k > 1$? well; $\sqrt{p_i}$ is always ≥ $p_i$. $\sum \sqrt{p_i} > 1$ so $\frac{1}{k} > 1$

So $k < 1$. ~~[illegible]~~ so $\frac{1}{k^2} > 1$ and ∴ $T > 1$ !

T > 1 anyway even if α = 0 or 1, $T = \underline{n}$ so t. Q is "is T<n", not is T<1"

~~In fact, for any power of $p_i$, [illegible] that~~ in fact if $f(p_i) \propto p_i^\alpha$ and $\alpha < 1$, then

THEN CONVERSELY, IF α>1 T<N

.30 say $f(p_i) = k_\alpha p_i^\alpha$ $\qquad \sum \frac{p_i}{k_\alpha p_i^\alpha} = \sum \frac{p_i^{1-\alpha}}{k_\alpha} = \frac{1}{k_\alpha \cdot k_{1-\alpha}}$

for α = 1 = $\frac{1}{k_1 \cdot k_0} = \frac{n}{1}$

" α = 0 = $\frac{1}{k_0 \cdot k_1} = \frac{n}{1}$

$\sum p_i^\alpha = \frac{1}{k_\alpha} = \sum_{i=1}^{n} 1 = n$; could = ∞

.35

If α < 1, $k_\alpha > 1$

" α > 1 $k_\alpha < 1$

" α = 1 $k_\alpha = 1$

" α = 0 ($k_\alpha$ = 0 – could = $\frac{1}{n}$ or 0 if α = 0)

If 0 < α < 1, then $k_\alpha$ & $k_{1-\alpha}$ are both < 1

so $\frac{1}{k_\alpha k_{1-\alpha}} > 1$ ∴ T > 1.

If α > 1 $k_\alpha < 1$, $k_{1-\alpha}$ is ? 1 − α < 0.

So maybe $k_\alpha < 0$ if ~~[illegible]~~ k < 0 or k > 1.

3.14.97 TM

Say $P_i = 2^{-i}$ $\quad i = 1\,60\infty$. $\quad P_i^{\alpha} = 2^{-i\alpha} \cdot k_\alpha \quad i = 1/\infty$

.02 $\sum_{i=1}^{\infty} (2^{-\alpha})^i = 2^{-\alpha} \sum_{i=0}^{\infty} (2^{-\alpha})^i = 2^{-\alpha} \cdot \frac{1}{1-2^{-\alpha}} = \frac{1}{2^{\alpha}-1} = \infty$ if $\alpha = 0$!

so $k_0 = 0$

Clearly a play problem; but I does seem to indicate that I was ~~wrong~~ very wrong about this matter!

No! see .26

from 42.35: $\frac{1}{k_\alpha} = \sum P_i^{\alpha}$ : If $\alpha = 0$ $\sum P_i^{\alpha} = \sum_{i=1}^{n} 1 = n = \infty$ if $n \to \infty$

$\frac{1}{k_0} = \infty$

$= \sum P_i^{\alpha} = 1$ if $\alpha = 1$.

If $\alpha < 0$? $\quad \sum (2^{-\alpha})^i$ diverges.

so I guess $\frac{1}{k_\alpha \cdot k_{1-\alpha}} > 0$ unless

I think I still have this wrong: $T = \frac{1}{k_\alpha \cdot k_{1-\alpha}}$ if

. $T = n$ at $\alpha = 0$ & $\alpha = 1$ ! ~~It's likely to have a min~~

for $\alpha$ outside $(0, 1)$, it diverges: It's likely to have a min for $0 < \alpha < 1$ !

in .02 $\frac{1}{k_\alpha} = \frac{1}{2^{\alpha}-1}$ : $\frac{1}{k_\alpha k_{1-\alpha}} = \frac{1}{2^{\alpha}-1} \cdot \frac{1}{2^{1-\alpha}-1} = \frac{1}{2^{\alpha}-1} \cdot \frac{1}{2 \cdot 2^{-\alpha}-1}$

$= \frac{2^{\alpha}}{2-2^{\alpha}} \cdot \frac{1}{2^{\alpha}-1}$ ← empirically, this seems to have a min. at $\alpha = .5$!

.26 Well, it would have to be at .5 since the funct. is symm about $\alpha = .5$!

so all is needed is to show that the funct at .5 is < those near by.

It's rather hz near $\alpha \approx 0$ & $\alpha \approx 1$.

In .29, $\frac{1}{k^2}$ was $> 1$, but that was not the problem!

T. p. 13 was it $< \underline{n}$!

So, did I prove this for all forms of $P_i$?

$\sum P_i^{\alpha} \cdot \sum P_j^{1-\alpha} \equiv$ max $\quad$ perhaps $\alpha$ has to be betw. 0 & 1.

---

Actually, the ~~divergence~~ danger of divergence of $\sum P_i^{\frac{1}{2}}$ is not so disturbing.

If we use $\alpha = 1$ or $0$ (as is normally done) we are concerned w. the divergence of $\sum_{i=1}^{n} \frac{P_i}{P_i} = \sum_{i=1}^{n} 1 \to \infty$, which is much worse!

3.14.97 TM

from 42.34 ; $T = \frac{1}{k_\alpha \cdot k_{1-\alpha}}$ which is symmetric about $\alpha = .5$, so if there is a min, it must be at $\alpha = \frac{1}{2}$ : This will be true, indep of how $p_i$ is a function of i.

So if we confine $f(p_i)$ to powers of $p_i$, $p_i^{\frac{1}{2}}$ is the rite soln;

But are other functional forms possl.?

Look at .42.34: ~~say~~ we have $\sum p_i = 1$;

.10 $\sum f(p_i) = 1$ $\sum \frac{p_i}{f(p_i)} = s$? Say we had finite n: use Lagrange's mults.

so, ~~take~~ looser: lets use an arby $f_i$ (not nec'ly a funct of $p_i$) Rollo1

$\sum f_i = 1$ s: $\sum \frac{p_i}{f_i}$ = min.

$$\frac{\partial s}{\partial f_i} + \lambda \frac{\partial \sum f_i}{\partial f_i} = -\frac{p_i}{f_i^2} + \lambda = 0 \qquad \frac{-p_i}{f_i^2} + \lambda = 0 \qquad \text{so } \lambda f_i^2 = -p_i \therefore f_i \propto \sqrt{p_i} \;!$$

So this proves it for all forms of $p_i$ as a funct of i! So the Theorem is stronger than the one I was trying to prove.

Great! [It's quite General]

So; (1) T. thm is true iff $\sum_{i=1}^{n} p_i^{\frac{1}{2}}$ converges : or if n is finite

(2) T. convergence of $\sum_{i=1}^{n} p_i^{\frac{1}{2}}$ is a weaker cond than $\sum_{i=1}^{n} 1 = n$, which is needed for the "usual" method to converge — e.g. "choosing i w. proby $p_i$"

Paradox: While .10 ff seems to prove optimality for $\sqrt{p_i}$, clearly,

.25 Choosing the i in $p_i$ order is better, or at least as good. ~~T. reason its~~

Reason of "Paradox": .10 assumes problistic soln. .25 is a non problistic soln.

Also: no replacement.

Just 108.tex
is arc. in 016.tex

3.15.98 TM: Admin: PATH:

Much work on this on 11.01ff:

Again an impt. idea: When I was in College, I would often find interesting, potentially important problems that I wanted to work on, but I couldn't spare t. time at that pt. — So I wrote them down, but didn't work on them then.

Again, I have a situation: I often find interesting problems, but I more often work on them — ~~and~~ to t. detriment of t. Main Goal (TM).

At ~~t.~~ present, I am even more pressed for time than when I was at U of C, but I haven't really taken it w/ adequate seriousness!

Present tentative emphasis:

.12 1) Workout details of Lsrch for 11 input mixer: ~~I want to see if it~~ I think there may be some impt. theoretic. diffys that I have to look at.

.15 2) Workout the details of Lsrch for t. SM problem pool! In particular, how t. corpus length of cands ↑ w. time spent. Also in ll w. this, one looks at shorter corpi for less likely stocks or comns. of stocks. [stuff leading to 40.26 is reasonable!]

Next: t. Q. of how to incorporate observd. reggys into t. "compressn." of t. p.d." — i.e. How do we modify t. P.d. in view of t. observd reggys thus far?

3) Think about J. Schmidhu's ideas. He is being less el. than I am — Can I use his ideas at all? Also, it may be that J. is rather naive about probty, & is still using random search & calling it "Lsrch".

4) 32.17ff13 a kind of summary of probs in .12 & .15

SN) Re: Lsrch on 11 input mixer: putting and, or, not in t. function pool will make sense in t. first generation or 2, but after that, they will be given low pc's because t. "If" will be almost entirely used.

If ≡ 3 input mixer (3 = 1+2')

SN2) On t. problem of putting MANY solutions of previous problems in Memory & then getting low pc's for those solns, because there are so many: Humans normally associate ~~each~~ each prob. solving method w. a certain area of science — this ↑ pc a lot. We can extend t. relevance of a prob. soln. to other areas by using analogy & other coding constructs.

Also people learn to apply (P.S. methods) (theorems) rather narrowly: They devise methods to suggest which thms to try: t. "methods" can be based on Gross statistics of

3.15.97 TM

.01 Can be based on very carefully designd "discriptors" → (50.01 for how to take advantage of This.

---

.02: $\boxed{40.40 \text{ spac}}$ ! ($\frac{\mu}{\sigma} \ll 1$) so even tho $\frac{\mu}{\sigma} \cdot \frac{1}{\sqrt{T}}$ may be large, there's a v.g. chance that t. observed $\mu$ is really not > 0.

I think tho "$\frac{\mu}{\sigma}\Big|_T$" & "T" params are quite difrnt. Just what Gore do I want?

If T >> 1, then $\frac{\mu}{\sigma}\frac{1}{\sqrt{T}}$ is a good Gore. If T << 1, I'm not sure what t. Gore is. I guess somethng like Expected value of $\frac{\mu}{\sigma}\Big|_{yr}$. — But How to Obtain This?

.08 One way; assume an apripd for $\frac{\mu}{\sigma}\Big|_{yr}$.

---

Quick impressn: Unless $\frac{\mu}{\sigma}\Big|_T$ somewhat >1, one can't tell much. — so if $\frac{\mu}{\sigma}\Big|_{yr}$ is, say 1 or < 1

Then I think T must be >> 1 before one has much confidence in its being > 0.

Note $\frac{\mu}{\sigma}\Big|_{yr} \sim .25$ can still be very useful! $\frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2 = \frac{1}{32}$ : So you'd need ~ 16 of these & to get yearly yield of $e^{.5}$ but only 8 gives yield of $e^{.25} = \times 1.284/yr$.

---

T. above remarks about $\frac{\mu}{\sigma}$ for SM are also true for more general "Action algm" Evalns. — i.e. in Non-SM applicns

---

.20 Still, I need a Gore for $\frac{\mu}{\sigma}\Big|_T$ w. T << 1. — I mite be possl. to get it using t. apripd of $\frac{\mu}{\sigma}\Big|_{yr}$ (⟨.08⟩).

What is wanted! $\frac{\mu}{\sigma} \cdot \frac{1}{\sqrt{T}}$ for small T. or about as small a T as one would ever use! for SM, This would be t. smallest T ∋

.25 $\underbrace{\left[\frac{1}{2}\frac{\mu^2}{\sigma^2_T}\frac{1}{T}\right]}_{(.25)}$ (t. yield) is useful. I actually want t. value of

Tho value can be obtaind by assuming a apripd for $\mu$. — is it necy to have an apripd for $\sigma$ or $\sigma^2$ also? Probably: We can know $\mu$ w. much accuracy w.o. knowing $\sigma$. Usually, hvr, Its easier to get $\sigma^2$ w. usable accuracy than $\mu$ (when $\mu$ is centerd about $\phi$ — as is usually t. case.)

ado disk
roall.zip
sound.
spchtool.
D:\disk\sound\
spchtool.zip
$\frac{133+2}{3}$

So one mite assume $\sigma$ is known fairly well, & one wants to know t. distribution of $\mu$ in view of t. data & t. apripd of $\mu$.

Say t. apripd is normal, mean $\phi$, varc = $\sigma_\phi^2$. Suppose we actually know t. $\sigma^2$ of t. d.f. of which $\mu$ is an estimate of mean. So we are only interested in t. apripd of $\mu$.

Maybe not so good to assume $\sigma^2$ is known; This screws up $\mu$ discovery, Since $\mu$ is normally given by an apsi for a given value because that value & its varc!

So start w. d.f. for $\mu$ & $\sigma$. At first let them be indep. — let $\mu$ be Gaussn. & let $\sigma^2$ be distributed like $x e^{-\frac{x}{\lambda}}$ or $x^2 e^{-\frac{x}{\lambda}}$ or $x^{\text{large}} e^{-\frac{x^{\text{large}}}{\lambda}}$

Another tack! ($\frac{\mu}{\sigma}$) is some apripd. Its most likely that $\mu \approx \phi$ & $\sigma$ is fairly large.

In this case $\frac{\mu}{\sigma}$ is not ≈ normally distributed. (there is "some talk" about t. d.f. for $x^2$ having an integrable singularity at t. origin.)

3.17.97 TM

.01 Say we are doing prediction: The mean of t. $\sigma^2$ d.f. is about t. $\sigma^2$ of t. signal itself. So, with an assumed $\sigma_0^2$, we ≈ have t. apripd for t. ensemble. So we have this [redacted] apsipd for $\mu, \sigma$.

---

SM (SN) If t. SM ~~trades~~ trades are not Poisson in time, Then t. time betw. successive trades is correlated. This is an easy hypoth to test. We can do a scatter in 2 & 3 dims of $T_n$ as a funct of $T_{n-1}$ ; as a funct of $T_{n-1}$ & $T_{n-2}$.

In fact, we can study t. time differences betw. trades as a t.s. & try to find its dimension using that olde pgm, from LeBaron et al!

We can also do this by plotting no of trades per 5 or 10 min / or any other size interval. bar. (directly available from "tick data" s.w).

Or t. times of successive "n trade sets".

T. relationship of latest $\Delta t$ (say) can be tried as ~~—~~ linear to start, ~~—~~ then try simple non-linear (multi-dim Taylor series), then ~~—~~ neural nets, ~~then~~ ModelQuest, then etc.

T. goal here is to try to predict varc, & ∴ option values.

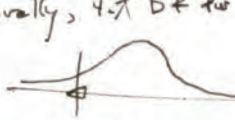
"Start of day" and "end of day" would have to be treated in special ways

---

- one way: do correlnt $T_n$ w. $T_{n-1}$ as a function of time since start of day. So one gets one value each day. Then one does some "pooling" — depending on how t. data looks.

---

Actually, t. apriori Df for $\sigma^2$ is not critical: probly one could use a gaussian! ☺



we may ~~—~~ cut off $\sigma^2<0$ & renormalize.

*.TRE subdir, file
*.DIR subdirs only
*.TXT subdirs, files & file dumps.

Use whatever is mathicly most tractable!

.26 So apsip: $\underbrace{P_\mu(\mu)\cdot P_{\sigma^2}(\sigma^2)}_{\text{Apripd}} \cdot \left[\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{\Sigma(\mu-s_i)^2}{2\sigma^2}\right)\right]$ also to $n^{th}$ power

This part proply has a peak at $\mu = \bar{s_i}$
$\sigma^2 =$ ms error of [redacted] $\mu = \bar{s_i}$,
& not $\frac{\Sigma \text{varc}}{n-1}$

$\Sigma(\mu-s_i)^2 = \sigma_{0B}^2 + (\mu-\mu_{ob})^2$ [illegible]

$\mu_{ob} \equiv$ [redacted] $\frac{1}{n}\Sigma s_i$

$\sigma_{0B}^2 \equiv \frac{1}{N}\Sigma s_i^2 - \mu_{ob}^2$.

.30 $$\frac{1}{\sqrt{2\pi}}\cdot\frac{1}{\sigma}\exp-\left[\frac{\sigma_{0B}^2+(\mu-\mu_{ob})^2}{2\sigma^2}\right]$$

if $\mu=\mu_{ob}$; t. peak wrt $\sigma$ is $\sigma^2 = \sigma_{0B}^2$

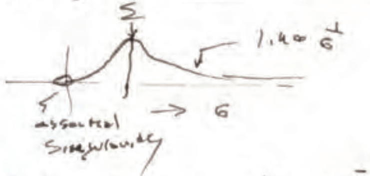
for any $\sigma$; t. peak wrt. $\mu$ is $\mu_{ob}$.

If $\mu \neq \mu_{ob}$; The peak for $\sigma^2$ at $\sigma^2 = \sigma_{0B}^2+(\mu-\mu_{ob})^2$

[illegible] say $\mu_{ob}=0$. The d.f. wrt $\mu$ is Gaussn ~~Gamma~~ w. varc $= \sigma^2$.

.36 ~~the d.f.~~ T. peak wrt $\sigma^2$ is at $\sigma_{0B}^2+\mu^2$ (if $\mu_{ob}=0$).

.37 look at d.f. $\frac{1}{\sigma}\exp-\frac{\Sigma^2}{2\sigma^2}$ error! should be $\frac{1}{\sigma^n}$ See 49.09



what is width? it is ~~always~~ ∝ Σ.

can be found by numerical analysis of t. curve $\frac{1}{x}e^{-\frac{1}{2x^2}}$

or $xe^{-\frac{x^2}{2}}$. we can find second moment about $x=1$ ↑ also find first & cent'l moments.

3 1797 TM

Not so easy! They all diverge at $x \to \infty$ (including $\phi^{th}$ moment) 48

first & 2nd moments can be found analytically.

is $\int_{-\infty}^{\infty} \frac{1}{x} e^{-\frac{1}{2x^2}} dx$ integrable?

say $y = \frac{1}{x}$ $dy = -\frac{dx}{x^2}$ = ~~[illegible]~~

$dx = -x^2 dy = -\frac{1}{y^2} dy$

so $S = -\int_{\infty}^{0} \frac{y}{y^2} e^{-\frac{y^2}{2}} dy \overset{?}{=} \int_0^{\infty} \frac{1}{y} e^{-\frac{y^2}{2}} dy$.

$\underbrace{\qquad}_{\text{diverges at t origin}}$

OH! The $\int$ ~~[illegible]~~ of 47.37 does diverge: $\frac{1}{x} e^{-\frac{1}{2x^2}} \to \frac{1}{x}$ at big ~~[illegible]~~ $x$.

Hm, maybe ~~[illegible]~~ t. d.f. for $\sigma^2$ doesn't diverge.

But one could just approximate this by using "½ power pts" — where prob. goes down by ½, say.

~~[illegible]~~ $S = \sigma^2$ $ds = 2\sigma d\sigma$ $d\sigma = \frac{ds}{2\sigma} = \frac{ds}{2\sqrt{s}}$.

$\frac{1}{\sigma} \exp - \frac{\varepsilon^2}{2\sigma^2} d\sigma$ : ~~[illegible]~~ $\frac{1}{\sigma} \exp - \frac{\varepsilon^2}{2\sigma^2} \frac{ds}{2\sigma} = \frac{1}{2} \frac{1}{\sigma^2} \exp - \frac{\varepsilon^2}{2\sigma^2} ds$

$= \frac{1}{2} \frac{1}{s} \cdot e^{-\frac{\varepsilon^2}{2s}} ds$. which ~~[illegible]~~ also diverges at $s \to \infty$!

Thurs:
Yusef
→ Grill Seafood
Beach st.
5 pm.

Well OK.: the $P_{\sigma^2}(\sigma^2)$ factor can take care of t. divergence at $\sigma$ or $\sigma^2$ at $\infty$.

We could approx. t. d.f. by its widths in $x$ & $\sigma^2$ directions at t. peak.: Then use a 2 dim. Multivariate d.f.

.19 So in $\mu$ direction t. var is $\sigma_{ob}^2$.

.20 " $\sigma^2$ " " is $\propto \sigma_{ob}^2$. some fixed fraction of $\sigma_{ob}^2$.

~~[illegible]~~

$$\approx \boxed{\propto \exp - \frac{1}{2}\left(\frac{(\mu-\mu_{ob})^2}{\sigma_{ob}^2} + \frac{(\sigma^2-\sigma_{ob}^2)^2}{\alpha(\sigma_{ob}^2)^2}\right)}$$

t. d.f. of .19 – .20 is modified by t. apriori $P_\mu(\mu) \cdot P_{\sigma^2}(\sigma^2)$ of 47.26.

$\alpha$ is some empirical constant < 1 which we will chose/determine.

.25 say $P_\mu(\mu) \cdot P_{\sigma^2}(\sigma^2) \overset{\propto}{\approx} \frac{1}{\sigma_{\mu ap}} \exp - \frac{\mu^2}{2\sigma_{\mu ap}^2} \cdot \frac{1}{\sigma^2_{\sigma^2 ap}} \cdot \exp - \frac{1}{2} \frac{\sigma^2 - \sigma^2_{mean\,apri}}{(\sigma^2_{\sigma^2 ap})^2}$

This is $\propto$ by some constant $\alpha$

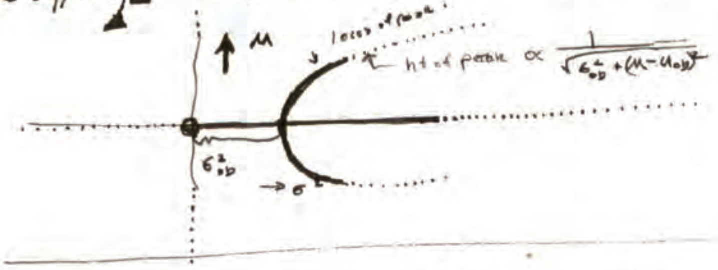
So t. problem is to get a ~~[illegible]~~ $\sigma^2_{\mu\,apriori}$ & a $\sigma^2_{\sigma^2 mean\,apri}$

$\sigma^2_{\sigma^2 mean\,apri}$ can be t. observed $\sigma^2$ of t. signal wrt $\phi$.

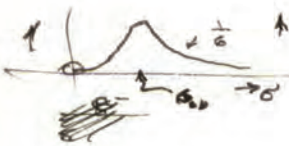
$\boxed{\sigma^2_{\mu\,apri}}$ is not so easy to guess: try various values & see how things look! I may [redacted] try a large value, then reduce it in view of empirical result.

T. output params are t. mean, $\mu$ ① | & its width ② | The $\sigma$ of t. ensemble [redacted] ③

.32 : 47.36 Gain, Back to 47.36. say ~~[illegible]~~ $\mu_{ob} = 0$



How this ↗ surface looks: along t. line $\mu = 0$



$\propto \frac{1}{\sigma}$ ↑ $\frac{1}{\sigma} \exp\left(-\frac{\sigma_{ob}^2}{2\sigma^2}\right)$

Hm. on 47.37: t. d.f. $\frac{1}{\sigma} \exp - \frac{\varepsilon^2}{2\sigma^2}$ diverges at $\sigma^2 = \infty$ ~~[illegible]~~ —

if ~~[illegible]~~ $y = \sigma^2$; $\frac{1}{\sqrt{y}} \exp \frac{y_0}{2y}$

this diverges even worse at $y \to \infty$:

At $\sigma$ or $y = 0$ there's no divergence because of essential singularity at origin

This sounds very strange! W.o. strong Apriori, t. a psip for $\sigma$ or $\sigma^2$ is meaningless!

Along line $\sigma = \sigma$ cont's gaussian width $\sigma$

I don't know what = value contours are.

3.17.97 TM

Since all $\geq 0$ moments of $\sigma$ ($\approx \sigma^2$) are infinite, t. mean's value of $\sigma^2$ is to a very impt. extent, determined by t. <u>apriori</u> !

---

Could it be poss/. to find $\frac{M_1}{M_0}$ & $\boxed{\frac{M_2}{M_0^2}}$ & $\boxed{\frac{M_2}{M_0^2}} - \left(\frac{M_1}{M_0}\right)^2$ . ? | $M_i = i^{th}$ moment.

Prbly should be $\boxed{M_0}$

$M_0 \to \infty$ as $\ln x$ ; $M_1 \to \infty$ as $x$ : $M_2 \to \infty$ as $x^2$

$\frac{M_2}{M_1^2} \to$ constant as $x \to \infty$ : The quantity $\frac{\mu}{\sigma}$ <u>may</u> converge .

$\mu = \frac{M_1}{M_0}$ : $\sigma^2 = \frac{M_2}{M_0^2} - \frac{M_1^2}{M_0^2}$ $\frac{\mu^2}{\sigma^2} = \frac{M_1^2}{M_2 - M_1^2}$ which does converge .

$\frac{1}{\sigma} e^{-\frac{\sigma_0^2}{2\sigma^2}}$ for large $\sigma \simeq \frac{1}{\sigma}$ <u>indep of $\sigma_0^2$</u> so this would be

$M_1 \to \sigma$ : $M_1^2 \to \sigma^2$ ; $M_2 \to \frac{\sigma^2}{2}$ so this $= -\frac{1}{2}$ : <u>not at all reasonable</u> !

---

.09 Important <u>Grossanor</u>! The ap. should be $\propto \frac{1}{\sigma^n} e^{-\frac{\Sigma^2}{2\sigma^2}}$ (for n data pts).

for <u>$n \geq 2$, $M_0$ converges</u>

for <u>$n \geq 2+i$ , $M_i$ converges</u> . : so for $n > 4$ $M_2$ converges .

$\int_0^\infty \sigma^{-n} e^{-\frac{1}{2\sigma^2}} d\sigma$ is prbly a ~~known~~ well known $\int$ . $= \int \sigma^{-n} e^{-\frac{1}{2\sigma^2}} \frac{2\sigma d\sigma}{2\sigma}$

$\sigma^{-2} = x$ $\sigma^2 = \frac{1}{x}$ $2\sigma d\sigma = -\frac{dx}{x^2}$ $= -\int_\infty^0 x^{\frac{n}{2}} e^{-\frac{x}{2}} \frac{dx}{2\sigma}$

$\sigma = \frac{1}{\sqrt{x}}$ $\int_0^\infty \sigma^{-\frac{n}{2}} \cdot \frac{dx}{2\sigma} e^{-\frac{1}{x}} d$ $\left[ \frac{1}{2\sigma} = \frac{1}{2}\sqrt{x} \right.$ | $\int_0^\infty \frac{1}{2} x^{\frac{n+1}{2}} e^{-\frac{x}{2}} dx$ which is "like" $\boxed{n!}$

$$\boxed{\int_0^\infty x^n e^{-x} dx = n!}$$ p255 B.S.

$\to \int_0^\infty x^m e^{-x^{-2}} = \underbrace{-\frac{1}{2}\Gamma\left(\frac{m+1}{-2}\right)}_{\substack{\text{This may be rite anyway!}\\ \text{perhaps check it.}}}$ « no! woops! for $m > -1$ ; $n > 0$ B.t $-2 < 0$

$m = m$, $n = -2$

Hvr. for $n = -2$, I $\int$ <u>may</u> work: I! $n\Sigma^2$ ← should be

$\beta = \frac{\Sigma}{\sqrt{2}}$

$\int_0^\infty \frac{1}{\sigma^n} e^{-\frac{\boxed{\Sigma^2}}{2\sigma^2}} d\sigma = \int \frac{d\sigma}{\sigma^n} e^{-\left(\frac{\beta}{\sigma}\right)^2}$ | $\boxed{\frac{\Sigma^2}{2} = \beta^2}$ $\Sigma = \beta \cdot \sqrt{2}$ $\boxed{\frac{\sigma}{\beta} = x}$ $d\sigma = dx \cdot \beta$ $\sigma = x\beta$

.29

$= \int_0^\infty \frac{dx\,\beta}{x^n \beta^n} e^{-x^2} = \beta^{n+1} \int_0^\infty \frac{dx}{x^n} e^{-x^{2}}$ ~~$x^2 = y$~~ ~~$2x\,dx = dy$~~ $\boxed{\frac{1}{x^2} = y}$

~~$x = y^{\frac{1}{2}}$~~ $-\frac{2dx}{x^3} = dy$

$dx = -\frac{1}{2} x^3 dy$

$= \beta^{-n+1} \int_0^\infty y^{\frac{n}{2}} e^{-y} \cdot \frac{1}{2} x^3 dy$

$x = y^{-\frac{1}{2}}$ | $dx = -\frac{1}{2} y^{-\frac{3}{2}} dy$

$x^3 = y^{-\frac{3}{2}}$

.32 $\frac{\beta^{-n+1}}{2} \int_0^\infty y^{\frac{n-3}{2}} e^{-y} dy = \frac{\beta^{-n+1}}{2} \left(\frac{n-3}{2}\right)!$

o.k. ~~doesn't seem rite~~ $1! = 1$ for $n \geq 2$ it converges , for $n = 1, 0, -1 \ldots$ it diverges

for $n = 1$, $\frac{1-2}{2} = -1$ $(-1)!$ diverges as it should | $n = 2$ $\frac{2-3}{2} = -\frac{1}{2}$ $\left(-\frac{1}{2}\right)!$ $\frac{1}{2}!$ converges as it should

$n = 3$ $\frac{0}{2} = 0$ $0!$ converges

$n = 4$ $\frac{4-3}{2} = \frac{1}{2}$ $\left(\frac{1}{2}\right)!$ converges .

.37 $= \frac{\Sigma^{-n+1}}{2^{\frac{-n+1}{2}} \cdot 2} \left(\frac{n-3}{2}\right)! = \boxed{\frac{\Sigma^{-n+1}}{2^{\frac{-n+3}{2}}} \left(\frac{n-3}{2}\right)!}$

$\left(-\frac{1}{2}\right)! = \sqrt{\pi}$

$\Gamma(n+1) = n!$

$\Gamma(n+\frac{1}{2}) = \frac{1 \cdot 3 \cdots (2n-1)}{2^n} \cdot \sqrt{\pi}$

38 $M_0 = .32$

$M_1 = .32 \cdot$ for $n-1$

$= \boxed{\frac{2^{\frac{n}{2}} \Sigma^{1-n}}{2^{\frac{3}{2}}} \left(\frac{n-3}{2}\right)!}$

$+\frac{1}{2}! = \frac{1}{2}\sqrt{\pi}$

$\Gamma(n-\frac{1}{2}) = (n-\frac{1}{2})!$

~~$M/M_0 = \frac{\sqrt{2}}{2}\Sigma \left(\frac{n-4}{2}\right)! / \left(\frac{n-3}{2}\right)! = \frac{\sqrt{2}}{2} \left(\frac{n-3}{2}\right) = \frac{\Sigma\,2\sqrt{2}}{(n-3)}$~~

$+\frac{3}{2}! = \sqrt{\pi} \cdot \frac{3}{4}$

→ 50.09 spec

3.18.97 TM Part

7:~~46.01~~ (SN) Part: Perhaps just have TM work induction problems in various fields: Say it starts in one area, learns to solve a lot of probs: Then we give it probs in a new field that it learns to solve. —— And so on w. many fields:

At first the concs for t. various fields are kept separate, but as time goes on, correspondences, analogies, are found betw. t. various fields — so t. p.d.'s for t. various fields get "Mixd".

Hvr. It would be well to chose Problem fields that were related closely, so as to encourage "Mixing" & actually speed up lrng.

---

.09:49.40: So Now, I want to get exact equs. for Gorc for search, w. ideas on just how to run t. search: ~~to[illegible]~~ how fast do we ↑ corpus length v.s. try new corpi?

Going back to 47.26: T. expressn should be; $\left(\frac{1}{\sqrt{2\pi}}\cdot\frac{1}{\sigma}\right)^n e^{-\frac{1}{2}\left(\frac{\Sigma(\mu-s_i)^2}{\sigma^2}\right)}$

$$= \left(\frac{1}{\sqrt{2\pi}}\cdot\frac{1}{\sigma}\, e^{-\frac{1}{2}\frac{s_0^2+(\mu-\mu_{ob})^2}{\sigma^2}}\right)^n \leftarrow$$ width in μ direction is $\frac{\sigma}{\sqrt{n}}$. should be $\left(\frac{n-4}{2}\right)!$ 2 k lines ~ 40 pp.

Mixing

$$\frac{M_1}{M_0} = \frac{2^{\frac{n}{2}}}{2^{\frac{n-1}{2}}}\;\frac{\boxed{\left(\frac{n-2}{2}\right)!}\;\Sigma^{-n+1}}{\left(\frac{n-3}{2}\right)!\;\Sigma^{-n}} = \sqrt{2}\,\Sigma\,\frac{\left(\frac{n-2}{2}\right)!}{\left(\frac{n-3}{2}\right)!}$$ $\sqrt{\ }\ n = \cdot\,\Sigma^2$

$$\frac{M_2}{M_0} = \frac{2\,\Sigma^2}{\cancel{2}}\,\frac{\left(\frac{n-5}{2}\right)!}{\left(\frac{n-3}{2}\right)!} = 2\,\Sigma^2\cdot\left(\frac{n-5}{2}\right) = (n-5)\,\Sigma^2$$

Go back to 49.32: $\frac{\beta^{1-n}}{2}\left(\frac{n-3}{2}\right)! = \underline{M_0}$

~~to~~ which is $M_0 \nearrow$ ~~$M_2 = \frac{\beta^{1-n-2}}{2}\left(\frac{n+2-3}{2}\right)!$~~ $n \to n-2$

~~$= \frac{\beta^{-1-n}}{2}\left(\frac{n-1}{2}\right)!$~~

~~$\frac{M_2}{M_0} = \frac{\beta^{1-n}}{\beta^{1-n}}$~~ $M_2 = \frac{\beta^{1-n+2}}{2}\left(\frac{n-2-3}{2}\right) = \frac{\beta^{3-n}}{2}\left(\frac{n-5}{2}\right)$

.26 $\frac{M_2}{M_0} = \beta^2\cancel{\left(\frac{n-5}{2}\right)}\frac{2}{n-5} = \frac{\Sigma^2}{n-5}$ $n \to n-1$

$M_1 = \frac{\beta^{1-n+1}}{2}\left(\frac{n-1-3}{2}\right)! = \frac{\beta^{2-n}}{\cancel{2}}\left(\frac{n-4}{2}\right)!$

.28 $\frac{M_1}{M_0} = \beta\cdot\frac{\left(\frac{n-4}{2}\right)!}{\left(\frac{n-3}{2}\right)!} \approx \frac{\Sigma}{\sqrt{n-4}}$ ← 51.25

$$\Gamma(n+\tfrac{1}{2}) = (n-\tfrac{1}{2})! = \frac{\overset{2}{1}\cdot\overset{4}{3}\cdot 7\cdots\overset{2n}{(2n-1)}}{2^n}\sqrt{\pi} = \frac{(2n-1)!}{(2^n\,n!)\cdot 2^n}\cdot\sqrt{\pi} = \frac{(2n-1)!}{n!\,2^{2n}}\sqrt{\pi}$$

$= \left(\frac{2n-1}{2}\right)!$

$\left(\frac{2n-1}{2}!\right)/\left(\frac{2(n-1)-1}{2}!\right)$

$\frac{\frac{n-4}{2}!}{\frac{n-3}{2}!}$ case ① : n is even! $n = 2m$ $\frac{2m-4}{2}! = (m-2)!$

$\frac{2m-3}{2}! = (m-1-\frac{1}{2})!$

$$(m-1-\tfrac{1}{2})! = \sqrt{\pi}\,\frac{(2(m-1)-1)!}{2^{m-1}(m-1)!\,2^{m-1}} = \frac{(2m-3)!}{(m-1)!\,2^{2m-2}}\cdot\sqrt{\pi}$$ damn

$$\frac{(m-2)!\,(m-1)!\,2^{2m-2}}{(2m-3)!\,\sqrt{\pi}} = \frac{1}{\sqrt{\pi}}\;\frac{2m-2}{2m-3}\;\frac{2m-4}{2m-5}$$ ≈ $e^{\frac{1}{2}}$

Better use Stirling: $x! \approx \sqrt{2\pi x}\cdot\frac{x^x}{e^x}$

$$\frac{\frac{n-4}{2}!}{\frac{n-3}{2}!} = \sqrt{\frac{n-4}{n-3}}\;\frac{\left(\frac{n-4}{2}\right)^{\frac{n-4}{2}}}{\left(\frac{n-3}{2}\right)^{\frac{n-3}{2}}}\;\boxed{\frac{e^{\frac{n-3}{2}}}{e^{\frac{n-4}{2}}}}$$

3 1997 TM

$$\frac{x+\frac{1}{2}!}{x!} : \sqrt{\frac{x+\frac{1}{2}}{x}} \cdot \frac{(x+\frac{1}{2})^{x+\frac{1}{2}}}{x^x} \cdot \boxed{\frac{e^{x}}{e^{x+\frac{1}{2}}}} - e^{\frac{1}{2}}$$

$$\frac{(x+\frac{1}{2})^{x+\frac{1}{2}}}{x^x} = \left(\frac{x+\frac{1}{2}}{x}\right)^x \cdot (x+\tfrac{1}{2})^{\frac{1}{2}} = \left(1+\tfrac{1}{2x}\right)^x \cdot \sqrt{x}\left(1+\tfrac{1}{2x}\right)^{\frac{1}{2}}$$

for large x ! $\sim e^{-\frac{1}{2}} \cdot \sqrt{x}$

$\sqrt{1+\frac{1}{2x}}$

$\cdot \left(1+\frac{1}{2x}\right) \cdot \sqrt{ex}$

$\sqrt{4.5}$

$\frac{5}{\sqrt{x+1}} =$

$\sqrt{x(1+\frac{1}{x})} =$

.15 $$\frac{x+\frac{1}{2}!}{x!} = \frac{(x+\frac{1}{2})^{\frac{1}{2}}}{x^{\frac{1}{2}}} e^{-\frac{1}{2}} \cdot \frac{(x+\frac{1}{2})^x}{x^x} (x+\tfrac{1}{2})^{\frac{1}{2}} e^{\frac{1}{2}} = \frac{(x+\frac{1}{2})}{x^{\frac{1}{2}}} \left(\frac{x+\frac{1}{2}}{x}\right)^x$$

$= \left(1+\frac{1}{2x}\right)^x \approx e^{\frac{1}{2}}$

$$= \frac{(x+\frac{1}{2})^{\frac{1}{2}}}{x^{\frac{1}{2}}} e^{-\frac{1}{2}} \cdot \frac{(x+\frac{1}{2})^x}{x^x} \cdot (x+\tfrac{1}{2})^{\frac{1}{2}} = e^{\frac{1}{2}} \frac{x+\frac{1}{2}}{x^{\frac{1}{2}}} \cdot \left(\frac{x+\frac{1}{2}}{x}\right)^x$$

$\approx e^{-\frac{1}{2}}$

$\sqrt{x+1}$ is closer

$$\boxed{\frac{(x+\frac{1}{2})!}{x!} \approx \frac{x+\frac{1}{2}}{x^{\frac{1}{2}}} = \sqrt{x} + \frac{1}{2\sqrt{x}}}$$

P 272 B.S. for tables of $\Gamma(x)$, $\Gamma(x+.5)$.

$\equiv \sqrt{x}\left(1+\frac{1}{2x}\right)$

$\approx \sqrt{x(x+1)}$

2·3·4

$\sqrt{x+1}$ is better approx. than

~~$\Gamma(30)$~~ $\Gamma(26.5)$ $\Gamma(5) = 24$ $\Gamma(5.5) = 52.3.427$

$4! = 24$ $4.5! = 52.3427$

ratio = 2.1809 ; $\sqrt{4} + \frac{1}{2\sqrt{4}} = 2.25$ ! not too bad!

.07

9! $\Gamma(10) = 3.6288\text{E}5$ ) ratio 3.1238 $\sqrt{9} + \frac{1}{2\sqrt{9}} = 3.166$ better

9.5! $\Gamma(10.5) = 11.3327\text{E}5$

.04

$= \sqrt{\frac{n-2}{2}}$

50.25 : $$\frac{M_1}{M_0} = \beta \frac{(\frac{n-4}{2})!}{(\frac{n-4}{2}+\frac{1}{2})!} \approx \beta \Big/ \left(\left(\tfrac{n-4}{2}\right)^{\frac{1}{2}} + \frac{1}{2(\frac{n-4}{2})^{\frac{1}{2}}}\right)$$

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

$\sqrt{x+1}$

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

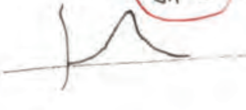
$\approx \sqrt{x}\left(1+\frac{1}{2x}\right)$

$\sqrt{x}\,(1+$

.25. $\frac{M_1}{M_0}$ : for large n its $\sim \beta \cdot \sqrt{\frac{2}{n-4}} = \frac{S}{\sqrt{n-4}}$ since $\beta \cdot \sqrt{2} = S = \frac{M_1}{M_0}$

I think $\frac{S}{\sqrt{n-2}}$ is better $\rightarrow$ 52.02

well : $\int_0^\infty \frac{1}{\sigma^n} e^{-\frac{S^2}{2\sigma^2}} d\sigma$ :



as $n \uparrow$ t. S gets smaller and smaller $\Rightarrow$

because t. mag. rend for hy $\sigma$ is smaller.

$$\frac{M_2}{M_0} - \left(\frac{M_1}{M_0}\right)^2 = \frac{S^2}{n-5} - \left(\frac{S}{\sqrt{n-4-2}}\right)^2 = S^2\left(\frac{1}{n-5} - \frac{1}{n-4-2}\right)$$

.32 $$\rightarrow S^2\left(\frac{n-4 \quad -n+5}{(n-4)(n-5)}\right) = \frac{S^2}{(n-4)(n-5)} = \boxed{\frac{3\ S^2}{(n-4-2)(n-5)}} \rightarrow 52.04$$

which is t. value at ~~to~~ 8 $\sigma$ in the $\sigma$ direction.

.32 doesn't seem right! ~~The first moment,~~ $\frac{M_1}{M_0}$ should $\rightarrow S$ as $n \rightarrow \infty$.

also, I left out a $\left(\frac{1}{\sqrt{2\pi}}\right)^n$ factor

also, use $\frac{x+\frac{1}{2}!}{x!} \approx \sqrt{x+1}$ is better approx (slightly better). $\sqrt{x+1-\epsilon}$ actually

Well ! One/serious error : in 49.29 " $S^2$ " should be " n $S^2$ " , also

3.19.97 TM

That last seems to fix it up <u>just fine</u>! replace $\Sigma$ w. $\sqrt{n}\,\Sigma$!

02:51.25 $\frac{M_1}{M_0} = \sqrt{\frac{n}{n-2}} \cdot \Sigma$ <u>good</u>! it <u>should</u> approach $\Sigma$ as $n \to \infty$.

04:51.26 $\frac{M_2}{M_0} =$ [illegible] $\cdot \frac{n\,\Sigma^2}{n-3}$ ; well, ~~this seems wrong. It should → $\Sigma^2$ as $n\to\infty$~~.

I think this is wrong: it leads to estimate $\Sigma^2 \to 3\,\Sigma^2$ as $n \to \infty$.

---

The expression for $\frac{M_2}{M_0}$ is probably wrong

~~[illegible]~~

49.32 should be $\left(\frac{\sqrt{2\pi}}{[illegible]}^{n}\right) \beta^{-n} \cdot \frac{\beta}{2} \left(\frac{n-3}{2}\right)!$ we get values for $M_0$ $M_1$ $M_2$: $n$, $n-1$ & $n-2$

Then in $M_1/M_0$ & $M_2/M_0$ we substitute $\beta^2 = \frac{n\,\Sigma^2}{2}$

<u>Not quite</u>! The $\sqrt{2\pi}$ factor doesn't change for the 3 cases — so it cancels, as does the $\frac{\beta}{2}$

~~$$\text{so } \frac{M_1}{M_0} = \beta \cdot \left(\frac{n-1-3}{2}\right)! \Big/ \left(\frac{n-3}{2}\right)! = \frac{\beta}{[illegible]} \times \frac{n-3}{2} = \frac{2\beta}{n-3}$$~~

$$\frac{M_1}{M_0} = \beta \left(\frac{n-4}{2}\right)! \Big/ \left(\frac{n-3}{2}\right)! \approx \frac{\beta}{\sqrt{\frac{n-3}{2}}} = \frac{\sqrt{2}\cdot\beta}{\sqrt{n-3}} = \frac{\sqrt{n}\,\Sigma}{\sqrt{n-3}} = \boxed{\sqrt{\frac{n}{n-3}} \cdot \Sigma}$$

.19

.

$$\frac{M_2}{M_0} = \beta^2 \left(\frac{n-5}{2}\right)! \Big/ \left(\frac{n-3}{2}\right)! = \frac{\beta^2}{\frac{n-3}{2}} = \frac{2\beta^2}{n-3} = \frac{n\,\Sigma^2}{n-3}$$

.20

<u>Very bad</u>: we get $\frac{M_2}{M_0} - \left(\frac{M}{M_0}\right)^2 = \phi$. 3.0822

$$\frac{x!}{(x-1)!} - \left(\frac{x!}{(x-\frac{1}{2})!}\right)^2 = ?$$

so how much does $\left(\frac{x!}{(x-\frac{1}{2})!}\right)^2$ differ from $x$?

from 51.5 ff: $\frac{x+\frac{1}{2}!}{x!} = \frac{(x+\frac{1}{2})^x \cdot (x+\frac{1}{2})^{\frac{1}{2}}}{x^x} \cdot \frac{e^x}{e^{x+\frac{1}{2}}} \cdot \frac{\sqrt{x+\frac{1}{2}}}{\sqrt{x}}$ | $\frac{(x+\frac{1}{2})^{x+\frac{1}{2}}}{x^{x+\frac{1}{2}}} \cdot e^{-\frac{1}{2}}$ $= \left(1+\frac{1}{2x}\right)^{x+\frac{1}{2}} \cdot e^{-\frac{1}{2}}$

$$= \left(1+\frac{1}{2x}\right)^x \cdot e^{-\frac{1}{2}} \left(\sqrt{x}\sqrt{1+\frac{1}{2x}}\right) \cdot \sqrt{1+\frac{1}{2x}}$$

$$= \left[\left(1+\frac{1}{2x}\right)^x \cdot e^{-\frac{1}{2}}\right] \cdot \sqrt{x} \cdot \sqrt{1+\frac{1}{2x}}$$ [illegible]

.30

& correction to make it almost exact: $e^{\frac{1}{12x+\frac{1}{2}} - \frac{1}{12(x+[illegible])}}$

$\frac{1}{12}\left(\frac{1}{x+\frac{1}{2}} - \frac{1}{x}\right) \approx \frac{1}{12}\left(\frac{\frac{1}{2}}{(x+\frac{1}{2})x}\right) = -\frac{1}{24x(x+\frac{1}{2})}$

$\approx x\left(1 - \frac{1}{24x(x+\frac{1}{2})}\right)$

so $\sqrt{x+\frac{1}{2}}$ but corrections ↑ & : → <u>Empirically</u>, $\frac{x+\frac{1}{2}!}{x!} \approx \sqrt{x+\frac{3}{4}}$

$$\ln (1+\epsilon) = \epsilon - \frac{\epsilon^2}{2} + \frac{\epsilon^3}{3} - \cdots$$

$$\ln + \tfrac{1}{2x} = \frac{1}{2x} - \frac{1}{8x^2} + \frac{1}{24x^3} \cdots$$

≈ cancels.

.36 $$\left(1+\tfrac{1}{2x}\right)^x = \exp\left(\tfrac{1}{2} - \frac{1}{8x} + \frac{1}{24x^2}\right)$$

So correction to $\sqrt{x+\frac{1}{2}}$ is $\approx \left(1 - \frac{1}{8x}\right)$ !! which is in <u>wrong direction</u>!

$\cdot\ x+\frac{3}{4} \big/ x+\frac{1}{2} = \frac{1+\frac{3}{4x}}{1+\frac{1}{2x}} \approx 1+\frac{1}{4x}$ ; $\sqrt{\ } = 1+\frac{1}{8x}$ so its <u>nice</u> but <u>wrong sign</u>!

3.19.57 TM

It would be possl. to check t. Algebra step by step, for $n = 4$, say, since t. approxns. are very ~~[illegible]~~ close!

$$= \sqrt{x} + \frac{1}{2\sqrt{x}} = \sqrt{x}\left(1 + \frac{1}{2x}\right)$$

$$\frac{(x+\frac{1}{2})^{x+\frac{1}{2}}}{x^{x+\frac{1}{2}}} \cdot (x+\tfrac{1}{2})^{\frac{1}{2}} \; e^{-\frac{1}{2}} = \left(\frac{(x+\frac{1}{2})}{x}\right)^{x} \cdot \frac{x+\frac{1}{2}}{\sqrt{x}}$$

$$\sqrt{x} \cdot \sqrt{1+\frac{1}{2x}}$$

$$\sqrt{x} \cdot \left(1+\frac{1}{2x}\right)$$

$$\simeq \sqrt{x}\sqrt{1+\frac{1}{x}}$$

$$\simeq \sqrt{x+1}$$

$$\approx \sqrt{x}\left(1+\frac{1}{2x}\right)$$

---

O.K. so that's it! T. error was on 52.30!

$\sqrt{x} \cdot \sqrt{1+\frac{1}{2x}}$ instead of $\sqrt{x} \cdot \left(1+\frac{1}{2x}\right)$: t. first order approxn:

$1+\frac{1}{2x} \approx \sqrt{1+\frac{1}{x}}$ : so this is $\simeq \sqrt{x+1}$.

$$x\left(1+\frac{1}{8x}\right) =$$

$$\sqrt{x} \cdot \left(1+\frac{3}{8}x\right)$$

$$\sqrt{x} \cdot \sqrt{1+\frac{3}{4x}}$$

The higher order correction of .36 is $\times \exp\left(-\frac{1}{8x}\right) \simeq 1 - \frac{1}{8x}$

or $\times \sqrt{1-\frac{1}{4x}}$ $\quad \sqrt{x+1} \cdot \sqrt{1-\frac{1}{4x}} = \sqrt{x+1-\frac{1}{4}-\frac{1}{4x}}$

$= \sqrt{x+\frac{3}{4}-\frac{1}{4x}}$ which is $\simeq \sqrt{x+\frac{3}{4}}$ to first order:

I seem to have lost much precision somewhere, but empirically, its v.g.

.18 $\boxed{\frac{x+\frac{1}{2}\,!}{x\,!} \approx \sqrt{x+\frac{3}{4}}}$ → see $\boxed{79.15}$ for more exact formula.

is within .2% for $x = 2$ and rapidly gets better for large $x$.

.5% error for $x=1$
2% error for $x=0$
1% " " $x=.5$
error is $\propto \frac{1}{n^2}$.

$\frac{n}{2}-2$ ; $\frac{n}{2}-1\frac{1}{2}$ : $\frac{n}{2}-1\frac{1}{4}$

.20 $\quad$ 52.19 $\boxed{\frac{M_1}{M_0}} = \beta\left(\frac{n-4}{2}\right)! \Big/ \left(\frac{n-3}{2}\right)! = \frac{\beta}{\sqrt{\frac{n}{2}-\frac{3}{2}+\frac{1}{4}}} = \frac{\beta}{\sqrt{\frac{n}{2}-\frac{5}{4}}} = \frac{\sqrt{2}\,\beta}{\sqrt{n-\frac{5}{2}}} = \frac{\sqrt{n}\,\Sigma}{\sqrt{n-\frac{5}{2}}}$

52.20 $\boxed{\frac{M_2}{M_0} = \frac{n\Sigma^2}{n-3}}$

$$\boxed{\frac{M_1}{M_0}} \approx \boxed{=\sqrt{\frac{n}{n-\frac{5}{2}}} \cdot \Sigma}$$

$$\sqrt{\frac{n}{n-\frac{5}{2}}} \simeq \left(1+\frac{5}{4}\cdot\frac{1}{n}\right) \text{ for large } n.$$

$$\frac{M_2}{M_0} - \left(\frac{M_1}{M_0}\right)^2 = \Sigma^2\left(\frac{n}{n-3} - \frac{n}{n-\frac{5}{2}}\right) = \Sigma^2 \cdot n\left(\frac{\frac{1}{2}}{(n-3)(n-\frac{5}{2})}\right)$$

$$\frac{M_2}{M_0} = \Sigma^2 \cdot \frac{n}{n-\frac{5}{2}} \cdot \boxed{\frac{1}{2(n-3)}} \leftarrow \text{ratio of } \downarrow \text{ from } \approx \Sigma^2$$

---

Is $\frac{n}{n-5/2}$ .20 consistent? should not t. estimate of mean be unbiased? perhaps not:

It's just a normzd first moment! The first moment Mult by $\sqrt{\frac{n-\frac{5}{2}}{n}}$ would perhaps be unbiased.

---

for large $n$, t. d.f. converges to a narrow peak in the $\mu, \sigma$ plane:

width in $\sigma$ direction is $\Sigma \cdot \sqrt{\frac{n}{n-\frac{5}{2}}} \left(\frac{1}{\sqrt{2(n-3)}}\right)$

width in $\mu$ " " $\sim \Sigma \cdot \frac{1}{\sqrt{n-1}}$ ≈ $\quad$ so for large $n$, its wider in $\mu$ than $\sigma$ direction, by $\sqrt{2}$ factor.

For smaller $n$, we want $\exists\, \frac{\mu}{\sigma}$, which is known for Gaussian 2 dim d.f.s.

The Gauss. in t. $\sigma$ direction can cause trouble w. $\sigma \doteq \phi$ (maybe the [illegible] may have [illegible] integrable $\infty$).

T. d.f. I'm using has an essential zero (singularity) at $\sigma = \phi$, so this would cause no trouble in that case, but t. integration ~~[illegible]~~ may be analytically diff't.

.38 $\quad$ 47.26

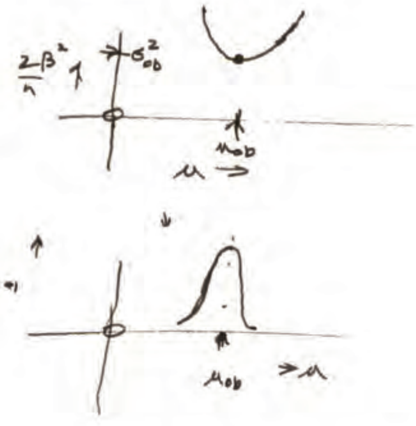
$$\int_{\sigma=0}^{\infty}\int_{\mu=-\infty}^{+\infty} \frac{\mu}{\sigma}\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{n} \exp - \frac{n(\sigma_{ob}^2 + (\mu-\mu_{ob})^2)}{2\,\sigma^2}\, d\mu\, d\sigma$$

t. factor $\frac{1}{\sigma}$ causes no/chang impt. in t. expressn. " " $\mu$ seems to change things a lot.

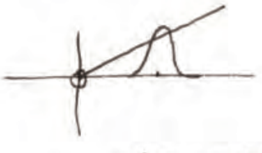
we can easily integrate in the σ direction.

.02 $\frac{2\beta^2}{n} = (\sigma_{ob}^2 + (\mu-\mu_{ob})^2)$ ; via 49.32 $53.38 = \int_{\mu=-\infty}^{+\infty} d\mu\, \mu \cdot \beta^{-n+1}$ | $\cdot \frac{1}{2(\sqrt{2\pi})^n} \cdot \left(\frac{n-2}{2}\right)!$

→ should be $\beta^{-n}$; See 55.20-.23



since t. function is symmetric about $\mu = \mu_{ob}$, the integration is trivial! — it's as if all mass were concentrated at $\mu_{ob}$! — Not so trivial! [See 58.01-.08 for exact argument]



The first moment is related to the zeroth moment by the factor $\mu_{ob}$.

— But we still have to find the zeroth moment.

just set $\mu$ to $\phi$ and integrate $\int_{-\infty}^{+\infty} \beta^{1-n} d\mu$ $\left(\frac{2\beta^2}{n} \to (\sigma_{ob}^2+\mu^2)\right)$

.20 $\int_{-\infty}^{+\infty}\left(\frac{1}{\sigma_{ob}^2+\mu^2}\right)^m d\mu = \frac{\sigma_{ob}}{(\sigma_{ob}^2)^m}\int_{-\infty}^{+\infty}\left(\frac{1}{1+\frac{\mu^2}{\sigma_{ob}^2}}\right)^m \frac{d\mu}{\sigma_{ob}} = \sigma_{ob}^{1-m}\int_{-\infty}^{+\infty}\left(\frac{1}{1+x^2}\right)^m dx$ No!

probably a known S.

$\sigma_{ob}^{1-2m}$

$\to \frac{2\sqrt{\pi}}{\sqrt{n-1}}$ for large n, Not appropriate for n=1

.23 G&R p245 3.252.2 Also p294 3.249.1

$\int_0^{\infty} \frac{dx}{(x^2+c^2)^n} = \frac{(2n-3)!!\,\pi}{(2n-2)!!\,c^{2(n-\frac{1}{2})}} \simeq \frac{\sqrt{\pi}}{\sqrt{n-1}}\cdot\frac{1}{c^{2n-1}}$

should be 2 here

Note

a=1; b=0 c=c

$z!! \equiv z(z-2)(z-4)\cdots$ (1 or 2)

$= \frac{\pi(2n-3)!!}{(2n-2)!!}\cdot\frac{1}{c^{2n-1}} \sim \frac{\pi}{c^{2n-1}}\cdot\frac{1}{\sqrt{\pi(n-1)}}$

2 4 6 8 10
1 3 5 7 9

.26 Approx'ly: $\exp\left(\frac{1}{2}+\frac{1}{4}+\frac{1}{6}+\frac{1}{8}\cdots\frac{1}{2J}\right) = \exp\left(\frac{1}{2}\ln j+\gamma\right) = k\cdot\sqrt{J}$

$\left(\frac{1}{2}\cdot\frac{3}{4}\cdot\frac{5}{6}\cdot\frac{7}{8}\cdots\frac{2J-1}{2J}\right)\cdot\sqrt{\pi J} \to 1$

see my note bottom of p190 of Jolly

It could be obtained from Stirling's, but not easily.

It may be that $\frac{1}{2}\cdot\frac{3}{4}\cdot\frac{5}{6}\cdots$ is related to $(n-\frac{1}{2})!$ ? No!

$\frac{2}{1}$ $\frac{4}{3}$ | $\frac{6}{5}$ $\frac{48}{15}$ | $\frac{8}{7}$ | $\frac{10}{9}$ | $\frac{12}{11}$

1.03 1.025 1.02

$\frac{8}{3}$

$\frac{8}{3}/\sqrt{2\pi}$ = $(.94)^{-1}$ = 1.063

$\frac{48}{15}/\sqrt{3\pi}$ = 1.04

2π

.28 $$\boxed{\int_{-\infty}^{+\infty}\frac{dx}{(x^2+c^2)^m} \approx \frac{2\sqrt{\pi}}{\sqrt{m-1}}\cdot\frac{1}{c^{2m-1}}}$$ ← formula is completely wrong for m = 1.

---

I don't understand the $\sigma_{ob}^{1-2m}$ factor in .20 — actually $\beta^{1-2m}$

4.063952

omit this "2" factor

$-m = -n+1$ ; $m = n-1$

$\int_{-\infty}^{+\infty}\frac{d\mu}{(\mu^2+c^2)^{n-1}} \approx \frac{2\sqrt{\pi}}{\sqrt{n-2}}\frac{1}{c^{2n-2-1}} = \frac{2\sqrt{\pi}}{\sqrt{n-1}}\cdot\frac{1}{c^{2n-3}}$ ; c = β

No. → I expect it to be of dimension 1). (no "dimensionless") No! — But Meanwhile of dimension $\frac{1}{\sigma}$, so when I mult it by μ, I got dim $\frac{\mu}{\sigma}$.] → Soln. 55.07

Telix
C:\Sim...
\kp41
\KPUSF.Bat
C:\Bn\TY.Bat

Look at 53.38 w.o. t. $\frac{\mu}{\sigma}$ factor; A purely dimensional form!

.38 $\int_{\mu=0}^{+\infty}\int_{\sigma=0}^{+\infty} \frac{1}{\sigma^n} \exp-\frac{n(\sigma_o^2+(\mu-\mu_{ob})^2)}{2\sigma^2}\, d\sigma\, d\mu$

form of a product:

T. integrand is in t. form of a product: $\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma}(S_i-\mu)^2}$

3.20.97 TM

Eq. 54.38R would seem to be dimensions $\frac{1}{\sigma^n}$ : When integrated in $\sigma$ direction; $\frac{1}{\sigma^{n-1}}$

When integrated w. $d\sigma$ & $d\mu$, dimension $\frac{\mu}{\sigma^{n-1}}$. or $\frac{1}{\sigma^{n-2}}$. (since $\mu$ & $\sigma$ have same [redacted] dimension

Consider a simpler problem: A variable has zero mean and unk. $\sigma^2$ & is gaussian.

[$S_i$] are observations: $\sigma$ is a possl. s.d.

.07 Proby of $S_i$ if $\sigma$ is true s.d. is: woops! proby of observation betw. $S_i$ & $S_i+\Delta$ is integrand·$\Delta$

So, we have this factor $\Delta^n$ whenever we make n observations! That explains t. dimensionality!

\+ The $\frac{1}{\sigma^n}$ factor gives t. probability density in an n dim space (of n observations).

---

.10 O.k. so lets get back to t. problem of 54.02 & 53.38

$$\int_{-\infty}^{+\infty}\beta^{1-n}\,d\mu = \qquad \beta=\left(\frac{n}{2}\left(\sigma_{ob}^2+\mu^2\right)\right)^{\frac{1}{2}} \quad \text{— (i.e. } \mu_{ob}=0)$$

$$\left(\frac{n}{2}\right)^{\frac{1-n}{2}}\int_{-\infty}^{+\infty}\frac{d\mu}{\left(\sigma_{ob}^2+\mu^2\right)^{\frac{n-1}{2}}} = \left(\frac{n}{2}\right)^{\frac{1-n}{2}}\cdot\frac{2\sqrt{\pi}}{\sqrt{\frac{n-3}{2}}}\cdot\frac{1}{\sigma_{ob}^{n-2}}$$

in 54.28 $m=\frac{n-1}{2}$

$m-1=\frac{n-3}{2}$

$2m-1=n-2$

I'm not sure about dimensions again: may be off by factor of $\sigma$.

.19

---

.20 In 54.02 I. dim after integration by $\frac{d\sigma}{\sigma}$, was $\frac{\beta}{\beta^n}$ : or $\sigma^{-n+1}$

I think in $\int$ I forgot the extra $\frac{1}{\sigma}$ factor: I. $\int$ should be of dim $\frac{1}{\sigma^{n+1}}\cdot d\sigma$ or $\frac{1}{\sigma^n}$

The integrand of (w.o. $d\sigma$) 54.02 is of dim. $\sigma^{-n+1}$ so I did forget t. $\frac{1}{\sigma}$ factor.

.23 the factorial part is correct, but only t. exponent for $\beta$ is wrong. should be $\beta^{-(n+1)+1}=\beta^{-n}$

So t. integrand is $\beta^{-n}=\left(\beta^2\right)^{-\frac{n}{2}}$

So

in 54.28 $m=\frac{n}{2}$

$m-1=\frac{n-2}{2}$

$2m-1=n-1$

So $$\int_{-\infty}^{+\infty}\beta^{-n}\,d\mu = \int_{-\infty}^{+\infty}\left(\beta^2\right)^{-\frac{n}{2}}d\mu = \left(\frac{n}{2}\right)^{-\frac{n}{2}}\cdot\frac{2\sqrt{\pi}}{\sqrt{\frac{n-2}{2}}}\cdot\frac{1}{\sigma_{ob}^{n-1}}$$ ← Mult by $\frac{1}{2(\sqrt{2\pi})^n}\cdot\left(\frac{n-2}{2}\right)!$ to get $\int$ of 54.02

.30

---

For t. final result, we mult this by $\mu_{ob}$ to obtain t. expected value of $\frac{\mu}{\sigma}$ (!)

Clearly there is something wrong! The dimension is $\frac{\mu\cdot\sigma}{\sigma^n}$ :

w. the integrand of dim $\left(\frac{\mu}{\sigma}\right)\frac{1}{\sigma^n}$ we integrated in $d\mu$ & $d\sigma$ directions to get $\frac{\mu\cdot\sigma}{\sigma^n}\to\frac{\sigma^2}{\sigma^n}$.

dim zero

But: we need a result w. dim zero: i.e. $\frac{\mu}{\sigma}$. Well; multiply by $\Delta^n$ to get t. proby of being in t. $\Delta^n$ n-cube assoc. w. t. n. observations, [$S_i$].

---

Maybe for some small $\Delta$ size, the integral of 54.38 gives t. relative distribution of $\sigma$, $\mu$ in the $\sigma$, $\mu$ plane. T. values do not sum to 1. T. values are $\propto\Delta^n$, $\Delta$ being edge of hicube assoc.

3.20.97 TM.

/ so maybe we have to normalize by integrating ~~[illegible]~~ 53.38 w.o. t. $\frac{\mu}{\sigma}$ factor

This new integration is only slightly difrent from t. previous one;

T. integral corresponding to 54.02 (via 49.32) is 53.38 (w.o. $\frac{\mu}{\sigma}$) ; )

.04 $$\underbrace{\int_{-\infty}^{+\infty} d\mu\ \beta^{-n+1}} \cdot \left(\frac{n-3}{2}\right)! \cdot \frac{1}{2} \cdot \left(\frac{1}{2\pi}\right)^{\frac{n}{2}}$$

evaluated at 55.10 - 19

$$\boxed{\left(\frac{n}{2}\right)^{\frac{1-n}{2}} \frac{2\sqrt{\pi}}{\sqrt{\frac{n-3}{2}}} \cdot \frac{1}{\sigma_{ob}^{\,n-2}}}$$

I want 55.30 divided by 56.04

.12 do it factor by factor $$\frac{\left(\frac{n}{2}\right)^{-\frac{n}{2}}}{\left(\frac{n}{2}\right)^{\frac{1}{2}-\frac{n}{2}}} \quad \frac{\sqrt{\frac{n-3}{2}}}{\sqrt{\frac{n-2}{2}}} \quad \frac{\sigma_{ob}^{\,n-2}}{\sigma_{ob}^{\,n-1}} \quad \frac{2(\sqrt{2\pi})^{n}}{2(\sqrt{2\pi})^{n}} \frac{\left(\frac{n-2}{2}\right)!}{\left(\frac{n-3}{2}\right)!}$$

$$\left(\frac{n}{2}\right)^{\frac{1}{2}} \sqrt{\frac{n-3}{n-2}} \cdot \frac{1}{\sigma_{ob}} \sqrt{\frac{n}{2}-1+\frac{1}{4}} = \frac{1}{\sigma_{ob}} \sqrt{\frac{n-3}{n-2}} \cdot \sqrt{\frac{\frac{n}{2}-\frac{3}{4}}{\frac{n}{2}}}$$

$$\uparrow$$

$$\sqrt{\frac{n-1.5}{n}}$$

$$= \frac{1}{\sigma_{ob}} \cdot \sqrt{\frac{(n-3)(n-\frac{3}{2})}{(n-2)\,n}} \quad \text{which} \to \frac{1}{\sigma_{ob}} \text{ as } n \to \infty$$

so t. expected value of $\frac{\mu}{\sigma}$ is $\frac{\mu_{ob}}{\sigma_{ob}} \underbrace{\sqrt{\frac{(n-3)(n-\frac{3}{2})}{(n-2)\,n}}}_{\text{which is always } < 1}$ $\approx \frac{\mu_{ob}}{\sigma_{ob}} \cdot \left(1-\frac{1.25}{n}\right)$ for $n \geq 6$

Try it out for n = 1, 2, 3 .... 20

.20

| n | exact $\sqrt{\frac{(\ )(\ )}{(\ )n}}$ | approx $1-1.25/n$ | $\boxed{1-1.25/(n-1)}$ |
|---|---|---|---|
| 4 | .559 | .69 | |
| 6 | .75 | .79 | |
| 10 | .86 | .875 | |
| 15 | .911 | .916 | |
| 20 | .931 | .934 | |
| 25 | .95 | | |
| 30 | .957 | | |
| 35 | .963 | | |
| 40 | .968 | | |

is very ~~very~~ close ! See $\boxed{56\frac{1}{2}}$

~~[illegible]~~ TM7-56.Bas

See 56½ for output of n = 1 | 40.

for large n: $\sqrt{1-\frac{3}{n}-\frac{3}{2n}+\frac{2}{n}}$

$$1\left(-1-\frac{3}{2}\right)\cdot\frac{1}{n} = 1-\frac{2.5}{n}$$

$$\approx 1-\frac{1.25}{n}$$

$$\simeq 1-\frac{1.25}{n}$$

$$\boxed{1-\frac{1.25}{n-1}}$$ is much closer, even for n = 5 its only off by × 1.0064 ; .64%

for n = 10 its off by .15% ; for n = 6, its exact.

for n = 100 its off by .0022% | for n = 20, off by .013%

T. forgo results may be much modified by t. apriori. T. forg. assumes uniform. apriori for σ & μ.

That t. ~~[illegible]~~ E value in t. forgo is not exactly $\frac{\mu_{ob}}{\sigma_{ob}}$ means it may not be "unbiased",

— so ~~[illegible]~~ possibly ~~mult it by~~ divide it by ? to "unbias" it (?).

| n | $\sqrt{\frac{(n-3)(n-\frac{3}{2})}{(n-2)n}}$ | $1-1.25/n$ |
|---|---|---|
| 4 | .559017 | .6875 |
| 5 | .68313 | .75 |
| 6 | .75 | .7916667 |
| 7 | .792825 | .8214285 |
| 8 | .8228507 | .84375 |
| 9 | .8451542 | .8611111 |
| 10 | .8624094 | .875 |
| 11 | .8761716 | .8863636 |
| 12 | .887412 | .8958333 |
| 13 | .8967696 | .9038461 |
| 14 | .9046836 | .9107143 |
| 15 | .9114654 | .9166667 |
| 16 | .9173428 | .921875 |
| 17 | .9224859 | .9264706 |
| 18 | .9270248 | .9305556 |
| 19 | .9310602 | .9342105 |
| 20 | .9346716 | .9375 |
| 21 | .9379228 | .9404762 |
| 22 | .9408652 | .9431818 |
| 23 | .9435408 | .9456522 |
| 24 | .9459843 | .9479167 |
| 25 | .9482248 | .95 |
| 26 | .9502867 | .9519231 |
| 27 | .9521905 | .9537037 |
| 28 | .9539536 | .9553571 |
| 29 | .9555912 | .9568965 |
| 30 | .9571162 | .9583333 |
| 31 | .9585398 | .9596774 |
| 32 | .9598719 | .9609375 |
| 33 | .9611211 | .9621212 |
| 34 | .9622946 | .9632353 |
| 35 | .9633995 | .9642857 |
| 36 | .9644413 | .9652778 |
| 37 | .9654254 | .9662162 |
| 38 | .9663566 | .9671053 |
| 39 | .9672388 | .9679487 |
| 40 | .968076 | .96875 |
| 99 | .9872666 | .9873737 |
| 100 | .9873950 | .9875 |

N.B. $1-\frac{1.25}{n-1}$ is a much better approxn. —

Its exact at $n=6$ & error = .15% at $n=10$

" = .002% at $n=100$.

at $n=1000$ error = $2.18 \times 10^{-5}$%

$= x(1+ 2.18\times10^{-7})$

Looks like error is like $\frac{1}{n^2}$.

A $\boxed{\text{Review of} \approx 46.20 - 56.40:}$

T. problem: I have a lot of data strings & a lot of strategies for predicting Them. I try t. strategies on some strings a/o sub-sets of strings, for various ssz's ($\equiv$ string lengths $\equiv n$). For each, I get

.04 a $\mu_{ob} = \frac{1}{n}\sum_{i=1}^{n} S_i$ and a $\sigma_{ob}^2 = \frac{1}{n}\sum(S_i - \mu_{ob})^2$.

Here $S_i$ is the "yield" of t. $i$th try. I'm interested in finding strats that have a large $\left.\frac{\mu}{\sigma}\right|_{yr}$. This is $\left.\frac{\mu}{\sigma}\right|_{per bet} \cdot \sqrt{\text{no. of bets/yr.}}$ or.

Max $\left.\frac{\mu}{\sigma}\right|_{\text{Time } T} \cdot \frac{1}{\sqrt{T}}$ (see 46.20-.25 for discn).

---

$\boxed{SN}$ This (46.20-56.40) investigation is a bit diffrnt from prediction. Its involvd w. $\boxed{\text{"yield"}}$ which is perhaps more generally useful.

---

Consider a particular data sequ. $[S_i]$ of yields. [illegible]

If the generator was a Gausn d.f. of mean $\mu$ & s.d. $\sigma$, t. proby of obtaining t. data in t. n-cube $(S_i, S_i + \Delta)^n$ (so to speak) is

.20 $$\prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi}\,\sigma}\cdot e^{-\frac{(S_i-\mu)^2}{2\sigma^2}}\right) = (2\pi)^{-\frac{n}{2}}\ \frac{\Delta^n}{\sigma^n}\ e^{-\frac{n(\sigma_{ob}^2 + (\mu_{ob}-\mu)^2)}{2\sigma^2}}$$ [See .04 for defns of $\sigma_{ob}^2$, $\mu_{ob}$]

This considers t. apripd of $\mu, \sigma$ to be uniform. For a poss. non-uniform apripd: see 48.25.

.22 If we let $\beta^2 \equiv \frac{n}{2}(\sigma_{ob}^2 + (\mu_{ob}-\mu)^2)$ we get

.23 $$(2\pi)^{-\frac{n}{2}}\Delta^n\ \frac{1}{\sigma^n}\ e^{-\frac{\beta^2}{\sigma^2}}.\quad \equiv f(\mu,\sigma)\cdot\Delta^n$$

Given that t. data occured in t. n-cube $\Delta^n$, the Bayesian a psi distribn on $\sigma, \mu$ is given by $f(\mu,\sigma)$

This is not, hvr a normalized d.f., & only gives t. relative densities in t. $\mu,\sigma$, plane.

To get t. expected value of $\frac{\mu}{\sigma}$:

.29 $$\left.\frac{\mu}{\sigma}\right|_{per bet} = \frac{\int_{-\infty}^{\infty} d\mu \int_0^{\infty} d\sigma\ \frac{\mu}{\sigma} f(\mu,\sigma)}{\int_{-\infty}^{+\infty} d\mu \int_0^{\infty} d\sigma\ f(\mu,\sigma)}$$ ← integral in denom. is indip of $\mu_0$, so set it to $\phi$ in denom.

Since the $(2\pi)^{-\frac{n}{2}}$ factor occures in both numerator & denominator, we will not consider it henseforth.

33 Consider t. denominator: $\int_0^{\infty} d\sigma\, f(\mu,\sigma) = \int_0^{\infty}\frac{d\sigma}{\sigma^n}e^{-\frac{\beta^2}{\sigma^2}} = \frac{\beta^{-n+1}}{2}\left(\frac{n-3}{2}\right)!$ see 49.29-.32 for this last derivation.

.34 next, consider $\int_{-\infty}^{+\infty} d\mu\ \beta^{(1-n)} = \int d\mu\,(\beta^2)^{\frac{1-n}{2}}$ [illegible]

say n=3 m=

$$\left(\frac{n}{2}\right)^{\frac{1-n}{2}} \int d\mu\cdot(\sigma_{ob}^2+\mu^2)^{\frac{1-n}{2}} \overset{\text{via } 54.28}{\approx} \left(\frac{n}{2}\right)^{\frac{1-n}{2}}\cdot\frac{2\sqrt{\pi}}{\sqrt{\frac{n-1}{2}-1}}\ \frac{1}{\sigma_{ob}^{n-1-1}}$$

$m = \frac{n-1}{2}$

$m-1 = \frac{n-3}{2}$

$2m-1 = n-2$

This is (see .22 for defn of $\beta^2$)

.38 $\boxed{\text{Denom}} \approx \boxed{\left(\frac{n}{2}\right)^{\frac{1-n}{2}}\cdot\frac{2\sqrt{\pi}}{\sqrt{\frac{n-3}{2}}}\cdot\frac{1}{\sigma_{ob}^{n-2}}}\boxed{\cdot\frac{1}{2}\left(\frac{n-3}{2}\right)!}$

from (.33R)

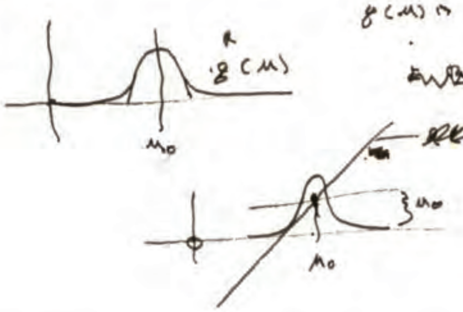
almost cancels

cancels

cancels.

→ 58.09

.01 $\boxed{SN}$ Re: ~~the~~ argumentB of 57.02 – .20 on integrating $(\beta^2)^{\frac{1-n}{2}}$ !



g(μ) is symm about $\mu_0$. I want $\int g(\mu)\,d\mu$.

~~[illegible]~~ It is $\mu_0 \int_{-\infty}^{+\infty} g(\mu)\,d\mu$ & reason is:

say ~~[illegible]~~ $l_{sl}(\mu)$ is a st. line thru $(\mu_0, \mu_0)$ w. slope sl.

Then the $\int$ of ~~[illegible]~~ $l_{sl}(\mu)\cdot g(\mu)$ is indep of sl.

So let slope $= \phi$ ∴ This is true

.08

---

.09 ! 57.40 ! Next, look at numerator of 57.29:

see 49.29 –.32 for derivation !

.10 first do $\int_0^\infty d\sigma \frac{1}{\sigma} f(\mu,\sigma) = \int_0^\infty \frac{d\sigma}{\sigma^{n+1}} e^{-\frac{\beta^2}{\sigma^2}} \uparrow\ \frac{\beta^{-n+1+1}}{2}\left(\frac{n-1-3}{2}\right)! = \beta^{-n}\cdot\frac{1}{2}\left(\frac{n-2}{2}\right)!$

$$\int_{-\infty}^{+\infty} d\mu \cdot \mu\,\beta^{-n} \overset{\uparrow}{=} \mu_{ob}\int_{-\infty}^{+\infty} \beta^{-n}\,d\mu = \mu_0 \int (\beta^2)^{-\frac{n}{2}} d\mu = $$

with $\mu_0$ set $= 0$

See 58.01 –.08 for why this is true.

$$= \mu_{ob}\ \frac{1}{2}\left(\frac{n}{2}\right)^{-\frac{n}{2}} \int d\mu\,(\sigma_{ob}^2 + \mu^2)^{\frac{n}{2}} = \left(\frac{n}{2}\right)^{-\frac{n}{2}} \frac{2\sqrt{\pi}}{\sqrt{\frac{n-2}{2}}}\ \frac{1}{\sigma_{ob}^{n-1}}$$

$m = \frac{n}{2}$
$m-1 = \frac{n-2}{2}$
$2m-1 = n-1$

.19 (.10) = entire numerator $= \boxed{\left(\frac{n}{2}\right)^{-\frac{n}{2}} \frac{2\sqrt{\pi}}{\sqrt{\frac{n-2}{2}}} \cdot \frac{1}{\sigma_{ob}^{n-1}} \cdot \frac{1}{2}\left(\frac{n-2}{2}\right)! \cdot \mu_{ob}}$

$$\frac{58\ .19}{57\ .38} \approx \left(\text{see 56.12 for actual calcs}\right) \approx \frac{\mu_{ob}}{\sigma_{ob}}\left(\frac{n}{2}\right)^{\frac{1}{2}}\sqrt{\frac{n-3}{n-2}} \cdot \underbrace{\frac{\left(\frac{n-2}{2}\right)!}{\left(\frac{n-3}{2}\right)!}}_{} \to \approx\sqrt{\frac{n}{2}-1+\frac{1}{4}} \leftarrow \text{from 53.18}$$

.25 $$\boxed{\approx \frac{\mu_{ob}}{\sigma_{ob}}\cdot\left(\frac{(n-3)(n-\frac{3}{2})}{(n-2)\cdot n}\right)^{\frac{1}{2}}}$$ This is the expected value of $\frac{\mu}{\sigma}$ for n observations → see 61.25 for comments on validity.

.26 $\approx \frac{\mu_{ob}}{\sigma_{ob}}\left(1-\frac{1.25}{n}\right)$ for $n > 10$: see $\binom{56.20}{55\frac{1}{2}}$ for values as a funct of n for (.23); for approx of (.26).

" $\cdot\left(1-\frac{1.25}{n-1}\right)$ is a much better approxn. hrr.

N.B. The approxn. of $\frac{(X+\frac{1}{2})!}{X!}$ is very good even for X as small as 2 & improves much as X ↑.

The approxn of $\int\frac{d\mu}{\beta^n}$ of 54.28 is also v.g. even for m as small as 2, it's only off by ~ 6%. ; 2% error for m = 6.

→ The $\int$'s diverge for n = 1 or 2; For n = 3 the formula for $\int_{-\infty}^{+\infty}\frac{d\mu}{(\sigma_0^2+\mu^2)^m} = \frac{2m-3!!}{2m-2!!}\cdot\frac{\pi}{c^{2m-}}$

Is correct, but the approxn $\frac{\sqrt{\pi}}{\sqrt{m-1}}$ is incorrect: m = 3/2 yields m = 1 :

$\int_{-\infty}^{+\infty}\frac{d\mu}{(\sigma_0^2+\mu^2)} = \frac{\pi}{\sigma_0}$ The correct form when n = 3 of equ. .25 is

$$\frac{\mu_{ob}}{\sigma_{ob}}\cdot\left(\frac{n-\frac{3}{2}}{n}\right)^{\frac{1}{2}} = \boxed{\frac{\mu_{ob}}{\sigma_{ob}}\cdot\sqrt{\frac{1}{2}}}$$

For nice way to deal w. n=3 & Also correction for 58.25

→ $\boxed{\text{See 65.01}}$

3.21.97 JM:

Another Q one mite ask is: If t. "true" μ & σ are μ & σ, what are t. expected values of $\mu_{ob}$ & $\sigma_{ob}$ for n observations.? What is t. d.f. of these estimates.?

Well, ~~t.~~ t. expected value of $\mu_o$ is μ. I think ~~n~~ t. expected value of $\sigma_{ob}^2$ is $\frac{\sigma^2(n-1)}{n}$

Fidelity Criterion.

T. result of 58.25 was obtaind assuming uniform apriori for σ: What would it look like if we assume uniform apriori for $\sigma^2$.?

$d(\sigma^2) = 2\sigma d\sigma$. It has t. effect of using $n \to n-1$ in t. defn. of $f(\mu,\sigma)$ of 57.23

so t. integrands in both num & den of 57.29 are mult by $2\sigma$.

The result should be easy to compute, using t. (almost) exactly same eqns. as 57.01–58.40:

— ~~But I'm not sure it makes is much~~ But my impressn. is that this is not t. main present problem.

At first Glance it seems to do $n \leftarrow n-1$ in 58.25

.13 So $\approx \frac{\mu_{ob}}{\sigma_{ob}} \left(\frac{(n-4)(n-\frac{5}{2})}{(n-3)(n-1)}\right)^{\frac{1}{2}}$ : But this should be checked more carefully, if it to be actually used.

For "large" n it is $\frac{\mu_{ob}}{\sigma_{ob}}\left(1-\frac{1.25}{n}\right)$ — same as 58.26, but n must be larger than in 58.26 before this becomes a good approxn. It may be ~~shown~~ an impt. difrnce say for n=5.

2.5

−1 −1.5

Look at table of 56.20: t. formula of .13 amounts to $n \to n-1$

T. difrnce betw. 56.25 & 59.13 is best studied by looking at $1-\frac{1.25}{n}$ v.s. $1-\frac{1.25}{n-1}$:

No dramatic difrnce! See, listing of 56½

If formula (.13) is wrong, it is most likely to be of form of changing certain n values (in both num & denom) by 1 — producing little change ~~[illegible]~~ & no change in approxn., $1-\frac{1.25}{n}$.

---

Actually, what I probly got in 58.25 was t. result of finding t. expected value of $\frac{1}{\sigma}$.

— Not quite! Integrating wrt. μ did produce an extra $\sqrt{\frac{(n-3)}{(n-2)}}$ factor;

~~Integra~~ I'm not sure this is meaningful: to get ~~E~~ E $\frac{1}{\sigma}$, one has to integrate wrt. μ, any way — So I guess t. result of finding E $\frac{1}{\sigma}$, then multiplying by $\mu_{ob}$ would be identical to 58.25.

---

O.K. Returning to hos Houtens: I want to find a strat. w. hy $\left(\frac{\mu}{\sigma}\right)^2 \cdot \frac{1}{T}$ : This is t. exponential yield per unit time. So perhaps I should have computed E $\frac{\mu^2}{\sigma^2}$.

Most likely, it will ~~[illegible]~~ be of form $\left(\frac{\mu_{ob}}{\sigma_{ob}}\right)^2 \cdot \left(\frac{(n-a)(n-b)}{(n-c)(n-d)}\right)^{\frac{1}{2}}$.

Now: say all ~~systems~~ strats considerd have 1 bet/day: — so t. want strat w. max

$$\left(\frac{\mu_{ob}}{\sigma_{ob}}\right)^2 \cdot \underbrace{\left(\frac{(n-a)(n-b)}{(n-c)(n-d)}\right)^{\frac{1}{2}}} \cdot \frac{1}{n}$$

"like T"

$\mu_{ob}$ & $\sigma_{ob}$ will ~~be~~ be around $\mu_{true}$ & $\sigma_{true}$, resp.

In general, $\sigma_{ob}^2 \cdot n$ will be about constant, so for various values of n, this expression will not change much; ~~Ex~~: this part will ↑ a bit w. n.

So from this discussn. there would seem to be little motivation to use larger n values in trials! — not at all reasonable!

3.21.97 TM

Perhaps what I want is a D.f. for $\left(\frac{1}{n}\cdot\frac{\mu^2}{\sigma^2}\Big|_n\cdot\right)$ induced by t. data.

To get this, integrate $f(\mu,\sigma)$ (57.23) along t. line $\frac{\mu}{\sigma} = k$ = constant

for a first approxn: Get $\sigma_{ob}^2$ using some reasonable n; then we have $\mu$ centered about $\mu_{ob}$ w. var $\sim \frac{\sigma_{ob}^2}{n}$. — or $\mu_{ob}$ w. var $\sim \sigma_{ob}^4$.

So $\frac{\mu^2}{\sigma^2}$ has $\approx$ mean $\frac{\mu_{ob}^2}{\sigma_{ob}^2}$ w. var $\frac{\sigma_{ob}^4}{(\sigma_{b}^{4})^2}\cdot\frac{1}{n} \approx \frac{1}{n}$.

$\tilde{\sim}$ mean $\left(\frac{\mu_{ob}}{\sigma_{ob}}\right)^2$ " " $\frac{1}{n}$

I suspect that $\left(\frac{\mu_{ob}}{\sigma_b}\right)^2 \div \left(\frac{1}{n}\right) = n\left(\frac{\mu_{ob}}{\sigma_{ob}}\right)^2$ has to be $>1$ before one can get much expected yield out of it.

.15 One trouble: $E\,\mu^2$ is always $>0$. The actual yield is not $\mu^2$, but true $\mu$ times what one thinks $\mu$ is. For small $\mu$, this product can be $>0$ or $<0$: On t. average, for small $\mu$, it's close to $\emptyset$. So t. yield may be $\propto \mu_{ob}\cdot\mu$. I think n has to be large enuf so that one is fairly sure of t. sign of $\mu$ (!) — I think this is $\approx n\left(\frac{\mu_{ob}}{\sigma_{ob}}\right)^2 > 1$ i.e. n must be large enuf so that t. $\left(\frac{\mu_{ob}}{\sigma_{ob}}\right)_n$ for that n, is $\approx > 1$. i.e. $n\left(\frac{\mu_{ob}}{\sigma_{ob}}\right)^2 \approx > 1$.

.19

---

Which suggests that I can't learn much from small n! In general, it would seem unlikely that I'd get $\left(\frac{\mu_{ob}}{\sigma_{ob}}\right)^2\cdot n > 1$ for $n < 2$ yrs!

One possy. is that w. $n << 2$ yrs, I get $\frac{\mu_{ob}}{\sigma_{ob}}\cdot n$ of same sign for a very large no. of stocks.

---

Woops! t. Idea of .15 – .19 is wrong! even if $\frac{\mu_{ob}}{\sigma_{ob}}$ is very small, t. expected value of $\mu_{ob}\cdot\mu$ is still $\mu_{ob}^2$! Assuming mean is $\mu_{ob}$ & betting that way, very often has neg. yield (i.e. $\mu<0$ & $\mu_{ob}>0$.) — but very often, yield is $>> \mu_{ob}^2$ (i.e. $\mu > \mu_{ob}$). — so on t. Average, yield is $\propto \mu_{ob}^2$. For low $\frac{\mu}{\sigma}$, Kelly bet is $\propto \frac{\mu}{\sigma^2}$ : yield is $1+\frac{\mu^2}{\sigma^2}$ or [illegible] mult yield $\approx \exp\left(\frac{\mu^2}{\sigma^2}\right)$ per bet. If $\mu$ has a gauss

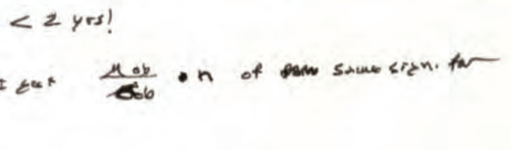
Actually if $\mu_{ob}$ is mean of broad Gauss d.f., one bets an amount $\propto \mu\cdot e^{-\frac{(\mu-\mu_o)^2}{2\sigma_o^2}}$ (assuming $\sigma_o = \sigma \equiv$ actual s.devn.)

---

Defs $\mu_o \equiv \mu_{observed} \equiv \mu_{ob}$; $\sigma_o \equiv \sigma$ observed $\equiv \sigma_{ob}$.

---

total bet = $\int_{-\infty}^{+\infty} \mu\, e^{-\frac{(\mu-\mu_o)}{2\sigma_o^2}}\,d\mu$



$\propto \boxed{\mu_o}$.

So it would seem that value of x-strat was $\propto \left(\frac{\mu_o}{\sigma_o}\right)^2$ (per bet ≡ "per day") — If t. $\left(\frac{\mu_o}{\sigma_o}\right)^2$ was obtained w. large n, t. fact that $\mu_o$ is known to high precision, does not seem to ↑ yield!

.01 Seems Unreasonable! Say t. data is random. We will have a broad distribn. of $M_0$'s for t. various strategies, ~~[illegible]~~ Our "expected" ~~[illegible]~~ (s) yield will be $\propto \frac{1}{n}$ maybe. I.e. we will Think $M^2$ is large when n is small.

So: in Case (.01) how would we realize that there were no winning strats"? ——

Well, we would notice that t. no. of apparently winning strats is consistent w. t. hypoth of randomness of data., & that this hypoth continues to fit for larger values of n.

This way of looking at things is easy to justify if t. "strat" is a ~~[illegible]~~ Pem — in which case t. random "data" hypoth. is ~~[illegible]~~ of high apriori, & gives t. same code length [redacted] (or no longer) than other models give.

87

Well, actually, if we multd. 57.20 (≐57.23) by t. apripd, we would have t. rite answer.

T. fact is, a uniform prior is unlikely & "unreasonable", & its not surprizing that it gives "Unreasonable" results.

On t. other hand, w. a reasonable prior to mult. by, 57.20 is correct, & unarguably so. Bayes idea is a "Theorem" [redacted] not a "suggestion".

So, a reasonable prior is t. Only way out of the "Unreasonableness" of ~~[illegible]~~ (.01)

We can start out by obtaining a prior via large n for maybe 10 trials: Have n large and so that $n\left(\frac{M_0}{\sigma_0}\right)^2 > 1$ — say ~ 2 or 4.

After one has begun to get useful results, t. apripd will change! it will be a funct. of t. params used to define t. strategy. .24

→ 66.01

66.20 spec

.25 (SN) Formula 58.25 is clearly wrong for n = 2 or 3; for n = 1 it gives imaginary values ∴ wrong.

Hvr., for $n = \overset{\text{no}}{1, 2}, 3$, t. exact Bayesian treatment gets reasonable results. The trouble is that t.

approxn : $\frac{(x+\frac{1}{2})!}{x!} = \sqrt{x+\frac{3}{4}}$ works o.k. for $\overset{.2\%\text{ error}}{x=2}$ or even $\overset{.5\%\text{ error}}{n=1}$

$-\frac{1}{2}! = \sqrt{\pi}$

~~[illegible]~~ $\frac{1}{2}! = \frac{1}{2}\sqrt{\pi}$

n = 1 $\frac{(1+\frac{1}{2})!}{1!} = \frac{3}{4}\sqrt{\pi} \doteq$ [redacted] 1.32934

$\sqrt{1+\frac{3}{4}} = \sqrt{\frac{7}{4}} = 1.322$ ) .5% error

$(\frac{3}{2})! = \frac{3}{2}\cdot\frac{1}{2}! = \frac{3}{4}\sqrt{\pi}$

$\frac{\frac{1}{2}!}{0!} = \frac{\sqrt{\pi}}{[illegible]}$

n = 2 $\frac{(2+\frac{1}{2})!}{2!} = \frac{5}{2}\cdot\frac{3}{2}\sqrt{\pi} = 6.646$

÷2! = 3.32335

$\sqrt{2.75} =$

$\frac{(0+\frac{1}{2})!}{0!} = \frac{1}{2}\sqrt{\pi} = .886227$

$\sqrt{.75} \doteq .8660$ — 2% off

n = 1, t. Σ of 54.23 converges, but I don't know what (2n−3)!! means or (2−2)!! means. −1 !! may be φ

From 58.26 : $\frac{(2n-3)!!}{(2n-2)!!} \sim \frac{1}{\sqrt{\pi(n-1)}}$ : for ~~the~~

for n = 2 : $\frac{1}{2} = \frac{1}{2} = .5$ ; $\frac{1}{\sqrt{\pi}} = .564$ ; which is only 13% off.

for n = 1 $\frac{-1!!}{0!!} = 1$ ; $\frac{1}{\sqrt{0}} = \infty$

3.22.77 TM

I. $\int$ converges for $m=1$.

T. approxn 54.28 is clearly wrong for $m=1$. | — The t. $\int$ <u>would</u> <u>diverge</u> for $m=\frac{1}{2}$.

I have note that $(-1)!!=1$; and of course $(2n)!!=2^n n!$

↑ I don't know where I got this!

, 03 $$\int_{-\infty}^{+\infty}\frac{dx}{(x^2+a^2)^1} = \frac{-1\,!!}{0\,!!}\cdot\frac{\pi}{a} = \frac{\pi}{|a|}$$ so t. <u>approxn</u> is <u>wrong</u> & diverges for $n=1$.

But for $n=2$ & 3 its o.k. — so 58.25 <u>must</u> <u>be</u> <u>wrong</u>!

In 58.25 the $\left(\frac{n-\frac{3}{2}}{n}\right)^{\frac{1}{2}}$ factor(s) are o.k. for $n \geq 1$. $\frac{n-3}{n-2}$ comes from $\int_{-\infty}^{+\infty}\frac{dx}{(1+x^2)^{n-k}}$ ($k=1,2,3$, perhaps,)

for $n=1$ : in 57.34. $\int du\,(\beta^2)^{\frac{1-n}{2}}$ diverges for $n=0,1,2$; but converges for $n=3$.

Th. main problem is t. denom. of 57.29 ("normal factor").

$$\int\int\frac{d\mu\,d\sigma}{\sigma^n}\,e^{-\frac{\sigma_0^2+\mu^2}{2\sigma^2}} = k\cdot\frac{1}{\beta^{n-1}}$$

$\beta=\left(\sqrt{\sigma_0^2+\mu^2}\cdot\sqrt{\frac{n}{2}}\right)$: $\beta^2=\frac{n}{2}(\sigma_0^2+\mu^2)$
let $\mu_0\equiv 0$.

$\int d\mu\,\frac{1}{\beta^{n-1}}$ diverges for $n=1$ or $n=2$, but converges for $n=3$.

---

(SN) The express for t. ratio of $\int\beta^{n-k}$ for t. ~~denom~~ numerator/denom. is obtained in an approxt. way: But Actually, if I wrote out the exact $\frac{a\,!!}{b\,!!}$ factors for num & denom, they would <u>mainly cancel</u> out & give a very simple <u>exact</u> ~~[illegible]~~ express'n!

say ~~[illegible]~~ Den = $\frac{2n-3\,!!}{2n-2\,!!}$ $\left(\approx\frac{\sqrt{\pi}}{\sqrt{n-1}}\right)$ $m=\frac{n}{2}-1$

$n-1$

$\equiv$ num

num: $m=\frac{n}{2}$

Num = $\frac{2n-5\,!!}{2n-4\,!!}$

~~num~~ den $\propto \int\frac{1}{(\beta^2)^{\frac{n-1}{2}}} = \frac{n-1-3\,!!}{n-1-2\,!!} = \frac{n-4\,!!}{n-3\,!!}$

num $\propto \int\frac{1}{(\beta^2)^{\frac{n}{2}}} = \frac{n-3\,!!}{n-2\,!!}$

So <u>nothing</u> <u>cancels</u>!

If t. numerator were for Expected value of $\frac{1}{\sigma^2}$, they <u>would</u> cancel a lot.: Id get ~~what happens is by~~ a factor like $\frac{n-3}{n-2}$! but by doing expected value of $\frac{1}{\sigma}$, I only go $\frac{1}{2}$ way, so $\sim\sqrt{\frac{n-3}{n-2}}$.

---

Looking at 57.33 & .34: for $n=1,2$. $\int\frac{1}{\beta^{n-1}}$ diverges:

for $n=3$, t. $\int\frac{1}{\beta^{n-1}}$ converges, but t. approxn. I'm using is invalid for $n-1=2$ | though. if $n-1=3,4,5\cdots$

for $n=3$ make special calcn.

denom $\int_{-\infty}^{+\infty}\frac{dx}{(c^2+x^2)^{\frac{n}{2}-\frac{1}{2}}}$, $\int_{-\infty}^{+\infty}\frac{dy}{(c^2+x^2)^{\frac{3}{2}}}$

for $n=3$ denom ~~[illegible]~~ $\int_{+\infty}^{-\infty}\frac{dx}{(c^2+x^2)^1} = \frac{\pi}{c}$ | for <u>numerator</u> ↗ $\int_{-\infty}^{+\infty}\frac{dy}{(c^2+x^2)^{\frac{3}{2}}} \approx \frac{\sqrt{\pi}}{\sqrt{2}}\cdot\frac{1}{c^2}$
$= \frac{3-3\,!!}{3-2\,!!}\cdot\pi\cdot\frac{1}{c^2}\quad\frac{1}{1}\cdot\frac{\pi}{c^2}$

So, for $n=3$ the $\left(\frac{n-3}{n-2}\right)$ factor is replaced by $\boxed{1}$.

3.22.97 TM

Maybe bad error on 54.28:

$n=3$ : $\frac{2n-3}{2n-2}$ : $\frac{6-3}{6-2} = \frac{3!!}{4!!} = \frac{3}{8}$

Actually, the formula I should be using is

.05 Turns out $\boxed{\frac{n-1!!}{n!!}} = \boxed{\frac{1}{\sqrt{\pi}} \frac{\left(\frac{n}{2}-\frac{1}{2}\right)!}{\frac{n}{2}!}}$ exactly! (prob'ly for n = integer or halfintegers).

so $\mapsto \approx \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{\frac{n}{2}+\frac{1}{4}}}$

.13 $$\int_{-\infty}^{\infty} \frac{dx}{(x^2+c^2)^m} = \frac{\pi}{c^{2m-1}} \cdot \frac{2m-3!!}{2m-2!!} \approx \frac{\pi}{c^{2m-2}} \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{m-1+\frac{1}{4}}} \approx \frac{\sqrt{\pi}}{c^{2m-1}} \frac{1}{\sqrt{m-\frac{3}{4}}}$$

Contrast w. 54.28 in which we have $m-1$ instead of $m-\frac{3}{4}$.

How good is the approx when $m=1$? $\frac{2-3!!}{2-2!!} = 1$ ± think.

.17 $\frac{-1!!}{0!!} \approx \frac{1}{\sqrt{\pi}\sqrt{\frac{1}{4}}} = \sqrt{\frac{4}{\pi}} = 1.128$ not great but tolerable. (13% error).

$n=2$: $\frac{1!!}{2!!} = \frac{1}{2}$ $\frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{1+\frac{1}{4}}} = .5046$

.20 So change 58.25 to $\frac{\sigma_{ob}}{\sigma_{ob}} \cdot \sqrt{\frac{(n-2.75)}{(n-1.75)} \frac{(n-1.5)}{(n)}}$

.375

This formula gives imaginary or $\infty$ for $n = 0, 1, 2$

for $n=3$, it gives a real value — but small

TM97-63.Bas

I think I did that conversion wrong. 58.25 was originally

$\sqrt{\frac{\frac{n-3}{2}}{\frac{n-3}{2}} \frac{\frac{n-1.5}{2}}{\frac{n}{2}}}$ | The correction is $\sqrt{\frac{\frac{n-3}{2}+\frac{1}{4}}{\frac{n-2}{2}+\frac{1}{4}} \cdot \frac{\frac{n-1.5}{2}}{\frac{n}{2}}}$

$\rightarrow \sqrt{\frac{n-3+\frac{1}{2}}{n-2+\frac{1}{2}} \frac{n-1.5}{n}} = \boxed{\sqrt{\frac{n-2.5}{n}}}$ quite simple, if true!

.29

$\sqrt{\frac{n-2.5}{n}} \approx 1 - \frac{1.25}{n}$. would $1-\frac{1.25}{n-1}$ be better?

$\left(1-\frac{2.5}{n}\right)^{\frac{1}{2}} = 1 - \frac{1}{2}\frac{2.5}{n} + \frac{1}{6}\left(-\frac{1}{2}\right.$ $\frac{3}{4}$

$(1+\epsilon)^{\frac{1}{2}} \sim 1+\frac{\epsilon}{2}$

$\left(1+\frac{\epsilon}{2}\right)^2 = 1+\epsilon+\frac{\epsilon^2}{4}$

$\left(1+\frac{1.25}{n-1}\right)^2 = 1 + \frac{2.5}{n-1} + \frac{25}{16} \cdot \frac{1}{(n-1)^2}$.

$\frac{2.5}{n-1} = \frac{2.5}{n}\left(\frac{n}{n-1}\right)$ $\frac{1}{1-\frac{1}{n}} \simeq 1+\frac{1}{n}+\frac{1}{n^2}+\frac{1}{n^3}$

$= \frac{2.5}{n}\left(\frac{1}{n}+\frac{1}{n^2}+\frac{1}{n^3}\right) + \frac{25}{16}\left(\frac{1}{n-1}\right)^2$.

3.22.97 TM:

The formula $\int_{-\infty}^{+\infty} \frac{dx}{(x^2+c^2)^n} = \frac{2n-3!!}{2n-2!!} \cdot \frac{\pi}{c^{2n-1}}$ is true only if n is an integer

If it's a $\frac{1}{2}$ integer (& not an integer), the formula is wrong by a factor $\frac{\pi}{2}$ (!).

~~[illegible]~~ $\underline{\underline{\text{Now}}}$: $\int_0^{\frac{\pi}{2}} \sin^n x\, dx = \int_0^{\frac{\pi}{2}} \cos^n x\, dx$

$= \frac{n-1!!}{n!!} \cdot \frac{\pi}{2}$ if n is ~~[illegible]~~ even int

$=$ " (1) " n " odd integer.

$= \frac{1}{2}\sqrt{\pi} \frac{\Gamma(\frac{n}{2}+\frac{1}{2})}{\Gamma(\frac{n}{2}+1)}$ ) in all cases.

} so, prob'y

$\int_{-\infty}^{+\infty} \frac{dy}{(x^2+c^2)^n} = \frac{\pi}{c^{2n-1}} \sqrt{\pi} \frac{\Gamma(\frac{n}{2}+\frac{1}{2})}{\Gamma(\frac{n}{2}+1)}$

for ~~[illegible]~~ even and odd n.

& prob'y $\underline{\text{any}}$ n.

say $\int_{-\infty}^{+\infty} \frac{dy}{(x^2+1)^n} = g(n)$ : $g(n+1) = \frac{2n-1}{2n} = (1-\frac{1}{2n}) g(n)$.

for $\underline{\underline{\text{integral}}}$ n. $= (\frac{2n-1}{2n}) g(n)$.

Most likely: My approx formula for $g(n) \approx \frac{\sqrt{\pi}}{\sqrt{n-.75}}$ is $\underline{\text{prob'y}}$ good for both ~~[illegible]~~ integral & $\frac{1}{2}$ integral values of n.

$g(1\frac{1}{2})$

---

$\frac{1-3}{2} = -1$

.20 57.37 ff: for n = 3 we want $\int dm\, (\sigma_0^2+m^2)^{\textcircled{\frac{1-n}{2}}}$ which =

$\int_0^{\infty} \frac{dx}{(x^2+1)} = \tan^{-1}x\Big|_0^{\infty} = \frac{\pi}{2}$ $\therefore$ $\boxed{\int_{-\infty}^{+\infty} \frac{dy}{(1+x^2)} = \pi}$ so see also 62.03

$\boxed{n=1}$ compare w. 63.13 $\frac{\sqrt{\pi}}{\sqrt{1-\frac{3}{4}}} = \sqrt{\frac{\pi}{4}} \cdot \sqrt{4\pi} = 2\sqrt{\pi} = 3.5449 = \pi \times 1.1284$ (see 63.17)

13% off. ($\underline{\underline{\text{high}}}$)

This is about as bad as it gets, how. mult or divide by $\frac{\pi}{2}$.

$\int_{-\infty}^{\infty} \frac{dx}{(x^2+1)^{\frac{3}{2}}} = \frac{0!!}{1!!} \cdot \pi$ ← $\underline{\underline{\text{no!}}}$ | $\int_{-\infty}^{\infty} \frac{dx}{(x^2+1)} = \pi$ also!

This is $\underline{\text{impos!!}}$ $(x^2+1)^{\frac{3}{2}}$ is ~~[illegible]~~ always >> $(x^2+1)$

$\boxed{n=1.5}$ so maybe $= \pi \cdot \frac{2}{\pi} = 2$. $\frac{\sqrt{\pi}}{\sqrt{1.5-.75}} = \sqrt{\frac{\pi}{.75}} = 2.0466$

12% high if .75 → .7

only $\underline{\underline{2\%}}$ off $\underline{\text{high}}$

$\boxed{n=2}$ $\int \frac{1}{(p^2)^2} = \frac{1!!}{2!!} \cdot \pi = \frac{\pi}{2}$ $\sqrt{\frac{\pi}{2-.75}} = \sqrt{\frac{\pi}{1.25}} \cdot \frac{2}{\pi} = \sqrt{\frac{4}{\pi}}$

1% off ($\underline{\text{high}}$)

~~[illegible]~~

1% low via "2 − .7" rather than 2 − .75

$\sqrt{\pi}$ v.s. $\frac{\pi}{2}$

$\frac{\sqrt{\pi}}{\pi} \cdot 2$

-1!! seems to be 1

on the other hand, it = $\phi$ from $\underline{\underline{63.05}}$

so:

$\frac{-1!!}{0!!} = \frac{1}{\sqrt{\pi}} \frac{(-\frac{1}{2})!}{0!}$

$-\frac{1}{2}! = \sqrt{\pi}$

so $\boxed{\text{if } 0!! = 1}$

$\boxed{-1!! = 1}$ also!

$\sqrt{\frac{\pi}{1.25}} \cdot \frac{2}{\pi} = \sqrt{\frac{\pi \cdot 4}{\pi^2\, 1.25}}$

$= \sqrt{\frac{4}{\pi \cdot 1.25}} = \sqrt{\frac{16}{5\pi}}$

$= 4\sqrt{\frac{1}{5\pi}}$

$\textcircled{.73}$

$\underline{-.73}$ instead of $\underline{-.75}$

gives 8.5% error on n = 1 & 8.8% error for n = 1.5

& .128 " " n = 2

Fooling w. .X to give better fit at n = 1. & 1.5 seems pointless, since I will prob'ly never use n that small.

3.23.97 TM $\boxed{\text{Review}}$ 65

.01: 58.40: The result of t. review pp 57 – 58 is ≃ correct, but t. formula of 58.25 is a bit off.

.02 Should be expected value of $\frac{M}{\sigma}$ is $\boxed{\frac{M_{ob}}{\sigma_{ob}} \sqrt{\frac{n-2.5}{n}}}$

The work for this correction is 61.25 — 63.29:

The error in t. previous work resulted from t. formula

$$\int_{-\infty}^{+\infty} \frac{dx}{(1+x^2)^m} = \frac{2m-3\,!!}{2m-2\,!!}\,\pi$$

being true only if m was an integer. If m was of form $\overset{\text{integer}}{m'}+\frac{1}{2}$, The formula had an extra $\frac{2}{\pi}$ factor!

for " $\frac{1}{2}$ integral m it's $\frac{2m-3\,!!}{2m-2\,!!}\cdot 2$

More generally $\int_{-\infty}^{+\infty} \frac{dy}{(1+x^2)^m} = \sqrt{\pi}\,\frac{(m-\frac{3}{2})!}{(m-1)!}$ is exact for all $m > \frac{1}{2}$
for $m \le \frac{1}{2}$ t. integral diverges.

This means I can use my ~~form~~ good approx for $\frac{x+\frac{1}{2}\,!}{x\,!} \approx \sqrt{x+\frac{3}{4}}$ (good for any x, in, half ints or whatever).

.18 So $\int_{-\infty}^{+\infty} \frac{dy}{(1+x^2)^m} = \frac{\sqrt{\pi}}{\sqrt{m-\frac{3}{4}}}$ This formula is 13% high for m = 1, 2% " " m = $1\frac{1}{2}$, 1% " " m = 2 } 64.20ff.

( instead of $\frac{\sqrt{\pi}}{\sqrt{m-1}}$ )

In 57.38 and 58.19 we got t. factors $\sqrt{\frac{n}{2}-\frac{3}{2}}$ & $\sqrt{\frac{n}{2}-1}$, resp:

t. final formula of 58.25 has their ratio; $\frac{\sqrt{\frac{n}{2}-\frac{3}{2}}}{\sqrt{\frac{n}{2}-1}}$ ; & 65.18 changes this ratio to

$$\sqrt{\frac{\frac{n}{2}-\frac{3}{2}+\frac{1}{4}}{\frac{n}{2}-1+\frac{1}{4}}} = \sqrt{\frac{n-3+\frac{1}{2}}{n-2+\frac{1}{2}}} = \sqrt{\frac{n-2\frac{1}{2}}{n-1\frac{1}{2}}}.$$

In 58.25, the two "$n-1\frac{1}{2}$" factors

$$\sqrt{\frac{n-2.5}{\cancel{n-1.5}}\;\frac{\cancel{n-1.5}}{n}} \rightarrow \sqrt{\frac{n-2.5}{n}}$$

cancel out and we get $\boxed{\frac{M_{ob}}{\sigma_{ob}}\cdot\sqrt{\frac{n-2.5}{n}}}$

)

3.23.97 ·TM:

.d: 61.24:

(SN) Actually, the 65.02 ~~wan~~ shouldn't work for $n=3$ anyway: this is because while

.01 $\int \frac{d\mu}{(\sigma_0^2+\mu^2)^2}$ does converge, $\int \frac{\mu d\mu}{(\sigma_0^2+\mu^2)^2}$ diverges: (the ~~—~~ t. ∞'s cancel out ☹).

~~whatever, we will want $E\frac{\mu^2}{\sigma^2}$ and~~ Well, no: the divergence trouble occurs in both Num. & denom of 57.29 for t. same values of n.

· In t. denom, divergence is for $n=0,1,2$:

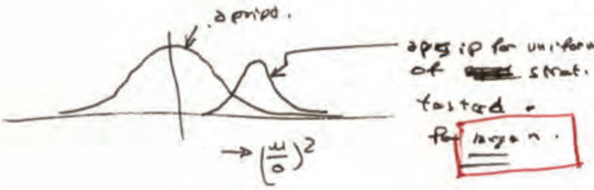
In numerator, when we integrate w $\frac{\mu}{\sigma}$, we increase t exponent in (.01) ~~search~~ by ½ so it converges better, then mult by $\mu$ so it converges worse. The ~~•~~ 2 effects leave t. fact of convergence invariant: so it converges for $n \geq 3$.

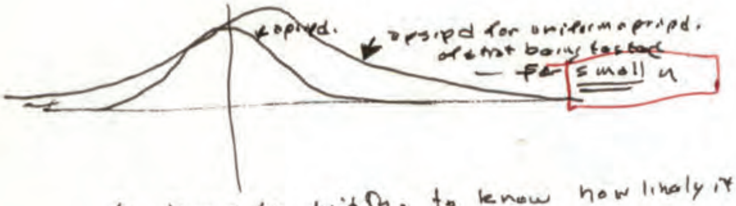
---

Similarly, when we want $E\frac{\mu^2}{\sigma^2}$; the effects of $\mu^2$ & $\frac{1}{\sigma^2}$ on convergence, cancel out, so it converges for $n \geq 3$.

Note, hvr, one must integrate wrt dσ then dμ. If one does it in ~~a~~ inverse order, I don't know what happens! Actually, it may be a much easier integration! Or, on t. other hand, it begins to look about t. same!

.20 ⁝ spec ⁝ 61.24 : Say we decide on an apripd for $\left(\frac{\mu}{\sigma}\right)^2_{\text{for bet}}$: we could then find the $E\left(\frac{\mu}{\sigma}\right)^2$ of a particular strat. If t. "true" $\left(\frac{\mu}{\sigma}\right)^2$ were largish (say >> the s.D. of t. apripd for $\left.\frac{\mu}{\sigma}\right)^2$) then, as n↑, the $E\left(\frac{\mu}{\sigma}\right)^2$ will ~~d~~ ↑, because ~~its~~ its d.f. gets narrower

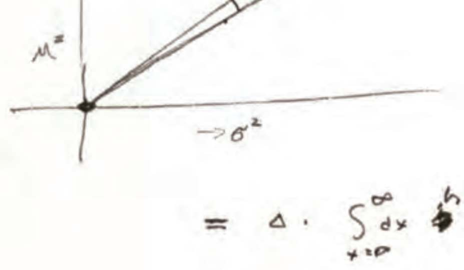




While $E\left(\frac{\mu}{\sigma}\right)^2$ is of importance, we are also much interested in t. d.f. of $\left.\frac{\mu}{\sigma}\right)^2$ — ~~its — the~~ t. width of this D.f. —

The reason is: If we are selecting t. "Best" of a bunch of strats, we need to know t. width, to know how likely it is that our "Best" is spurious.

One approach: as a func of $\mu^2$ & $\sigma^2$ plot in z direction, t. apsip of density at $\mu^2, \sigma^2$. as a function of [Si]. Then integrate in t. narrow angle Δθ.



Say $f(\mu^2,\sigma^2)$ is t. apsip. $k < \frac{\mu^2}{\sigma^2} < k+\Delta$

$\sigma^2 = x$); $\mu^2 = kx$ to ~~kx+Δx~~ ~~then~~

$(k+\Delta)x$

$= kx + \Delta$ ~~Kx~~ $+\Delta x$

Then we want $\int_{x=0}^{\infty} h(\mu^2, \overset{x}{\overset{\shortparallel}{\sigma^2}}) \cdot \int^{\cdot \Delta x} dx$

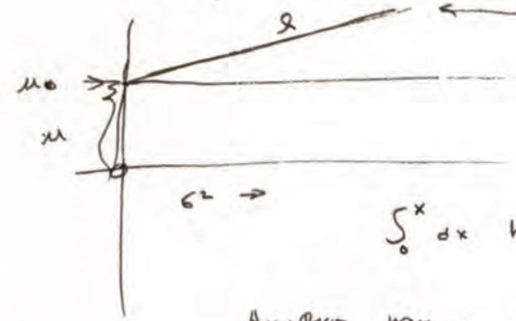
$h(kx, x)$

$\int_{x=0}^{\infty} dx \left( h(kx,x) \cdot x \cdot \Delta \right)$

$= \Delta \cdot \int_{x=0}^{\infty} dx \; h(kx,x) \cdot x$ ~~—~~ we want this as a function of $\left(k = \frac{\mu^2}{\sigma^2}\right)$.

$h(\mu^2, \sigma^2) = 57.20$ : not so easy to do, since in 57.20, $\mu^2$ isn't expressed as such.

A way out! Integrate along the line "l"



Essential part of h:

$$h(\mu, \sigma^2) = \frac{1}{\sigma^n} e^{-\frac{n(\sigma_0^2+(\mu_0-\mu)^2)}{2\sigma^2}}$$

That would make the integration easy now — but it's not what I want!

$$\int_0^x dx\ h\ (\mu_0 + \cdots$$

Another way, let $\sigma$ be the integration param. $\mu = k\sigma$, but we have to mult. the integrand by something because of the $\frac{\sigma^2}{\mu^2} \leftrightarrow \frac{\mu}{\sigma}$ conversion. This may not change form of integrand, hwr.

---

→ Well, if I had the d.f. for $\frac{\mu}{\sigma}$, I could get the d.f. for $(\frac{\mu}{\sigma})^2$ from it directly, so get the d.f. of $\frac{\mu}{\sigma}$.

---

constant factor

.20 $$h(n, \mu_0) = \int_{\sigma=0}^{\infty} \sigma \frac{d\sigma}{\sigma^n} e^{-\frac{n(\sigma_0^2+(\mu_0-k\sigma)^2)}{2\sigma^2}} = \int_0^\infty \frac{d\sigma}{\sigma^{n-1}} e^{-\frac{n(\sigma_0^2+\mu_0^2) - 2n\mu_0\sigma k \boxed{+nk^2\sigma^2}}{2\sigma^2}}$$

Hwr. this integrand, as 13.13 rather simpler.

See 78.15 for a practical resoln. of this problem! just set $\mu_0 = 0$ (!) →

No! Hwr, for $k \geq \sqrt{2}$ the integral diverges at hy $\sigma$. $k=2$ means $(\frac{\mu}{\sigma})^2 > 2$: It's unlikely that we'll ever get near that region! $\frac{\mu}{\sigma}$ is for a single beat!

for any k, the $\int$ converges for $n > 3$ or 4, say.

$$\int_{\sigma=0}^{\infty} d\sigma\ \sigma^{-n}\ e^{-\frac{A}{\sigma^2} + \frac{z}{\sigma}}$$ as a funct of z. ($0 < z$)

$x = \frac{z}{y}$

$$\int_0^\infty dx\ x^{-n}\ e^{-\frac{A}{x^2} + \frac{z}{x}}$$

Say we let $y = \frac{z}{x}$ That may do it!

$dy = -\frac{z}{x^2}dx$ We may end up w.

$$\int_0^\infty y^m\ e^{-ax^2+bx}$$ which is simply a moment of an offset Gaussian & is probably well known.

$dx = -\frac{x^2 dy}{z}$ $x = \frac{z}{y}$

$dx = \frac{z^2}{y^2 \cdot z}dy = \frac{z}{y^2}dy$

$$-\int_\infty^0 \frac{z}{y^2} \cdot \frac{y^n}{z^n}\ e^{-\frac{Ay^2}{z^2}+y}\ dy$$

$$= z^{1-n} \int_0^\infty y^{n-2}\ e^{-\frac{Ay^2}{z^2}+y}\ dy$$

So we need $\int_0^\infty y^m\ e^{-\frac{(y-\alpha)^2}{2\beta^2}}$ which is obtainable from $\int_0^\infty y^m\ e^{-\frac{(y-\alpha)^2}{2}}\ dy$

=

$z = y - \alpha$
$Y = z + \alpha$

I think I usually want $\alpha > 0$.

$$g(m, \alpha) = \int_{\alpha}^{\infty} (z+\alpha)^m\ e^{-\frac{z^2}{2}}\ dz$$

40 $$\frac{\partial g(m,\alpha)}{\partial \alpha} = +\int_{+\alpha}^{\infty} m\,(y+\alpha)^{m-1}\ e^{-\frac{y^2}{2}}\ dy = +\ m\,g(m-1, \alpha)$$

3.24.97 · TM

$\rho(m,\alpha) = \int_0^\infty dx\ x^m e^{-\frac{(x-\alpha)^2}{2}} dx$ : $\rho(0,\alpha) = \int_0^\infty e^{-\frac{(x-\alpha)^2}{2}} dx$ is like erf(α)

$$\frac{\partial}{\partial\alpha}\rho(m,\alpha) = \int_0^\infty dx\ x^m \left(+(x-\alpha)\right) e^{-\frac{(x-\alpha)^2}{2}}$$

$$= \rho(m+1,\alpha)\ \overset{+}{-}\ \alpha\rho(m,\alpha)$$ — which seems to differ from 67.40 ! → 69.09

.03

condon, odrshaw p. 1-141 col 1 seems relevant ! / Relation of moments about origin & "central moments"
some expansion of moments see ibid p 1-140 col 2.

I think $X_1^{a_1} X_2^{a_2} \cdots\cdots X_n^{a_n}$ means $X_1^{a_1} \cdot X_2^{a_2} \cdots X_n^{a_n}$.

For this integral, the d.f. that I'm obtaining the moments of is $e^{-\frac{(x-\alpha)^2}{2}}$ / for x>0 but is 0 for x<0

It is normed so $\int_0^\infty f(x)dx = \int_{-\infty}^{+\infty} f(x) = 1$.

in cond. & od; m is the mean of the df. I think.

p 1-141 col 1 eq 12.72, 12.73 : In 12.72: shouldn't $\binom{n}{j}$ be binomial coefs: so they are sym. in j → n-j ? ← (see ibid p I-10 col 1.)

$$\mu_2 = \binom{2}{0} -1^0 \alpha_0 m^2 + \binom{2}{1} -1^1 m\alpha_1 + \binom{2}{2} -1^2 m^0\alpha_2$$

Maybe some terms are zero.

$$\overset{+}{1}\quad \underset{\overset{\shortparallel}{1}}{\alpha_0} m^2 \ \overset{-}{-}\ 2\ m\cdot \underset{\overset{\shortparallel}{\alpha_1}}{m}\ \overset{+}{+}\ 1\cdot\alpha_2 \qquad 1 + 6 + 1$$

$$= \alpha_2 - m^2$$ so 2 terms partly cancel.

$$\mu_3 = \overset{3}{-}\alpha_0 m^3 + 3\alpha_1 m^2 - 3\alpha_2 m + \alpha_3 = +2m^3 - 3\alpha_2 m + \alpha_3$$

$\binom{3}{0}$ 1 3 3 1

so its o.k. In general, the terms in 12.72 are the result of cancelation: I think that 2 terms always combine: i.e. $\alpha_0 m^n$ and $n\alpha_1 m^{n-1}$ giving $\pm(n-1) m^n$.
So in 12.72, remove last term on r.h.s.: the other terms are the first "n-1" binomial coef terms.

$$\begin{array}{c|c|c} n & & \\ 2 & 1 & 2\ 1 \\ 3 & 1-3 & 3\ 1 \\ 4 & \underbrace{1-4+6}_{\text{as is.}} & \underbrace{4\ 1}_{\text{partly cancel}} \end{array}$$

So, O.K. :

I need m, the first moment which is $m(\alpha) = \int_0^\infty x\ e^{-\frac{(x-\alpha)^2}{2}} dx$ which is sort of directly integrable.

$$\frac{d}{dx} e^{\frac{-(x-\alpha)^2}{2}} = -(x-\alpha) e^{-\frac{(x-\alpha)^2}{2}}$$

$$= \int_0^\infty (x-\alpha) e^{-(\ )} dx + \alpha\int_0^\infty e^{-\frac{(x-\alpha)^2}{2}} d$$

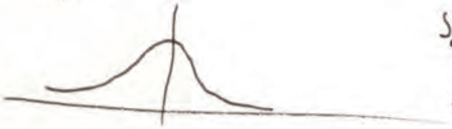
$$\propto e^{-\frac{(x-\alpha)^2}{2}}\Big|_0^\infty + \alpha \sim \mathrm{erf}\ \alpha$$

$\therefore$ like $e^{\frac{\alpha^2}{2}} + \alpha\,\mathrm{erf}\alpha$

3.29.97 TM

So I can compute $m(\alpha)$:

next



$\int_0^\infty x^n e^{\frac{-x^2}{2}} dx$ is probably well known.

$\int_0^\alpha x^n e^{\frac{-x^2}{2}} dx$ I'm no so sure how well known.

anyway cond, Od p 1-1A1 col I eq. 12.73 tells how to get Moments from central moments:

To get $m^{\text{th}}$ moments, I need all central moments up to & including m.

---

.09 try setting 67.40 = 68.03: $\frac{\partial}{\partial \alpha} g(m,\alpha) = +m\, g(m-1,\alpha) = -g(m+1,\alpha) + \alpha\, g(m,\alpha)$

so $g(m+1,\alpha) + m\, g(m-1,\alpha) = \alpha\, g(m,\alpha)$ or $g(m+1,\alpha) = +\alpha\, g(m,\alpha) - m\, g(m-1,\alpha)$.

If this is true, we have $g(m,\alpha)$ for all m, if we have

.12 $\boxed{g(m+1,\alpha) = \alpha\, g(m,\alpha) - m\, g(m-1,\alpha)}$

$g(0,\alpha)$, $g(1,\alpha)$ (or any 2 consecutive (maybe not necly consecutive) $g(m,\alpha)$)

$$\boxed{g(0,\alpha)} = \int_0^\infty e^{-\frac{(x-\alpha)^2}{2}} dx$$

$$= \int_{-\alpha}^\infty e^{-\frac{x^2}{2}} dx \sim \boxed{\text{erf}(\alpha)}\text{''}$$

$m\, g(m-1,\alpha) = -g(m+1,\alpha) + \alpha\, g(m,\alpha)$

$g(m+1,\alpha) = \alpha g(m,\alpha) - m g(m-1,\alpha)$

α has to be > 0; since m > 0: If m > 0 & m < 0

Then since $g(0,\alpha)$ & $g(1,\alpha)$ are > 0, $g(2,\alpha)$ would be < 0.

Which, ~~[illegible]~~ is impossl, since integrands are all always > 0.

$$g(1,\alpha) = \int_0^\infty (x-\alpha) e^{-\frac{(x-\alpha)^2}{2}} dx$$

$$= \underbrace{\int_0^\infty x\, e^{-\frac{(x-\alpha)^2}{2}} dx} - \alpha \underbrace{\int_0^\infty e^{-\frac{(x-\alpha)^2}{2}} dx}_{\sim \text{''erf}(\alpha)}$$

cancel?

$$\int_{-\alpha}^\infty (x+\alpha) e^{-\frac{x^2}{2}} dx = \underbrace{\int_{-\alpha}^\infty x\, e^{-\frac{x^2}{2}}}_{e^{-\frac{\alpha^2}{2}}} + \alpha \underbrace{\int_{-\alpha}^\alpha e^{-\frac{x^2}{2}}}_{\sim \text{''erf}(\alpha)\text{''} = g(0,\alpha)}$$

so $g(0,\alpha)$ & $g(1,\alpha)$ are expressable in terms of ~ erf(α) & $e^{-\frac{\alpha^2}{2}}$.

I still don't know the sign of $\underline{\underline{\alpha}}$.

$g_{m+1} = \alpha g_m - m g_{m-1}$ ; $g_m = A^m$

$A^{m+2} = \alpha A^{m+1} - m A^m$

$A^2 = \alpha A - m$ $A^2 - \alpha A + m = 0$

.33 $A = \frac{\alpha \pm \sqrt{\alpha^2 - 4m}}{2}$ for fixd $\alpha^2$, 4m will eventually be > $\alpha^2$ so A is complex.

$$\boxed{g(1,\alpha)} = \int_0^\infty (x-\alpha)\, e^{-\frac{(x-\alpha)^2}{2}}$$

$$= \cdot e^{-\frac{(x-\alpha)^2}{2}} \Big|_0^\infty$$

$$= \boxed{+ e^{-\frac{\alpha^2}{2}}}$$

$\frac{d}{dx} e^{-\frac{(x-\alpha)^2}{2}} = (x-\alpha) e^{-\frac{(x-\alpha)^2}{2}}$

Equ. (.12R) seems unlikely to have g > 0 for all large m!

3.24.97 TM

70

I got the signs in .03 wrong!

should be $\overset{\frac{\partial}{\partial\alpha}g(m,\ )}{6(m+1,\ )} - \alpha g(m\cancel{\beta},\ )\ \Big| = m g_{m-1}$

.03 so $\boxed{g_{m+1} = \alpha g_m + m g_{m-1}}$ — much more reasonable!

from 69.33 $A^2 = \alpha A + m \qquad A^2 - \alpha A - m = 0$

$A = \frac{\alpha \pm \sqrt{\alpha^2 + 4m}}{2}$ so one positive A, one negative.

Still, if .03 is correct, $g_m$ will be always > 0 since $g_0$ & $g_1$ are > 0.

✓

Well, the best thing to do is Evaluate 67.20 by this recursion method.

For 67.20, $\frac{\partial h(n, \mu_0)}{\partial \mu_0} = \int_0^\infty 6 \frac{d\sigma}{\sigma^n}\left(-\frac{n(\mu_0 - k\sigma)}{\sigma^2}\right) e^{-\frac{n(\sigma_0^2 + (\mu_0 - k\sigma)^2)}{2\sigma^2}}$

$= -n\mu_0\, h(n+2, \mu_0) + \underline{nk}\, h(n+1, \mu_0)$

$\frac{\partial}{\partial\mu_0} \quad -\frac{n(\sigma_0^2 + (\mu_0 - k\sigma)^2)}{2\sigma^2}$

$= \frac{-n}{2\sigma^2}\ \underbrace{\frac{\partial}{\partial\mu_0}(\mu_0 - k\sigma)^2}_{2(\mu_0 - k\sigma)}$

$= \frac{-n\mu_0}{\sigma^2} + \frac{nk}{\sigma}$

---

change exponent to $(k\sigma - \mu_0)^2$

~~$\mu_0 - k\sigma = z \ ; \ k\sigma = \mu_0 - z$~~

$k\sigma - \mu_0 = z \qquad k\sigma = z + \mu_0$

$\sigma = \frac{z}{k} + \frac{\mu_0}{k}$ ;

$d\sigma = \frac{dz}{k}$

$= \int_{-\mu_0}^{\infty} \frac{1}{k} \frac{dz}{\left(\frac{z}{k} + \frac{\mu_0}{k}\right)^{n-1}} \cdot e^{-\frac{n(\sigma_0^2 + z^2)}{2\left(\frac{z}{k} + \frac{\mu_0}{k}\right)^2}}$

This isn't working out well!

C ✓
E ✓
6781
8621
SE (1040) ✓
Selfamp

ES est. tax.
ē Pub 505 income

---

Woops! .03 is not a linear eq. in m ← non-constant coeffs. (But linear). Simpler.

~~$g_{m+1} = \alpha g$~~ $\boxed{g_{m+1} = \alpha g_m + m g_{m-1}}$

So this is not a simple eq. !

~~If $\alpha = 0$ $g_m = (m-1)!$ ($g_m = (m-1)$~~

However, it is linear: If $x_m$ & $y_m$ are solns,

$x_m + y_m$ is a soln.

in $x_{m+1} = m x_m$

$x_m = (m-1)!$ ; if $x_1 = 1$

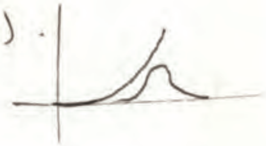
$x_1 = 1$

$x_2 = 1 \cdot x_m$ ; $x_3 = 2$

$x_2 = 1$
$x_3 = 2$
$x_4 = 6$

Since m is usually large, $g_n$ increases rather rapidly w. m:

& even if α had a large neg value, $g_m$ would for large m be $\uparrow \sim \boxed{m!}$ (!).



This surprizes me; I would think that $g_n$ would not ↑ very much as n ↑ !

look at $g_n$ if $g_1 = 0$ ; $g_0 = 1$

$g_2 = 1, \quad g_3 = 0$

| | G |
|---|---|
| 0 | 1 |
| 1 | 0 |
| 2 | 1 |
| 3 | $\alpha$ |
| 4 | $\alpha^2 + 3$ |
| 5 | $\alpha^3 + 3\alpha + 4\alpha = \alpha^3 + 7\alpha$ |
| 6 | ~~$\alpha^4 + 3\alpha^2 +$~~ $\alpha^4 + 7\alpha^2 + 5\alpha^2 + 15 = \alpha^4 + 12\alpha^2 + 15$ |
| 7 | |

If $\alpha = 0 \quad g_{m+1} = m g_{m-1}$ ~~Say $g_0 = 1$~~

Say $g_1 = 1 \quad g_3 = 2, \ g_5 = 2 \cdot 4, \ g_7 = 2 \cdot 4 \cdot 6 = (7-1)!!$

.35 similarly if $g_0 = 1, \ g_2 = 1, \ g_4 = 1 \cdot 3, \ g_6 = 1 \cdot 3 \cdot 5 = (6-1)!!$

3.24.97 TM

71

to P(m): $\boxed{p_{m+1} = \alpha g_m + m g_{m-1}}$  use a circular file: A 1 $\boxed{g_m}$ B 2 $\boxed{g_{m+1}}$ state 1  $\boxed{g_{m-1}}$ $\boxed{g_m}$ state 2

2 states state 1 $\boxed{B = \alpha A + B m : m=m+1}$

" 2 $\boxed{A = \alpha B + A m : m=m+1}$

TM7.71.B05

Start w. A=1, B=0
Then A=0, B=1

```
  A =    : B = :    m = 1
5 B = AL*A + B*m : m=m+1 : print m, B
                            print m, A : if m<40 then 5
  A = AL+B + A*m : m=m+1 :
```

P(m) seems to work reasonably

---

~~Per~~ m=41: 1.3 E+25  A=0 B=1 AL=.5 ) for AL>0 its faster than m!!
m=41: 2.1 E25  A=1 B=0 AL=.5 ) for AL=0 its m-1!!
~~m=apho~~ 2.55 E24  A=1 B=0 AL=0

---

m=41 3.27 E26 " " AL= -1  output alternates in sign!
A is always >0 ) or contrary wise
B " " <0

---

Γ?

[redacted] wrong eq:
For α=0, A=1; B=0

| m | # |
|---|---|
| 2 | 0 |
| 3 | 2 |
| 4 | 0 |
| 5 | 8 |

If α is -.01 alternate values are negative.

---

If both A & B start at 1 w. α = -.01  $g_m$ remains >0 (!).

[redacted]

---

$$\int_0^\infty x^n e^{-\frac{x^2}{2}} dx = -\int_0^\infty x^{n-1}\left(-x e^{-\frac{x^2}{2}}\right)dx = \int x^{n-1}\frac{d}{dx}f(x)\,dy$$

$$\int\left(x^{n-1}\,df(x) + (n-1)x^{n-2}f(x)\right) = x^{n-1}f(x)$$ so a recursion relation.

fir p 337 $\int_0^\infty x^{2n} e^{-px^2} = \frac{2n-1!!}{2(2p)^n} = \frac{2n-1!!}{2}$ for $p=\frac{1}{2}$

$\rightarrow = \frac{n-1!!}{2}$ for even n ; which is observed in 70.35

I don't see why it should ↑ very rapidly w. n — but maybe ↑ in n is strong enuf so that t. effect of $e^{-\frac{x^2}{2}}$ is dealt w. strongly!

~~AL~~ V
ALTV

---

ctrl D dialer

Cont stuff dialer

/4 → 43/50

/1 132 col.

---

ALT2 stuff
Baron?N.

or F3

Debug < Dialer. Pa

Simtel\List 90H\Dialer.
Pat

3.25.87 TM

Well, using t. method up to 71.40 I probably could get t. expected d.f., but it would be an enormous amt. of work & fraught w. error. I mite do some numerical runs to ~~[illegible]~~ conform my conjectures on what t. soln. looks like

Conjectures:

.08 1) t. mean of t. $\left(\frac{\mu}{\sigma}\right)^2$ d.f. is about $\left(\frac{\mu_0}{\sigma_0}\right)^2$.

.09 2) Its Varc is about $\frac{1}{n}$ (t. no. of data pts). || assumes uncorrelated bets. → see 78.20! its $\boxed{\text{exactly } \frac{1}{n}}$ for small $\frac{\mu}{\sigma}$ (which is what we want, because its $\frac{\mu}{\sigma}$ per bet that we're after

3) If I assume $\left(\frac{\mu}{\sigma}\right)^2$ has a Gaussn. D.f. & t. apripd of $\left(\frac{\mu}{\sigma}\right)^2$ is also Gaussian, but centered about zero; Then its easy to ~~combine~~ multiply these 2 d.f.s. to get a new Gaussn.

$$(\mu_1,\sigma_1^2);(\mu_2,\sigma_2^2)\rightarrow\overbrace{\left(\frac{\mu_1}{\sigma_1^2}+\frac{\mu_2}{\sigma_2^2}\right)\Big/\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right)}^{\text{new }\mu}\;;\;\overset{\text{new }\sigma^2}{\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right)^{-1}}.$$

The idea is that w. each $\sigma_i^2$ we associate a ssz $\propto \frac{1}{\sigma^2}$, so t. results follow easily.

4) ~~[illegible]~~ ONE Q is: What ssize ($\sigma^2$) do we assign to t. apripd? We assign ssz (n) to t. [Si] data ~~[illegible]~~ ~~[illegible]~~ If we assign Varc = $\sigma^2_{ap}$ to t. aprip $\frac{\mu}{\sigma}$ d.f. ~~[illegible]~~ (in "units" of 1) Then its equivt ssz is $\frac{1}{\sigma_{ap}^2}$.

Say the typical strat has a yield of ~~[illegible]~~ ~~[illegible]~~ 10%/yr (kelly betting): so $\frac{1}{2}\frac{\mu^2}{\sigma^2}\Big/_{yr.}=.1$

or $\frac{\mu^2}{\sigma^2}\Big/_{yr}=.2$ or $\frac{\mu}{\sigma}\Big/_{yr.}\sim\sqrt{.2}=.44$ We can use this as a ruff "Ball Park" estimate.

After we've made lots of studys of strats w. hy n, we ~~can~~ will have a better idea of what t. d.f. is.

.25

.26 5) The $\sigma^2$ of $\left(\frac{\mu}{\sigma}\right)^2$ enters in follg. way: If we have lots of strats, ~~s~~ each w. its own Gore $\left(\frac{\mu}{\sigma}\right)^2$ & a varc for each gore. —— Then we "pick t. best" Gore. If there are many uncorr. strats to pick from — all w. ~~[illegible]~~ about t. same gore, & ~~[illegible]~~ about same S.D. = $\sigma_{emp}$ for that Gore, Then t. peak will tend to be about $\sigma_{emp}$ hyer than its true value.

If we have only 1 or only 2 few strats anywhere near t. hy peak, Then t. hy pard has a better chance of being near an "unbiased estimate" of t. Gore.

~~A~~ In general, a conservative estimate of the Gore of a strat that was picked as a peak, will be that peak minus its S.D. If ~~there are~~ t. ~~[illegible]~~ "best strat" was ~~[illegible]~~ chosen out of $\left(\begin{matrix}\text{few}\\ \text{many}\end{matrix}\right)$ strats, t. correction will be $\left(\begin{matrix}\text{smaller}\\ \text{greater}\end{matrix}\right)$ than t. S.D.

( I once did a study of Re/ expected increase in max of a set of ~~#~~ indep normal vars, as a funct of t. no. of vars ) —— I don't know where it is, hwr. → 73.01

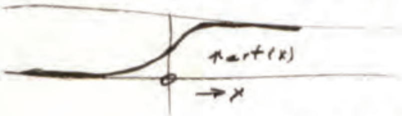
.30 The results are on 75.27 : A Big problem is 74.18 : How to estimate "how many" strats were around t. peak one chose, due to corrns betw. strats. → 74.22 discusses ~~a~~ theoretical Goal. Soln.

3.25.97 TM

SPac:
.01 : 72.36 : Say we have n random vars of zero mean & var = 1. What is p.d. that t. max of t. n is X ?

Or. What is t. expected value of t. max ?

define erf(X) as t. probty that one of t. vars will be < X : $erf(-\infty)=0$ ; $erf(+\infty)=1$, $erf(\phi)=\frac{1}{2}$.



$\therefore\ erf(x) = \int_{-\infty}^{x} f(x)\,dx \qquad f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ .

probty that peak is ~~≠~~ x is ∝ $n\cdot(erf(x))^{n-1}\cdot f(x)$ . $\left(= \frac{d}{dx}\,(erf(x))^n\right)$

.07 E(n) ≡ expected value = $\int_{-\infty}^{+\infty} x\cdot n\cdot(erf(x))^{n-1}\cdot f(x)\,dx$ .

perhaps get a version wo t. ∫ vix int by parts. Mite be possl. for normzn factor,

.10 $\int_{-\infty}^{+\infty} n\,(erf(x))^{n-1}\cdot f(x)\,dx \qquad \frac{d}{dx}\underbrace{(erf(x))^n} = n\,f(x)(erf(x))^{n-1}$.

So t. normzn factor is $\underline{\underline{1}}$.

$g(x)\equiv$

$$\int_{-\infty}^{+\infty} x\cdot\frac{d}{dx}g(x)\,dx + \underbrace{\int_{-\infty}^{+\infty} g(x)\cdot\frac{dx}{dx}\,dx}_{=\infty} = \left(x\,g(x)\right)\Big|_{-\infty}^{+\infty} = \infty\,(!) .$$

Well, I could simply do it numerically — but its certainly better to have an analytic Expressn!

I already have a pgm that may do something like $\int_0^x e^{-\frac{x^2}{2}}dx$ .
I could decide on $n_{max}$ in advance, then have it do t. integration for all ~~[illegible]~~ values up to n, simultly. — would take much less time than doing them sequentially .

Strange! As n↑ $erf(x)^n$ gets smaller for larger x. | so $\int x\,erf(x)^{n-1}f(x)\,dx$ should ↓ !

on t. other hand, how come $\underbrace{\int_{-\infty}^{+\infty} n\,erf(x)^{n-1}f(x)\,dx}_{\neq}$ $\overset{=1}{\text{is}}$ indep of n!

oh! $\int erf(x)^{n-1}f(x) = \frac{1}{n}$

.25 If I do it numerically, I can test t. pgm by having it do ~~■~~ : t. pgm will be same as actual pgm but (X ← 1.) .

.07 gives E(1) = 0 since $\int x\ erf(x)^0 f(x)\,dx = 0$ .
as it should.

I could just do a quick & dirty "Riemann" ∫. Do it for an adjustable Δx, so I can quickly test t. pgm for large Δx then do ~~■~~ long runs w. small Δx.
Also adjustable "~~[illegible]~~ "Max n".
shouldn't take long w. Δx = .01, ~~x = [illegible]~~ x = -5 | .01 | +5 (~5 ~~■~~ 6)
start out w. ~~[illegible]~~ x from -3σ to +3σ.
Test t. pgm on (.10) first. = (.25)
The starting value of erf(x) at -3σ or -5σ can be obtained from some asymptotic formula
um B.S. Or. I can just look it up. in a table ~~known~~ ! ~~Bc~~ its a value you insert befor running t. pgm.
say erf for -5, -4, -3, ~~■~~

∴ Pgm has ~~[illegible]~~ A = .00015 (TLU for erf -5
B = .001 " " " -4
C = .01 " " -3
we set starting value for integration at A, B, or C.

3.25.97 TM

74

$f = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

$f'(x) = -x\, e^{-\frac{x^2}{2}}$

$$\underbrace{\int d\left(n\, \text{erf}^{n-1} f(x)\right)}_{\text{zero?}} = \int n(n-1)\, \text{erf}^{n-2} f^2(x) + \underbrace{\int n\, \text{erf}^{n-1} f'(x)}_{\text{want } \int \text{ of this}}$$

$$d\ \ n\, x\, \text{erf}^{n-1} f =$$

$$-\int n\, \text{erf}^{n-1} f'\, dx \quad + \quad \int n(n-1)\text{erf}^{n-2} f^2$$

$$\text{so}\quad \underbrace{\int n\, \text{erf}^{n-1} f'(x)}_{=\text{ what I want}} = \int n(n-1)\, \text{erf}^{n-2} f^2(x)\cdot dx$$

---

Well, o.k. say I <u>know</u> E(n):

---

Another <u>very easy</u> (& not so accurate) way would be straight monte carlo. I have this pgm that generates Gaussian vars. Just take the peak of n of them, again and again. Maybe ~~reset~~ reseed random ~~nos~~ every 10,000 times by output from typing timer, so I type 10 k characters: this gives, say 10 k random nos, that I use as 2° seeds.
Another way is to ~~do~~ use that big "1 time pad" pgm to put random nos from zip files by xoring many of them!

---

.18 ~~Another~~ → A <u>Dif'lt Q</u>: To estimate $\boxed{\text{"how many"}}$ "uncorrelated" strategies are in the pool one picked the peak of!
There is also the case of an ∞ of strats (like in linear regressn) in which one picks the peak. In this case, Maxm may tell what the ~~expected~~ unbiased expected error is.
Say one has 10 strats: look at vare of "sum" of strats to get idea of how correlated they are. ($\boxed{\text{A formal Soln}}$: We assume that <u>joint</u> p.d. for all strats yields is <u>known</u> — then $\boxed{83.30 - 84.15}$ has <u>good theoretical soln</u>. → optimized
.22

---

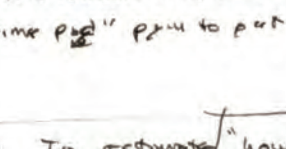
Maybe try <u>Maple</u>: ① to try to do symbolic integrations ② To do numerical integrations

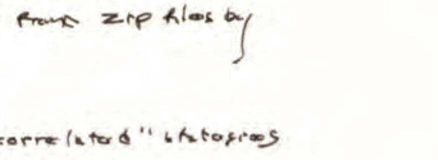
int(4(erf(x))^④ * exp(-x^2)), -infinity..infinity)/(Pi)^(1/2) → $\frac{1}{5}$

$\int_0$

If I use erf + ½ I got $1\frac{61}{80}$ for result! I expect $\frac{1}{5}$.

erf(x) → $\boxed{\frac{1}{2}\text{erf}() + \frac{1}{2}}$ ← correct.





O.K.: I got $\frac{1}{5}$ as I should

with x* here I got .433242 1579  n=5

with using exponent I got .6994711402  n=2

" " 0 ! !

(-1)°

wrong!

I used (erf(x)/2+½)^n

n=1 ~~putting~~ putting x in ~~=~~ pit a. n=0 gave result ∅ as it should

.36

| | | | |
|---|---|---|---|
| n = 2 | ~~(.1994711402)~~ x2 = $\boxed{.3989422804}$ ~ 9" for calcn. | 40 | 1.5279 |
| 2.5 | .511368 | 60 | ~~[illegible]~~ 1.659977 |
| n = 3 put in mos | .5984134206 | 80 | 1.715988 |
| 4 | .7278783067 | 100 | 1.773136 |
| 5 | .82234006 | 150 | |
| 7 | .9561345 | | |
| 10 | 1.08806 | | |
| 15 | 1.22747 | | |
| 20 | 1.320504 | | |

~~66~~

3.25.97 TM Int n M-1

Evalf/( 10*~~X*~~(erf(x)~~/2~~ + 1/2 )^g * exp(-X*X)/Pi^(1/2), x=-infinity..infinity));

Evalf(Int( ← no space

TMZ-75.ms

This ↗ was saved as $\boxed{\text{TM 7-75.ms}}$ ~~problem~~ c:\maple:V2\bin

One can get it by asking for TM Z5-75.ms. ~~If one types~~

Works ok for up to N=0, but N=150 **Bombs** it!

7" to do n = 4

7" for n=4
13" " n=80

---

Note the results of 74.36 — .40 are for $e^{-x^2} = e^{\frac{-x^2}{2(\frac{1}{\sqrt{2}})^2}}$

I.e. s.d. = $\frac{\sqrt{2}}{2}$ = .707. $\sigma^2 = \frac{1}{2}$.

So the results should be mult by $\sqrt{2}$ for σ = 1 !

This gives ~ 1 for n = 4. ~ 1.4 for n ≈ 10 . 1.86 for n = 20 | 2.15 for n=40

~ 2.5 for n = 100

It would be interesting to know if continued doubling eventually gives ~ same ↑ in S.

.16

| | ρ | σ | Δ | Δ² | Δ³ |
|---|---|---|---|---|---|
| 4 → | 2.5 | .72 | .44 | .06 | → ~1 |
| | 5 | 1.16 | .38 | .06 | .00 |
| | 10 | 1.54 | .32 | .03 | .03 |
| | 20 | 1.86 | .29 | .02 | .01 |
| double( | [redacted] | | | | |
| double( | 40 | 2.15 | .27 | | |
| | 80 | 2.42 | | | |
| | 100 | 2.5 | | | |

---

TM7-752.ms

$\boxed{\text{TM 7-752.ms}}$ Prettyprinting of formula.

The system may pgm about as easy as Basic: It has a simple do, for, od loop structure.

$\boxed{\text{No sign of statement lables !}}$

Also it's interpreted lang.

The "sum" operator simplifies many "for next" loops.

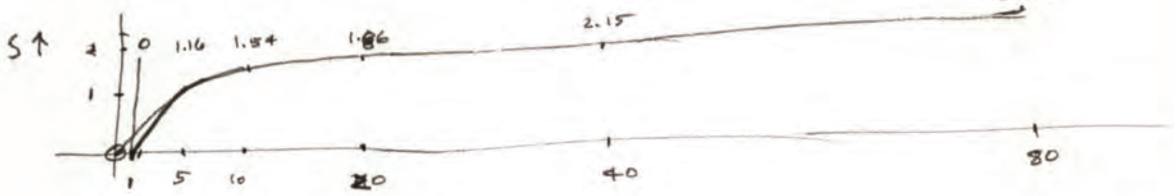
---

.27 O.k., so I evaluated 73.07 w. [redacted] $\sigma^2 = \frac{1}{2}$, $\sigma = \frac{1}{\sqrt{2}} \approx .707$

erf in Maple is $\frac{2}{\sqrt{\pi}} \int_0^x e^{-x^2}$ : its -1 for x→-∞ ; +1 for x→+∞.

To get my erf of 73.07 I did $\frac{1}{2}(\mathrm{Erf}(x)+1)$ : (still wrong σ, tho.).

The values for 73.07 that I got over on 74.36 : Multiply them by $\sqrt{2}$ to get the results for σ = 1 ; See (75.16) ff.



See $\boxed{\text{82.01- 82}\tfrac{1}{2}\text{.40}}$ for n = 1, 2, 4, 8, 16, 64, 128 : giving both mean and σ of "overestimate of σ"

3.26.97: TM

Got Maple to plot3D :

$\sim$ 47.26 $$\left(\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right) \exp -\frac{\sigma_0^2+\mu^2}{2\sigma^2}\right)^n = \left(\sqrt{2\pi}\,\sigma \exp\left(\frac{\sigma_0^2+\mu^2}{2\sigma^2}\right)\right)^{-n}$$ as a funct of $\mu$ & $\sigma$

TM7-76.ms

Say $\sigma_0^2 = 1$. $\quad ((\sqrt{\ } \ \boxed{\exp((1+\mu{:}\mu)/(s\times s))})^{-1}$

A xes ~~framed~~ Boxed

No perspective.

to 5 (

so I plotted $\left(\sqrt{2\pi}\cdot\sigma \exp\left(\frac{1+\mu^2}{\sigma^2}\right)\right)^n$

.09 — .55 is $\mu$ $\frac{1}{2}$ width. 1.1 width $\mu$. ) $n=5$ 2σ ~ $\frac{1}{2}$ value pts.

.10 3.65 to ~ 1 2.65 width σ ) wrong a/o acceptire.bas 7701-.25

$n = 10$

Lost it!

rerun saved!

$\theta = 90$ $Phy = 180$ = top view.

Got it to plot the funct for $\mu = 0$ for various n:

$n = r$ $\left(\frac{1}{\sigma} e^{-\frac{1}{\sigma^2}}\right)^n$ peak is indep of n; its at:

$$\frac{d}{dz}\left(z e^{-z^2}\right) \quad z = \frac{1}{\sigma} \quad = \frac{d'}{dz} = e^{-z^2}\cdot(1 - z^2)$$

so peak is at $z=1$

$1 \quad \cdot e^{-1} = .367879$ $z_{min}$ $z e^{-z^2}$ $\quad 1\cdot e^{-z^2} + z\cdot -2z\, e^{-z^2}$

$1.01^2\ e^{-1.01^2}$ $\quad 1 - 2z^2 = 0$

$1.01 \rightarrow .364$ $\quad z = \frac{1}{\sqrt{2}} = \sqrt{.5}$

$.99 \quad .371$

---

$\sqrt{.5} \rightarrow$ ~~.2144~~ .42888 So peak is at $z^2 = .5$

$\sigma^2 = 2$.

$1+.01 \quad .42879$

$1-.01 \quad .42879$

$\frac{1}{\sigma} e^{-\frac{1}{2\sigma^2}} = z e^{-\frac{z^2}{2}}$

$z e^{-\frac{z^2}{2}}$ so 1 is peak. $\quad 1\cdot e^{-\frac{z^2}{2}} + (z\cdot -z)e^{-\frac{z^2}{2}}$

$1 - z^2 = 0 \quad z = 1. \quad \sigma = 1$

| z | |
|---|---|
| 1 | .60653 |
| 1.01 | .6064021 |
| .99 | .606469 |

.30

$(1-x^2)^n$

$2^{\frac{1}{n}}$ for large n

$e^{\ln 2\cdot\frac{1}{n}} = \left(e^{\frac{1}{n}}\right)^{\ln 2}$

$(1-x^2)^n = .5$

$1 - x^2 = .5^{\frac{1}{n}}$

$x^2 = 1 - .5^{\frac{1}{n}}$

$x = \sqrt{1 - .5^{\frac{1}{n}}} = \sqrt{1 - \frac{1}{2^{\frac{1}{n}}}}$

$= \sqrt{\frac{\ln 2}{n}}$

$2^{\frac{1}{n}} = 1 + \frac{\ln 2}{n}$

$2^{-\frac{1}{n}} = 1 - \frac{\ln 2}{n}$

.34 for $\frac{1}{2}$ width $\sqrt{\frac{\ln 2}{n}}$

3.27.97 TM

.01. Well that clears up t. problem of t. d.f. of $\left(\frac{\mu}{\sigma}\right)^2$ to a large extent!

T. d.f. for $\mu, \sigma$ (w.o. apriori) is (47.30) 7( )7

$$\left(\frac{1}{\sqrt{2\pi}\cdot\sigma}\, e^{-\frac{\sigma_0^2+(\mu-\mu_0)^2}{2\sigma^2}}\right)^n$$

The peak is indep of n. the $\mu$ peak is $\mu_0$ ; at varc $\sigma_0^2$

the $\sigma$ " " $\sigma_0$ at varc $\sim(2.5)^2\sigma_0^2$ (from 76.09-.10)

The varc's in both directions are $\propto \frac{1}{n}$ (76.30-.34)

The s.d.'s for $\sigma$ is ~ 2.5 times that for $\mu$ : the d.f. for $\mu$ is sym - Gaussian.

t. d.f. for $\sigma$ is  sharp on low end, slowly on hy end.

Simplify to $\frac{1}{\sigma}\, e^{-\frac{1+\mu^2}{2\sigma^2}}$ let $\mu=0$ i get $\frac{d}{d^2}$ at $\sigma=1$.

$= \frac{d}{d^2}\ \frac{1}{\sigma}\, e^{-}\ \frac{1}{2\sigma^2}$ or just use this take $\frac{d}{d^2}$ of ln for both $\sigma$ & $\mu$

$$f = -\ln\sigma - \frac{1+\mu^2}{2\sigma^2} \left|\ \frac{\partial^2 f}{\partial \mu^2} = -\frac{1}{\sigma^2}\ ;\ \text{at } \sigma=1 = 1 \qquad \frac{\partial f}{\partial \sigma} = -\frac{1}{\sigma} + \frac{1+\overset{=0}{\mu^2}}{\sigma^3} = -\frac{1}{\sigma} + \frac{1}{\sigma^3} = 0 \text{ at peak.}\ \right| -\frac{1}{2}\sigma^{-2} + \frac{1}{\sigma^3}$$

since t. peak is at $\sigma=1$, $\mu=0$ ; $\frac{\partial^2}{\partial\sigma^2} = \frac{1}{\sigma^2} - \frac{3}{\sigma^4} \to -2$ at peak.

.20 This is $1-3=-2$ so $\sigma$ peak is only $\sqrt{2}$ times as wide as $\mu$ peak : not 2.5 times as wide! Woops! it's t. other way around!

. large $\frac{\partial^2}{\partial(\ )^2}$ at t. peak means narrow peak! $\sigma$ is $\sqrt{2}$ times narrower than $\mu$.

Plot of $\frac{1}{\sqrt{2\pi}\,\sigma}$ ~~exp~~ $\exp -\frac{1}{2\sigma^2}$ width at level .2 (peak is ~.242) $\frac{32}{32}=1$

" " $\frac{1}{\sqrt{2\pi}}\ \exp-\frac{\mu^2+1}{2}$ " " $\frac{.50}{.40} = 1.25$

.25 so ratio is 1.25 rather than 1.414. I think as we go down on t. curve, t. width in t. $\sigma$ direction ↑ more rapidly than in t. $\mu$ direction: Even going from .242 down to only .2, t. curve in $\sigma$ is very assymetrical.

---

I got 2 second derivs from Maple! wrt. $\mu$ ; $-.24197072 44.$ $\mu$

wrt. $\sigma$ I got $-.4839419488$ = exactly twice $\sigma$ that of $\mu$

~~so~~ $\sigma$ may, indeed have $\frac{1}{2}$ varc of $\mu$ as in .20

These values are for

$\boxed{\sigma}\ -\frac{\sqrt{2}}{\sqrt{\pi}}\ e^{-\frac{1}{2}}\left(\underbrace{\frac{3}{2}-\frac{5}{2}+\frac{1}{2}}_{-}\right) = -\frac{\sqrt{2}}{\sqrt{\pi}}e^{-\frac{1}{2}}$

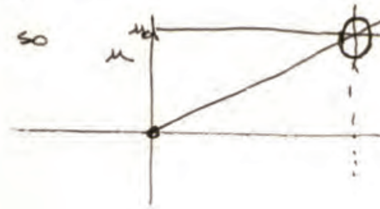
ⓜ $-\frac{1}{2}\ \frac{\sqrt{2}}{\sqrt{\pi}}\ e^{-\frac{1}{2}}$

See TM7-76.ms for how I was finally able to get these evaluations.

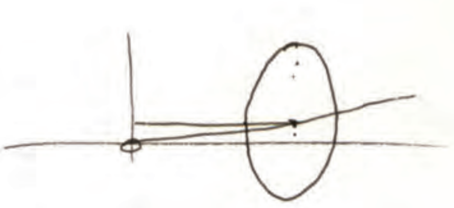
evalf(subs( x=1 , diff( $e^{-\frac{x^2}{2}}$ ,x,x )); does second derivs.

3.27.98 TM

so

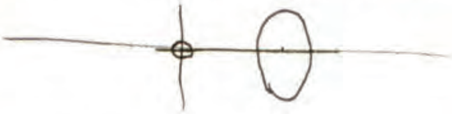


or more like:



In fact, we would prob'ly get a nice picture by setting $\frac{\mu}{\sigma}$ to φ and studying t.

Variation of ~~vs~~ S w. angle
as a funct of n.



Apparently, this makes t. (analytical) interpretation easy!

The stuff from 66.20 ff is on this: I think I got a way to solve it: t. eq. of 70.03:

But t. algebra is formidable!

By assuming $\mu_0=0$, I think it becomes very simple!

.15 I think t. S is 67.20: $\int \sigma \frac{d\sigma}{\sigma^n} e^{-n\frac{(\sigma_0^2+(\mu_0-k\sigma)^2)}{2\sigma^2}}$ set $\mu_0$ to 0, set $\sigma_0$ to 1!

$$\int \frac{d\sigma}{\sigma^{n-1}} e^{-n\frac{(1+k^2\sigma^2)}{2\sigma^2}} = \int \frac{d\sigma}{\sigma^{n-1}} e^{-\frac{n}{2\sigma^2}-\frac{nk^2}{2}}$$

which makes t. S indep of k, ~~which is ridiculous!~~ ~~[illegible]~~

.20 dependence on k in a trivial way! i.e. $\propto e^{-\frac{nk^2}{2}}$ i.e. $\boxed{\text{of vare} = \frac{1}{n}}$ — for $(\frac{\mu}{\sigma})^2$ d.f.

Which is strange since it's indep of $\sigma_0^2$! (not so strange! — see 72.09)

So check on ~~—~~ expected values of μ, σ & their vares.

$E\,\mu$ is $\mu_0$ (because of symmetry).

$E\frac{\mu}{\sigma}$ was obtained on 65.02 ($\frac{\mu_0}{\sigma_0}\sqrt{\frac{n-2.5}{n}}$ $n \geq 3$.)

$E(\frac{\mu}{\sigma})^2$ maybe a bit difrnt than t. sq. of ⟶.

---

| n | $n!/((n-\frac{1}{2})!\cdot\sqrt{n+\frac{1}{4}})$ | |
|---|---|---|
| 1 | 1.009253008 | — 3.0805 |
| 2 | 1.003003704 | 3.5 |
| 3 | 1.001459433 | |
| | .000858136 | 1.0 |
| 4 | 1.000563896 | ÷ 4 |
| 5 | 1.000229072 | |
| 8 | 1.000148 | |
| 10 | | |
| | 1.000067 | |
| 15 | 1.000059139 | |
| 16 | 1.000038 | |
| 20 | | |

3.75 (4–8)

3.87395 (8–16)

So error ~ $\frac{1}{n^2}$

2 + ¼

3.5

$n \to 2n$

~~error~~ error ~~4~~ = $\epsilon/(4-2^n)$

$\frac{\epsilon}{n^2}$

| n | $\frac{1}{n^2}$ |
|---|---|
| 1 | .0152 |
| | .0092 |
| | .006 |

$\frac{.006}{n^3}$

$\frac{.0152}{n^2} - \frac{.006}{n^3}$

~~.0008~~

0

.0038

I got it to do a simple pgm to print out this approximation.

There may be a way to get it to do tables: see index.

T. way it looks! What I've been observing is part of error $\propto \frac{1}{n^2}$, part $\propto \frac{1}{n^3}$ or $\frac{1}{n^4}$

The $\frac{1}{n^2}$ part is ~ ¼ for doubling:

1
2 ) 3
4 ) 3.5
8 ) 3.75
16 ) 3.885

$\frac{1}{n} \to \frac{1}{n^2}$ T. formula?

This no. is ~~rather~~ not so accurate, hvr.

error ≈ $\frac{.0152}{n^2} - \frac{.006}{n^3}$ is not bad.

~~[illegible]~~ $\frac{1}{n^2} - \frac{5}{n^2}$

16 64

3.28.97 TM

$\frac{x_{16}}{x_{32}} = 3.937088$ $\Delta = .063$

$\frac{x_{32}}{x_{64}} = 3.968560186.$ $\Delta = .032$

$x_n \equiv \frac{n!}{(n-\frac{1}{2})!}$

$x_{\frac{1}{2}} = \frac{\sqrt{\pi}}{2}$ : $\sqrt{\frac{1}{2}+\frac{1}{4}}$

$\rightarrow \sqrt{\frac{\pi}{4}}$

$x_1 = .009253008$

$x_{32} = .000015021$ $\times 32^2 = .015381504$

$x_1 : .009253008$

$= 6.128496 \text{ E-}3$

1:023326

so try $\boxed{\text{arr} \approx \frac{.015381504}{n^2} - \frac{.006129496}{n^3}}$

$x_{16}$ should be: $5.85875 \text{ E-}5$ according to formula

$16^3 = 2^{12} = 4096$

but it actually is $5.9139 \text{ E-}5$ so, not bad, but not perfect

.15 $$\boxed{\text{So } \frac{x!}{(x-\frac{1}{2})!} \approx \sqrt{x+\frac{1}{4}} \cdot \left(1 + \frac{.015381504}{x^2} - \frac{.006129496}{n^3}\right)}$$

I probably don't need so many figures of precision — prob'y 4 is enuf.

$\frac{.01538}{x^2} - \frac{.00613}{x^3}$ ratio of coeffs = 2.508977268

$\rightarrow 5.85815 \text{ E-}5$

www.aimslab.com

66 ween

4# 56

62

38

Solidox 25

pxx.org.

68

7

Solidox : Contains $KClO_4$

$x_1 = \frac{1!}{\frac{1}{2}!} = \frac{2}{\sqrt{\pi}}$

.01298

Doesn't seem to work for $n = \frac{1}{2}$: $x_{\frac{1}{2}} = \frac{\sqrt{\frac{\pi}{4}}}{\sqrt{\frac{3}{4}}} = 1.023326$

, formula gives 1.01248 about $\frac{1}{2}$ as much error as previous.

Still, corrected formula then gives $\frac{x!}{(x-\frac{1}{2})!}$ for x as low as $\frac{1}{2}$, w. max error ~ 1.25%

& error ~ $\frac{1}{n^4}$ perhaps, for large n.

So maybe try $A\left(\frac{1}{x^2} - \frac{\frac{2}{5}}{x^3} + \cdots \frac{\frac{4}{7}}{x^4}\right.$ ?

$2 \cdot \frac{2}{10}$

$4 - \frac{16}{5} + \frac{8}{35} \times 64$ $+ \frac{.8}{16}$ $\frac{8}{10 \cdot 16} = \frac{1}{20}$

---

So: Now I'm at t. pt. at which I can evaluate simple (say daily bet) strats., & tell how much spurious yield I get from picking t. apparently best of several strats.

Next Q is how to search over strats, : A good lang for strat Generation.

In particular, I think t. problem I've solved is t. Q of when to ↑ ssz of old strats, as opposed to trying new strats w. small ssz's.

.35 A Q. that I haven't really solved yet, is t. just how a priori of a strat enters into t. entire system: Perhaps t. idea is that if one tries a large amount no. of strats, even if t. data is pure noise, one will get apparently good strats — so we want to subtract, or divide out, t. amount of yield we get in trying $\frac{1}{P}$ strats, in considering strats w. a pri'y, P. ← I did do some earlier discussn. of this Q: Not sure if I had any good results: Maybe look. At any rate, t. allover system should be such that w. random noise ~~input~~ data, we should have very small likelyhood of thinking we have a good strat. or that t. expected value of a "best" strategy found should be very low. The stuff

3.28.97 TM

? 73.01 or 72.35 seem like start of this "Thread". Yes

around 71.20 – 75.40 is very relevant. It does help solve t. problem — maybe even good enough for t. present problem area — but I still don't feel I have a general idea of just how to define a soln.! ——— 88.07 .03

✓ Would a narrow approx/around the origin, for x = [illegible] do it? I don't think so. I'd guess t. approx would be concern w. n, t. data length, rather than m, t. no. of stars being ~~com~~ searched.

Perhaps I want t. formulae of 71.20 – 75.40 for large "m".

A useful inequality is B&S. P298 7.1.13!

.08
$$\frac{1}{x+\sqrt{x^2+2}} < e^{x^2}\int_x^{\infty} e^{-t^2}dt \le \frac{1}{x+\sqrt{x^2+\frac{4}{\pi}}}$$

$\frac{4}{\pi} = 1.27323954$

Apparently t. erf as defined in Maple is same as Matlab [redacted] B.S. Var[redacted] = $\frac{1}{2}$ : $\sigma = \frac{1}{\sqrt{2}}$

My modified erf may be ~~wrong~~ wrong, hvr.

.12 They define $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$. $\operatorname{erf}(\infty) = 1$ ; $\operatorname{erf} -\infty = -1$

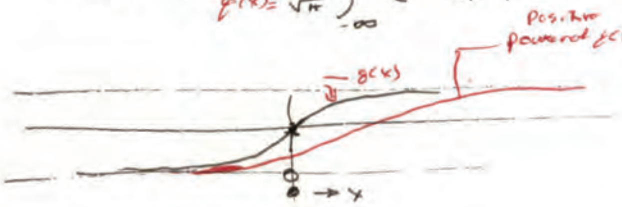
So it ~~is~~ increases by 2 from $-\infty$ to $+\infty$: So its not normalized as I thought! ~~Its twice as big as~~ It it twice as fast as it should, & it starts at ∅ w. x=0 — I want it to be ½ at x=0

So $\frac{1}{2}\operatorname{erf} + \frac{1}{2}$ is correct ( except for t. varc. — which I also corrected for

So it looks like I did t. right thing.

Actually, .08 & .20 are close!

.20 I was using $g(x) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{x} e^{-x^2}dx$.



$\int_0^{\infty}$ v.s. $\int_{-\infty}^{x}$

In fact $\int_x^{\infty} e^{-t^2}dx = \int_{-\infty}^{-x} e^{-t^2}dt$.

$$\frac{e^{-x^2}}{x+\sqrt{x^2+2}} < \int_{-\infty}^{-x} e^{-t^2}dx \le \frac{e^{-x^2}}{x+\sqrt{x^2+\frac{4}{\pi}}}$$

probably this approxn ~~was~~ was meant only for x > 0.
-- ~~But I don't know~~ but text said → x ≥ 0.

So $\int_{-\infty}^{-x} e^{-t^2}dy$ is close to $\frac{e^{-x^2}}{2x}$ for large |x|.
for |x| < 1 say, its not so close.
— but other approxns can be used.

Work on 79.35: maybe 38.01 ff ) 31.30 31.08 is on that problem.

10297 9½ f

.32 Another apparently related problem; Cover's universal portfolio: Using $\vec{b}$ w. max yield is probably best, because yield is ~~u~~ usually a very smooth function of $\vec{b}$. ~~Also~~ ~~For~~ Each $\vec{b}$ corresponds to a diffrnt strat. — Hvr, strats w. close $\vec{b}$ values are hyly correlated.

.35 What is expected "over optimism" involved in chosing $\vec{b}$ w. ~~max~~ max past yield?

.36 So; 3 ~~[illegible]~~ apparently related problems:

① T. problem of n uncorr Gaussians of ~~[illegible]~~ var = 1, means all zero: What is expected value of max? — also what is t. varc. of this max? New problem. Easy to solve numerically: Just put $x^2$ instead of x in 73.07. I can get Maple to do this for a reasonable set of ~~[illegible]~~ "n" values: Try to find way to get it to print it up in a more acceptable way! This would also solve problem of relative goodness of posn vs. mean |x| or mean($x^2$) as a way to get $\sigma^2$ of a dist. (73.01 – 75.40 ~~ff~~ is recent work on this)
82.01 – 82½.40 is numerical work on offset & ~~[illegible]~~ σ of that offset.

3.29.97 ΣM

.01 ② In Cover's set of strats (80.32–.35): what's expected (~~spuriousness~~ optimism) of pool of a continuum of highly correlated strats? — also t. vare of this (optimism param).

③ I Guess t. **General Problem** of estimating expected future yield of a strat. or "action Algm" — and t. vare. of that expected yield: More generally, t. d.f. of expected future yield. ← This is a "harder problem" but ~~it~~ stating it that way may make it easier to understand/solve it.

79.35 is some (perhaps) ~~[illegible]~~ recent relevant remarks.

④ The work ~~[illegible]~~ reviewed on 57.01–58.40, 65.01–.40: Given a ~~[illegible]~~ strat. w. a set of empirical yields, [$S_i$], estimate $\frac{\mu}{\sigma}\Big|_{yr}$ & t. vare of it. ↳ Also 67.20 & 78.15–.20

.08

{ 25% /4 mo / **Rad Rvo;** this → $(1.25)^3 = \times 1.95/yr$ : very smooth 1, $\times 1.953125/yr$.

.12 in .01 (2) Cover's problem: to get insite: assume t. underlying stocks are completely uncorrelated. Each has its own $\mu_i, \sigma_i$. It may be possl. to give them all same $\mu$ or all same $\sigma$, hvr. In any case, I can get d.f./of yield of any linear comb. of them.

⟶ Try it w. linear combs of just 2/ uncorr stocks.

Well, going back to t. indip stocks: It think linear combinations have a spurious gain → like t. "indip strategies". It looks like **Makm** deals w. this [by properly ① considering that t. coifs have certain expected error, also ② that fitting a "Best set of coifs" itself has a certain amt. of A. Hrassm in it. ← N.B. These 2 effects are exactly "same" as ~~This second effect is same as spuriously good~~ Gove when optimizing.

⟶ So it does 2 things: Any way to genz. this to the general "Action Algm" problem?

Well, take any strat: express its set of yields as a $\mu, \sigma^2$ model. If there are no other strats being considerd: Compare its $\mu\sigma$ w. that of some default model. — Say same $\sigma$ but $\mu = 0$. In this latter case, t. $\frac{\mu_0}{\sigma_0}$ of t. strat will probly give us its expected future yield! (also n is relevant — t. no of bets in t. data) ⟶ $\sigma_0^2 \approx \frac{n+k}{n-k}\sigma^2$ ← $\sigma^2$ of Makm is relevent here.

See 88.2 for simple, exact, Explain, treatment

If we have several strats — each has its own $\frac{\mu_0}{\sigma_0}\Big| = \alpha_i$ & its $n = n_i$. well $n_i\alpha_i^2$ will be of some interest, but also $\alpha_i^2 \cdot \frac{1}{T_i}$ ($T_i$ is t. time in which t. yields occurd) is of import. To first approxn, we want stock w. best $\left(\frac{\mu}{\sigma}\right)^2_{yr}$. To get this, we first Combine t. stocks $\frac{\mu}{\sigma}\Big)_{observed}$ & its $n_i$ w. t. apriori $\left(\frac{\mu}{\sigma}\right)^2$ of all strats, & it's equiv. n (= ssz) from that into we get an equit $\left(\frac{\mu}{\sigma}\right)^2_{yr}$ for t. stock, & that's our Gove for that stock

.27 If we have k strats, we can find k linear combs of them that are statistically indip. — The ($\sigma^2$'s / vares) of these "linear combs" are t. squares of t. eigvalues of t. original cov. matrix.

( of course, obtaining cross correln. of non-synch. strats. is non-trivial, but I think what I want

.30 is a set of wts ∋ w. ~~[illegible]~~ rebalancing at every bet, t. mult yield is max.

(SN) In .27–.30 t. ssz's of t. individual strats in t. mix can be diffrnt. — which makes it unclear as to what t. overall $\frac{\mu}{\sigma}$ really is!

1800 Collect.

So just consider K indip strats: each w. own known $\left(\frac{\mu}{\sigma}\right)^2$ ← No not known, but of known ssz & known no. of params

If we assume stationaryness, the $\mu$ & $\sigma$ of each strat is clearly defined (for ssz = ∞). (T. successive data pts may, hvr, be correlated ~~[illegible]~~ But then its auto corr is clearly defined).
If t. data is correlated, use Makm to get t. D.f. of t. next Bet.

Th. original problem I may have been working on: That one has t. ~~[illegible]~~ M. yield for a strat for a year, say (a single number): One has t. P.C. of t. strat. . From this info, it's hard to get much idea of reliability. If one is given t. individual trades for t. year, or monthly or weekly yields, 83.01 Spec

3.29.97 TM Maple:

Attempt to get it to print output in nice format: See "Tables" 83. ff on Tables & arrays

I'm going to try to devise an array, ~~and~~ write a pfm to fill it, then get maple to "print it" (P88 of Comos)

```
a := array ( 1..3, 1..~~20~~ 5 )   ← hor. index
              3 vertical index
```

$3 \times \overset{5}{\cancel{20}}$ array

```
a [1, n] = n
a [2, n] = expected value of x
a [3, n] = expected value of x^2
```

first ~~[illegible]~~ xx:= ~~[illegible]~~ a [3, n] - a[2., n]^2
a [3, n] = ~~[illegible]~~ sqrt ( xx ):

First then n:= 0

```
. ~~[illegible]~~ while n < ~~25~~ 5 do
   nn = n+1 :  n = nn :
   a [1, n] = n :      (n,1)
   od  :   ← s-uppresses printing.
   Print (a);
```

5 , 3

Tab creates extra blank line.

End  P 122 FL ~~[illegible]~~

Print; expression FL P35

l print: P35 FL

L is not ~~[illegible]~~ for "printer"

(no) 76 visual ~~[illegible]~~
(no) 78
452 has formula.
2 5 in property pany.

Name of Maple pfm.

.~~[illegible]~~ .20 TM 7-81.ms.   .583819

$\sqrt{\quad}$ 2nd − (1st)² = .34

| | first moment | Second moment ( | zeroth moment. |
|---|---|---|---|
| n = 1 | 0 | | |
| n = 2 | .3989422804 | .5 exactly! | |

.16

| | J | $2^J=n$ | first moment | σ | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0 | | 0 | ~~8~~.7071067 = $\sqrt{\frac{1}{2}}$ | | | | |
| | 1 | 2 | .3989422<br>~~.3989422~~ | .5838193701 | .12329 | | | |
| .7 | 2 | 4 | .727878 | .49584 | .0880<br>.0640 | .0353<br>.024 | 13<br>8 | 5 |
| | 3 | 8 | 1.00663 | .4307569106 | .0478 | .0162 | 5 | 3 |
| | 4 | 16 | 1.24874449 | .38406367 | .0364 | .0134 | 4 | 1 |
| | 5 | 32 | 1.4634768 | .34757 | .0287 | .007 | 2 | 2 |
| | 6 | 64 | 1.657269 | .318972 | .023 | .005 | | |
| 1.9 | 7 | 128 | 1.834657 | .296010091 | | | | |

1 2 4 8 16 32
64
7

quite interesting! σ ↓ slower than $\ln \frac{1}{x}$ !!   2.8

Note: .20 ff is ~~[illegible]~~ ≈ a "upper bound" on SOY: If the strats has correlnl 1, then This ↓ ~~[illegible]~~ effective no. of strats. If we have 2 strats & corr = −1, t. SOY is ≈ same as for N = 1 It amounts to integrating on Abs value of one strat.

For several strats. w. difrnt. $\mu_i$ & $\sigma_i^2$, it's not clear as to what extent this is an upper bnd; If all $\sigma_i$ are the same, then assuming all $\mu_i$ = largest $\mu_i$ gives simple upper bnd.

It would be interesting to know how σ goes as ~~[illegible]~~ n → ∞. Does it → an asymptote > 0)

If all ~~[illegible]~~ $\mu_i$ are same (≡ 0, say) then ~~[illegible]~~ we can get upper bound on Soy by assuming all $\sigma_i$'s are = max $\sigma_i$.

3.29.97 Maple

TM7-82.ms

82½

```
> a := array(0..7,0..3);
```

$$a := \text{array}(0\,..\,7, 0\,..\,3, [\ ])$$

```
> n := 0;
```

← Didn't seem to work w/ n := -1;

$$n := 0$$

```
> while n < ~~6~~7 do
> nn := n + 1: n := nn: m := 2^n:
> a[n,1] := m ;     a[n,2] := evalf(int(m*x*(erf(x)/2+1/2)^(m-1)*exp(-x*x)/Pi^(1/2),x=-infinity..infinity));
> z := evalf(int(m*x*x*(erf(x)/2+1/2)^(m-1)*exp(-x*x)/Pi^(1/2),x=-infinity..infinity));
> a[n,3] := evalf(sqrt(z - a[n,2]^2));
> od:
>
```

: or ; end each command.

: , hvr. means that results of that command should not be printed out.

```
> j := 0:
> while j < 6 do
> jj := j +1 : j := jj:
> print (a[j,1], a[j,2], a[j,3]);
> od
> j := 2;
```

$$j := 2$$

```
> print (a[j,1], a[j,2], a[j,3]);
```

4, .7278783067, .4958403153

```
> for j from 1 by 1 to 7 do
> print (evalf(a[j,1]), a[j,2], a[j,3]);
> od;
```

m ↓ first moment ↓ σ ↓

1 φ .7071067781 ← $=\sqrt{\frac{1}{2}}$

m first moment σ

2., .3989422804, .5838193701

4., .7278783067, .4958403153

8., 1.006637430, .4317969106

16., 1.248744490, .3840636718

32., 1.463476863, .3475755328

64., 1.657269826, .3189721555

~~$a_{[7,1]}, a_{[7,2]}, a_{[7,3]}$~~

Note These values are for $\sigma^2 = \frac{1}{2}$ ; $\sigma = \frac{1}{\sqrt{2}}$.

For σ = 1, the values for first moment & σ must be mult. by $\sqrt{2}$.

```
>
>
```

128 1.834657394 .29601009

3.30.97 TM : AAE (.22) Action Algm. Evaln. 83

.01: (see 81.40) Than its possl. to get t. vare. of t. stat to some extent. I was thinking that if one knew t. stat itself, one could know varc of yield (about 40) for random, "MKK-like" data — & this could be very useful for comparison — but if one had ~~[illegible]~~ enuf info to find varc. of stat. for "random data", one prob'ly ~~[illegible]~~

.04 had enuf data to get varc. of ~~[illegible]~~ actual stat. itself.!

---

About Maxm: Say t. stat. has k continuous params. + $l$ bits of descr. — Then Sol78Th3 corr. —

.05 gives convergence rate as ~~[illegible]~~ $\in \Sigma$ err $< k + \underbrace{\tfrac{1}{2} l \ln n \cdot \ln\sqrt{2}}$ ← I'm not sure of this: the "$\frac{1}{2}$" may have already been included in $\sqrt{2}$. | No!

Bits.

$5 l \sim .57$

$< \frac{k}{2\ln 2} + \frac{1}{4} l \ln n$

$\boxed{\frac{k}{2} + \frac{1}{2} \cdot \frac{l}{2} \ln n}$

If ~~[illegible]~~ err$^2 \propto \frac{\alpha}{n}$ : $\Sigma$ err$^2 = \alpha \ln n + \gamma$

$\alpha = \frac{1}{2} l$ but k is wrong! it is $\gamma$, (c.l.) indip of k!

say err$^2 \propto \frac{\alpha}{n - 2l}$ so $\Sigma$ err$^2 = \alpha \ln (n - 2l)$ : not much larger!

So err$^2 = \frac{\frac{1}{2} l}{n}$ + a part that sums to $< k$ : say $\frac{1}{n^{1+\epsilon}}$. This second part is ~~[illegible]~~ negligible for ~~[illegible]~~ sufficiently large n.

---

So: Maxm actually has much to add to Sol78Th3 — by getting t. expected error for linear regressn., & / $\overset{\text{most}}{\text{n.l.}}$ regressn at hy enuf n so t. system is "locally linear"

→ Hvr. I don't see how t. $\approx$ ~~[illegible]~~ $\left(\frac{n}{n-k} \cdot \frac{n+1}{n+1-k}\right)$ of $\overset{\text{Maxm}}{\text{Maxm}}$, is related to t. stuff in Sol78Th3! (see .20–.21)

~~$\frac{n}{n-k}\ \frac{n+1}{n+1-k}$~~ $\frac{n}{n-k}\ \frac{n+1}{n+1-k}$

Well, in t. case of MDL approx to ALP, one can find the no. of bits needed in each linear (or N.L.) coif, but a more exact value is obtaind by integration

.20 (in t. param. space — Has to do w. t. Hessian at t. peak. So this integration obtains t. coif of $l$ in (.05)

.21 It also obtains t. "$\approx \frac{n}{n-k}\ \frac{n+1}{n+1-k}$" of Maxm.

---

.22 So maybe we now have a ~~[illegible]~~ usable / "Rufflé dirtily"? soln. to t. ("Action Algm Evaln.") Problem (AAE)

AA

~~[illegible]~~ — for a stationary seq.:

We get ~~[illegible]~~ Yields at successive bets ~~[illegible]~~ $\{S_i\}$ — we then use a mean + $\sigma$ or several linear params or N.L. params, to get (w. Maxm) ~~[illegible]~~ a Pd for yield of next bet. I think this takes all impt. biases into account.

4.1.97: I'm not sure its really relevant! ; see also some stuff up to 80.03 | 88.17

.28 → Not quite! There is no mention of t. pc of t. AA itself! T. forgoing techniques will give a nice P.d. (MSE improvement) for t. next bet of t. AA. If we have several AA's, & pick t. "best", we have to modify. results.

.30 Hvr., say we have several (k) objects & we have a d.f. for t. score of each:

(or, if t. scores are not indip. a d.f. for t. joint d.f. of their scores)

.32 maybe not! see (.36) So / we have P($g_1, g_2 \cdots g_k$): what we want is t. expected value of (t. apparent peak minus t. true peak). T. true peak is, in t. case of indip scores, t. max of t. expected values of ~~[illegible]~~ t. scores of each of t. k objects.

What it means, if t. scores are not indip, is unclear! ~~[illegible]~~

.36 Well, what I want is expected value of error in picking peak rather than mean: So, for each possl. peak, we have an assoc. true Expected value of t. Score of t. object that to be peak.

Given P of .32 We can get t. expected $G_j$ of each of t. k objects. ~~[illegible]~~ at each pt. in t. k dim space of params: we have a choice of one of t. k as peak, & we have t. difference betw. that peak & t. ~~[illegible]~~ $EG_j$ of t. assoc. object. We want t. ~~expected~~ value

3 30 97 TM AAE

C: windows Command.com.

C:\windows\system\PIF\MGR.DLL

CTRL ALT Esc

.01 of Mat differences. → (.15)

If we have a Continuum of AA's: (~~[illegible]~~ Maneuverable linear): ~~[illegible]~~ & we select a "best" one: This is like n.l. curve fitting w. Non-Gaussian Gore. — & I think its Needed like Maxim. → (.15)

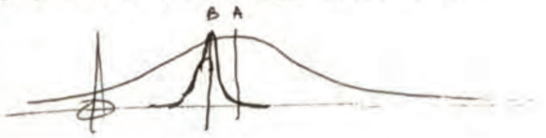
.105 2 ① I Think t. AAE prob is what was formerly called SMA (Stock Mkt Advisors) problem. T. difficulty was that if there were very many SMA's, There would be a spurious ↑ in t. peak, even if mean were zero.

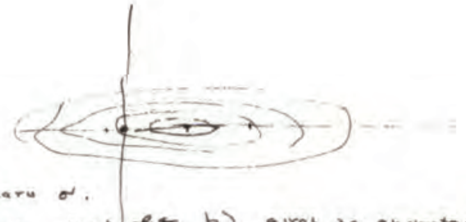
2 ② In t past, I may have been trying to solve SMA w. SSZ = 1. While This is of interest, it may not occur so often, & it may give what look like unreasonable results, because usually one has SSZ ≫ 1.

3 ③ I think I did get some impt. progress on SMA: Maybe at SAAB (≈ 1990) or more recently.

4 ④ In some of work of last few days, a propd ~~[illegible]~~ wasn't explicitly Considerd: This may be justified in that we may normally only consider strats w. mean values near that of the peak.

(.15) → 5 ⑤ Criticism of analysis of 83.36 – 84.01: It assumes that the peak was selected as "Best"! This may not be so. A peak of a hy varc. strat may not be as good as a much lower low varc. strat.





Overiz.

At 09 Greyhound.

.20 Consider 2 d.f.s. a) One about zero w. zero σ.
b) One about .5 w. s.d = .707. Picking a peak of b) gives an expected value
.22 of ≈ .5 – .707 = –.207 v.s. zero for d.f. (a).

Here, t. analysis of .20 – .22 is incorrect: But what would a correct analysis say?

→ It may be nec. to consider at least 3 d.f.s: c), d) are zero mean varc = 1; a) is mean = –.1, σ = 0.

.26 Its Important That I really understand what's Going on! → 89.26

6 ⑥ Just exactly what was t. AAE problem in TM? ~~However~~ Seems Related to Jergen's Very Greedy algm. for evaluating "Pgm Modifn. strats".

1 ⑦ T. recent discn. of AAE has been for stationary systems using prediction from ~~essentially~~ S; values alone, & very little about t. Structure of t. strat. (perhaps a prior value involves Structure of strategy to some extent). For t. more general non-stationary case we would use non-stationary predn. — with or wo. consideration of "structures" of strategies.

3.31.97 TM SMA/AAE problems:

I want to first review what t. problems are:

80.36 — 81.08 Lists 4 aspects of t. problem: A 5th aspect: we are given n strats (SM advisors).
.03 w. yearly yields of $\{Y_i\}$ $i=1/n$. How should we deal w. this? If n is large, chosing t. ~~best~~ apparently
.04 best one, gives a spuriously hy yield. If we only have a single $S_i$ for each SMA, do we really have much of a basis for decision?

Say we have n mutual funds or stocks, we have more detaild data: for some funds monthly, others daily, others hourly! (E.g. Sector funds might have hourly data available).

Or, we have SMAS that make buy/sell recomendations: They vary considerably in
.08 t. no. of bets/yr. Can we assign Cores to these SMA's? One concern is that t. set of SMA's "in business" is selected, so that the ones w. empirically poor
.10 records ~~have~~ tend to drop out.

.03 - .04 & .08 - .10 are major concerns. I think I was looking for a "simpler" soln. involving apriori pd., because using a large no. of SMA's did ↓ pc of each. The recent work of ≈ 80.06 ff really didn't use pc very critically (or simply): [The apriori pd was impt. in some of t. work.] Note 83.28

---

(SN) Some impt. reviewing ideas: 1) t. "conjectures" of 72.08 — 2) t. discns of 81.12 ff. 3) t. results of 73.01 – 75.40; 82.01 – 82½.40. on Bias due to choice of "Best" of several strats. 4) Reviews of 57.01 – 58.01, 65.01 on expected ~~μ~~ σ of a stat & expected value of that estimate

---

.20 → One very impt. aspt. re. SMA problem: A person ("SMA") make 10 hy vari. choices in a year — under $2^{10}$ difrnt names: T. choice sequences w. hy yields are published. Under t. assc. names. / — one doesn't hear about t. low yield SMA's. A thing ~ to this occures in "SM pickers" contests.
I felt that somehow t. pc of t. difrnt systems would somehow recognize what was going on. To make money, t hy yield SMA's simple use hy vari, random choices, using hy management & hy performance fees. Some of them will do very well! So its a useful technique for making Money (by t. SMA's))

---

One Big thing about TM's SM perambulations: t. specific difties of (.20) do not occur!: TM invents all of its own Strats — it knows how many of them faild/succeeded.

---

On t. other hand: A reasonably good apprxt. soln. to .20 would enable one to select SMA's a/o pool them optimumly — certainly a very Valuable Skill! — If there were any really Good SMA's one could find them: But there may not be any V.G. SMA's!

---

.33 I did have one (initial) soln. to t practical SMA problem: Take all Mvtt yield data of all of t. SMA's: This can be modeld as a single μ, σ, or as a μ, σ for each SMA — or some mixture of t. 2 models: See which has Most PC.
One difty is the dropping out of SMA's w. low yield, (≈ .20) so one gets a spuriously hy yield for t past. (SMA's w. poor past yields tend to drop out). One could, here, make that part of t. model — Not clear just how to do this — But I think I did figure out a reasonable way.

Almost all SMA's do worse than S&P!

Actually, one could study t. "dropout" problem by using data (if available) of t. list of all SMAS (say of a certain type) & their yearly (or preferably monthly) yields.

3.31.97 TM SMA problem : AAE prob.

One could make a (~ to Sim. Anneal) model in which t. dropout proby in each year was some simple function of yield for last k yrs.
By selecting only SMA's w. ≥ 5 yrs. "track record", one gets rid of many that started & dropped out because of low yield. Hvr. this "5 yr rule" is partly, simply a rule for "skimming t. top off." SMA's —
So it gives a set of SMA's w. spuriously hy yield.

If t. best model was t. same μ & σ for all SMA's, w. a dropout proby based on near past yield, then this
.08 Model could deal w. several SMA's w. large mean yields — (tho these yields are "spurious".) 87.18

.09 So, it looks like there are "Gradations" of t. SMA problem: Depending on how much info one has about t. creation of t. strategies.

One one extreme, one has one or just a few SMA's & only a little data on their performances.

Then in t. middle one has a larger no of SMA's & more data on their strats — including some info on rate of exit of SMAs.

On t. other extreme, one has, in TM, a complete knowledge of how
.19 all t. strats were created & all of their known trades. ≡AAE problem

ALP does not seem to offer any simple tricks to solve it all — (as far as I can tell.... but I may some day discover such a trick): Use of strats to code t. corpus (or partially code) may be part of such a trick. — Tho I'm not sure that that's so relevant; What I'm doing now is just using the yield sequences of t. strats, & using whatever models I can to fit t. data & predict it. Presumably, knowledge of t. structure of t. strats would help devise better predictive models — & this "Yield as coding" seemed to be a step in that direction.... but now I'm beginning to think, it can't have any more info than t. yield sequence itself.!

.27 A possible direction of trying to use str. of strats to get probys of yields:
Consider Carlos TK. Some basic ideas ① That behavior of MKT at difrnt times of day is difrn. ② That behavior in one section could tell one about behavior in another section.

(.28) ③ The "Occurrence" model: That if a certain condition obtains we have an "occurrence of j". When j occurs, t. Mkt will run $\left(\substack{dn \\ up}\right)$ a distance ρ(j) in ½ time τ(j). From ρ(j) & τ(j) we can construct an optimum betting strategy.

④ Time is important: Just knowing t. up down (at some level of granularity) history tends to be RW. — But knowing how long it took ~~of up or dn step~~ betw. up v.s. dn steps could give very Non-Random walk.

⑤ Slope in t. sense of ~~#~~ less square error line can be useful for pred'n: Maybe as Thresholded an "occurrence" (.28) criterion.

(Actually, I'm not sure about how .27 ff is supposed to work....)

3.31.97 TM / ~~$\frac{s}{p}$~~ MA prob. | **AAE Prob.** ← Maybe ~~final solution~~ → See (TM : LHL 5.19.97) 141.01 for Complaints ⑥ 87

.01 So 86.09 – .19 characterizes difrent degrees of t. SMA problem, depending on how much info one has on how t. ~~info was gene~~ strats were generated.

T. extreme case where all info is known, as in TM is t. "AAE" case

---

(SN) Analysis of clinical trials is a rather bad SMA problem – since t. conds under which t. trials were made are not very well known, ì lots of cases get rejected that one doesn't know about.

---

One kind of soln of t. SMA problem is 85.33 – .40 – to model t. process of creation of t. strats. This can be done in various ways ì ~~#~~ they depend much on just what info – ì how much info one has about t. whole process. Hvr ALP definitely is relevant to all of this.

---

.14 As for the AAE problem, I may have it more or less under control, ~~[illegible]~~ 83.22 was about to write a review of this, but we got sidetracked at 83~~[illegible]~~.28 ff

Try to complete that review! (81.12 – 83.04) is very relevant (~~[illegible]~~).

---

.18: 86.08: **On SMA!** Anyway: 85.33 → 86.08 is a reasonable soln. to t. SMA problem.

Dropout probly is a funct of $\mu_i$ ì $\sigma_i$. ("i" is ~~1st~~ SMA). Perhaps use $\mu_i - \frac{1}{2}M_{sip500}$ as a variable.

Either $\tau$ (= ½ life) ~~[illegible]~~ or $\rho$ = dropout probly per unit time. $\tau$ has possl. range ~~[illegible]~~ $0 \le \tau \le \infty$; $0 < \rho < 1$;

or use ~~$\frac{1}{\tau}$~~ or $\frac{1}{\rho}$. ~~$\tau \overset{?}{=} e^{-\frac{1}{\rho}}$~~ or $\tau = \frac{k}{\rho}$. ? Probty of length $T$ is $e^{-\frac{T}{\tau}}$ $= \lim_{\epsilon\to 0}\left(1-\epsilon\rho\right)^{\frac{T}{\epsilon}} = e^{-T\rho}$

Expected value of ~~so~~ $\rho = \frac{k}{\rho}$ $\rho$ = probly ~~[illegible]~~ per unit time, for small times

of dropout. So it's ~~[illegible]~~ $(1-\rho\epsilon)^{\frac{T}{\epsilon}}$ ~~as proby e~~ $= e^{-T\rho}$ as proby of no dropout after time $T$.

Hvr, it also depends on $\sigma$. If $\sigma$ is large, we can have a dropout in which a large (invisible to us) loss occurred during t. last month & a few dropout. ← so this is

Anyway, t. fitting of $\tau$ as a funct of $\mu_i$ ì $\sigma_i$ can be done separately (first). ~~This can b~~ A prelimy study of a plot of $\tau_i$ v.s. $\mu_i$ ì $\tau_i$ v.s. $\mu_i$ ì $\sigma_i$..

Then, the second part, is to either assign a $\mu_i$ to each SMA$_i$ ì a common $\sigma$, (or perhaps individual $\sigma_i$'s.).

The fitting of data to $[\mu_i]$ or $[\mu_i, \sigma_i]$ ì to $\sigma, \mu$ is a very direct, once that, $\tau$ as a funct of $\mu_i$ or $\mu_i, \sigma_i$ is obtaind.

I think its a chain type fit, and so we need t. Hessian of each fit.

The "$\tau$" fit removes t. bias obtaind by dropout of low $\mu_i$ SMA's.

The fitting "factors" into t. 2 separate parts. ① $\tau$ ì ② $[\mu_i, \sigma_i]$ ì $\mu, \sigma$. (or $[\mu_i], \sigma$ ì $\mu$)

.37 We want t. rel. wts. of t. 2 models: $[\mu_i, \sigma_i]$ or $[\mu_i]\sigma$ v.s. $\mu, \sigma$.

Another possly model (seems ~~to~~ combinatorial difft. to compute) is a given SMA$_i$ can have either ~~a common $\sigma$~~ or common to a large group at low r cost or, its own $\sigma_i$ (at hyer r cost).

88.01

4.1.97 TM $\boxed{\text{SMA}}$ solution ! ≈ Final Soln. $\boxed{\text{MEXM}}$ (.23) : A simple explnn for linear Regressn

.01: 87.40! To do this study, I need several yrs. of data! SMA's yields! preferably Monthly, maybe quarterly or yearly. There is a co. that sells such data. I mite be able to get it free from them, if I promise to publish results! because ~~[illegible]~~ such results would make their data more valuable, Sellable. ~~[illegible]~~

First fit ? to [$u_i$] or [$u_i$, $\sigma_i$]; The second part, ~~[illegible]~~ 87.37 ~~[illegible]~~ (≈ fitting [$u_i$, $\sigma_i$]) is perhaps easy, but I would want to do a more complete ALP ~~[illegible]~~ analysis to make sure its all correct. This would involve a study like that ~~[illegible]~~ reviewed in 57.01 – 58.40; 65.01 … also the "Bias due to Optimization": 83.36 – 84.01, 84.15 – .26 (impt!)

.10

→ see 141.12 for Criticism

---

T. General AAE problem is same as ≠SMA, but w.o. dropout.

A prelimy review of ~~[illegible]~~ General Soln. to t. AAE problem see 87.14;

83.22 was start of review. ~~[illegible]~~ also see 81.12 – 83.04

---

17: 80.03 There is also t. ~~[illegible]~~ of "How relevant is t. Eprrp of a strategy?" presumably its then related to t. no. of stats. considerd, so this could be relevant to t. amt. of "spurious score" one obtains by Optzn. 83.28R

Superficially, it would seem not directly relevant. Aprrp ~~[illegible]~~ only gives a ≈ upper bound on t. effective no of stats one is considering. Hvr, in any real problem, one knows how many stats one is optimizing over. Of more imprtnc, is t: amt. of ~~[illegible]~~ correlation betw. t. stats & their distribn of $\mu_i$'s & $\sigma_i$'s.

↑ This may be an impt thing to understand, but I don't have time now! (look at stuff leading to 80.03 again)

---

(.23) → (SN) Re: $\boxed{\text{Mexm}}$: I did this analysis of a 1 param linear fit — find t. mean of $n$ measurements: If t. mean were known (say = 0) t. best estimate of $\sigma^2$ is ms. deviation from that mean. Hvr, one obtains in fact t. empirical ms deviation from $\mu_0$, which ≠ $\mu_{true}$!

Moving from $\mu_0$ to $\mu_{true}$ adds to var. ~~[illegible]~~ by $(\mu_0-\mu_t)^2$: $E(\mu_0-\mu_t)^2 = \sigma_T^2 \cdot \frac{1}{n}$.

So $\sigma_T^2 \approx \sigma_0^2 + \frac{1}{n}\sigma_T^2$ or $\sigma_T^2(1-\frac{1}{n}) = \sigma_0^2$ | $\sigma_T^2 = \sigma_0^2/(1-\frac{1}{n})$ Error in next bet, has

$\sigma_0^2$ error² plus ~~[illegible]~~ error due to $E(\mu_0-\mu_t)^2 = \frac{1}{n}\sigma_T^2 \neq 0$. so error is $\sigma_T^2(1+\frac{1}{n}) =$ ~~[illegible]~~

$$\sigma_T^2 = \boxed{\sigma_0 \cdot \left(\frac{1+\frac{1}{n}}{1-\frac{1}{n}}\right) = \frac{n+1}{n-1} \cdot \sigma_0}$$

.24

In t. case of linear regressn: we have t. same situation: error in coefs causes spuriously small $\sigma_0^2$: $\sigma_T^2 = \sigma_0^2(1+\epsilon)$: Error in next bet is $\sigma_0^2\left(\frac{1+\epsilon}{1-\epsilon}\right)$.

I guess $\frac{1}{1-\epsilon} = \frac{n}{n-k}$ where $k$ is no. of coefs; so $\epsilon = \frac{k}{n}$ ! $\hat{\sigma}_T = \boxed{\sigma_0^2 \, \frac{n+k}{n-k}}$ $\circ\!\!\!\!\!\bigcirc$ $= \sigma_T^2$

→ (actually if one variable has $E\epsilon^2 = \frac{1}{n}$, k vars have $E\Sigma\epsilon_i^2 = \frac{k}{n}$ ! $\frac{1+\frac{k}{n}}{1-\frac{k}{n}} = \frac{n+k}{n-k}$)

I should be able to get this result using t. techniques of 57.01 – 58.40, 65.01 r .90, — i.e. the integrations leading to these results.

ABCDE

I think this result (.23 – .29) generalizes to linear curve fitting, & to ~~[illegible]~~ (for large n) locally ≡ linear $\boxed{\text{n.l.}}$ curve fitting. Hvr, I'm not exactly sure of t. results for linear curve fitting (& n.l. curve fitting).

AMD address

→ $\boxed{\text{spec } 92.01}$ →

4.1.97: TM SMA, AAE REV

ABCDEFG
HIJKLMNOP

Another apparently Impt Q: Mexm takes care of "overfitting errors" (as 88.23): It does this in a natural, general way. If I use Mexm for getting t. d.f. for t. "nonexpert" of a strat. it would seem that this would not "interfer" w. t. next step of correcting for t. "Optzn effect" of selecting "Best" of several strats. It would seem that these are difrnt, "separate" effects — so I would not be "double counting" by making t. Second Correction.

---

.08 So: In review of A.A.E. (Action Algm. Evaln):

.09 One has a bunch of AA's: $[A_i]$: For each one one obtains t. d.f. of yield or G or whatever, for t. next bet, trial, or next year or whatever it is one is concernd w.
.11 in t. future. In most cases: these d.f.'s are defined by $[\mu_i, \sigma_i]$

They are obtaind by using t. empirical yields of t. $A_i$ in t. past: $Y_i^j$ say is t. yield of $A_i$ i jth trial. The d.f. ($\mu_i, \sigma_i$, say) can be obtaind from $Y_i^j$ using any predn method one likes. One could just take $\mu_i$ = mean of all $Y_i^j$ or use linear predn on past $Y_i^j$'s or use data from other sources a/o use data from other $A_k$'s. At any rate, in t. most genl. case one has obtaind t. d.f. for t. yield of t. next bet (.09–.11) This is t. most impt. desired output.

Anyway; In t. simplest case one has this ~~same~~ set of indep $[\mu_i, \sigma_i]$.

.23 From this info, using t. ideas of ≈ 72.26 – 75.40; $\underbrace{82.01 - 82\tfrac{1}{2}.40}$ ($\mu$ & $\sigma$ of effect via MAPLE.), one finds effect of picking "Best" of several strats.
.24 For t. case of Correlated d.f.s see discn. of 83.22 – 84.26 ← Also raises another impt. Q.

---

A review of where various ideas, derivations, are written: A "Bibliographic Review:

.26: 84.26: The Discn. of 83.30 – 84.26, on t. General Case of "Spurious G & assoc. w. Optzn" is Very impt.! I think I have all of t. relevant Ideas there, but I then suspect t. details may not be correct. I really should write an exact formulation of what t. problem is, as well as an
.29 out line ~~of~~ a Theoretical Soln.

---

→ listing Impt. ideas, & the Thread in t. notes.

Some impl. ideas: 1) → **MEXM:** simple way to get it; b) relation to Sol 78 Th 3 cor. I.

2) SMA prob. v.s. AAE prob. ("Eliminal trials" etc = SMA): stationary v.s. non stationary

3) Spurious of G assoc. w. "optzn." — size of effect & also its error

4) ~~Analysis of simple~~ Simple analysis of a strat. (or AA) in terms of. $\left.\frac{\mu}{\sigma}\right)^2_{yr.}$ — expanded to more Genl. case.

5) Mehe Integrals: How to use MAPLE. ©. $\frac{x!}{(x-\frac{1}{2})!}$ approxn. also, related to J's obtained pp review 57, 58, 65.

4.2.97 TM

Review of recent work: (Bibliog. review): starts ~ 31.01

.02 1) MEKM: 88.23-.40: A simple/point way to understand $\frac{n+k}{n-k}$ factor! for linear regressn w. probable extension to linear curve fitting & lin. regressn & curvefitting:

83.05-.21: Relation of MEKM; to C + Abl n n in Sol78 Thm 3 (P427 col 1)

Also note: continn. of 88.23-.40 suggested by 92.01ff

.05 2) The SMA problem & t. AAE problem (SM Advisors; Accum Algm Evaln).

Most recent work starts ~ 31.01. [up to that pt. was t. "PATH" Thread]

Relation of 2 probs: SMA was most recently reorganized on 84.05: Its relation to AAE was made clear in 86.09-.19 (Difnt "o's of SMA": AAI is extreme case w. all info on how strats were created).

.09 3) General Soln. of SMA problem in which one has some (but imperfect) knowl. of creation of strats: 85.33 (start.) - 86.08; 87.18 - 88.10: This is probly a quite practical soln. 72.36 is t. start of this thread: Some prelimy results on 75.27. This discusses only t. part related to SMB, But not AAE. To solve SMA, we also need a soln to AAE problem.

.15 4) The AAE problem: 89.08-.29 reviews results: a general Approach. 89.23-.24 is biblio on work on Bias that is introduced When one selects t. "Best" of several strats → we estimate t. magnitude & varc. of this spurious ↑ in "Yield".

.18 5) The AAE Prob. Another aspect: The work summarized in 57.01-58.40; 65.01-.40. ← corrected version.

In this work, I was interested in t. d.f. of $\left(\frac{\mu}{\sigma}\right)^2$ of various strats.

Say $\mu_0$ & $\sigma_0$ are observed empirical $\mu$ & $\sigma$. for "$\frac{\mu_0}{\sigma_0} \gg 0$" this is hard to compute, but if $\frac{\mu_0}{\sigma_0}=0$, (which is t. usual case, since $\left(\frac{\mu}{\sigma}\right)^2$ is for a single bet) then t. d.f. is easy to obtain: its mean is of course 0,: its $\boxed{\text{varc} = \frac{1}{n}}$ (Discovered on 78.20) which simplifies things tremendously. It makes t. main conjecture of 72.08-.25 correct — and so 72.08-.25 does tell how to get an apsip for each strat in a very simple way! — using some reasonable a priori about zero. (t. width to use is unclear ... try various values).

.27 6) Some Mathmcl results: Analysis of strategy w. yield data $\{x_i\}$ of empirical. $\mu$ & $\sigma$ of $\mu_0$ & $\sigma_0$ resp: n data pts. We find d.f. of $\mu$, $\sigma$, $(\mu, \sigma)$ also $\frac{\mu^2}{\sigma^2}$: eventually we find t. d.f. & varc. of d.f. of $\frac{\mu^2}{\sigma^2}$ for d.f. centered about $\left(\frac{\mu}{\sigma}\right)^2=0$ (which is most interesting case). Results on this are summarized on 57.01-58.40; 65.01-.40.

The part about d.f. of $\left(\frac{\mu}{\sigma}\right)^2$ about $\left(\frac{\mu}{\sigma}\right)^2$ peaked at zero is 78.20

One discovery: $\boxed{\frac{x!}{(x-\frac{1}{2})!} \approx \sqrt{x+\frac{1}{4}}}$ : 79.15 has makings of even better aprxm.

W.o. further correction error is
2% for x = .5
1% for x = 1
.5% " x = 1.5
.2% " x = 2.5
} error $\propto \frac{1}{n^2}$ for large n

.35 7) MAPLE: How to use: see 82½.01-.40 for one listing; See TM7-( ).ms files in c:\maplev4\bin for examples of how to pgm. MAPLE.

4.2.97 TM

Going thru 31.01 to present: some random remarks!

32.17 puts t. present SMA/AAE problems in t. context of t. Genl T.M. problem.

One of t. problems was to have a basis/ (in Lsrch) for deciding between increasing lengths of data sequences evaluated w. current stats. v.s. creating new stats ε evaluating them w. short/ data sequences.

What <u>looks</u> like a soln. was obtained by putting t. formulas for expected $\left(\frac{M}{G}\right)^2$ ε its uncertainty "δ". as a function of t. data for t. stats considered, as well as their <u>ssz</u>. <u>and</u> t. assumed a prior <u>width</u> of $\boxed{\cancel{[illegible]}}$ t. $\frac{M}{G}$ about t. origin.

---

T. main results re: Using Lsrch on a large data base of, say, SM data, is

90.02, .15, .18
① ④ ⑤

For later return to this stuff: Read 90.01 –.40 ε refs to get main ideas, then review

$\boxed{31.01 \text{ to } 92.40}$ to get some <u>other</u> <u>impt</u> <u>ideas</u>.

4-3-97 TM : MEXM

.01: $\boxed{\begin{array}{c}\text{Spec}\\ 88.40\end{array}}$: **Mexm:** The discn of 88.23-.40 seems relevant to Portfolio Management!
Also, it gives an example of a continuous ∞ of choices. In the present case, say we have k SM's.
We can find k linear combns. that are linearly indep. So we diagonalize the set.
Each comb. has its own $\mu_i, \sigma_i$ & each $\mu_i, \sigma_i$ has its own uncertainty.
We then pick (within a constant factor) a set of wts ∋ $\left(\frac{\mu}{\sigma}\right)$ of sum is max. What is expected
.06 bias in the resultant $\left(\frac{\mu}{\sigma}\right)^2$? The ideas of 88.23-.40 certainly seems relevant!

As a "Study problem": we have k measurements of the same quantity.
They were of different precision & are indep! $[\mu_i, \sigma_i]$ $i=1/k$. To make a linear combn. of these
measurements w. min $\sigma^2$: Hmm.: what is $\sigma^2$ for next measurement, — well its unclear.
depends on how one measures it! (≃ which one of the "k" ways). I'm not sure I'm phrasing it the
rite Question!

---

Going back to .01-.06: since I optimize $\frac{\mu}{\sigma}$; both $\mu$ & $\sigma$ will tend to be biased.

Well: This is an Impt. problem & promising approach! But I don't have time just now!

---

See 141.31-.40 (+ SMA problem) for More on this.
147.01ff

4·16·97 : TM : "50% solution"

A couple of related probs. probly having about same solns. :

.01 1) Given a system w. a $TM_1$ & a $TM_2$ : How much relative time (& just when to switch) betw. $TM_1$ & $TM_2$ ?

.03 2) for $RTM, h$ = 100 say (horizon) : How much Time to spend on predicting t. future, v.s. how much time spent on trying to solve its problems.

---

.06 In Barc (latest version) I desbd t. **RLP** problem as finding provable largest poss. lower bound for $\sum 2^{-l_i}$ : Given a certain available cc.

.06 was an explication of a more el. version in which we wanted to find as

.10 many codes as possl. so as to maximize $\sum 2^{-l_i}$ in t. available cc : We wanted to generalize to continuous codes (equivt to discrete codes).

Anyway, using .10 as a (prelimy) model : The spirit of the Thing is this :

consider ① (.01) : We list various $TM_1$, $TM_2$ models : T. order of listing we try to approach is $\sum \frac{T_{TM_1} + T_{TM_2}}{pc_{TM_1} \cdot pc_{TM_2}}$ — so its a single problem for optzn.

for ② (.03) again t. order of listing is $\frac{T_{\text{predicting future}} + T_{\text{solving future probs}}}{pc_{\text{predicted future}} \cdot pc_{\text{solution to that future prob}}}$.

Actually ② is more complex : It involves expected future Gorc based on codes of t. past Gorc values.

T. formg (.01 ff) seems to me to have great (likely hood of working)!

It certainly needs a lot of work : The details have to be worked out : I have to draw up some examples.

T. reason its impt. is That ~~it is an~~ They are examples for fairly complex problem types That I want to be able to formalize so TM can usefully work on Them.

---

In .01 perhaps there is really only 1 goal : t. $TM_2$ goal. What would soln. be if it had an enormous amt. of time (≡ cc)? Well, it would take a finite amt. of time to solve t. "present problem" of $TM_1$ : T. rest of t. time would be used by $TM_2$.

Another possy : we work on $TM_2$ only : in t. course of this work, $TM_2$ gets $TM_1$ to work on its probs.

Looking at it another way : Say we have a TM that works on a bunch of INV & OZ probs ("problem pool approach"). One of t. probs in t. pool is "self improvement". It normally is given rewards (R) for solving or working on various probs, but no direct reward for working on S.I. So how much time does it spend on S.I.?

1) dealing w. induction problems w.o. noise (≡ MTM) is an adequate form of t. general induction problem

Relevance of this stuff to AAE problem
See Review 90.05
$\boxed{90.05}$ ff

~~W~~Windows\Tips

4.16.97 TM

TMC problem ≈ IR problem (.07)

pp 95 thru 100 do not exist — jump to 101

Well, t. approach of TM: LHL ~~3.~~29.01 – 30.40: (29.24ff in particular) seems relevant.

We first chose a time, $T_2$ in t. future, then try to optimumly allocate time between direct "self. improvement," tasks & ~~[illegible]~~ S.I.'.

{≡ X}

[illegible] TM₁ TM2 .

On t. Q of How much time to spend making good models of t. future! One way to look w. This is t. "Problem Pool" approach: TM knows what future "Big Ticket" items are:

---

.07 Def **TMC** (≡ Too many concepts ≡ IR problem): Say we have ~~[illegible]~~ put into "memory" every Program we've proved, & every sub-function (in our function trees) that has proved "useful". We can associate Unconditional pc's w. these concs, but for use in building new concs & solns to problems, these uncondl. pc's are ~~[illegible]~~ usually too small. david

.14 The way out is conditional pc's: A particular conc. will have ~~arrisen~~ (been discovered) in
.15 a particular envt. This envt. will suggest ~~[illegible]~~ where to use it next. (see .26 – .27) Normally, ~~various~~ academic areas are divided up into special areas like Math, Physics, Chemistry... also sub & sub...classes. A tree structure. ~~[illegible]~~ Any Conc. is ~~[illegible]~~ normally closer to concs. & applications in its own branch of t. tree. To get to other branches & sub branches requires a certain amt. of traveling, & each step in this travel will have its special [illegible] of pc. (Academic Darling) (WolfGang's "Directory's" ~~[illegible]~~ is one way of dealing w. this "compartmentalization of knowledge") Solved by the Academics.

.26 T. idea of .14 – .15 is that how t. conc. was invented (where / when), how long it took to be found, will give some constraints
.27 (problistic) on how it is expected to be used in t. future.

The TMC problem is closely related to t. general IR (info retrieval) problem.

---

~~Her Ther~~ There are several impt ideas here:

① TMC ≈ IR

② ~~other~~ = Before TMC arises we have to generate/discover t. relevant concs. T. process of discovery — how, how long it takes, techniques used || are related to t. IR params used in t. later retrieval of that conc. Furthermore, t. techniques used in IR mite be somehow used to discover ~~originals~~ concs. in t. first place.

ABCde

③ Human Heuristic: "State problem precisely, formally"

inverse

Machine heuristic "take precisely stated problem: generalize it, related probilistically to ~~[illegible]~~ other probs, that will help solve it. State problem in various languages; State problem vaguely, ~~[illegible]~~ w. params that make it exact. (Vague statement is then a [illegible])