1.2.99
~~1.2.98~~ TM INV

.987M156.40
.01

Extra fine
Abcdefghijk. Continuity ABCdefg. ABcdefg
Fine Abcdefghijklmnopqrstuvwxyz

1. On solving equs ~~[illegible]~~ as Inv. probs! Consider: Eq. $3x^2+7x+3=0$.
To be solved w. closest fit w. ~~[illegible]~~ resoln. .01 on x.
With a very simple P.d. (~~[illegible]~~ No experience in eq. solving), we just try all nos. of form $\frac{n}{100}$ in order of pc ($\approx$ bc $\approx$ log n). (— Not really! — see .20)

2. When we "know" about continuity, we mite try finding ~~[illegible]~~ ~~next~~ $X_{1,2}$ ∋ $f(X_{1,2})$ are of opposite sign, Then try for values betw. them at .01 resoln.

1) Send 597 to Bergman.
2) Try COR-R first.
3) PS cc's no hurry

3. if we knew how to solve quad. equs., we could ~~ed~~ do it much faster

---

4. [illegible] This seems related to Q. of "probly that 10^{100} digit of π is 3."

In case of π, we have to define "π" ~~[illegible]~~ to T.M. If we define it via an infinite series, early, (non-sophisticated) trials will involve experiments ~~[illegible]~~ finding many digits of π via t. series. w. large enuf CB, t. 10^{100} digit of π can be found exactly, via t. series.

LLF ABCDEFG $\int_{-\infty}^{+\infty} e^{-x^2/2\sigma} dx$ ABCDEFG abcdefg

On th. Otherhand.

.20 5. On t. other hand, we can solve equs. by using t. function "Solve(,)". — This seems to be quite difrnt from solving t. eq. as an Inv. prob.!

Maybe not so much! — Remember that in .01, t. Inv. problem gives rise to a conditional p.d. — t. condition being t. eq. "$3x^2+7x+3=0$". — which is t. "inversion constraint" defn.".

It would seem that t. "Solve" function would be easier to ~~train~~ train/teach than t. INV. ~~[illegible]~~ P.D. of .01: But presumably, a Brite T.M. would eventually discover that they were identical!

1399 TM: ~ CHESS.   maybe Great Break Thru! (25)

Expo: Title of Book, Paper, Talk: T. Science of Discovery, ALP & MCT!

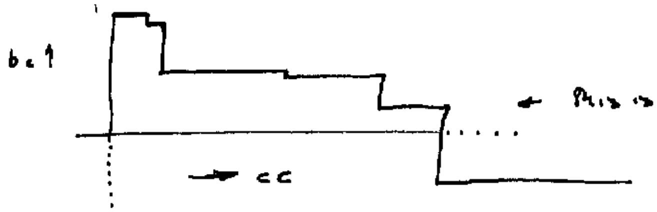
This is GREAT.

.01 Say one is spending cc on finding short codes for a corpus: One notes t. bc of t. corpus (thus far) as a function of cc expended. While, in general, one can't legally extrapolate, to what extent can one, using finite CB for t. Extrapolns?

One can note that t. σ^2 becomes larger as we project further into t. future. So in general, we don't expect much neg accuracy for cc ∞ extrapoln. A st. line, parabola, exponential w. horizontal asymptote, ... or other forms can all be used for Least Σ sq. fit.

This is the one

ABC



← This is ~~more~~ what it's likely to look like.

Note that even at cc = ∞, bc can never be zero, but it can be close to zero!

Similar extrapolns can be made as corpus size ↑ w. time and cc ↑ w. time.

This discussion may be going in direction of soln. of "Chess" problem!

An area where we may have some idea of how bc varies w. cc is looking for
.23 models in SM. I try a few, then on basis of results, I may have ideas as to what to try next, & expected bc↓. —— (.30)

(.25) → see (98)145.09 for previous work, discn.

→ In fact, in t. Lsch soln of t OZ problem, one normally has quantitative idea as to what G to expect w. a gn. cc. [In t. present case G = − bc.].
So, actually, while this problem is unsolvable for CB = ∞, it probably is quite solvable ~~[illegible]~~ for finite CB .... In fact I hope to solve it every time I solve an OZ problem via Lsch!

.30 (.23) May be related to t. "What to Work on Next" problem.

1.6.99 TM : Updating: Sumrzng Machine · "Summarizing Machine"

01: [redacted](98)146.90 ; So it looks like I can mainly think about a sequential corpus, [redacted] w. impunity
Hvr. [redacted] coding ~~a~~ "corpus increment" can be a "very many flavored task" ;
The varieties of how this can be done are quite large. They all, hvr, have a common "dimension" & so are interconvertible.

[SN] T. "Summarizing Machine" is but one way of expressing all or part of a P.D.

So: List of a few P.D. forms: t. output of an "Updater":

.09 1) A uncond proby ~~mono~~ machine (unidir IO ∄ — no hardy umc) : short random inputs give d.f. on outputs.

2) ~~a~~ stoch operators: [redacted] Input is augmentation of corpus ; output is P.D. on continuations of corpus. Its output can be realized ~~by~~ as .09 or by listing (outputs in order of p.c. along w. p.c.'s: (for 1 stoch)) ~~stop~~ or by having possl. output presented, & pc of that output is given

.14 3) So 1) & 2) can be varied ⓐ condl. or uncondl p.c. ⓑ output can take various forms ; α Monte carlo ~~w~~ usually w.r.t. random input to stoch operator β given possl. output string, we get pc of that string. γ (output, pc) pairs in ↓ pc order or simply output in Lsrch order. δ [redacted] One (or more) control
.17 objects & an eqn. giving density distribution about those control objects. → 9.20

Very Probably .14 — .17 doesn't cover all cases ··· but leave it for a while.

Next: Ways to get ~~these~~ t. P.D.'s of .14 — .17 (How to do updating) :

.20 First "look at" t. corpus: at all times we have a bunch of definitions that we have been told about or discovered to have been useful in coding. As we scan t. corpus, we look for instances of all of these defns. (This is a process that is very ~~[illegible]~~ parallelizable) Many of these defns. correlate w. methods of corpus analysis, coding. These "defns" are "obs" (observations): They somehow imply "ops" (operations on t. corpus) — that yield new obs & suggest new ops, (etc). [the ob, op algebra]

One perhaps commonest op: that certain obs occur w. certain frequency that differs "Signifly" from that expected in random corpus.

How does t. stoch lang srch of ± SGA fit into this?

ABC

(SN) .20 ff. looks pretty Messy! Contrast w. my recent ideas on "how to lrn. practically any tsq." ! perhaps go into this simple stsq idea ··· See if it works & how far it can go.

Tsq? Simple tsq

Mixd Corpus Thrm.

Actually t. early stuff in (98)88.01 & even before has a lot of nice ideas about MCT & its applicns in learning TM!

1.7.99 TM TEXTWORLD:

.01: 1) On "Novels": We can let T.M. read a novel ... coding it sequentially, then perhaps it can reread it & recode it again & again. We can then ask Q's <u>about that novel.</u> The text for t. novel will be labeled w. ⟨Author, Name of Novel.⟩. Similarly label

.04 for Q's about that novel. After TM has read several novels, it will find common abs. (& concs.) in them, thus making future coding ("Reading") of novels easier.

Reading novels will help TM answer Q's about "R.W." {R.W. = novels in which info is from either "reliable text" a/o TM's inputs (sensors) in R.W.

Also, in addition to using Q's & A's for ~~more~~ learning about novels, TM could, be asked to predict contins. of text. — Hvr., this may be redundant. — TM normally trys to <u>compress</u> <u>any</u> input text that it is given.

2) <u>Wolff's</u> methods of analyzing text to discover CFG's <u>can</u> be <u>very useful</u>. — For Text, certainly — but ~~for also~~ also for discovering abs. of <u>all</u> <u>kinds</u> — by "Chunking". Start w. simple lang; small vocab, small grammar, simple ideas, meanings, statements.
<u>I could ask</u> Q's about a specific, word, sentence or paragraph.

3) At first, I was uncertain how TM would treat "info" from a novel .... how it would know that t. "info" in t. novel was <u>specific</u> to <u>that</u> <u>novel</u> — & not others. Hvr., if we <u>label</u> Q's as .01–.04, T.M. <u>will</u> learn to answer Q's about a gn. novel, using <u>mainly</u> t. text of that novel.

4) Could I teach TM to read ~~a~~ a "Translated to Logic" version of that book on <u>Contract Bridge?</u> {William [redacted] Cohen; 8th Nat. Conf on A.I. AAAI-90 I pp 743—748

5) On lng. Grammar: I could <u>help</u> T.M. learn grammar by defining nouns, verbs, subject, predicate. ("Defining" means have T.M. learn these <u>defns.</u> as ~~e~~ <u>functions</u>: name of function
e.g. equation (2x+5=1) = "True" : These Defns can be considered to be "HINTS".

6) An early "Grammar": A <u>very</u> simple sentence str. Subject verb object. ||6PM||
"Subject" & "object" are nouns & can have adjectives attached. Verbs <u>may</u> have adverbs attached. There is one time sense for verbs: The sentences tell a story sequentially: (no way for 2 sentences to be simultaneous. Can we find a path to ~~me~~ slowly modify this grammar into a fuller CFG, a/o add temporal forms of verbs, ~~s~~ use sing, plural, male, female, etc to ~~assoc~~ associate nouns, verbs w. proper adjectives, adverbs .....

Subject Verb object can be translated as "Subject", "Object" are in relation "Verb."

1.8.99 **TM: PLAN**

I think I want to get TM to work a bunch of disparate problems: So I can see if there ~~wa~~ is a simple, general, way to get it to work all of them.

.03 (SN) General Remarks ① TM can use info if it is "Told" that info: just as a human can. — But (just as a human), T.M. will get much more out of t. info if it "understands" t. info: i.e. it has compressed it — preferably if t. compression also uses cones in t. rest of t. corpus. — i.e. If TM is able to connect t. new info w. ideas of t. past corpus.

② As TM Matures, it gets many more Cones: But this does not expand t. such Space, because assoc. w. t. new Cones. are also conditional P.D.'s that Narrow down when t. cones. should be used.

③ ① (.03) is impt., because it means we can "tell" TM things — like how to solve a problem or set of problems: This is easy on t. "Teacher". TM then ~~can~~ will (when it has time — just as a human) try to compress t. things it's been told" so as to fit into t. Macro corpus — & to compress t. Macro corpus.

Adobe Distiller & SS\ keeps.

j

k

④ We can perhaps "Give" TM t. set of heuristics that Lenat used in AM (if we could figure out what they are! — perhaps get them from Ken Haas).

⑤ I need to write down just why I felt that I was much much closer to finishing TM than ever before. This will enable me to focus on remaining unsolved probs (if any!). Sol 86 & 89 have much of t. outline of how TM is supposed to work; At those times, I had rather narrow ideas on what kinds of TSQ's to write. This has been much expanded since then. Also, while I had this idea of "Compressing t. Macro Corpus" I didn't ~~know~~ really understand it in a sufficiently exact way .... ~~also~~ in particular, how to Mix info from OZ & Inv probs, & to Mix info from several seq. extrapln. & finite set extrapoln. probs .... I didn't know how on any exact level.

⑥ One trick was "Working backwards" from a particular problem soln's to obtain a conc. net [redacted] starting [redacted] at t. primitive conc. level.

⑦ My fear is that I will begin work on a specific tsq. ~~area~~; get involved, get stuck, spend a yr. or 2 on it, then forget where I was going — i.e. forget the "Grand Picture". So I want to first make a clear, somewhat detailed desc. of t. "Grand Picture" — then get down to details. See → 7.01

7.01

1.8.99 TM: SPEECH Generation, ~~[illegible]~~ also Speech to text conversion **ANALOG Lrng** (.12)

(SN) Just how does TM code errors ("exceptions") when its doing a corpus of unordered finite objects? Say "operator induction". Well, in Op.ind., we have input & we get a pd. on output strings. There is no actual "error" involved. We might, hvr, have a stoch. operator that is ~~[illegible]~~ deterministic & almost always right. One way to do error correction:

.05 It has 2 output classes of object: ① (w. probly $(1-\epsilon)$ exact) it does its deterministic

.06 output. ② w. pc = ε it has a ~~[illegible]~~ general way of describing objects. The pc of such a descn is rather low, hvr. & it's multiplyd by ε, to make it worse.)

Another possy is that the ε of .05–.06 varies w. each output — so we are more

.10 certain of certain responses than others. (29)

TM lrns to speak by close emulation of a given person's spectrum, saying certain text; It then can translate to any new text. Then it travels (in reverse) (speech to text), see previous stuff (conservatively) for stuff on analog lrng ((98)95.24)

.12 : (98)95.24 : More on Analog Lrng: T.M. is given set of pairs (Q_i, A_i)

The Q_i's are English texts: words:

" A_i's are Time varying spectrum (or a better, perhaps filter params of supposed human vocal tract) (using maybe Kalman filters for tracking.) corresponding to a particular person (α) ~~saying~~ pronouncing Q_i.

From this, we want to induce ~~that person~~ α's pronouncing a new Q_j.

For some words, there is necessary ambiguity on pronunciation (like "read") This could be resolved by context. In fact, we can have this be a separate problem; to resolve various word meanings (≡ pronunciation) by context of text ... This is a purely "digital" problem.

T. ~~Gen~~ General Inverse of .12 ff is t. problem of "Voice Recognition" It fits in well w. TM's usual approach — i.e. given t. beginning of a sentence, ... "What is t. P.d. of t. next word?"

.29 : (10) (6.9.00) In unordered set of finite strings induction: t. code for a given finite object can be in 2 parts; ① code for an approximate object ② instructions on how to change approxn. into exact object. If α is ~~[illegible]~~ approxn. & β is desired object, we want condl. complexity of β w.r.t. α i.e. $\frac{pc(\alpha,\beta)}{pc(\alpha)}$ This condl. pc will be developed as TM matures: It will find that in certain environments, diffnt α → β xfmns are more likely.

Karby

~~[illegible]~~ 2rtinc.com

kjsmith@ballatlantic.net

1.11.99 **TM** : Grand Plan!

.01: 5.40 : 1) So, I start by solving a simple problem using Lsrch, & a very simple ~~[illegible]~~ unconditional P.D. I may solve several (~ or disparate) problems in this way.

2) Next : Updating I now have this Macro corpus of examples : I have to compress this Macro corpus. (It may or may not be a "Mixed corpus").

[Actually, does t. corpus need to consist ~~[illegible]~~ only of examples of problems that TM has solved?

— Not nec'ly ! Solved probs are nice because they are already compressed, parsed, "factored", — but at a certain stage in TM's education, I can put unfactored, uncompressed data into t. Macro corpus & have TM compress t. new Macro corpus. TM will be able to do this well, only if its able t. "recognize" t. ~~[illegible]~~ new data, so it ~~[illegible]~~ has ideas as to what compression methods to use on it → [see 3.20 on t. "op, ob algebra" idea]. "Recognition" can usefully occur only when ~~[illegible]~~ TM has accumulated an adequate set of obs (data) & ops (ragy types) & correlations between them.

3) Superficially, it looks like t. main problem is Updating.

4) An apparently difrnt problem: (a) Just how do we express various heurs as ~~[illegible]~~ (C.)P.D's? Conditional P.D.'s? How do we update ~~[illegible]~~ these ~~[illegible]~~ ~~[illegible]~~ c.P.D's? (i.e. change their pc's)
(b) Also, are all heurs expressable as CPD models? (e.g. "Quick Abort")

5) Well, re (4) It would seem that Heurs would always be of t. form: "I am in a certain situation & ~~[illegible]~~ my long term or short term goal is α" What is t. cPD of my trial actions? This may include "Quick abort": since one's goal is always "fast" soln. [On t. other hand, ~~[illegible]~~ Many prob solving methods make very few trials (usually only 1), but its a rather hy cc trial!]

6) (SN) Rereading [Sol 86] is not a bad way to organize my thots. I think I had the idea of very carefully constructed TSQ's — T. idea of "Perfect TSQ's & t. costs of corruption" was still ~~[illegible]~~ lingering in my mind. NOW, t. idea is that Order of learning is not very critical.

Another idea was that TM_1 & TM_2 were somewhat different — that t. "self improvement" of TM_2 would not occur untill TM matured enuf. Now, my impression is that TM_2 is "updating" & is done even by t. newborn TM!

7) One thing I haven't workd on ~~[illegible]~~ much (?): How do we get TM to work problems
.36 in ~ t. ~~[illegible]~~ way people do by concious deductive reasoning? → 9.10 — I did do work on that "Planning" heuristic in Sol 86: (Planner) It looks at problem, assigns probys to 4 (or more) difrnt ways of solving it. This is normally what I do now using condl. proby of problem, I get probys for various approaches: Each of these approaches can call on t. "Planner", which is t. condl. P.D. We can get potentially ∞ depth trees this way. → 8.01

1.13.98 probably "9" about TM Grand plan.

Does "RM" mean "Real Mathematician"

8

(Probably: AM (Lenat) may have meant "Artificial Mathematician".)

.01 : 7.40 So this begins to look like a common type of human problem solving: We have this very branched tree to search. If we are familiar with problem area some nodes will have only 1 path w. much proby.)

→ 8) So I'd like to make a list or tree or some kind of structure w. many kinds of problems in it — so I can see how TM would work on each type.

9) [redacted] RM doesn't have to do an actual problems when it is being trained. We can give it solved problems & then to "compress" soln it has to find how it would solve t. prob — so it uses Lsrch — same as if it had been given t. unsolved problem to solve → (27)

(.09)

We would want to give it [redacted] real probs. in its trng. to be ~~sure~~ sure that it was really learning useful stuff!

10) from ⑨ with this hyly branched tree: Ordering trials by pc or cc (Learn cst?) can be a dfflt problem. → It's a special case of t. general problem of getting [redacted] outputs of a P.D. in a (ruff pc order.) (for Lsrch)

At 2005: I think I know how to solve this!

838 9037 SouthCarrot. Cellular

11) The **General Plan** of RM: ⓐ It [maintains / has] at all times this "C.P.D." (also "Corpus"). This cpd can take several forms & it may be ~~mixed~~ a ~~mixture of these forms~~. (See 3.14–.17 for a small set of forms it may take.) (Also 3.20) It can (to an impt. extent) convert betw. difnt. forms →

.20

ⓑ It is able to use t. CPD of ⓐ to solve Inv. & OZ. probs. also to extrapolate, interpolate, w. probys, "get most likely set of (extra interpolations.

ⓒ It is able to (usual CPD) ① Improve t. CPD wrt t. present ~~corpus~~ corpus

② When t. corpus is augmented, it is able to ~~modify t. old CPD~~ to optimumly modify t. old CPD to ~~accommodate~~ accomodate t. augmentation: It will do this w. whatever cc is given to it.

(.27)(.09) 12) In general, RM need not have a TSQ: It can just have a big corpus that it continues to work on to "compress", to find regys in it, to obtain a better & better [redacted] CPD wrt that corpus. Hvr., if we do give TM a TSQ that's carefully thot out, TM should be able to get a v.g. CPD w. much less cc.

13) [redacted] So, t. history of TM is th. history of its CPD: T. corpus history is like this & t. corpus may be (well) regarded as a special form of t. CPD. (It has all of t. info, but much of it is at very hy cc.).

"updating" is one form of conversion betw CPD forms: like t "corpus" form to other forms

↑ Anomaly here (April 2005)

1.15.99 TM: HEURISTICS

Heuristics

[3/12/01: True esp. if we are able to lrn by trials. Slightly true if t. "conds" of t. CPD include traces of previous trials w/ t. present problem — so it's a "static" CPD, But it acts like an "adaptive" ~~CPD~~ CPD!

9

This will be a discussion of F.N. of PS of Sol 89. The idea is that "Blind Srch" guided by pc's, [essentially Lsrch] can be equivt. to any kind of Heuristic Srch.

Actually, I don't want to develop just what "Blind Srch" is & how it differs from "Heur srch"

— What I want to do is show how RM goes about solving ~~a~~ various specific kinds of probs.

I think this is prob'ly an adequate justifn. of FN PS Sol 89.

.05 I had t. idea that any valid heur. must be based on past experience a/o ~~a~~ (post or present reasoning (analysis). As such, it should be possl. to take t. experience a/o reasoning that gave rise to t. heur ~~& use it as a step~~ & get t. same heur in more logical, probabilistic form. See Bulg. 4.31 (1.3.01 pagenos) for a different, justifn. of FN PS of Sol '89 Hvr, also noted diffy in this ibid 4.13

very IMPT idea

.10+7.36 How logical reasoning ("analysis") fits into this must be examined. My impression is that t. laws of reasoning are discoverable as extremely hy pc (#1) rules for generating new ("valid") strings from old. — That RM could easily lrn such rules from a few

.14 examples of [redacted] adequate info content (bits cost). → (10.01)

T. FN of PS Sol 89 ~~&~~ said we'd have to know "traces of previous problems" to get t. desired heur. Well, normally, t. traces of previous probs have already been "abstracted" & they've been used to construct t. Cond. P.d. that Guides t. Lsrch that ~~selects~~ seeks ways to solve t. current problem.

(3/12/01) → [I'm afraid this unnaturally vaunted way "lrng by trials: some of this can be obtained by a CPD in which t. "conds" are t. problem. and previous trials their "traces"

.20:3+17 (SN) On forms a C.P.D. can take: Occasionally, a C.P.D. can take t. form of a set of representative examples, (w. perhaps (wts frequencies) [redacted].

~~&~~ Another(?) way: Storing solutions to problems that one expects will be ~ to future problems to be solved: (This facilitates OSL): One can Genz. t. example on t. spot (as in Expln. Based Genzn) or keep t. example ungenzd, but index it so we will know when it's relevant to a future problem.

A "CPD" is any data structure from which one can (using avbl cc) obtain conditional pc's.

T. corpus itself may be regarded as one (rather inefficient) form of C.P.D. Since one can (w. much cc) get ~~[illegible]~~ C.P.D's out of it.

.20 is also related to t. idea that a "P.D." can (& usually is) be a rather complicated object, that gives ~~out~~ out ~~[illegible]~~ pc's of various goodness, depending on how much cc ~~[illegible]~~ one ~~[illegible]~~ wants to spend. T. meaning of "Improving t. P.D." becomes unclear, because P.D's are only partially ordered w.r.t. "Goodness".

But, in general, one has this ~~[illegible]~~ Macro (usually Mixd) corpus, & one has a C.P.D. that one has computed, assoc. w. it. Then 2 things can occur.

1) One wants to improve t. CPD by using available cc: How best to do this?

2) One has augmentation of t. corpus w. a partial (= incomplete) or complete object.

How best to improve CPD wrt. t. augmented corpus (complete object)

" " " " " " " " " incomplete object & oriented toward ~~[illegible]~~

P.C. over continuations of t. incomplete object

1·15·99 TM

This is a purely Natural Domain

·01: 9.14: Logic has 3 aspects for TM: ① Learning to work problems in logical calculus ② Using Logic to solve problems of all kinds in Many domains. ③ learning common before ① & ②

6/9/00: Also, Logic can give a very short code to things that "follow logically" & very long codes to things that don't. Logic can be viewed as a constraint, that makes the search space much smaller. It is not ~~[illegible]~~ always of low cc, hvr, & if cc is too hy, the [illegible] reduction of search space size may not materialize w.t. available CB!

1.25.99 T/M. **OSL**

.01 : (98)154.40 : I guess what Wallace would say: That when α arrives, we will try to find t. best short code for "old corpus + α".

Rebutl So "t. best code for OC = "old corpus" does not ~~[illegible]~~ have all t. statistical info that we'd like it to contain: i.e. info on less short codes would be valuable in dealing w. "unexpected situations" (which usually eventually occur).

From a practical standpt. — t. shortest code is a very incomplete descrn of t. corpus. A more complete descrn would involve telling about many of t. less short codes that were found (& ignored) during t. srch for t. "shortest" one.

T. difference betw. t "SCIENCE" problem & OSL, is that in "Science" we usually knew what "not so good" (ideas, codes) to remember (retain). In t. case of OSL, it is usually not possbl.: There are too many (cones, sequences) "α" that have occurd only once, (or twice but w.o. enuf bits in them to warrant a "definition").

2.27.99 **TM** GENL: →UTILITIES←

GREAT BREAKTHRU!!

For some time, I've been concernd that ALP didn't solve t. ultimate problem of human behavior; i.e. "What to do next?": I assumd that one could devise a suitable "Utility funct", & that one would then try to Max. one's expected Utility wrt that function.

→ How one obtaind such a function was unclear. ←

The new idea is this! That t. above "Utility" Model, is merely one possible way to dscrb. human (or living creature) behavior. As such, ALP would say "Consider all possl. Utility funcs & behaviors they induce. From corpus, get pc of future behavior as a wtd mean over all these Utility funcs: Wt. being ∝ pc of past predn of corpus"

.12 —— i.e. Utility funcs are just another way of dscrbg (coding) behavior for predn. → (.20)

T. foregoing is a discn. of ~~[illegible]~~ Utility funcs' use to predict other people's (not myself) behavior. How can I use this idea (or any other method) to help me make ~~[illegible]~~ decisions for myself?

Is this Bordering on Q of asking TM. for "Whatever it is that I would like most" — & asking TM to figure out just what this is. (?). — A very ill defind Question! Also, I probably wouldn't want TM to do that!

(20) (.12) — That t. defn. of Utility is that which, (w. pc calcns) enables one to predict a person's behavior.

1/7/04 Some reasons that I don't want TM "finding out what I really want"

1) It could then use this model of me to manipulate me to further its ends.

2) "What I want" is based on ~~[illegible]~~ my limited power, intelligence: When these are increased much, "What I want" of t. past, becomes not very appropriate!

3) I could have TM. take (2) into account ~~[illegible]~~ my deciding what I would want w. new power, intelligence, but I'm uncertain of t. wisdom of this!

1·4·99 SMFt .

.01: (98) 181.40 : This previous stuff was very much like ANN. One way to do it: Say we have 10 good drivers for a gn. driven. excitation $F = \sum_{i=1}^{10} a_i \cdot P_{dri}$, (P_{dri} is t. precedty./ith driver) we then excite driven w.  . So we have 10 "wts".

We can update t. wts. in t. usual ANN fashion, since we can get t. slopes from t. derivatives — Hvr, t. amt. to move them is unclear. If we get t. second derivatives also, we could get a very exact optimum soln!

Examine this in detail: This second div could be applicable to speed up ANN in general, by a tremendous factor. (Tho already, using standard N.L. optzn., there is an enormous speed up.)

One effect is that I will probably want to use many fewer wts than normal in ANN exp.

Another possib. model: Excitation for driven = $\sum a_i F(b_i P_{dri})$

Here we have 20 params, but we may be able to solve for t. wts. in t. same way as before, w. first & second derivs.

Jaspers's "Hessian" approach. = Fisher Info

.01 ff needs a lot of work! First I have to review ANN — get a clear idea how back-prop. works. — Then, perhaps modify it suitably.

In estimating α & γ  of various DD's t. ideas in Stein's Paradox are critical; t. ssz's will be small, so I'll have to pool data.

SMFT

1.6.98 ~~TM~~: Stein's Paradox in Statistics: Efron, Morris, Sci Am May 77 pp 119 – 127 1½

~~[illegible]~~ Say we have ~~[illegible]~~ k collections of data; $i=1 \mid k$ Each w. its own mean, $\bar{y}_i$.

~~[illegible]~~ Say $\bar{Y}$ is grand mean ($= \overline{\bar{y}_i}$) ? (is $\bar{Y} = \frac{1}{k}\sum_{i=1}^{k}\bar{y}_i$?) ← for t. S.E. analysis, it's not nec'y to know: l = constant = 45 for all players.

.04 Then Stein uses ~~[illegible]~~ $\bar{Y} + c(\bar{y}_i - \bar{Y})$ as ~~an better estimator of~~ $\bar{y}_i$. $z = \bar{Y}(1-c) + \bar{y}_i \cdot c$

$= (1-c)\bar{Y} + c\,\bar{y}_i$

.05 $c = 1 - \frac{(k-3)\sigma^2}{\sum(\bar{y}_i - \bar{Y})^2}$ (= "shrinking factor") Maybe σ^2 is ~~[illegible]~~ or about $\bar{Y}$? —] ~~[illegible]~~ ~~[illegible]~~

(estimator of $\bar{y}_i$) $= \frac{(k-3)\sigma^2}{\sum(\bar{y}_i - \bar{Y})^2}\cdot\bar{Y} + \left(1 - \frac{(k-3)\sigma^2}{\sum(\bar{y}_i - \bar{Y})^2}\right)\bar{y}_i$ σ^2 must be $\sum_i\sum_j(\bar{y}_i - y_{ij})^2 / (k\cdot 45)$ ← data pts per player

.06 They got $c \sim .2$ for one case of batting averages.

They then divided data into 2 parts; made ~~[illegible]~~ c & mean estimates from first ½ of data — (in this case, $c < 0$ for large k)

Then checked deviations from means in second ½, & it worked much better → (No!)

than using $c = 1$ (= no shrinking factor)

My impression is that this is ≈ t. SMA problem. One reason it took to 1955 to discover it was reluctance of Statisticians to use much a priori info.

t. eqns. .04 – .05 would be a nice "rough & dirty" soln. to t. SMA problem.

.20 I'd need to know exact defn of σ^2, hvr. [It may be that σ^2 is the mean var. of t. players w.r.t. their averages, so $\sigma^2 = \frac{1}{k}\cdot\frac{1}{45}\cdot\sum_i\sum_j(\bar{y}_i - y_{ij})^2$: the average var. of t. players.]

On t. other hand, in my normal SMA soln., I really know what's going on,

& so I have more confidence in it. In fact, t. .04 – .05 eq. is not v.g. if

t. data sets vary very much in size. (e.g. E & M discuss size 2 for one baseball player — p 124 col. III)

In general, t. art. is rather vague, I'm not really sure what $\bar{Y}$ is — if each player has diffnt no. of cases.

I'm sure I know what $\bar{y}_i$ is, but I'm uncertain about $\bar{Y}$ & σ^2.

⟹ Is $\bar{Y}$ a wtd average of t. $\bar{y}_i$ or a simple mean of t. $\bar{y}_i$?

---

to get ~~[illegible]~~ c we want ratio of σ^2 to $\frac{\sum_{i=1}^{k}(\bar{y}_i - \bar{Y})^2}{k-3}$

Go to Harvard: Get that issue again — see if there are references (usually there are) — try last p. of t. issue] ← Got them! Got it!

Or: $\boxed{\sigma^2 = \frac{1}{k}\sum_{i=1}^{k}(\bar{Y} - y_i)^2}$ ← No! that's like t. denominator of (.05)

.08 R is ≈ nice defn. of σ^2: They assume σ^2 is t. same for each i.

In other papers → E & M 1975 p 53 p 314 – 316 they deal w. σ_i^2 varying w. i. My impression;

.35 $\sigma^2 = \frac{1}{k^2\cdot 45}\sum_i\sum_j(\bar{y}_i - y_{ij})^2$: this σ^2 is t. expected err² in the $\bar{y}_i$ — } I think t. coeff should be $\frac{1}{k\,l^2}$ not $\frac{1}{k^2 l}$

$\frac{1}{k\cdot l}\sum_i\sum_j(\bar{y}_i - y_{ij})^2$ would be indep of k, l for large k, l.

Hvr, ~~[illegible]~~ (.35) is $\propto \frac{1}{k}$ for large k, l as desired. ($\propto \frac{1}{l}$ is more likely)

1.16.99: SMFT ~~TM Gain~~ : Stein's Paradox: ("T. Stein Effect")

Often in statistics, t. ssz is so small that σ dominates t. mean, — So if one takes t. average of N observations, $\sqrt{\frac{\sigma^2}{N-1}} > \mu$: so one would do better using ϕ as t. mean, than using μ! "Doing better" means that using μ as a predictor has more MS error than using ϕ as a predictor.

.05 I observed this, using 30 window for averaging & I was much confused, disturbed. Now it seems clear!

.07 T. way one can deal w. this effect is by pooling "related" distributions. In the "Batting average" example in Sci Amer (May 77 pp119ff), we pool data for .09 difrnt players. A good model for this is t. one I used in my SOY analysis; Say t. μ_i's of t. various $\boxed{N}$ players has mean M and s.d. = S. Then each player has his own μ_i & σ_i. So before, in our analysis, we had those 2N params, μ_i & σ_i. Now we have 2N + 2 params.

Omitting pc's of t. 2N+2 params; t. pc of t. data is

.20 $$\prod_i \prod_j \frac{1}{S} e^{-\frac{(\mu_i - M)^2}{2S^2}} \frac{1}{\sigma_i} e^{\frac{(d_{ij}-\mu_i)^2}{2\sigma_i^2}} \; \longrightarrow \; \boxed{4.01}$$

We end up w. convolutions of d.f.'s w. s.d.'s of S & σ_i.

(.20 has to be worked out more carefully; It may have been in SOY work in .24 SMFT of previous yr (or 2).) → $\boxed{\text{See 3.13 for refs}}$

Anyway, this Stein effect is ~~very~~ impt. I was concernd in linear regression, that the coefs in t. cov. matrix could be of poor accuracy, so t. predns would be very poor. — In fact, this is so. If t. ssz isn't large enuf, t. default predn (say zero) will be better than t. predn. obtaind via linear regressn. In linear regressn, on can compare $\boxed{(N+k)\frac{\sigma^2_{\text{obsrvd w. l.r. model}}}{(N-k)}}$ (for k params) w. σ^2 of t. raw data (σ^2 about ~~zero~~ or about mean), to see if t. linear regressn helpd.

In SM, one can effectively ↑ ssz by "pooling" in at least 2 ways.

1) As in t. Soy analysis, (.09 –.20) pool data from related stocks, securities.
2) ↑ ssz by using longer smoothing time — even tho t. longer more distant past gives data w. somewhat diffrnt characteristics.

1.16.99 SMFT ~~[illegible]~~ : T. "Stein Effect".

An probably impt application is in SMFT! The determination of the params of the (Price of driver v.s. Price of driven) curve.

A possl. other applin: Could it be relevant to th. fact that giving driver drivers difent wts. didn't ↑ yield of FT?

AH! Is t. Stn. effect relevant to determination of whether a seq. is ~~[illegible]~~ low dim chaotic? (≈ Grassberger) felt that w.o. enormous SSZ, one ~~[illegible]~~ could get no useful info. Is this really true? — or is it an example of Statisticians not wanting to make predns unless they are 95% certain?

.13:1.24 Early work on SOY: In 97SMFT!
231.01 — 222.32 — maybe earliest ref.
232.37—.40;
233.30
237.22—239.21
242.23 : Ref. to older work using MAPLE ⓤ.
251.01—40 is a BIG Bibliog. review of SOY.
(98) 98.01 is also a sort of Review ... partly on SOY.

I really have to look at this Soy stuff & write a review.

Also there was t. SMA problem! May be closer yet to Stein

Q: Could Stein be relevant to t. even/odd diffy of (97 SMFT 223.33)?

---

(98)98.08—30 ff) Bibliogy on aspects of SOY!
also (credibly large RMS error in predns using x windo smoothing/ 87.01—89.4

Also Much work on ~~[illegible]~~ SOY in 97TM : 141.01—40; 147.01—40
153.01 to 166.40 ; 170.01—171.40; 200.01 ff
Backwards
~~→ 147.40 — 147.01; 141.40—[illegible]~~
92 etc

1997: 88.. said to be "final soln." to SMA problem. Also see 85.23—86.08
We can trace backwards from 83.01: (84, 82 are Maple pgm. & outputs.
(97) 90.05: Bibliographic review of SMA/AAE problem. / recent Thread starts on 31.01
Also see other reviewish ideas on 90.01—.40

---

97 SMFT
D1797 : 254.21
Driver w. hy σ ~~[illegible]~~ may be good drivers because they respond to "News" quickly

looking for Soy in SMFT!
(97) 242.23

Ibid (237.22 — 239.21): early work.

232.37—.40 is t. beginning of "reasonable" "Soy" work.
233.30 is not bad

237.22 looks good

INDEX Note: SMFT How to get Median of a set.
(98) 35.01 : This is for making √x windo median filters

1.18⁹⁹ SMFT "Stein Effect."

.01 Re eq. 2.20: Suppose we have a Grand d.f. for th. μs w. known s.d. = s
& known Mean = 0. We observe a bunch of data for a particular ball player:
[dj] its mean is μ, its s.d. = σ, K cases. $\mu = \frac{1}{k}\sum_j d_j$.
Viewed this way as f. pc of data is $\rho \propto \frac{1}{s}\cdot\exp -\frac{1}{2}\left(\frac{\mu^2}{s^2}\right)\cdot\frac{1}{\sigma}\exp -\frac{1}{2}\left[\left(\frac{\sum_j (d_j-\mu)^2}{\sigma^2}\right)\right]$ ↑ wt.

~~We can view μ, now as not merely, but just a variable & we want the functio~~ controversial

Now let us move the ~~[illegible]~~ of the players to μ+x

.10 $\rho \propto \frac{1}{s}\exp -\frac{1}{2}\left(\frac{(\mu+x)^2}{s^2}\right)\cdot\left(\frac{1}{\sqrt{\sigma^2+x^2}}\right)^k \exp -\frac{1}{2}\left(\frac{\sum d_j-\mu)^2+x^2}{(\sigma^2+x^2)}\right)$.

k σ² — Bug! This factor may be indep of x.

.11 $\ln\rho = -\ln s - \frac{1}{2}\frac{(\mu+x)^2}{s^2} - \frac{k}{2}\ln(\sigma^2+x^2) - \frac{1}{2}\frac{\sum(d_j-\mu)^2+x^2}{\sigma^2+x^2}$.

.12 $\frac{\partial\rho}{\partial x} = -\frac{1}{2}\frac{1}{s^2}\cdot 2(\mu+x) - \frac{1}{2}\frac{2xk}{\sigma^2+x^2} - \frac{1}{2}\frac{d}{dx}\left(\frac{k(\sigma^2+x^2)}{\sigma^2+x^2}\right)$

$-\frac{1}{2}\frac{d}{dx}\frac{k\sigma^2+x^2}{\sigma^2+x^2} = +\cancel{\frac{1}{2}}\frac{k\sigma^2+x^2}{(\sigma^2+x^2)^2}\cdot\cancel{2}x - \cancel{\frac{1}{2}}\frac{\sigma^2+x^2}{(\sigma^2+x^2)^2}\cdot\cancel{2}x$

$= \frac{(k-1)\sigma^2 x}{(\sigma^2+x^2)^2}$

$\frac{d}{dx}\frac{u}{v} = -\frac{u}{v^2}\frac{dv}{dx} + \frac{1}{v}\frac{du}{dx} = \frac{+uv' - vu'}{v^2}$

.20 $\frac{\partial\rho}{\partial x} = -\frac{\mu+x}{s^2} - \frac{x}{\sigma^2+x^2} + \frac{(k-1)\sigma^2 x}{(\sigma^2+x^2)^2} = -\frac{\mu+x}{s^2} + \frac{(k-2)\sigma^2 x + x^3}{(\sigma^2+x^2)^2} = 0$

$-\frac{\sigma^2 x + x^3}{(\sigma^2+x^2)^2}$

$-\frac{\mu+x}{s^2} + \frac{(k-2)x}{\sigma^2+x^2} = 0$ for >1 player $\sum\left(\frac{-(\mu+x_j)}{s^2} + \frac{(k_j-2)x_j}{\sigma_j^2+x_j^2}\right) = 0$

$(\mu+x)(\sigma^2+x^2)^2 = +s^2(k-2)(\sigma^2 x+x^3)$ which means if k > 2, then $\mu+x \geq 0$; ~~[illegible]~~ Always! —

Unlikely! Maybe Get Maple to do derivative! (It's on the third disk — Big book.)

It may be necy to reinstall it! Woops! This term can be < 0. $\frac{(\sigma^2+x^2)x}{\mu+x} > 0$.

.27 So $\frac{x}{\mu+x} > 0$: $\frac{1}{\frac{\mu}{x}+1} > 0$ $\frac{\mu}{x}+1 > 0$ so $\frac{\mu}{x} > -1$ $\frac{\mu}{x}$

if μ = 1, x > -1 so μ+x > 0 so if μ > 0, μ+x > 0

.29 if μ = -1, x < 1 so μ+x < 0 $\frac{1}{x} > 1$ if μ < 0, μ+x < 0

μ > 0 ⇔ μ+x > 0
Always μ < 0 ⇔ μ+x < 0

$\frac{x}{\mu+x} > 0$ if μ+x > 0; x > μ+x 0 > μ ⇐ This is reasonable!

$(\mu+x)(\sigma^2+x^2)^{\cancel{2}} = s^2(k-2)\cancel{(\sigma^2+x^2)}x$

.35 $\boxed{(\mu+x)(\sigma^2+x^2) = s^2(k-2)x}$ That if s² & if k-2 should both have same effect on soln. is ~~[illegible]~~ reasonable..

large k means small x, small s² means large x

$\left(\frac{\mu}{x}+1\right)(\sigma^2+x^2) = s^2(k-2)$ as $s^2(k+2)\to\infty$, $x\to 0$.

If ~~.12~~ .11 is correct $\frac{\partial\rho}{\partial s} = -\frac{1}{s} - \cancel{\frac{1}{2}}(\mu+x)^2\cdot -\cancel{2}\frac{1}{s^3}$ ∴

$\cancel{-1 = \left(\frac{\mu+x}{s}\right)^2}$ $= -\frac{1}{s} - \left(\frac{1}{2}(\mu+x)^2\right)\left(-\frac{3}{s^3}\right) = -\frac{1}{s} + \frac{3}{2}\frac{(\mu+x)^2}{s^3} = 0$

$1 = \frac{3}{2}\left(\frac{\mu+x}{s}\right)^2$ $\frac{\mu+x}{s} = \sqrt{\frac{2}{3}}$

1·18·99 SMCT Scatch effect

If quess last form of 4.10 & 4.11 is indip of x!

$$\frac{\partial \rho}{\partial x} = -\frac{\cancel{1}}{\cancel{2}} \frac{1}{s^2} \cdot \cancel{2}(\mu+x) - \frac{\cancel{1}}{\cancel{2}} \frac{2x\cdot k}{\sigma^2+x^2} \; ; \; \frac{\mu+x}{s^2} + \frac{x}{\sigma^2+x^2} = 0$$

we may need a factor "k" here: $\frac{\mu+x}{s^2} + \frac{kx}{\sigma^2+x^2} = 0$; $(\mu+x)(\sigma^2+x^2) = -ks^2x$ ~~which is~~

~~almost identical to 4.35~~ !

~~so~~ ~~$\mu+x < 0$ always~~ !

~~Unlikely~~ !

i.e. in 4.10 $\left(\frac{1}{\sqrt{\sigma^2+x^2}}\right)^k$

$(\mu+x)(\sigma^2+x^2) = -ks^2x$ which is opposite in sign to 4.35! — ∴

Unlikely i.e. 4.27 – .29 looks like what I want.

$$\frac{\mu+x}{x} < 0 \qquad \frac{\mu}{x}+1 < 0 \; ; \; \frac{\mu}{x} < -1 \qquad \text{so if } \mu=1, \; x<-1$$

$$\rho \propto \frac{1}{s} \exp -\frac{1}{2}\left(\frac{(\mu+x)^2}{s^2}\right) \cdot \left(\frac{1}{\sqrt{\sigma^2+x^2}}\right)^k \cdot \underbrace{\exp -\frac{1}{2}\left(\frac{k(\sigma^2+x^2)}{\sigma^2+k^2}\right)}$$

$$\ln\rho = -\ln s - \frac{1}{2}\frac{(\mu+x)^2}{s^2} - \frac{k}{2}\ln(\sigma^2+x^2) \qquad e^{-\frac{k}{2}} \text{ = const}$$

$$\frac{d}{dx}\ln\rho = \cancel{\;} - \frac{\cancel{2}}{\cancel{2}s^2}\,\mu+x - \frac{k}{\cancel{2}}\frac{\cancel{2}x}{\sigma^2+x^2} = 0$$

$$\cancel{\;} - \frac{\mu+x}{s^2} - \frac{kx}{\sigma^2+x^2} = 0 \quad \text{Bad!} \qquad \left(\frac{\mu}{x}+1\right)(\sigma^2+x^2) = -ks^2$$

as $ks^2 \to \infty$

$x \to 0$ but from " – " side.

$$\frac{e^{-\frac{(x+\mu)^2}{s^2}}}{(\sigma^2+k^2)^k} \to \left((\sigma^2+k^2)^k \cdot e^{\frac{(x+\mu)^2}{s^2}}\right) \text{— min}$$

$$\cancel{e^{\frac{(x+\mu)^2}{s^2}}} \cdot 2x\cdot k\,\cancel{(\sigma^2+x^2)^{k-1}} + (\sigma^2+x^2)^{\cancel{k}} \cdot \cancel{e^{\frac{(x+\mu)^2}{s^2}}} \cdot \frac{2(x+\mu)}{s^2}$$

$$\cancel{2}xk + (\sigma^2+x^2)\cdot\frac{\cancel{2}(x+\mu)}{s^2} = 0$$

.27 $$s^2kx + (\sigma^2+x^2)(x+\mu) = 0$$ all cubes w. real coeff have at least 1 real soln.

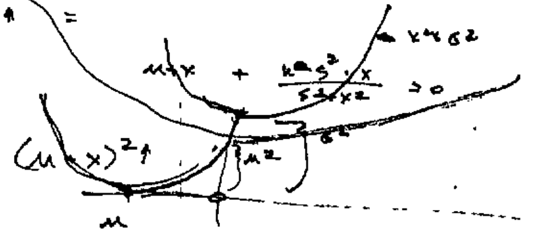
4.10 $$p = \alpha\cdot\left(\frac{1}{s}\, e^{-\frac{1}{2}\left(\frac{(\mu+x)^2}{s^2}\right)} \cdot \left(\frac{1}{\sqrt{\sigma^2+x^2}}\right)^k \right|\cdot\; e^{-\frac{1}{2}\frac{(\Sigma(d_j-\mu)^2)+kx^2}{k(\sigma^2+x^2)}} \Bigg)^{k}$$ $\alpha = \left(\frac{1}{\sqrt{2\pi}}\right)^{2}\cdot$

.30 $$\ln\rho = \ln\alpha - k\ln s - \frac{1}{2}\frac{(\mu+x)^2}{s^2} \Big| - \frac{k}{2}\ln(\sigma^2+x^2) - \frac{k}{2}$$

$$(-2\ln\rho)s^2 = \beta + (\mu+x)^2 + ks^2\ln(\sigma^2+x^2) \; ; \; = \beta + \boxed{(\mu+x)^2 + ks^2\ln(\sigma^2+x^2)}$$

$$\frac{d}{dx}\,\cdot = \qquad 2(\mu+x) + \frac{2ks^2}{\sigma^2+x^2}\cdot 2x = 0$$

$$\mu+x + \frac{ks^2\cdot x}{s^2+x^2} = 0$$



$$(\mu+x)(\sigma^2+x^2) + ks^2x = 0$$

slope = 0 must be at $x<0$, if $\mu>0$.

from the picture, x should be < 0 but $-x < \mu$

i.e. $x+\mu$ should be > 0.

$$\Big|\;\frac{\partial}{\partial s}\ln\rho = -\frac{k}{s} + \frac{2}{2}\frac{(\mu+x)^2}{s^3}$$

$$= 0$$

$$-k + \frac{2}{2}\left(\frac{\mu+x}{s}\right)^2 = 0$$

~~impossible~~

$$\cancel{\;} + \frac{2}{2}\,\Sigma\left(\frac{\mu_i+x_i}{s}\right)^2 =$$

< no "k" involved >

OH! its ok in .27 say μ = +1 ; x < 0 $(\sigma^2+x^2)(x+\mu) = -s^2kx > 0$; fine!

so x + μ > 0. as it should be!

$$\cancel{\frac{x+\mu}{x}} \quad \cancel{\frac{1}{s^2k}} \quad \cancel{= -\frac{1}{(\sigma^2+x^2)}}$$

1.18.99

$\sigma \frac{1}{2}$

$$\frac{(x^2+\sigma^2)(x+u)}{x} = -ks^2$$

$$\frac{x}{\sigma} + \frac{\sigma^2}{x} \quad (u+x) = \frac{-ks^2}{\sigma}$$

$$\frac{\sigma Y}{\sigma} = Y \qquad x = \sigma Y \qquad \boxed{(Y+\tfrac{1}{Y})(u+\sigma Y) = \frac{ks^2}{\sigma}}$$

$$(Y+\tfrac{1}{Y})(\tfrac{u}{\sigma}+Y) = \frac{ks}{\sigma^2}$$

5

2000 1

$$(Y+\tfrac{1}{Y})(\alpha+Y) = \beta$$

$$(Y^2+1)(\alpha+Y) = \beta Y$$

$$Y^3+\alpha Y^2+Y+\alpha = \beta Y$$

$$Y^3+\alpha Y^2+(1-\beta)Y+\alpha = 0$$

$$\frac{u+x}{s} = -\frac{ks^2x}{\sigma^2+x^2} \qquad \left(\frac{u+x}{s}\right)^2 = \frac{k^2s^2x^2}{(\sigma^2+x^2)^2}$$

Dr Chitany Haveitt.

1) colonoscopy — Whole Colonoscopy

2) Psych. tests ! Not Useful. Tested any kind of thing

3) dermitology o.k. — Low fluid → dizzy

4) Cough o.k.

Colonoscopy : chitany

church.

Blood pressure

~~ch~~ Geordinos Says 2 liters of fluid/d. ! ? Seems any hy! 140

fluid should have some Electrolytes in it. That Blood Press 132 / 78

110

Time was ~ 4 30 PM, none dizzyness him.

She said Hydra would not produce such peaks in Blood pressure. (?)

But fluids ~~wouldn't~~ ~~Become~~ likely Go to less Possible (any be no fluid at rate deficit?)

5.27: $-s^2 k x = (\sigma^2 + x^2)(x+\mu)$: $\frac{x+\mu}{s^2 k x} = -\frac{1}{\sigma^2+x^2}$: $1+\frac{\mu}{x} = -\frac{s^2 k}{\sigma^2+x^2}$

$1+\frac{\mu}{x} < 0$ $\boxed{\frac{\mu}{x} < -1 \quad \text{inconsis.} \quad \frac{x}{\mu} > -1 \quad x > -\mu \quad \mu < -x}$ $x+\mu > 0$ ← No! if $\mu < 0$

$\mu + x$ should be < 0.

~~Phone: S \$10K Pim w/ \$100 1" closer acct.~~

say $\mu = -1$ $1 - \frac{1}{x} < 0$ $-\frac{1}{x} < -1$ $\frac{1}{x} > 1$

$x < 1$ so $x - 1$ $(\equiv x+\mu) < 0$ $x + \mu < 0$.

I think my manipln. of inequalities is faulty, (inconsis).

$1+\frac{\mu}{x} < 0$

O.H. if $\frac{\mu}{x} < -1$ if I mult both sides by x I get $\mu < -x$ ← But only if $x > 0$! So Equs looks O.K.!

.11 eq. 4.10: using $\frac{\partial}{\partial x_i}$ (for the ith player) we get for each i, $\boxed{(\mu_i + x_i)(\sigma_i^2 + x_i^2) = -k_i x_i S^2}$

μ_i, σ_i known; x_i, σ_i^2, S^2 unknown

$-k_i x_i S^2$

~~Deliberately wrong~~

This can be configured as a 2 param. cubic eq. for x_i in terms of μ_i, σ_i, S.

---

we can do eq. 4.10 for many i — then take $\frac{\partial}{\partial S}(\ ) = 0$

.17 from (5.30 R) we (may) get $\boxed{\sum_i (\mu_i + x_i)^2 = \frac{2}{k} S^2}$ ($[k_i]$ are not involved) so $\boxed{\frac{1}{n}\sum_{i=1}^{n} (\mu_i + x_i)^2 = S^2}$ ← more reasonable

.19 ∝ So we can start w. initial S value! Solve .11 for all $(\mu_i + x_i)$; then get next approxn for S viz .17

I don't know if it converges, hvr. ☺ I could just try it; if it does converge — — in a few cases, ~~[illegible]~~ probly converges in all.

For eq. .11, perhaps use the formula in Sci Amer for first approxn.

For production runs, I'll update each day, using yesterday's values ~~for first approxn.~~ approxn.

On eq. (.11 R): I set $\frac{x}{\sigma} = y$: $y^3 + \alpha y^2 + \beta y + \alpha = 0$

Is this cubic particularly easy to solve? I certainly can remove 1 term from any cubic by linear subst.

$y = z + \Delta$ maybe get rid of const term? | gotten $\Delta^3 + \Delta^2\alpha + \Delta\beta + \alpha = 0$ is same as solving original eq! I think this is true for all poly eqns!

$$z^3 + 3z^2\Delta + 3z\Delta^2 + \Delta^3$$
$$z^2\alpha + 2z\Delta\alpha + \Delta^2\alpha$$
$$z\beta + \Delta\beta$$
$$+\alpha$$

z^2 or z is removable

$\boxed{\Delta^3 + 2\Delta\alpha + \alpha = 0}$

$y = z + \frac{\Delta}{z}$

$$z^3 + 3z\Delta + 3\frac{\Delta}{z} + \Delta^3$$
$$z^2\alpha \quad 2\Delta\alpha \quad + \frac{\Delta^2}{z^2}\alpha$$
$$z\beta + \beta\frac{\Delta}{z}$$
$$+\alpha$$

$z^3 \quad z^2 \quad z \quad \frac{1}{z} \quad 1 \quad \frac{1}{z^2}$

It may be possl. to solve .11 R by $x_i^{n+1} = -\frac{1}{k_i S}(\mu_i + x_i^n)(\sigma_i^2 + (x_i^n)^2)$

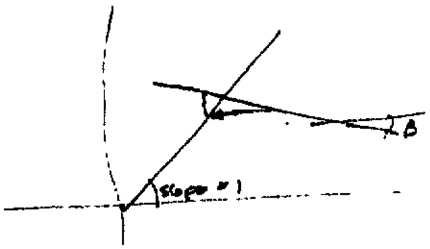
or $\frac{x^{n+1}}{\mu + x^n} = \frac{1}{\frac{\mu}{x^{n+1}} + 1} = f(x^n)$ so $x^{n+1} = \left(\frac{1}{f(x^n)} - 1\right)^{-1} \cdot \mu$

12099 SMFT Strain Effect.

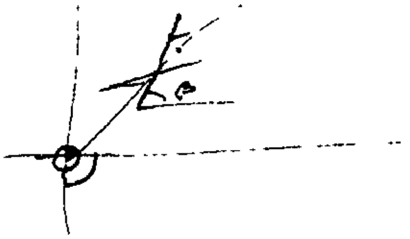
In general, if $x^{n+1} = f(x^n)$ doesn't converge, try $x^{n+1} = f^{-1}(x^n)$.

Also to speed convergence in  situation

.03 try ~~f~~ $x^{n+1} = \big(f(x^n) + ff(x^n)\big)/2$. Averages 2 last estimates.



if $\beta = -1$ doesn't converge
$\beta < -1$ diverges
if $\beta > 1$ diverges



So; if ~~f'(x) is small~~ $-1 < f'(x) < 1$ at the soln, the process will converge.

Since ~~f has~~ f^{-1} has slope $(f'(x))^{-1}$ if $f(x)$ doesn't converge, f^{-1} (will)!

If slope of f(x) is ±1 then we ~~only do anyway~~: can do .03.

In fact, even if the system diverges somewhat with slope < −1, .03 will get it to converge.

$38×2

Prepaid withdrawn on T. Coulter.

For slope > 1, we can also fix it, but it's a bit harder; we look
.21 at differences betw. $f(x)$, $ff(x)$ & $fff(x)$: They form a geom. series. from these, we extrapolate
.22 backward to expected $f^{-\infty}(x)$, & use it for next approx.
Then we do .21–.22 from that approx. to get next pts. etc.

.24 I'm quite uncertain as to the defn of $\bar{Y}$, $\bar{Y}_i$, & σ^2 (σ^2 in particular)

In general, I'd expect σ^2 to be $> \frac{1}{c}\sum_i (\bar{Y}_i - \bar{Y})^2$ so for large k, $c \ll 1$ — which is unreasonable. See (1½.20 R) for a more reasonable defn. of σ^2. → However (27 they say c can be < 0! which is not possl. for the (1½.20R) defn. of σ^2.

↳ Also 10.20–.23

Another Approach is suggested by (1½.04 R):

$Z = \bar{Y}(1-c) + \bar{Y}_i c$: which means Z is a wtd. mean of the 2 averages. $\bar{Y}$ & $\bar{Y}_i$ are both predns of the next data pt. of ~~a~~ player i.

Say ~~we [illegible]~~ s is the s.d. of the distribn of all data (w.o. player labels) & $\bar{Y}$ is ~~the~~ mean of same thing.: the pr. of derv. of the entire data set is $\left(\frac{1}{s}\right)^{TOT}$ using this model. (tot is total no. of players, trials)
↳ TOT = 45·k.

The pr. of derv. using individual $\bar{Y}_i$ & σ_i^2: $\prod_i \left(\frac{1}{\sigma_i^{45}}\right)$

The wt. ratio is $\left(\prod_i \left(\frac{\sigma_i}{s}\right)\right)^{45}$ — This Doesn't Seem Reasonable (~~[illegible]~~! i.e. Typical Σ^2 is always > the mean of the σ_i^2

I think the coding should use "Mixed corpus" formula

1.20.93 SMRT Stein. 8

.01 $\left(1\frac{1}{2} \cdot 05\right)$: $c = 1 - \frac{(k-3)\sigma^2}{\Sigma(\bar{y}_i - \bar{y})^2}$ $z = \bar{y} + c(\bar{y}_i - \bar{y}) = (1-c)\bar{y} + c\bar{y}_i$

$$\therefore \quad z = \frac{(k-3)\sigma^2}{\Sigma(\bar{y}_i-\bar{y})^2}\bar{y} + \left(1 - \frac{(k-3)\sigma^2}{\Sigma(\bar{y}_i-\bar{y})^2}\right)\bar{y}_i$$

$$= \underset{A}{\underset{\parallel}{\frac{(k-3)}{\Sigma(\bar{y}_i-\bar{y})^2}}} \cdot \left[\sigma^2\bar{y} + \underbrace{\left(\frac{\Sigma(\bar{y}_i-\bar{y})^2}{(k-3)} - \sigma^2\right)}_{\equiv \Sigma^2}\bar{y}_i \right.$$

.10 $$= A\left[\sigma^2\bar{y} + \Sigma^2\bar{y}_i\right] = A\sigma^2\Sigma^2\left[\frac{\bar{y}}{\Sigma^2} + \frac{\bar{y}_i}{\sigma^2}\right]$$

$A\sigma^2\Sigma^2$ prob'ly $= \frac{\Sigma^2\sigma^2}{\Sigma^2+\sigma^2}$: yes! – so/ its nothing but a normalizing facto.

so (01 R) & (.10) imply that $\bar{Y}$ & $\bar{Y}_i$ have/ rel' wts $\frac{1}{\Sigma^2}$ & $\frac{1}{\sigma^2}$ resp: ——

.16 This is equiv. to forming each product = wt. but using gaussian'd fs w. means $\bar{Y}$, $\bar{Y}_i$

.17 & vares: Σ^2 & σ^2 resp'l. We just taken the product of 2 Gaussn Df's w. difent μ's & σ^2.

→ 15.01 for more expect discussn.

Some Q's: ① ~~#~~ why k-3? (≠ k-2 via Akaike; but – maybe using 2 means, – but that would be k-4 .... so No!)

((② The result is equiv. to multiplying the 2 Gaussns of .16 together & obtaining a new Gaussian w. the mean, z. I don't see any way to justify this, tho. !

I really must ~~read~~ those papers! Try the Lastome Jun 1975 Journ Amer Stats Assoc.

Anyway: try recursion on 6.11 R!

.26

```
10  X = (m+x)*(Sgs+X*X)/K/Ss
20  input A:
30  Print X: Go to 10
```

"–"

$\sigma^2 = 2$, $S^2 = 4$, $k_i = 10$ $\mu = 1$ (mis irrelevant I think.)

Sgs = 2 Ss = ~~4~~ k = 10 m = 1

X = 1

$X = (\mu + x)(\sigma^2 + x^2)/(k s^2)$ ← should be "–". (!)

$x = -\frac{(\mu + x)}{2}$ $x(1+\frac{1}{2}) = -\frac{\mu}{2}$

$x = -\frac{\mu}{2(1+\frac{1}{2})} = -\frac{\mu}{2+1}$

Converged ~~x~~ x = 1, .05, .0525, .0527 ...... .0527089.

for k = 9 x = 1 .375, .1839, .1454 ... .1495658

k = 10, Ss = ~~4~~ → 3 : 1, .2, .0816 .... .07162441

Ss → 2.1 1, .285, .127, .108 .... .1059161

Ss → 4, m = 1 → 10 1., .825, .725, .677, ... .6415926 converged more slowly

m → 100 1 overflow 1, 7.5075, 159, 165k, 1.135E19 ... overflo diverges

m → 20 diverges

m → 15 diverges

m → 12 ~~conv~~ .975, .957 — .918

m → 13 diverges

m → 12.5 ~~diverges~~ conv

12.25 conv

12.3 conv

12.4 conv.

12.45 conv slower.

m → 12.47 conv.

12.48 conv

12.49 conv.

12.495 conv

12.499 conv

12.9 div

12.8 div slowly

12.7 conv

12.75 con

12.77 conv slowy

12.78 conv

12.79 conv. very slowly

12.795 diverges but slow to start.

12.8 div

1 20 99 SMPT STEIN ~~Stein~~

Try $\sigma^2 \approx s^2 = 2$ ⌐ other params of 8.26 conv

sgs = 4 conv. try sgs = 100 diverges rapidly.
Those changes (probably ↑) $f'(x)$.

sgs = 2, ss = 2, k = 1, m = 1 diverges. — so using large enough k + ss will usually converge.

k = 100 ; x → .206 ; k = 1000 : x → .02
k = 10 diverges.

Woops! $\mu = 1$, but $x > 0$! How come! OH! forgot "−" sign in ~~Eq.~~ 6.11

Convergence should be same here

sgs = 2, ss = ~~4~~, k = 10, m = 10 conv. sign > 0
sign "−" as it should be.

$x = -.542$

STEIN1.BAS

.17 $\boxed{x = -(\mu + x)(\sigma^2 + x^2)/(ks^2)}$ If $x^2 < \sigma^2$, then we have $x \approx -\frac{\mu + x}{z}$ (z large)

$x = -\frac{\mu}{z} - \frac{x}{z}$ $x(1+\frac{1}{z}) = -\frac{\mu}{z}$ $x \approx -\frac{\mu}{1+z}$ So $\frac{x}{\mu} = -\frac{1}{1+z}$: contraction fact.

.19 $\boxed{z = \frac{ks^2}{\sigma^2+x^2} \quad \mu \to \frac{\mu}{(1+\frac{1}{z})} \approx \mu(1-\frac{1}{z})}$ as per (10.10R) : So how did they get such a large contraction factor, w. k = 45 (which is large)? Maybe small s^2 & large σ^2?

Unless my eqns. are wrong! Maybe off by factor of k?

My contraction factor is ~ $\frac{\sigma^2+x^2}{\textcircled{k}\, s^2}$ (k: 45) Theirs is $\frac{(k-3)\sigma^2}{\Sigma(\bar{y}_i - \bar{y})}$ ← $\left(\frac{k-3}{\Sigma(\bar{y}_i-\bar{y})}\right) \approx \frac{1}{s^2}$

GOOD!

Well, its both same, their σ^2 is really $\propto \frac{1}{k}$ (see 18.35) ← → See 10.20–23

.25 So they have no "$\frac{1}{k}$" factor here

It could be part of exponent in 4.10 of $\left(\frac{1}{\sqrt{\sigma^2+x^2}}\right)^{k,l}$ should be l. (l).
k = 45 rather than no. of players. This inconsist notation is bad! I should use l instead of k ; l = no. of games per player.

---

75 : 6PM
1.20.99

Also, extending 4.10 to k players, we should get

$$p \propto \left(\frac{1}{s}\right)^k \exp -\frac{1}{2}\left(\frac{\Sigma(\mu_i+x_i)^2}{s^2}\right)\cdot\left(\frac{1}{\sqrt{\sigma_i^2+x_i^2}}\right)^{l}$$

This ↑ affects eq. ~~Eq~~ 6.17

$$\ln p \sim -k\ln s - \frac{1}{2}\frac{\Sigma(\mu_i+x_i)^2}{s^2} - \sum_{i=1}^{k}\frac{l_i}{2}\ln(\sigma_i^2+x_i^2)$$

$$\frac{\partial \ln p}{\partial s} = -\frac{k}{s} + \frac{2}{2}\frac{\Sigma(\mu_i+x_i)^2}{s^3} = 0$$; $ks^2 = \Sigma(\mu_i+x_i)^2$ $\boxed{s^2 = \frac{1}{k}\Sigma(\mu_i+x_i)^2}$

k

1.21.99 SMRT Stein

From Carter, Ralph paper (see SA. Bibl.)

X_i are mean values observed, at each of k locations ~~[illegible]~~ ($i = 1 \mid k$)

At each location the d.f. is known to be Gaussian of var = D_i. ($D_i = D$ = constant in this example)

True means θ_i, are unknown; to be estimated. usually $\theta_i \approx X_i$.

$\hat{\vec{\theta}} = (X_1 \ldots X_k)$ is max likelihood Estimator

$S = \sum (X_i - \bar{X})^2$ $\bar{X} = \frac{1}{k}\sum_{i=1}^{k} X_i$.

$\hat{B}(\vec{X}) = \min[1,$ ~~[illegible]~~ $\frac{k-3}{S} \cdot D]$ usually $= \frac{k-3}{S} \cdot D$ — & is ≤ 1 ~~[illegible]~~, but if $\frac{k-3}{S} D < 1$, then $\hat{B}(\vec{X}) = 1$.

So C & R get same as Efron, Morris, but

they define D more clearly! (and C/R discuss for a while) ;10

According to Tukey, the Stein effect is "usually very small."

Hvr E & M say Tukey was criticizing "Bradley's method only"

My own impression: If k is large, effect is small, & depends on k, σ^2, & S^2.

My interpretation of 1½.04–.05 is that if $k \to \infty$ & $\ell_i (\equiv \ell) \to \infty$,

that C would be about the same as it was for, say k = 20; ℓ = 20!

Hvr., this is completely unreasonable! If ℓ (the no. of times at bat (≥ 45 in Barrenaught!)) → ∞ then C should = 1; i.e. no "shrinking", since the $\bar{y}_i$ would be known for certain.

The only possy I can think of is that my interpretation of σ^2 is incorrect — that σ^2 is $\propto \frac{1}{\ell}$.

.20 i.e. σ_i^2 is, say, the uncertainty in $\bar{y}_i$. ! In which case $\sigma_i^2 = \frac{1}{\ell^2}\sum_{j}^{\ell_i}(\bar{y}_i - y_{ij})^2$

$$X = \frac{M}{1+Z}$$

$$X + M = \frac{M + ZM - M}{1+Z}$$

$$= \frac{ZM}{1+Z} = M\frac{Z}{1+Z}$$

$$= \frac{M}{1+\frac{1}{Z}} \approx \boxed{M\left(1-\frac{1}{Z}\right)}$$

Look at E&M's treatment of "Not all $\bar{Y}_i$ have same Var." §3: pp 314–316

p314 II: "Vars differ because of diffnt SSE's" (suggesting that (.20) is true!)

.23 — So it looks like (1½ .35) is the proper defn. of σ^2. $\boxed{\text{This Resolves the diffy of 9.19–.25}}$!

I could analyse their Baseball data from the batting average A, & var = ℓA(1−A) (= npq)

I could do an analysis of a system w known $\vec{\theta}$ & known true σ^2, etc,

& see what E & M's formula gets, v.s. what my formula gets.

Also, look at 8.16–.17: This is a very simple way to look at Stein! It's as if we had 2 sets of measurements for θ_i: One w. mean M & var Σ^2; another w. mean X_i & var σ_i^2.

Actually this is what occurs! We have 2 diffrnt ways to estimate M_i, & they each have their own mean & var. 8.16–.17 simply combines them by multiplying the 2 d.f.'s — & I think its pure Bayesian. Say $G_1(x)$ & $G_2(x)$ are the 2 d.f.'s obtained. The pc of x being the true mean is $\propto G_1(x) \cdot G_2(x)$. — i.e. Both measurements have to occur, & the product of these 2 (indep) pc's is the resultant pc.

So the Q is, "how do E&M get (k−3)"? — Akaike should be applied to both d.f.'s — Just exactly how is unclear.

12.01

1·24·99 SMFT <u>COVER</u> : (.01)

.01 Re Cover: When one looks at t. past, one can ① chose a single set of params for future bets. ② One can look at various params & chose t. single ~~[illegible]~~ set that gave max yield in t. past. ③ One can assume some d.f. for params in t. past & bet equal amt. on each (This is Cover's approach).

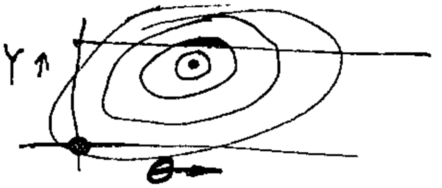
It <u>may be</u>, that w. "Multiplicative Strategies" (which I defined somewhere in t. last few yrs.), Cover's strat. is really best. T. major improvement may be that I now may be able to know how to best <u>update</u> t. d.f. based on subsequent behavior

It may be that t. main idea here is that Cover's method gives a v.g. estimate of yield, & t. others <u>don't</u>.

To start t. apriori d.f. for t. <u>expected</u> yield ~~of~~ for each param value is needed: $P(\vec{\theta}, Y)$. param, yield

~~[illegible]~~

Say we pick θ_0 ∋ Expected value of Y is max.



My impression is that Cover's method of playing may actually be <u>best</u>, if one uses only th. info that he uses. — But I have gone over this before: One objection that I had to Cover, was that <u>his</u> was a method of averaging out noise: so t. degree of goodness of his method, over simply picking t. "peak ō this far" should depend on t. noise level: & t. system doesn't seem to <u>deal</u> w. "noise level".

A big objection to Cover: As we ↑ no. of stocks in portfolio, his method ~~[illegible]~~ ~~starts to~~ converges more & more slowly to t. "best" param value, & highest yield.

It would be good if I could inject "side-info" in an optimal way. Cover's last SM paper was on "side info" but ~~[illegible]~~ my impression was that it was trivial: I don't ~~know how~~ remember ~~[illegible]~~ just what kind of "side info" he used.

1.25.99 SMFT : STEIN effect ;

.01 ; 10.40 : Start w. ≈ 4.10 & explain t. derivation in detail for future reference.

Do it for all σ_i^2's not t. same:

We have k players: The i^{th} player has been at bat l_i times.

~~[illegible]~~ At the j^{th} ~~[illegible]~~ "at bat" this player got a score of r_{ij}.

$\mu_i = \frac{1}{l_i}\sum_{j=1}^{l_i} r_{ij}$ is t. "batting average" of t. i^{th} player.

.06 $\sigma_i^2 = \frac{1}{l_i^2}\cdot\sum_{j=1}^{l_i}(r_{ij}-\mu_i)^2$ is ≈ the uncertainty in μ_i.

.07 ? → We will assume a Gaussian distribution for r_{ij} $(j=1,\dots l_i)$ about its mean, μ_i, with a variance of $\frac{1}{l_i}\sum_{j=1}^{l_i}(r_{ij}-\mu_i)^2 = l_i\,\sigma_i^2$.

.09 ? → We will also assume a Gaussian distribution for the means, μ_i, ~~but we will shift the origin so its at zero~~

$M = \frac{1}{k}\sum_{i=1}^{k}\mu_i$ is the grand mean.

— Its variance is $S^2 = \frac{1}{k}\sum(\mu_i - M)^2$.

~~Let us assume that the [r_{ij}] [illegible] were obtained in the following way!~~

Let us ~~[illegible]~~ make the following a priori assumptions about the ~~[illegible]~~ source of the data, $[r_{ij}]$

— That the observed r_{ij} were obtained by first choosing a μ_i' (to be determined) from a gaussian distribution of mean M and variance (to be determined) $(s')^2$ — then using μ_i' as mean and variance $l_i(\sigma_i')^2$ to be determined, the r_{ij} are obtained.

We will pick certain $(s')^2$ and $[(\sigma_i')^2]$ and $[\mu_i']$ and, as a function of those parameters, we will compute the probability of t. observed $[r_{ij}]$ being generated. — Call this $\tilde{P} = \tilde{P}((s')^2, [(\sigma_i')^2], [\mu_i'])$

Then via Bayes, this $\tilde{P}$ gives us the aposteriori probability distribution of these unknown parameters. If we pick the peak of this distribution, we will get a set of values for $(s')^2$, $[(\sigma_i')^2; \mu_i]$.

It turns out that these values are approximately what Stein got.

To get more accurate values, I would integrate $\tilde{P}$ in various directions to get the expected value of each of the parameters.

1.25.99 SMFT

Stein Effect

Formula for $\tilde{P}$:

.02 $\tilde{P} = \left(\frac{1}{\sqrt{2\pi}\, s'}\right)^{k} \cdot e^{-\frac{1}{2}\left(\frac{\sum (M-\mu_i')^2}{(s')^2}\right)} \cdot \prod_{i=1}^{k} \left(\frac{1}{\sqrt{2\pi}\,(\sigma_i')}\right)^{l_i} e^{-\frac{1}{2}\left(\frac{\sum(\mu_i'-r_{ij})^2}{(\sigma_i')^2}\right)}$

.03 $\ln\tilde{P} = -\frac{k}{2}\ln(2\pi) - k\ln s' - \frac{1}{2}\left(\frac{\sum(M-\mu_i')^2}{(s')^2}\right) + \sum_{i=1}^{k}\left(-\frac{l_i}{2}\ln(2\pi) - l_i\ln\sigma_i' - \frac{1}{2}\frac{\sum_{j=1}^{l_i}(\mu_i'-r_{ij})^2}{(\sigma_i')^2}\right)$

We will now take partial derivatives and set them to zero.

.10 $\frac{\partial\ln\tilde{P}}{\partial s'} = -\frac{k}{s'} + k\cdot\frac{\sum(M-\mu_i')^2}{k}\cdot\frac{1}{(s')^3} = 0$

$$\boxed{(s')^2 = \frac{\sum_i (M-\mu_i')^2}{k}}$$ ✓ (6.17 R)

.12 $\frac{\partial\ln\tilde{P}}{\partial\sigma_i'} = -\frac{l_i}{\sigma_i'} + l_i\,\frac{\sum(\mu_i'-r_{ij})^2}{l_i}\cdot\frac{1}{\sigma_i^3} = 0$

$$\boxed{(\sigma_i')^2 = \sum_j\frac{(\mu_i'-r_{ij})^2}{l_i}} = l_i\sigma_i^2 + (\mu_i'-\mu_i)^2$$

$\frac{\partial\ln\tilde{P}}{\partial\mu_i'} =$

$= +\frac{M-\mu_i'}{(s')^2} - \sum_{j=1}^{l_i}\frac{(\mu_i'-r_{ij})}{(\sigma_i')^2} = 0$

$$\boxed{\frac{M-\mu_i'}{(s')^2} = \frac{\sum_{j=1}^{l_i}(\mu_i'-r_{ij})}{(\sigma_i')^2}}$$

$(s')^2 = (M-\mu_i')\left(\frac{\sum(\mu_i'-r_{ij})^2}{l_i}\right) \Big/ \sum_{j=1}^{l_i}(\mu_i'-r_{ij})$

$= (M-\mu_i')\left(\frac{\mu_i'-\mu_i}{\;}\right)$

$\mu_i = \frac{1}{l_i}\sum_{j=1}^{l_i} r_{ij}$

$(s')^2\cdot(\mu_i'-\mu_i) = (M-\mu_i')\left(\frac{\sum(\mu_i'-r_{ij})^2}{l_i}\right) \equiv \frac{\sigma_i^2+(\mu_i-\mu_i')^2}{\boxed{l_i}} \Leftarrow$

.10R & .12R can be simplified, i.e. $(s')^2 \to s^2 + \frac{1}{k}\sum(\mu_i-\mu_i')^2$ I think

$(\sigma_i')^2 \to l_i\sigma_i^2 + (\mu_i'-\mu_i)^2$

$(s')^2\cdot\underbrace{(\mu_i'-\mu_i)}_{y_i} = \underbrace{(M-\mu_i')}_{(\mu_i+y_i)}\underbrace{(l_i\sigma_i^2 + (\mu-\mu_i)^2)}_{\sim\sigma_i^2+x^2}$

So the factor of l_i' in 6.11 is not there. There are serious differences in definitions betw. σ_i^2 & $(\sigma_i')^2$. (a factor of l_i)

Also possib. error by factor of l_i' in the work on 6.11, .17.

To simplify calcs: use 12.06: $\sigma_i^2 = \frac{1}{l_i}\sum(\ )^2$ not $\frac{1}{l_i^2}\sum(\ )^2$. We end up w. diffrnt eqns. ~~but the difference is a~~ from Stein, but the needed correction is obvious.

1·26·99 SMFT STEIN effect:

Better way: We have this game with k players: When the i^{th} player played his j^{th} game, he got a score of r_{ij} — he played a total of ℓ_i games.

problem is to obtain good estimate of player i's score on next game: $r_{i,j+1}$.

· $\frac{1}{\ell_i}\sum_{j=1}^{\ell_i} r_{ij}$ is one reasonable estimate. Stein has a better estimate.

Assume the $[r_{ij}]$ were obtained starting from a grand Gaussian distribution $N(M, s^2)$ which generates indep. $[\mu_i]$.

These $[\mu_i]$ are in turn used in indep $N(\mu_i, \sigma_i^2)$ which generate the r_{ij}.

M, s^2, $[\mu_i, \sigma_i^2]$ are all ~~un~~ initially unknown params. We will choose them so that the set $[r_{ij}]$ has max probability (density).

The probability density of $[r_{ij}]$ is then. (§ 13.02)

~~$\tilde{p} =$ [illegible]~~

$$\tilde{p} = \left(\frac{1}{s\sqrt{2\pi}}\right)^{k} e^{-\frac{1}{2}\left(\frac{\sum(M-\mu_i)^2}{s^2}\right)} \cdot \prod_{i=1}^{k}\left(\frac{1}{\sigma_i\sqrt{2\pi}}\right)^{\ell_i} e^{-\frac{1}{2}\left(\sum_{j=1}^{\ell_i}\frac{(\mu_i-r_{ij})^2}{\sigma_i^2}\right)}$$

we want $s, M, [\mu_i, \sigma_i] \ni \tilde{p}$ is max.

$$\ln\tilde{p} = \frac{k}{2}\ln(2\pi) - k\ln s - \frac{1}{2}\left(\frac{\sum_{i=1}^{k}(M-\mu_i)^2}{s^2}\right) + \sum_{i=1}^{k} -\frac{\ell_i}{2}\ln(2\pi) - \ell_i\ln\sigma_i - \frac{1}{2}\frac{\sum_{j=1}^{\ell_i}(\mu_i-r_{ij})^2}{\sigma_i^2}$$

$$\frac{\partial\ln\tilde{p}}{\partial s} = -\frac{k}{s} + \frac{\sum_{i=1}^{k}(M-\mu_i)^2}{s^3} = 0 \quad \text{so} \quad \boxed{s^2 = \frac{\sum_{i=1}^{k}(M-\mu_i)^2}{k}}$$

$$\frac{\partial\ln\tilde{p}}{\partial s} = -\frac{\ell_i}{\sigma_i} + \frac{\sum_{j=1}^{\ell_i}(\mu_i-r_{ij})^2}{\sigma_i^2} \quad \text{so} \quad \boxed{s_i^2 = \frac{\sum(\mu_i-r_{ij})^2}{\ell_i}}$$

$$\frac{\partial\ln\tilde{p}}{\partial\mu_i} = \frac{\cancel{\sum}(M-\mu_i)}{s^2} - \sum_{j=1}^{\ell_i}\frac{(\mu_i-r_{ij})}{\sigma_i^2} = 0 \quad \text{so} \quad \boxed{\frac{M-\mu_i}{s^2} = \sum_{j=1}^{\ell_i}\frac{(\mu_i-r_{ij})}{\sigma_i^2}}$$

$$\frac{\partial\ln\tilde{p}}{\partial M} = \frac{\sum_{i=1}^{k}(M-\mu_i)}{s^2} = 0 \quad \text{so} \quad \sum_{i=1}^{k}(M-\mu_i) = 0 : \quad \boxed{M = \frac{1}{k}\sum_{i=1}^{k}\mu_i}$$

1.22.99 SMPT Stein Effect

.01: ≈ 8.17 : A simpler view of Stein Effect: we have this data on k players:

$i = 1|k$, $j = 1|\ell_i$: r_{ij} is score of i^{th} player on his j^{th} game: The i^{th} player played ~~[illegible]~~ a total of ℓ_i games:

.04 Q: ① What is good estimate of $\underline{r_{i,\ell_i+1}}$: ? ② what is a good estimate of $\overset{\beta}{=}\ \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} r_{i,j+n}$

We can make estimates 2 ways:

Assumes ℓ_i is indep of i ∴ all ℓ_i's the same.

.06 If $\mathcal{M}_i' = \frac{1}{\ell_i}\sum_{j=1}^{\ell_i} r_{ij}$; $M' = \frac{1}{k}\sum_{i=1}^{k}\mathcal{M}_i'$ and $(S')^2 = \frac{1}{k}\sum(\mathcal{M}_i'-M')^2$ }

$(\sigma_i')^2 = \frac{1}{\ell_i}\sum_{j=1}^{\ell_i}(\mathcal{M}_i'-r_{ij})^2$

.07 ① $\mathcal{M}_i'$ is an estimate for : its expected error^2 for (new pred'n of r_{ij}) is $(\sigma_i')^2\frac{\ell_i+1}{\ell_i-1}$ (GA Keriko)

2.20.99 shouldn't $(S')^2$ include extra $(\sigma_i')^2$?
$(S')^2$ should be as error in predicting each r_{ij} score
so $(S')^2$ is more like

$\frac{1}{k}\sum_i(\mathcal{M}_i'-M')^2+\frac{1}{\ell_i}\sum_{j=1}^{\ell_i}(\mathcal{M}_i'-r_{ij})^2$

$\boxed{=\frac{1}{k}\sum_i(\mathcal{M}_i'-M')^2+(\sigma_i')^2}$

Its expected error for β is: $\approx \frac{(\sigma_i')^2}{\ell_j-1}$

.15 ② M' is also an estimate of β : Its expected error is $\approx \frac{(S')^2 \cdot k}{k-1}$

Using ~~[illegible]~~ method ①, the probability density of the value ~~[illegible]~~ β_i ($\equiv\beta$) for i^{th} player

$$P_1(\beta_i) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma_i'}e^{-\frac{1}{2}\frac{(\beta_i-\mathcal{M}_i')^2(\ell_i-1)}{(\sigma_i')^2}}$$

Using ~~[illegible]~~ method ②, the prob'y density, β_i

$$P_2(\beta_i) = \frac{1}{\sqrt{2\pi}}\frac{1}{S'}e^{-\frac{1}{2}\left(\frac{(M'-\beta_i)^2}{(S')^2}\frac{k-1}{k}\right)}$$

.25 If we assume
If we assume the 2 methods of estimate of β_i are indep.
If we assume

(they are not, since both contain data r_{ij} $(j=1|\ell_i)$ — but they are not
actually, this is a small part of ②'s data; we can subtract it

.29 out or subtract out its effect on the final estimate). —

~~[illegible]~~ $P_1(\beta_i)\cdot P_2(\beta_i)$ is the prob'y of having made both estimates, & this is a Bayesian d.f. for β_i.

Its close to Stein's estimator.

The present estimate is a wtd mean of $\mathcal{M}_i'$ & M' : The wts are $\propto \frac{\ell_i-1}{(\sigma_i')^2}$ & $\frac{k-1}{k(S')^2}$, resp.
Stein's results are on 7.01 – .17 his wts are somewhat ~~[illegible]~~ difrnt ! $\frac{1}{\sigma^2}$ & $\frac{k-3}{k(S')^2}$
relation of σ^2 & $(\sigma_i')^2$ is prob'y $\sigma^2 \approx \frac{(\sigma_i')^2}{\ell_i-1}$.

53.01

1.28.99 SMFT STEIN effect.

.01 : 15.01 – .40 can be a rather exact analysis!

We try to estimate β (of 15.04) in 2 ways. Both are, by themselves, legit ways. Estimate via 15.07 ① ; or via 15.15 ② — But in ②, to make the 2 methods indep., we should delete / direct info about the "i" to be predicted — this will give corrections to μ' & s'.

SN : An annoying problem: If ~~$~~ μ' & s' are to characterize the distribution of the [redacted] (β_i)'s (15.04R): Then ~~$~~ $(s')^2$ will be excessively large — by the mean error² in the (μ_i')'s. Also, μ' will have an expected error². So error² in estimating (β_i)'s will be due to errors in μ' & the (corrected) $(s')^2$.

I discussed this pt. much a long time ago in the Soy protein work..

If all the σ_i^2 were the same, the resultant d.f. would be convolution of the μ', s' d.f. & the μ_i', σ_i' d.f.'s. So if ~~Σ~~ $\tilde{\sigma}^2$ is the common $(\sigma_i')^2$

Then observed $(s')^2 = \tilde{\sigma}^2 + (s^2_{true})$.

If all n_i's were ~~[illegible]~~ the same $\equiv n$, but $k \to \infty$; ~~[illegible]~~ errors² in μ' and ~~$~~ $(s')^2 \to 0$: error in $\tilde{\sigma}^2$ is also zero, so we know s^2_{true} and μ'_{true}. In this case, 15.07 & 15.15 ~~would be the same~~ (if we delete proper data — tho since $k = \infty$, it would make no difference) can be regarded as 2 ways to estimate β_i, and we can then just take the product of the 2 Normal d.f.'s, — the mean of the resultant d.f. is the $\frac{1}{\sigma^2}$ & $\frac{1}{s^2}$ wtd mean ~~of~~: μ_i and μ' resp.

The Stein effect does occur if $k \to \infty$, but not if $n_i \to \infty$.

SN One of the main impt. ideas in this Stein discussion, is that in all pred'n cases, one must watch the SSZ & errors², to be sure they are not much larger than the errors² one would obtain using some other method of pooling data — or just using zero as pred'n.)

I did run into this problem in the past, & I didn't realize what was going on! I was using F winds to smooth a T.S.; the output I got, had larger M.S. error in prediction than using zero for prediction! If the true mean is close to zero, then the smoothed estimate has to be quite small, or else it will add its value² to the value² of the signal to give larger errors than the value² of the signal!

The large idea would seem to be very impt. in SM prediction — since the mean yield is usually quite small.

1·28·99 SMRT Stein effect

Hvr, if we are trying to tell which stock of several to bet on, these simple (erroneous) preds may all have about the same additive error (or multiplicative error) — so picking the apparently best predicted stock would be easy.

SN Another Q is that for baseball players, $0 \leq \mu_i < 1$, so it can't be Gaussian!

The distribution $p^{N_1}(1-p)^{N_2}$ is more reasonable. p = prob of "hit", N_1 is no of hits, N_2 is no of non-hits. N_1+N_2 = no. times at bat.

To normalize, [redacted] $\frac{N_1!N_2!}{(N_1+N_2+1)!}$: ($\int_0^1 p^{N_1}(1-p)^{N_2} = \frac{N_1!N_2!}{(N_1+N_2+1)!}$ is related to Beta funct of N_1, N_2. P258 Bureau of Standards

For the "Grand" distribn of the μ_i's: it also has to be between 0 & 1: we could try the 2 param. d.f. $p^{T_1}(1-p)^{T_2}$. A first approx: to T_1, T_2: upside down!

Say μ_i are the empirical means of players [redacted] (15.06).

.14 Then we want $T_1, T_2 \ni \prod_{i=1}^{k} \frac{T_1!T_2!}{(T_1+T_2+1)!} \mu_i^{T_1}(1-\mu_i)^{T_2}$ = max.

Take the k-th root: $\frac{T_1!T_2!}{(T_1+T_2+1)!} \cdot \underbrace{\left(\prod_{i=1}^{k}\mu_i\right)^{T_1/k}}_{\gamma} \cdot \underbrace{\left(\prod_{i=1}^{k}(1-\mu_i)\right)^{T_2/k}}_{\delta} = \frac{T_1!T_2!}{(T_1+T_2+1)!}\gamma^{T_1}\delta^{T_2}$

$x \equiv T_1$, $y \equiv T_2$

$x! \approx \frac{x^x}{e^x} \cdot \sqrt{2\pi x}$

$\ln x! = x \ln x - x + \frac{1}{2}\ln 2\pi + \frac{1}{2}\ln x$

.20 $\ln x! + \ln y! - \ln(x+y)! - \ln(x+y+1) = x\ln x + y\ln y - \cancel{x} - \cancel{y} + \ln 2\pi + \frac{1}{2}\ln x + \frac{1}{2}\ln y$

$- x\ln(x+y) - y\ln(x+y) + \cancel{x} + \cancel{y} - \frac{1}{2}\ln 2\pi - \frac{1}{2}\ln(x+y) - \ln(x+y+1)$

$= x\ln\frac{x}{x+y} + y\ln\frac{y}{x+y} + \underset{\text{omit}}{\left(\frac{1}{2}\ln(2\pi)\right)} + \frac{1}{2}\ln\frac{x}{x+y} + \frac{1}{2}\ln\frac{y}{x+y} - \ln(x+y+1)$

say $u \equiv x+y$; $y = u-x$

$= \boxed{(x+\tfrac{1}{2})\ln\frac{x}{u} + (u-x+\tfrac{1}{2})\ln\overset{\ln(1-\frac{x}{u})}{\frac{(u-x)}{u}} - \ln(u+1)}$

$(x+\frac{1}{2})\ln x - (x+\frac{1}{2})\ln u$

$-(u-x+\frac{1}{2})\ln u$

$+(u-x+\frac{1}{2})\ln(u-x)$

$\frac{\partial}{\partial u} = \frac{(x+\frac{1}{2})}{u} - \frac{(u-x+\frac{1}{2})}{u} + \frac{(u-x+\frac{1}{2})}{u-x} - u\ln u + u\ln(u-x) - \frac{1}{u+1} + \ln\delta = 0$

add in $\ln(\gamma^x \cdot \delta^y)$

$x\ln\gamma +$

$(u-x)\ln\delta$

$\frac{\partial}{\partial x} = \frac{x+\frac{1}{2}}{x} + \ln x - \ln u - \frac{(u-x)+\frac{1}{2}}{u-x} - \ln(u-x) + \ln\gamma = 0$

$= \frac{1}{2x} - \frac{1}{2(u-x)} + \ln\frac{x}{u(u-x)} = -\ln\gamma$

$\frac{1}{2}\ \frac{u-2x}{x(u-x)}$

This stuff may simplify, but I'm not at all certain of my algebra!

12999 SMFT Stem Effect. $\int_0^1 p^{u-1}(1-p)^{v-1}dp = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)} \equiv \beta(u,v)$ Beta function 18

.01 $F \equiv \int_0^1 p^x (1-p)^y = B(x+1, y+1) = \frac{\Gamma(x+1)\Gamma(y+1)}{\Gamma(x+y+2)} = \frac{x!\,y!}{(x+y+1)!}$

$\frac{\partial f}{\partial x} = \left(\frac{\Gamma'(x+1)}{\Gamma(x+1)} - \frac{\Gamma'(x+y+2)}{\Gamma(x+y+2)}\right) F(\ )$

$\frac{\partial f}{\partial y} = \left(\frac{\Gamma'(y+1)}{\Gamma(x+1)} \quad \frac{\Gamma'(x+y+2)}{\Gamma(x+y+2)}\right) F(\ )$

$\frac{d}{dx} xyz =$

$\left(\frac{1}{x}\frac{dx}{dx} + \frac{1}{y}\frac{dy}{dx} + \frac{1}{z}\frac{dz}{dx}\right) xyz$

$\frac{df}{\partial x} = \frac{\partial}{\partial x} \ln \Gamma(x+1) - \frac{\partial}{\partial x} \ln \Gamma(x+y+2) = \psi(x+1) - \psi(x+y+2)$

$\psi(\ ) \equiv$ digamma funct. $=$ Psi funct.

lots of stuff on ψ Bur Stand. pp 258 - 260

so we need to solve $\psi(x+1) - \psi(x+y+2) = A$, $\psi(y+1) - \psi(x+y+2) = \beta$ for various A, B points

~~$\psi(x+1) - \psi(y+1) = c$~~

$\psi(w) - \psi(w+v) = A$

$\psi(v) - \psi(w+v) = B$

$\psi(w) - \psi(v) = A - B (\equiv c)$

$\psi(w) = \psi(v) + c$

If $w = x+1$, $v = y+1$

NB $\psi(z+1) = \psi(z) + \frac{1}{z}$ (!!)

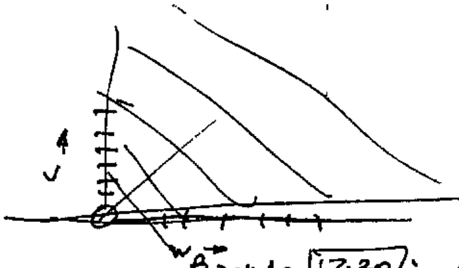
$\psi(1.5) \simeq 0$, so

$\psi(z)$ is "linear" ∴ $\ln(z-1)$ for large z

$\psi(1.5+n) = 0 \quad \frac{1}{1.5} + \frac{1}{2.5} + \frac{1}{3.5} \cdots + \frac{1}{n+.5}$

$\sim \ln n$

$\psi(1) = -\gamma$; $\psi(n) = -\gamma + \sum_{j=1}^{n-1} \frac{1}{j}$



Back to $\boxed{17.20}$: in terms of x, y :

$F = x \ln x + y \ln y + \frac{1}{2}\ln x + \frac{1}{2}\ln y - x \ln(x+y) - y \ln(x+y) - \frac{1}{2}\ln(x+y) - \ln(x+y+1) + x \ln \gamma + y \ln \delta$

$\frac{\partial f}{\partial x} = \textcircled{1} + \ln x + \frac{1}{2x} - \ln(x+y) \boxed{-\frac{x}{x+y} - \frac{y}{x+y}} - \frac{1}{2}\frac{1}{x+y} - \frac{1}{x+y+1} + \ln \gamma = 0$

$\ln x - \ln(x+y) + \frac{1}{2x} - \frac{1}{2(x+y)} - \frac{1}{x+y+1} + \ln \gamma = 0$

$\frac{\partial f}{\partial y} = \ln y - \ln(x+y) + \frac{1}{2y} - \frac{1}{2(x+y)} - \frac{1}{x+y+1} + \ln \delta = 0$

$\ln x - \ln y + \frac{1}{2}\frac{1}{x} - \frac{1}{2y} = \underbrace{-\ln \gamma + \ln \delta}$

~~$u = \frac{1}{2x}, v = \frac{1}{2y}$~~

$u = \frac{1}{2}\frac{1}{x}$, $v = \frac{1}{2}\frac{1}{y}$

$\ln \frac{v}{u} + (u - v) = c$

$\frac{v}{u} + (e^{u-v}) = e^c$

$\ln v - \ln u + u - v = c$

$(\ln v - v) - (\ln u - u) = c$

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

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

$\ln(x - \frac{1}{2})$

$\psi(x) \equiv \Gamma'(x)/\Gamma(x) \simeq \ln(x - \frac{1}{2})$

so $\boxed{\Gamma'(x) \simeq \Gamma(x) \cdot \ln(x - \frac{1}{2})}$

$\ln x - 1$	$\psi(x)$	x	~~$\ln(x - \frac{1}{2})$~~
4.595	4.60	100	4.600
3.89	3.90	50	3.90
3.178	3.198	25	3.198
2.197	2.25	10	2.25
1.386	1.506	5	1.504
	.922	3	.916

$\ln w - \ln(w+v) = A$

$\ln v - \ln(w+v) = B$

$\gamma = \left(\prod_{i=1}^{k} \mu_i\right)^{\frac{1}{k}}$ $\quad \delta = \left(\prod_{i=1}^{k} (1 - \mu_i)\right)^{\frac{1}{k}}$

.3 $\quad$ 1 - 3

$\frac{w}{w+v} \approx e^A = \gamma$ ∴ $e^A + e^B$ must $= 1$!

$\frac{v}{w+v} \approx e^B = \delta$

so is $\gamma + \delta \cong 1$? In a case I tried — ≈ true!

.2 .3 .5 → .31

.8 .7 .5 → .654

.964

Let's go thru 18.01 ff more carefully!

$c\int_0^1 p^{x'}(1-p)^{y'} = 1$! $\int_0^1 p^{x-1}(1-p)^{y-1} = B_{\text{also}}(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$

$x' \equiv x-1 \quad y' \equiv y-1$

so $c \equiv \frac{\Gamma(x+y)}{\Gamma(x)\Gamma(y)}$ we want $\overset{x',y'}{x,y} \ni \prod_{i=1}^{K} \frac{\Gamma(x+y)}{\Gamma(x)\Gamma(y)} \mu_i^{\overset{x'}{\overset{\shortparallel}{x-1}}} (1-\mu_i)^{\overset{y'}{\overset{\shortparallel}{y-1}}} = \max$] ← see 17.14

$$F = K(\ln\Gamma(x+y) - \ln\Gamma(x) - \ln\Gamma(y)) + (x-1)\overbrace{\sum_{i=1}^{K}\ln\mu_i}^{K\ln\gamma} + (y-1)\overbrace{\sum_{i=1}^{K}\ln(1-\mu_i)}^{} \equiv (\ln\delta)\cdot K$$

$$\frac{F}{K} = \ln\Gamma(x+y) - \ln\Gamma(x) - \ln\Gamma(y)) + (x-1)\underbrace{\frac{\sum\ln\mu_i}{K}}_{\equiv\ln\gamma} + (y-1)\underbrace{\frac{\sum\ln(1-\mu_i)}{K}}_{\equiv\ln\delta}$$

$$\frac{\partial}{\partial x}\frac{F}{K} = \frac{\Gamma'(x+y)}{\Gamma(x+y)} - \frac{\Gamma'(x)}{\Gamma(x)} + \ln\gamma$$

$= \psi(x+y) - \psi(x) + \ln\gamma$ | $\psi(z) \approx \ln z - \frac{1}{2}$

$= \ln(x+y-\frac{1}{2}) - \ln(x-\frac{1}{2}) + \ln\gamma = 0$

119 $\frac{(x+y-\frac{1}{2})\gamma}{x-\frac{1}{2}} = 1$; ① $\boxed{(x+y-\frac{1}{2})\gamma = x-\frac{1}{2}}$

120 $\frac{\partial}{\partial y}\frac{F}{K}$ gives ② $\boxed{(x+y-\frac{1}{2})\delta = y-\frac{1}{2}}$ ③ $\frac{x-\frac{1}{2}}{y-\frac{1}{2}} = \frac{\gamma}{\delta}$

Cf
$(x+y)\gamma = x$
$(x+y)\delta = y$
∴ $y = \frac{\delta}{\gamma}x$
$(x+\frac{\delta}{\gamma}x)\gamma = x$
$\gamma x + \delta x = x$
$(\gamma+\delta)x = x$
∴ $\delta+\gamma = 1$
∴ no soln for x.

From these, we get "exact" values of x & y: But usually x & y will be small ≪ 10: we may have to use corrections to $\psi(z) \approx \ln(z-\frac{1}{2})$

i.e. $\psi(z) \approx \ln(z-\frac{1}{2}) + \frac{1}{24z^2}$ which is much better than $\ln(z-\frac{1}{2})$: the $\frac{1}{z^3}$ term is zero; & next term is $\frac{1}{1202^4}$:

for z = 2, error in $\ln(z-\frac{1}{2})$ is only ~ .01.

$x+y-\frac{1}{2} = \frac{x}{\gamma} - \frac{1}{2\gamma}$

$x+y-\frac{1}{2} = \frac{y}{\delta} - \frac{1}{2\delta}$

$x + y(1-\frac{1}{\delta}) = \frac{1}{2} - \frac{1}{2\delta} = \frac{1}{2}(1-\frac{1}{\delta})$

$(x-\frac{1}{2}) + y = \frac{y}{\delta} - \frac{1}{2\delta}$

$Y = \frac{1}{\delta}(Y-\frac{1}{2})$

$\gamma Y - \frac{\gamma}{2} = \delta X - \frac{\delta}{2}$

$Y = (\delta X - \frac{\delta}{2} + \frac{\gamma}{2})/\gamma$

$Y = \frac{\delta}{\gamma}X - \frac{\delta}{\gamma 2} + \frac{1}{2} = \frac{\delta}{\gamma}(X-\frac{1}{2}) + \frac{1}{2}$

$X - \frac{1}{2} = (Y-\frac{1}{2})\frac{\gamma}{\delta}$ from ③

$x-\frac{1}{2} + y = \frac{1}{\delta}(y-\frac{1}{2})$ from ②

$(Y-\frac{1}{2})\frac{\gamma}{\delta} + Y = \frac{1}{\delta}(Y-\frac{1}{2})$

$Y(\frac{\gamma}{\delta} + 1 - \frac{1}{\delta}) = \frac{\gamma}{2\delta} - \frac{1}{2\delta}$

$$Y = \frac{\frac{\gamma}{\delta} - \frac{1}{\delta}}{2(1+\frac{\gamma}{\delta}-\frac{1}{\delta})} = \frac{1}{2}\,\frac{1}{1+\frac{1}{\frac{\gamma}{\delta}-\frac{1}{\delta}}} = \frac{1}{2}\,\frac{1}{1+\frac{\delta}{\gamma-1}}$$

$$X = \frac{1}{2}\,\frac{1}{1+\frac{1}{\frac{\delta}{\gamma}-\frac{1}{\gamma}}} = \frac{1}{2}\,\frac{1}{1+\frac{\gamma}{\delta-1}}$$

So $\frac{\gamma}{\delta-1}$ and $\frac{\delta}{\gamma-1}$ are the critical params. $\gamma \equiv \left(\prod_{i=1}^{n} \mu_i\right)^{\frac{1}{n}}$ $\delta \equiv \left(\prod_{i=1}^{n} (1-\mu_i)\right)^{\frac{1}{n}}$

δ & γ are both below 1 so [illegible] ~~[illegible]~~ these]ances.

$x = \frac{1}{2} \frac{1}{1-\frac{\gamma}{1-\delta}}$ so if $\gamma \simeq 1-\delta$ x & y are "large".

if $\gamma = 1-\delta$ $x + y = \infty$ — (as they should)

This result is reasonable, — but only if x & y are > 2, say.

Try E & M's baseball data.

$$\frac{\gamma}{\delta-1} - \frac{\delta}{\gamma-1} = \frac{\gamma^2-\gamma-\delta^2+\delta}{(\delta-1)(\gamma-1)} \quad \Big| \quad \frac{\gamma}{\delta-1} \cdot \frac{\gamma-1}{\delta} = \frac{\gamma^2-1}{\delta^2-1}$$

$1 - \frac{\gamma}{\delta-1}$ $\frac{\gamma}{1-\delta}$; $\frac{\delta}{1-\gamma}$ $1-\delta > \gamma$

$1 > \gamma + \delta$

$1 - \frac{\gamma}{1-\delta} = \frac{1-\delta-\gamma}{1-\delta}$ $x = \frac{1}{2} \frac{1}{1-\frac{\gamma}{1-\delta}} = \frac{1}{2} \frac{1}{\frac{1-\gamma-\delta}{1-\delta}} = \frac{1}{2} \frac{1-\delta}{1-\delta-\delta}$

$y = \frac{1}{2} \frac{1-\gamma}{1-\gamma-\delta}$; $x+y = \frac{1}{2}\left(\frac{1-\gamma-\delta+1}{1-\gamma-\delta}\right)$

$= \frac{1}{2}\left(1 + \frac{1}{1-\gamma-\delta}\right)$

$x+y-2 = \frac{1}{2} \frac{1}{1-\gamma-\delta} - \frac{3}{2}$

is like the SS2 for the Grand d.f.

data 1, 346, 400, 280

.26 $\gamma, \delta,$ x, y do soft [illegible] 19.19:20

.23

I got $\underset{\gamma}{.2569572}$, $\underset{\delta}{.7313832}$ for γ, δ using 18 Ball players; 45 times at bat each.

$\frac{1}{1-\gamma-\delta} = 85.76739$ $x-y-2 = \left(\frac{1}{2} \cdot 85.76\right) - 1.5 = 41.383695$ ← wt! checks,

δ

$x = \frac{1}{2} \frac{1-\delta}{1-\gamma-\delta} = 11.51927663$

$x-y-\frac{1}{2} = 42.883695$

$y = \frac{1}{2} \frac{1-\gamma}{1-\gamma-\delta} = 31.86442$

mean $\simeq \frac{x-1}{x+y-2} =$ ~~[illegible]~~ $.2541889$

$x+y =$

$\frac{1}{2}\left(\frac{1+(1-\gamma-\delta)}{1-\gamma-\delta}\right)$

$= \frac{1}{2}\left(1+\frac{1}{1-\gamma-\delta}\right)$

$\frac{45 \cdot \mu_i + 41.3836 \cdot .2541889}{86.383695}$

$.55093\ \mu_i + .12177$

1 → .342

9 ~~.2126~~ .256

10 .256

18 → .207

.244

My method gets Σ orr^2 of .032 v.s. .0216 for Stein & .07537 for raw means!!

.34

Maybe check on my arithmetic) — The $\bar{V}$ (grand mean) was .265 not

Also, Got mean of "true" means,

One Q is the "σ^2" I used for the "Grand" d.f. — it really should be "Sharper" —

I measured its variance from the sample directly: I should subtract off the var of the individual Players: One way to deal w. this: Add ~~[illegible]~~ mean wt. of Players to grand d.f.

So wt. of Grand d.f. is 86.383695

~~[illegible]~~ $42.88\overset{6}{3}95 \cdot .2569572 = 11.01927663$ ① ✓

$\left(x+y-\frac{1}{2}\right) * \delta\ 4.5 = y$ ✓ ②

30.7673

1309g SMRT STEIN. .2591585 .265

A worrysome thing! the mean of 20.25 & the arith mean of 20.34 : (I checked calcns: seem ok.)

.05 try $\frac{45 \cdot \mu_i + (41.383695 + 95)}{x + y - 2 + 45 + 45} \cdot + .65749 \cdot \mathcal{E}) = .16712\,7103$

34.2508 00000

.17449 for $\bar{\mu}$ = .26539

But

(.520931641 μ_i + .16712~~2~~103) gave ~~Σ~~ Σ err² = .0595 — (> better)

.08 Using Σ ~~#~~ prod mean of .26539: I got even worse ~~M~~ Σ err² = ~~.~~.0712 almost as bad as uncorrected data!

.26539 = $\bar{\mu}$

.09 .342508 μ_i + ~~[illegible]~~ .16712 7 Got Σ error² of .02407

This is .03 Slightly > Stein. (this is for extra wt. on grand mean | v.s. .0216 for Stein.

.10 Try $\bar{\mu}$ = .26539 : so $\bar{\mu}$ ≠ .16713 + .17449 Slightly less error: .02318 than

Still > error than Stein!

.13 STEIN2.Bas Got "Basic" to do calcns of coefs, [illegible], etc.

P	Q	1/(1-p-q)
[redacted] .2569572	.7313833	85,76739

~~Using logs, I got P = .2716366 Q = .7231656; [illegible]~~ ← using P=Q=1 initially!

Using logs instead of Mult. I got PQ same to 5 decimal accuracy and 1/(1-P-Q) to 5 decimals.

SN — ~~[illegible]~~ 2 ~~[illegible]~~ possibly serious errors:

1) If s_1 & S_2 are ssz's of 2 dfs: the ssz of their

Convolution is not $s_1 + s_2$, but $\left(\frac{1}{s_1} + \frac{1}{s_2}\right)^{-1} = \frac{s_1 s_2}{s_1 + s_2}$

.26 for differences $s_3 = \frac{s_1 s_2}{s_1 - s_2}$: Since s_1 & s_2 are large, and close to one another, is noisy; this is a big source of error)

P		
n	(An)	
x to 1	n	
V 2	Final Ave	Final
Z 3	Initial Ave (Naive)	
W d	Stein Ave	
Their		
+ 5	Times at bat in rest of season	

2) The ~~[illegible]~~ Gore ~~[illegible]~~ ~~[illegible]~~ should not be Σ ~~[illegible]~~ err² but

stein 4 Max apsip: i.e. $\prod_i \mu_{i\,\text{est}}^{\mu_{i\,\text{true}} N_i} \left(1-\mu_{i\,\text{est}}\right)^{(1-\mu_{i\,\text{true}})N_i}$ ← no times at bat

$\Sigma s_g - ss_g + \sum_i N_i \overset{V}{\mu_{i\,T}} \log \overset{U}{\mu_{i\,\text{est}}} + (1-\mu_{i\,T}) \log (1-\overset{U}{\mu_{i\,\text{est}}})$ (1-v)

Z or W

U = Actual or U = [illegible] pred'n

My est = .11

S1	naive est.	Z	— 3969,881		
S2	Stein	W	— 3909,794	2.5	
S3	my best est		— 3912.350	+ 17.6	77.6
S4	Actual rest of season	V	— 3892.172		

v. near close to Stein.

← Δ ln.

using .09 I got — 3915.202 even worse! ~~[illegible]~~

13199 SMFT Stein

$S_2 = x+y-z = 41.383695$ (20.21)
$S_1 = 45$

Try t. in wt by formula $\frac{S_1 S_2}{S_1 - S_2} = 514.9638$

.03 $\frac{45 \cdot u_i + 514.9638 \cdot \overset{\bar{u}}{.26539}}{45 + 514.938}$ $.08036\, u_i + .24405$

→ This gets − 3908.012 (1.7 ~~worse~~ Better than Stein — So it was Better!)

Bayesian for (other than std) for (Σ type Score.)
→ Try using $\bar{u}$ alone as estimate − 3908.837 ~~[illegible]~~ 3.4 worse than Stein)

Using Σ error as criterion, $\bar{u}$ got .02485; v.s. .02318 for my best
v.s. .021611 for Stein

my best for Σ error
→ using .03 Σ got for ($.08036\, \overset{v}{u_i} + .24405$) (Σ error = .02265) Better than

Try $\bar{u}$ of 20.25: .2541889 → .23376   Still slightly worse than Stein,
$.08036\, \overset{v}{u_i} + .23376$ Σ error = .024052 so > "my best" previous

NB What t. exp. suggests is that it is very impt to get t. true value of t. "Grand D.A.". Since it involves ~~[illegible]~~ small differences betw large, noisy numbers: it is often of much uncertainty! — Hr, one can always integrate over all possys in doing t. predn. — This can be very impt in such a noisy situation.

T. big problem was my ~~Σ~~ Grand $\bar{u}$ using the $\bar{u}$ of 20.25 did seem worse than t. simple arith mean.

→ In fact, the difference can be <0 ("var" <0) which is meaningless if taken by itself. — but if we integrate over all legal possys — then it's o.k. if t. "most likely var" was <0 (!)

Actually, if t. difference is positive but small, we'd prob. just use wt. = ∞ for $\bar{u}$ — or if difference is <0. | If we "integrate", I suspect that result (for large wts) will not be much different than for wt = ∞.

For t. product Score, we pick x & y to max. this Score, then we use it again, it would seem, for prediction!

T. BIG Q is why t. $\bar{u}$ I got via t. product Score (20.25, .2541889) differs ~~[illegible]~~ so much from t. (20.00, .26539 usual mean) (using t. Σ sq. Score). (.254 v.s. .265) 4% difference.

We could compare t. Σ error gave in product Score, by considering error per data pt.

For product, consider amount above perfect score: 3908.12 − 3892.17 = 15.95 = Δln. (my best using $\bar{u}$) → 17.62 for Stein
+1.67
v.s. .021611 for Σ error for Stein. (18 predictions for Σ error)

Many more for product Score. ~ 300×18 ≈ 6K! ~~[illegible]~~ ~ 30 = times as many $\frac{16}{300}$ = .053.
so within factor of 2?

Feb 1·99 SMFT ▽ Stein

23

Some random notes: 1) look at std'n mean of "restoff season" data; How does it compare w. mean of "first 45"?

.03 2) Note: $\frac{\partial}{\partial ?}$ of the $\overset{u}{p}(1-p)^{v}$ distribn is not exactly u+v. — it depends on u (say) also.

∴ the stuff about varc. of convolution being sum of varc's of convoluted df's doesn't hold.

.05 Also they don't "convolute anyway"! T. range of 2 d.f.'s each of range [0,1] → [-1, +3]! —

T. reason to got fairly good results is that perhaps t. effects .03 & .05 "cancelled"

Anyway, t. correct way to do this (like w. t. Gauss dfs!

.10 $\exp(G) = \prod_i \left(\mu_i^{x}(1-\mu_i)^{y} \cdot \mu_i^{\mu_i' \cdot 45} (1-\mu_i)^{(45-\mu_i' \cdot 45)} \frac{\Gamma(x+y+2)}{\Gamma(x+1)\Gamma(y+1)} \right)$

$G = \sum_i \Big(x \ln \mu_i + y \ln(1-\mu_i) + (45 \cdot \mu_i') \boxed{\ln \mu_i} + (45 - \mu_i' \cdot 45) \ln(1-\mu_i)$
$+ \ln\Gamma(x+y+2) - \ln\Gamma(x+1) - \ln\Gamma(y+1) \Big)$

$\int_0^1 p^x (1-p)^y dp = \frac{\Gamma(x+1)\Gamma(y+1)}{\Gamma(x+y+2)}$

$\frac{d \ln \Gamma(x)}{dx} = \frac{\Gamma'(x)}{\Gamma(x)} \approx \ln(x-\frac{1}{2})$

$\frac{d(\ln \Gamma(x))}{dx} \approx \ln(x - \frac{1}{2})$

$\mu_i^{45\mu_i'} (1-\mu_i)^{45-45\mu_i'} \cdot \frac{\Gamma(45+2)}{\Gamma(45\mu_i'+1)\Gamma(45-45\mu_i'+1)}$ = constant

.18 $\frac{\partial G}{\partial x} = \left(\sum_i \ln \mu_i + \ln(x+y+\frac{3}{2}) - \ln(x+\frac{1}{2}) \right) = 0 \therefore \left(\prod_i \mu_i\right) \cdot \left(\frac{x+y+\frac{3}{2}}{x+\frac{1}{2}} \right)^{k} = 1$

.19 $\frac{\partial G}{\partial y} = \left(\sum_i \ln(1-\mu_i) + \ln(x+y+\frac{3}{2}) - \ln(y+\frac{1}{2}) \right) = 0 \therefore \left(\prod_i (1-\mu_i) \right) \cdot \left(\frac{x+y+\frac{3}{2}}{y+\frac{1}{2}} \right)^{k} = 1$

$\boxed{\text{Def's}}$ A_i, B_i

.20 $\frac{\partial G}{\partial \mu_i} = \frac{x + 45\mu_i'}{\mu_i} - \frac{y + (45 - 45\mu_i')}{(1-\mu_i)} = 0$ ~~[illegible]~~ $\frac{x + \overbrace{45\mu_i'}^{\equiv A_i}}{\mu_i} = \frac{y + (45 - 45\mu_i')}{(1-\mu_i)}$ $\ni B_i' \equiv 45 - A_i$

$\frac{x + A_i}{\mu_i} = \frac{y+B_i}{(1-\mu_i)}$

.21 .20 → $\frac{\left(\prod_i (x + A_i)\right)}{\prod \mu_i} = \frac{\prod_i (y+B_i)}{\prod (1-\mu_i)}$ | .18 $\prod_i \mu_i = \left(\frac{x+\frac{1}{2}}{x+y+\frac{3}{2}}\right)^{k}$ | .19 ∴ $\prod_i (1-\mu_i) = \frac{(y+\frac{1}{2})^{k}}{(x+y+\frac{3}{2})^{k}}$

.21 → $\frac{\prod (x + A_i)}{(x+\frac{1}{2})^{k}} = \frac{\prod (y+B_i)}{(y+\frac{1}{2})^{k}}$ $\prod \left(\frac{x' + A_i'}{x'} \right) = \prod \left(\frac{y' + B_i'}{y'} \right)$

.28 $x' \equiv x + \frac{1}{2}$ $A_i' = A_i - \frac{1}{2}$ (~ for y) $\prod_i \left(1 + \frac{A_i'}{x'}\right) = \prod_i \left(1 + \frac{B_i'}{y'}\right)$

So $y = f(x)$ from (.28R)

$\frac{x' + A_i'}{\mu_i} = \frac{y' + B_i'}{1-\mu_i}$ so if we know x & y, all μ_i's can be obtained from this linear eq.

.20 $\frac{x + A_i}{\mu_i} = \frac{y + B_i}{1 - \mu_i}$ $\mu (y + B) = x + A - \mu(x+A)$

.35 $\mu(y + B + x + A) = x + A$ ∴ $\mu_i = \left(\frac{y + \overbrace{B_i + x + A_i}^{\Sigma = 45}}{x+A} \right)^{-1}$

$\boxed{\mu_i = \cancel{\frac{x+y+45}{x+A_i}} \frac{x + A_i}{x+y+45}}$

$1 - \mu_i = \cancel{\frac{x+y+45}{y+B_i}} \frac{y+B_i}{x+y+45}$

~~$1 = (x+y+45)$ [illegible] $\frac{(x+A_i)(y+B_i)}{x+y+(A_i+B_i) = 45}$~~

x

~~So $(x+A_i)(y+B_i) = 1$? unlikely! x, y, A_i, B_i are usually >1.~~

$1 - \frac{45}{x+y+45} = \frac{x+y+45-45}{x+y+45} = \frac{x+y}{x+y+45}$

$\frac{x + A_i}{x+y + 45}$ 45 wt. ↑ x+y wt.

$\frac{x + 45 A_i'}{x+y+45}$

$\mu_i = \frac{\frac{x}{45} + \frac{A_i}{45}}{\frac{x+y}{45} + 1}$

2·[illegible]·99 SMFT Stem

.01 $\mu_i = \frac{x+A_i}{x+y+45} = \frac{x+45\mu_i'}{x+y+45}$ is quite reasonable! x & y are actual "ssz's" not "ssz+1 or ssz−1"

.02 $(1-\mu_i) = \frac{y+B_i}{x+y+45}$

Is $\frac{x}{x+y} = \frac{\sum \mu_i'}{k}$ i.e. is $\sum\left(\mu_i' - \frac{x}{x+y}\right) = 0$?

I don't know, but out of fear have (23.18): $\prod_i \left(\mu_i \Big/ \frac{x+\frac12}{x+y+\frac32}\right) = 1$, Also $\left(\mu_i \Big/ \frac{x+\overset{A_i}{45\mu_i'}}{x+y+45}\right) = 1$

$\mu_i \Big/ \left(\frac{\frac{x}{45}+\mu_i'}{\frac{x}{45}+\frac{y}{45}+1}\right)$ So maybe "normalize" x & y to $\frac{x}{45}, \frac{y}{45}$?

.04 (α) Given x, y total; ~~[illegible]~~ we get to this from .01 then

we get $\frac{x+\frac12}{x+y+\frac32} = C$ & $\frac{y+\frac12}{x+y+\frac32} = D$ from 23.18, 19. from these, we get x & y & loop to (α) (.04)

say $\frac{x+\frac12}{x+y+\frac32} = C$ $\quad \frac{y+\frac12}{x+y+\frac32} = D \quad \frac{x+y+\frac32-\frac12}{x+y+\frac32} = (C+D)$ ~~$(x+y)(1-C-D) = -\frac12 + \frac12(C+D)$~~

$\frac{x+y+1}{x+y+\frac32}$ $\quad$ x

$Z = x+y+\frac32 \quad \frac{Z-\frac12}{Z} = C+D \quad 1-\frac{1}{2Z} = C+D$; $\frac{1}{2Z} = 1-C-D$ | $Z = \frac{1}{2(1-C-D)}$

$Z = x+y+\frac32 = \frac{1}{1-C-D}$! $x = C\cdot Z - \frac12 \quad y = D\cdot Z - \frac12$

$x = \frac{C}{2(1-C-D)} - \frac12$ $\quad y = \frac{D}{2(1-C-D)} - \frac12$

$x+y = \frac{C+D}{1-(C+D)} - 1$

$\boxed{x+y = \frac{1}{1-C-D}}$

.20 $C = \left(\prod_i \mu_i\right)^{\frac1k} \approx \left(\frac{x+\frac12}{x+y+\frac32}\right) = \boxed{\frac{\left(\prod_i (x+A_i)\right)^{\frac1k}}{x+y+45}} \rightarrow = \frac{x+\frac12}{x+y+\frac32} = \textcircled{C}$

.21 $D = \left(\prod_i (1-\mu_i)\right)^{\frac1k} = \left(\frac{y+\frac12}{x+y+\frac32}\right) = \boxed{\frac{\left(\prod_i (y+B_i)\right)^{\frac1k}}{x+y+45}} \rightarrow \frac{y+\frac12}{x+y+\frac32} = \textcircled{D}$

.24 either this → $\boxed{\text{Initial } \mu_i = \overline{\mu_i'} \; ; \; \frac{x}{x+y} = \overline{\mu_i'}}$ $\quad x+y = 500$ (Guess).

.25 or this → for starting $\boxed{\text{So } x = 500\,\overline{\mu_i'} \; ; \; y = 500(1-\overline{\mu_i'}) \; ; \; \mu_i = \overline{\mu_i'} = .26539}$

$500 \cdot .26539 \quad 500(1-.26539)$ $\quad$ not μ_i

.25 is easier to start with.

~~[illegible]~~

(Starting w. .24 $\quad C = \overline{\mu_i'}$, $D = 1-\overline{\mu_i'}$ gives $x+y = \infty$. (!)

.30 α ~~[illegible]~~ w. x & y approx.: from .01, .02 get μ_i, $1-\mu_i$! from 20, 21, get C, D.

then $x = \frac12\left(\frac{C}{1-C-D} - 1\right)$ $\quad y = \frac12\left(\frac{D}{1-C-D} - 1\right)$; $x+y = \frac12\left(\frac{C+D}{1-C-D} - 2\right) = \frac{3C+3D-2}{2(1-C-D)}$ $\quad \frac12\left(\frac{C}{1-C-D} - 1\right) \quad \frac{2C+D-1}{1-C-D}$

→ loop to α (.30)

$x+y+\frac32 = \frac{3C+3D-2}{2(1-C-D)} + \left(\frac{3-3C-3D}{2(1-C-D)}\right) = \frac{1}{2(1-C-D)}$ as it should!

$\frac12 \frac{C+D}{1-C-D} + \frac12 = \frac12\left(\frac{C+D}{1-C-D} + 1\right) = \frac12 \frac{1}{1-C-D}$

2.1.99 SOFT STEIN

Stein 5. After Array AC ( ) has been loaded,

$X = 500 \cdot .265 = 132.5$ $Y = 500(1-.265) = 368.5$; C=1 : D=1

~~C=1~~ [illegible] ~~.95~~

80 2495

1997
39.95 ~~#60~~

Model 1
[illegible]

RCN Guy said it would take local cable co. ≐ 2 yrs. to do it

u $\overline{\mu_i}$ μ_i μ_i^{stein} [illegible] 25

A 1 2 3 4 .5

~~[illegible]~~

.03α For ~~J=1 to 18:~~ Z = A(J, 3)*45/1000

V =	Z =	W =	T =
2	3	4	5
μ_i^T	μ_i'	μ_i^{stein}	et but.

Unray

Z

C = C*(X+Z) : D = D*(Y+(45−Z))

Next

~~a=a/(x+y+45)~~ C = C^(1/18) /(X+Y+45) D = D^(1/18)/(X+Y+45)

X = (C/(1−C−D) −1)/2 : Y = (D/(1−C−D) −1)/2

Print "X" = X , "Y" = Y

Input L L

Go To α (103)

It converged quickly to X = 2.370157 Y = 8.575965

$\boxed{X + Y = 10.947535}$

PGM Bug: "C" ;"G" was not reinitialized on successive passes.

This is clearly not working!

C = .2306065 D = ~~[illegible]~~ .7292203

.2306065
.7292203
1 − .9598268 =
1 − C − D = .0401732

$X + Y + \frac{2}{2} = \frac{1}{2}\left(\frac{1}{1-C-D}\right) \sim 12\frac{1}{2}$

X + Y ~ 11.

Π(y+

$\bar{\mu}$	Gore
.26	−3909.730
.27	−3907.071
.275	−3906.77
.28	−..07.142
.2755	−..06.779
w = .9	6.774
.95	6.973
.85	7.03
.89	6.783
.91	6.773

Bush w = .895 w = .895

So ($\bar{\mu}$ = .275 = wt of .895 to Bush)

X + Y + 2 ~ 13 only!

$$\mu_i = \frac{X + 45\mu_i'}{X+Y+45} = \alpha\rho + q\mu_i' \quad q = \frac{45}{X+Y+45} \quad \rho = \frac{X+Y}{X+Y+45} \quad : \alpha\rho = \frac{X}{X+Y+45}$$

so α = .2165 only = $\frac{X}{X+Y}$.

$$\alpha = \frac{X}{X+Y+45} \cdot \rho^{-1}$$

$$\boxed{\alpha = \frac{X}{X+Y}} = .2165 \text{ (Bad!)}$$

$$\mu_i = \frac{X}{X+Y+45} + \frac{45}{X+Y+45} \cdot \mu_0$$

= .042364 + .804325 · μ_i' — as expected, using STEIN4.BAS

— The Product Gore: this got −3950.585 : Bafflingly, Simple mean is much worse than Stain.

There is certainly something Grossly Wrong!

The Base was, as for X45 ≈ 10 × 45 = 450

For the product + Gore (Stein 4. bas) There was a local max of PC at $\bar{\mu}$ = .275, w = .895 for $\bar{\mu}$; wt .105 for μ_i

i.e. μ_i = .895 × .275 + .105 × μ_i so stein put very close to Best!

Later: Grand Mean $\frac{\Sigma A(I,2)*A(I,5)}{\Sigma A(I,5)}$ = .2744989

Gore	wt. on $\bar{\mu}$
−3909.79	
−3908.194	.85
3907.953	.90
3908.108	.85
3907.967	.88
3907.952	.89
3907.969	.91
3907.95	.895 Best

2.1.99 SMRT Stern, $\boxed{\text{Rpar}}$ also see 31

$x+y=k$ corresponds to what I had before — <u>w.o. subtracting intervariance.</u>

So look at formula for e^G (23.10): I think it's essentially correct, w. a constant factor omitted

$$e^G = \prod_{i=1}^{K}\left[\left(\frac{\Gamma(x+y+2)}{\Gamma(x+1)\Gamma(y+1)}\,\mu_i^{x}(1-\mu_i)^{y}\right)\cdot\left(\underbrace{\frac{\Gamma(45+2)}{\Gamma(45\mu_i'+1)\Gamma((45-45\mu_i')+1)}}_{\text{These are constant factors that don't influence the max on of } x,\, y,\, [\mu_i]\text{.}}\;\mu_i^{\overset{A_i}{\overset{\shortparallel}{45\mu_i'}}}\;(1-\mu_i)^{\overset{B_i=45-A_i}{\overset{\shortparallel}{45-45\mu_i'}}}\right)\right]$$

.10 $$G = \sum_{i=1}^{k}\left[\ln\Gamma(x+y+2)-\ln\Gamma(x+1)-\ln\Gamma(y+1)+(x+45\mu_i')\ln\mu_i+(y+45-45\mu_i')\ln(1-\mu_i)\right]$$

Road runner warner cable.

.12 $$\boxed{\frac{\partial G}{\partial \mu_i} = \frac{x+45\mu_i'}{\mu_i} - \frac{y+45-45\mu_i'}{1-\mu_i} = 0}$$ ←

.16 $$\frac{\partial G}{\partial x} = \sum_{i=1}^{k}\left(\frac{\partial\ln\Gamma(x+y+2)}{\partial x} - \frac{\partial\ln\Gamma(x+1)}{\partial x} + \ln\mu_i\right) = 0$$

.17 $$\frac{\partial G}{\partial y} = \sum_{i=1}^{k}\left(\frac{\partial\ln\Gamma(x+y+2)}{\partial y} - \frac{\partial\ln(y+1)}{\partial y} + \ln(1-\mu_i)\right) = 0$$

$$\frac{d\ln\Gamma(z)}{dz} \equiv \psi(z) \approx \ln(z-\tfrac12)$$

.21 ~~[illegible]~~ from .16 $k\left(\ln(x+y+\frac32)-\ln(x+\frac12)\right)+\sum_{i=1}^{k}\ln\mu_i=0$

.22 from .17 $k\left(\ln(x+y+\frac32)-\ln(y+\frac12)\right)+\sum_{i=1}^{k}\ln(1-\mu_i)=0$

.23 from .21 $\left(\prod_{i=1}^{k}\mu_i\right)^{\frac1k} = \frac{x+\frac12}{x+y+\frac32}$ call this "C"

.25 from .22 $\left(\prod_{i=1}^{k}(1-\mu_i)\right)^{\frac1k} = \frac{y+\frac12}{x+y+\frac32}$: call this "D"

Aside: $\psi(z) = \ln z - \frac{1}{12z^2}+\frac{1}{120z^4}\cdots$

← No $\frac{1}{z^3}$ term.

$\ln(z-\frac12) = \ln z + \ln(1-\frac{1}{2z}) = \ln z - \frac{1}{2z}+\frac{1}{8z^2}\cdots$

z	$\ln(z-\frac12)$	$\psi(z)$
1	-.693	-.5772
2	.405	.422
3	.916	.922
5	1.504	1.506
10	2.2513	2.2517
25	3.19867	3.19874

We want to solve the system .12, .16, .17. This is equivalent to .12, .23, .25

Plan: Pick an initial x, y: Use .12 to solve for all μ_i and ~~[illegible]~~ $1-\mu_i$ from these, use .23 and .25 to generate C and D. From C and D solve for x and y. This x and y is for the next recursion.

Re: Eq. .12: Let $45\mu_i' \equiv A_i$; $45-45\mu_i' \equiv B_i$; so $A_i+B_i=45$

.34 Then .12 is $\frac{x+A_i}{\mu_i} = \frac{y+B_i}{(1-\mu_i)}$ →: Given x and y: $\left.\begin{aligned}\mu_i &= \frac{x+A_i}{x+y+45}\\ 1-\mu_i &= \frac{y+B_i}{x+y+45}\end{aligned}\right\}$ from .12

So: <u>Program</u>: start with x and y:

from .23 $C = \left(\prod_i\mu_i\right)^{\frac1k} = \left(\prod_{i=1}^{k}\left(\frac{x+A_i}{x+y+45}\right)\right)^{\frac1k} = \frac{x+\frac12}{x+y+\frac32}$

$$D = \left(\prod_i(1-\mu_i)\right)^{\frac1k} = \prod_{i=1}^{k}\left(\frac{y+B_i}{x+y+45}\right)^{\frac1k} = \frac{y+\frac12}{x+y+\frac32}$$

$C = \left(\prod_i\left(\frac{y+A_i}{x+y+45}\right)\right)^{\frac1k}$; $D = \left(\prod_i\left(\frac{x+B_i}{x+y+45}\right)\right)^{\frac1k}$: since $\frac{x+\frac12}{x+y+\frac32} = C$ and $\frac{y+\frac12}{x+y+\frac32} = D$;

Given C and D we can solve for x and y for the <u>next recursion</u>: $x = \frac12\left(\frac{C}{1-C-D}-1\right)$; $y = \frac12\left(\frac{D}{1-C-D}-1\right)$

2.1.99 SMRT Stein.

x + y = 11 corresponds to what I had before — w.o. subtracting irrelevances.

So look at formula for e^G (23.10): I think it's essentially correct, w. constant factors omitted

.03 $$e^G = \prod_{i=1}^{K}\left[\left(\frac{\Gamma(x+y+2)}{\Gamma(x+1)\,\Gamma(y+1)}\,\mu_i^{x}(1-\mu_i)^{y}\right)\cdot\left(\underbrace{\frac{\Gamma(45+2)}{\Gamma(45\mu_i'+1)\,\Gamma((45-45\mu_i')+1)}}_{\text{These are constant factors that don't influence the max. of } x,\,y,\,[\mu_i]}\;\mu_i^{\overset{A_i}{45\mu_i'}}\,(1-\mu_i)^{\overset{B_i = 45-A_i}{45-45\mu_i'}}\right)\right]$$

.10 $$G = \sum_{i=1}^{k}\left[\ln\Gamma(x+y+2) - \ln\Gamma(x+1) - \ln\Gamma(y+1) + (x+45\mu_i')\ln\mu_i + (y+45-45\mu_i')\ln(1-\mu_i)\right]$$

Road runner warner cable.

.12 $$\boxed{\frac{\partial G}{\partial \mu_i} = \frac{x+45\mu_i'}{\mu_i} - \frac{y+45-45\mu_i'}{1-\mu_i} = 0} \quad \leftarrow$$

.16 $$\frac{\partial G}{\partial x} = \sum_{i=1}^{k}\left(\frac{\partial \ln\Gamma(x+y+2)}{\partial x} - \frac{\partial \ln\Gamma(x+1)}{\partial x} + \ln\mu_i\right) = 0$$

.17 $$\frac{\partial G}{\partial y} = \sum_{i=1}^{k}\left(\frac{\partial \ln\Gamma(x+y+2)}{\partial y} - \frac{\partial \ln(y+1)}{\partial y} + \ln(1-\mu_i)\right) = 0$$

$$\frac{d\ln\Gamma(z)}{dz} \equiv \psi(z) \approx \ln\left(z-\tfrac12\right)$$

.21 ~~from~~ from .16 $k\left(\ln(x+y+\frac32) - \ln(x+\frac12)\right) + \sum_{i=1}^{k}\ln\mu_i = 0$

.22 from .17 $k\left(\ln(x+y+\frac32) - \ln(y+\frac12)\right) + \sum_{i=1}^{k}\ln(1-\mu_i) = 0$

.23 from .21 $\left(\prod_{i=1}^{k}\mu_i\right)^{\frac1k} = \frac{x+\frac12}{x+y+\frac32}$ call this "C"

.25 from .22 $\left(\prod_{i=1}^{k}(1-\mu_i)\right)^{\frac1k} = \frac{y+\frac12}{x+y+\frac32}$: call this "D"

Asside: $\psi(z) = \ln z - \frac{1}{12z^2} + \frac{1}{120z^4} \cdots$ ← No $\frac{1}{z^3}$ term.

$\ln(z-\frac12) = \ln z + \ln(1-\frac{1}{2z}) = \ln z - \frac{1}{2z} + \frac{1}{8z^2}\cdots$

z	$\ln(z-\frac12)$	$\psi(z)$
1	−.693	−.5772
2	.405	.422
3	.916	.922
5	1.504	1.506
10	2.2513	2.2517
25	3.19367	3.19874

We want to solve the system .12, .16, .17. This is equivalent to .12, .23, .25

Plan: Pick an initial x, y: Use .12 to solve for all μ_i and $1-\mu_i$ from these, use .23 and .25 to generate C and D. From C and D solve for x and y. This x and y is for the next recursion.

Re: Eq. .12: Let $45\mu_i' \equiv A_i$; $45-45\mu_i' \equiv B_i$; so $A_i + B_i = 45$

Then .12 is $\frac{x+A_i}{\mu_i} = \frac{y+B_i}{(1-\mu_i)}$: Given x and y: $\mu_i = \frac{x+A_i}{x+y+45}$, $1-\mu_i = \frac{y+B_i}{x+y+45}$ } from .12

So: Program: start with x and y:

from .23 $C \equiv \left(\prod_i \mu_i\right)^{\frac1k} = \left(\prod_{i=1}^{k}\left(\frac{x+A_i}{x+y+45}\right)\right)^{\frac1k} = \frac{x+\frac12}{x+y+\frac32}$

$D = \left(\prod_i (1-\mu_i)\right)^{\frac1k} = \prod_{i=1}^{k}\left(\frac{y+B_i}{x+y+45}\right)^{\frac1k} = \frac{y+\frac12}{x+y+\frac32}$

$C = \left(\prod_i\left(\frac{x+A_i}{x+y+45}\right)\right)^{\frac1k}$; $D = \left(\prod_i\left(\frac{x+B_i}{x+y+45}\right)\right)^{\frac1k}$: since $\frac{x+\frac12}{x+y+\frac32} = C$ and $\frac{y+\frac12}{x+y+\frac32} = D$;

Given C and D we can solve for x and y for the next recursion: $x = \frac12\left(\frac{C}{1-C-D} - 1\right)$; $y = \frac12\left(\frac{D}{1-C-D} - 1\right)$

~2.4.94 sm fr.

.01 **Stein5:**... starting w. $R = 500$; $x = R \cdot .265$, $Y = R - x$

~~500~~ → $x, y = 1656.42, 4545.23$

			$X+Y$	$x/(x+y)$
$R =$ 1000 →	6073, 16845		22918	.2645
100 →	177, 327.	.351		
50 →	50, 140			
25 →	26, 76			
15 →	19, 56			
12 →	17, 50			
10 →	16, 46.			
8 →	14, 43.	$: \frac{x}{x+y} = .245$	s.t. : .01	

$R =$ 10.946 ; $\bar{\mu} = .2165$; gives $x = 15.77$, $y = 47.2$! , $x+y = 63.6$, $\bar{\mu} = .24822$!

C & D were not properly ~~set~~ re-initialized after each loop! !

.14 Same Bug in ΣGamL: **Fixed**:

starting w. $X = 132$, $Y = 368$; it diverges. ; The old prgm. w. no reinitialization of C, D ;

it did converge — maybe luckily! — apparently w/o, because the 6 values converged to,

the system did not give those values on repeat.

I haven't been able to find a soln for R & $\bar{\mu}$ "by hand"

I tried only .2165 & .265 for $\bar{\mu}$ & various R values.

Try to do a maxzn of 26.10 "By hand".

$\ln\Gamma(x) = (x-\tfrac{1}{2})\ln x - x + \tfrac{1}{2}\ln(2\pi) = x\ln x - \tfrac{1}{2}\ln x - x + c$ concavity. $\quad x^x \cdot \sqrt{x}$

$\alpha \equiv \ln\Gamma(x+y+2) - \ln\Gamma(x+1) - \ln\Gamma(y+1) = (x+y+2)\ln(x+y+2) - (x+1)\ln(x+1) - (y+1)\ln(y+1)$

$- \tfrac{1}{2}\ln(x+y+2) + \tfrac{1}{2}\ln(x+1) + \tfrac{1}{2}\ln(y+1) \boxed{\boxed{-x+y}+2 + \boxed{x+y} + } + 2 \cdot \tfrac{1}{2}\ln(2\pi)$

$A = x+1$, $B = y+1$, $C = x+y+2 = A+B$, $F = x+y+45$ $\quad$ ~~$(x+z_i)(y+45-z_i)$~~

$\alpha = A\ln A - B\ln B + C\ln C - \tfrac{1}{2}(\ln(A \cdot B/C))$ $\quad k=14$; $l = 45$

$G(26.10) = k \cdot \alpha + \sum^{k}\left((x+z_i)\cdot\ln\mu_i + (y+45-z_i)\ln(1-\mu_i)\right)$ $\quad z_i = 45\mu_i$ $\quad$ 26.34

$= k\alpha + \Sigma\; (x+z_i)\cdot\ln((x+z_i)/F) + (y+45-z_i)\cdot\ln((y+45-z_i)/F)$

$= k \cdot \alpha + \Sigma\; \overset{H}{(x+z_i)}\ln(x+z_i) + \overset{L}{(y+45-z_i)}\ln(y+45-z_i)$ $\quad (-\ln F)(x+z_i+y+45-z_i)$

$- F\ln F$ $\quad = F\ln F$

$A = x+1$: $B = y+1$: $C = A+B$:

```
For J = 1 to 14 ; Z = .045*A(2,J).
                  H = X+Z : L = Y+45-Z : F = H+L
G = 18*(-A ln A - B ln B + C ln C - 1/2 ln( A*B/C)
G = G + H*ln H + L*ln L - F ln F
Next.
```

2.5.99 SMFT Stein's oil.

Check the $\ln \beta(x,y)$ formula w. calculations.

~~.03~~ .03 $A = X+1$, $B = Y+1$: $E = A+B$ ~~-1~~ ~~$+\frac{1}{2}\ln$~~

.04 $G = (\cdot E\ln E - A\ln A - B\ln B$ ~~$-\frac{1}{2}\ln(A+B/\ldots+2\pi)$~~ $+\frac{1}{2}\ln(E/A/B/2/\pi))*18$

$$\frac{\ln \Gamma(x+1)\Gamma(y+1)}{\Gamma(x+y+2)} = \ln\left(\frac{x!\,y!}{x+y!}\right)$$

say $x=3$, $y=4$
$x+y=7$

$$\frac{3!\,4!}{7!} =$$

$.02857 = \frac{1}{35}$

$\ln\frac{1}{35} = -3.555348$

I want

$$\frac{7!}{3!\,4!} = 35$$

$\ln 35 = 3.555348$

$\ln X! =$

$(X+\frac{1}{2})\ln X + \frac{1}{2}\ln(2\pi)$

$-X$

$X\ln X + \frac{1}{2}((\ln X + \ln 2\pi)$

$-X$

I Got G = 3.59192 for A = 3, B=4, E= A+B

exp(3.59192) = 35.23. | Well 3! and 4! can be off via simple Stirling;

Try A = 10, B = 11, E = 21
fom got 2
→ G = 12.78535 ok.

~~[illegible] 1.5366 [illegible]~~

$$\ln\frac{21!}{10!\,11!} = \text{~~[illegible]~~ on calculator. } 12.7734$$

$\ln X! \doteq (X+\frac{1}{2})\ln X + \frac{1}{2}\ln(2\pi) - X = X\ln X + \frac{1}{2}\ln(X\cdot 2\pi) - X$

~~close for X=10~~

for $J = 1$ to 18 : $Z = .045 * A(J,3)$: $H = X+Z$: $L = Y+45-Z$: $F = H+L$

$G = G + H*\log H + L\ln L - F\ln F$
next.

£50 + 50 min
5£ 10£

Using R ≐ 450, U ≐ .265 I got G = −455.2558
500, " −496.2154
400 −444.1
300 . −.49.2
100 −479.9

R=10, .265 : ~~−457.2595~~
2, .265 : ~~−448.57~~
.001, .265 : ~~[illegible]~~, −444.5771
−10, .265 } negative R is illegal.

.001, .200 −444.5718
.001, .001 −444.5681
.001, −.1 −444.5661

looks suspicious(?)!

I find ! E = A+B+1 in .03

for U = .265 R = 400 G −758
.2 −365 450 −368
.3 −366 500 −366
1000 : .260 −360.457 1000 −360
.24 60.19 2000 −354
max. → .25 −359.9 4000 −348
.24 360.00 8000 342
.23 −360.5 100k −319345

} doubling R → G→G+6 ≐

2000. .25 → −353.885

So problem at $\bar{\mu}$ = .25 360 1000 back
w. R=∞. (!)
So very probably bug.

~~F(x+1+y+1)~~ U ≐ ~~.265~~ doubling w. 2 corrections "+1" → back to "+0"
"−" → "+"

doubling R still ↓ G by about 6 ! So it must be inside of the second order terms,

$\sum A_i \ln(X+A_i) - A_i\ln(X+Y+45) + B_i\ln(Y+B_i) - B_i\ln(X+Y+45)$

$\sum A_i\ln(X+A_i) + B_i\ln(Y+B_i) - (A_i+B_i)\ln(X+Y+45)$

$\sum (X+A_i)\ln X+A_i - (X+A_i)\ln(X+Y+45) + X+B_i\ln(X+B_i) - (Y+B_i)\ln(X+Y+45)$

$\underbrace{(X+A_i)\ln(X+Y+45) + \ldots + (Y+B_i)}_{(X+Y+45)\ln(X+Y+45)}$

2

2-5-99 SMFT STEIN.

For 26.10 ; the first part of G: we want $\ln\Gamma(x+y+2) - \ln\Gamma(x+1) - \ln\Gamma(y+1)$

since $A = x+1$, $B = y+1$, $E = A+B = x+y+2$: we want $\ln\Gamma(E) - \ln\Gamma(A) - \ln\Gamma(B)$

$\ln\Gamma(E) = E\ln E - \frac{1}{2}\ln E + \text{const}$ $\qquad \ln x! = x\ln x + \frac{1}{2}\ln x$

should be + also $\ln(2\pi)$

so $G_{\text{initial}} = \left[E\ln E - A\ln A - B\ln B \pm \frac{1}{2}\ln\left(E/(A\cdot B\cdot 2\pi)\right)\right]\cdot 18$ (is ~~correct~~ . I checked w. $A=10$, $B=11$, $E=21$.) actually wrong! see 30.03

.05 → Also for J=1 to 18 : | $Z = .045 \cdot A(J,2)$: $H = X+Z$; $L = Y + .955$; $F = H+L$
$G = G + H\ln H + L\ln L - F\ln F$) seems to be correct

Stop? check carefully.

Next

I may have gotten Γ ≠ ! mixed up! See 28.03 ; the "G" of 28.04 is for

$\frac{\Gamma(E)}{\Gamma(x)\cdot\Gamma(y)}$, but I tested it for $\frac{E!}{x!\,y!}$: That's o.k. : it's only

.12 ~~need~~ to change the sign of $\frac{1}{2}\ln\left(\frac{E}{x\cdot y}\right)$ to get from $x!$ to $\Gamma(x)$.
"+" is $x!$ "−" is $\Gamma(x)$.

(strange! for U = .25
As I double R , we usually
got $G \to G \pm 6$.
1000 → −360
but R = 2000 → −354
R = 4000 → −362
R = 8000 → −356] !!?
R = 16 k → −350
R = 32 k → −344

.14 $\boxed{2.6.99}$ Random notes: Maybe speed up, simplify STEIN7.BAS

1) By putting $F\cdot\ln F$ (which is indep of Z) in front of
initialization of G.

2) Perhaps use MAPLE to fit R's i.e. use Beta function

So see first $\left(\text{Beta}(x+1, y+1)\right)^{-1} \int_0^1 z^x(1-z)^y\,dz = 1$.

Then check that ↑

or just ~~100~~ for z = 0 to 1 step .01
s = s + z ^ x * (1-z) ^ y
Next
print s/100

$(x+A)\ln\left(\frac{x+A}{F}\right) + (y+B)\ln\left(\frac{y+B}{F}\right)$

$(x+A)(-\ln f) + (y+B)(-\ln f)$
$= -A\ln R$.

Print G initial : $\left(\int_0^1 \quad\right)^{-1} = 8.5\ E{+}7$ 85 M

G = Newman. const.

and (G = 20.1) : 5.391 E 8

using step .001 ; 8.4877 → 8.488E7

$E = x+y-1$ gives closer 7.6E7

$\frac{x!\,y!}{x+y+1!}$ $\qquad \frac{26!}{10!\,15!}$ = ~~8.4E7~~ 8.49877 ! closer

So S is o.k. $= \frac{x!\,y!}{x+y+1!}$

$\int_0^1 t^{x-1}(1-t)^{y-1}\,dt$
$= B(x,y) = \frac{\Gamma x\,\Gamma y}{\Gamma x+y}$

According to 28.04 the formula for G gives $\frac{E!}{A!\,B!}$ for $A = 10$, $B = 11$, $E (= A+B) = 21$.

x+y

---

2 " $+\frac{1}{2}\ln(\quad)$ " in line 28 gives $\frac{(A+B)!}{A!\,B!}$ × 22

I did $A = x$, $Y = B$, $E = x+y+1$

$G = \quad \ldots B\ln B - 1 + \frac{1}{2}\ln(E/A/B \ldots$ — No! It was 8498. vs 8590) only ~1.2% off

~~I got G within .1 of correct value.~~

G^G, G

$A = x$ $B = B$ $E = x+y$

using $G = \quad B\ln B + \frac{1}{2}\ln(E/A \quad) + \ln(E+1)$ & got very close for G = 5 in error by x same ",".

$A = 10$, $B = 15$; $E = 25$

How and "+" formula gave:
3303435 , 15.0047

$\boxed{3268762 \ , \ 14.99992129}$ exact
very close

So; $A = x$, $B = y$, $E = A+B$, $G = \ldots + \log(E+1)$; $\pm\frac{1}{2}\ln(E/$

2.5.99 start STEIN: 50

.01 So I "fixed" G !! Still for $U = .265$. Gorch by $\overset{N}{6}$ for each doubling of R ! — like before.

.03 So 1. correct G is:

$A = x,\ B = y,\ E = x + y.$

$G = (E\ln E - A\ln A + B\ln B + \underline{\ln(E+1)} + \frac{1}{2}(\ln E/A/B/2/\pi) \rightarrow$

Also ~~[illegible]~~ $A = x,\ B = y\ E = \underline{x+y+1}$

$G = E\ln E - A\ln A - B\ln B \underline{-1} + \frac{1}{2}(\ln E/A/B/2/\pi)\ \left(\frac{1}{2E}\right)$ I think

I think

$\ln\left(\frac{(x+y)!}{x!\,y!}\right) =$ $A = x, y = B, E = x+y$

$E\ln E - A\ln A - B\ln B + \frac{1}{2}\left(\ln\frac{E}{A\cdot B\cdot 2\pi}\right)$

$\log(E+1):$

$\log E + \log(1+\frac{1}{E})$

$\sim \ln E + \frac{1}{E}$

$(E+1)\log(E+1)$

$= E\log(E+1) +$

$\log(E+1)$

$= E(\ln E + \frac{1}{E})$

$= E\ln E + 1$

$+ 1$

$\frac{1}{2}\ln(E+1) =$

$\frac{1}{2}(\ln E + \frac{1}{E})$

$R =$ $U = .5$ (or .265)

$A = U\cdot R,\ B = (1-U)R$

$\frac{A+B+1}{A!\,B!}$ squares when R is doubled so it's $\propto \alpha^{R}$ for some α.

or $e^{\alpha R}$ " " "

For $U = .5$

$\frac{2x!}{x!\cdot x!} \rightarrow \frac{4x!}{2x!\,2x!}$

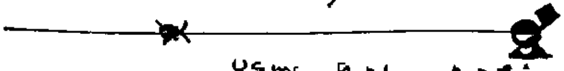
$\frac{ss}{2^{x}} \rightarrow 2^{2x} = (2^{x})^{2}.$

$2x! \simeq (x!)^{2}\cdot 2^{x}$

$2x$ ②

$\left(\frac{U}{\alpha(1-\alpha)}\right)^{1-U})^{R \to 2R}$

$X + 45U_i$ y



Using $R = 1$, $n = 4$.

Doubling R gives

$\Delta \simeq .3$ for $u = 1$

$\Delta \simeq 1.2$ " $u = 4$

$\Delta \sim 6$ " $R = 18$ $\simeq$ 18 ~ 5 A4

Joes/Donna.

Bede

(8 B

Leaving Prison

.22 So say $\delta \propto k$

Since present B.M. seems to have peaks at $R = \infty$ (maybe $U = .20$??), look at behavior of each

$\simeq \delta(U - \epsilon)$ for ma.var.

$$\prod_{i=1}^{u} \Bigg/ \left((B(x+1, y+1))^{-1}\ \mu_i^{x}(1-\mu_i)^{y} \right) \left(\mu_i^{\cdot l\cdot \mu_i'}\ (1-\mu_i')^{l(1-\mu_i')} \right)$$

$\beta(x,y)^{-1}$

$\mu_i^{l\mu_i'}$ $(1-\mu_i')^{l(1-\mu_i')}$ l^{l}

$\frac{\mu_i'^{\,x}\ (1-\mu_i')^{y}}{l^{x+y}}$

$z_i^{x}\ z_i'^{y}$ $z_i^{l\mu_i}$

$z_i \equiv \mu_i$ $z_i' = 1 - z_i\ .$

$z_0 = \frac{\mu_i'}{x+y+l}$ no: $\mu_i = \frac{x + l\mu_i'}{x+y+l}$

$A_i = l\cdot A_i'$

This result is independent of the Norman constant for x, y.

~~the $\delta()$ for~~ x, y. → Easy, but as $x, y \to \infty$, $\mu_i \to \frac{x}{x+y}$ (!).

I changed "45" to "100" $N=$ $\Delta = 6$ ~~of~~ .01 + $G = 4.5$ 4.5 4.2

" " "200" Δ → .7, 2.8, 5.5, 3.9 ~~(for R = [illegible])~~ approx. ΔR ~~[illegible]~~ = 5.7 (!)

G is $\overset{\sim}{\propto}$ N (= "45")

For $N = 400$, ~~for~~ $U = .265$; the plot of G v.s. R is peculiar

← local max

R	.	25	50	100	200	400	450	3600	7200	1M " 1e6
(G+4000)	-61	-59	-57	-59	-64	-71	-73	-85	-83	-405

so maybe for large enough R, ~~[illegible]~~ G begins to ↑ again.

(BEV) also see <u>26</u>

2394 SMFT: STEIN

So <u>A review:</u> The problem is Baseball batting Averages. See <u>Efron, Morris, Sci Amer, May 77 pp 109 ff.</u>

The model I used: Each of k players has a <u>true</u> batting average μ_i $(i = 1 \cdots k)$. Each time at bat he has probty of μ_i of making a "hit"; $(1-\mu_i)$ is probty of failure. I assume the μ_i are from a <u>monomodal distribn from 0 to 1</u>. Call this Distribution $P(\mu_i)$.

After μ_i is chosen, each player's performance is a Bernoulli sequence with parameter μ_i, so for N times at bat, the probability of the apparent average of μ_i' is just $\mu_i^{\mu_i' N}(1-\mu_i)^{N(1-\mu_i')}$.

Say we have k values of μ_i' for N times at bat for each player.

.11 The probability of this data is then $\prod_{i=1}^{k} P(\mu_i) \cdot \mu_i^{\mu_i' \cdot N}(1-\mu_i)^{N \cdot (1-\mu_i')}$.

.12 A reasonable form for $P(\mu_i)$ is $(B(x+1, y+1))^{-1} \mu_i^{x}(1-\mu_i)^{y}$.

The $B(\ ,\ )$ is the Beta function; this is a normalization factor for the integral.

$$B(x+1, y+1) = \frac{x!\,y!}{(x+y+1)!}.$$ [for info on Beta funct, see Bureau of Standards Math tables p 258]

The P form of .12 is also <u>very tractable mathematically</u>.

~~[illegible] Beta distribution, if we know k, y and then~~

The mean of the P distribution is $\frac{x}{x+y}$; the width of the distribution is $\approx \frac{1}{\sqrt{x+y}}$ maybe $(\sqrt{pqn})^{-1} = \left(\frac{x+y}{x \cdot y}\right)^{\frac{1}{2}}$

i.e. its standard deviation

maybe $\sigma^2 = \frac{x+y}{x \cdot y}$

Consider the cost of coding the $k \cdot N$ bits of data on whether each player had a hit or not on his N times at Bat.

Given x and y we code the data in 2 parts: First we describe the μ_i values.

.22 The probability cost is $\prod_{i=1}^{k} P(\mu_i) = \left(\frac{(x+y+1)!}{x!\,y!}\right)^{k} \prod_{i=1}^{k} \mu_i^{x}(1-\mu_i)^{y}$

.24 Given the μ_i, the probability of the observed data is $\prod_{i=1}^{k} \mu_i^{\mu_i' \cdot N}(1-\mu_i)^{N \cdot (1-\mu_i')}$

Given x and y; the total probability of the data is (.22) times (.24) or

.27
$$\left(\frac{(x+y+1)!}{x!\,y!}\right)^{k} \prod_{i=1}^{k} \mu_i^{x}(1-\mu_i)^{y} \cdot \mu_i^{\mu_i' N} \cdot (1-\mu_i)^{N(1-\mu_i')}$$

.30
$$\text{let } G = \ln(.27) = k \ln\left(\frac{x+y+1!}{x!\,y!}\right) + \sum_{i=1}^{k}\left[(x + \mu_i' \cdot N)\ln \mu_i + (y + N(1-\mu_i'))\ln(1-\mu_i)\right]$$

If we are to pick the μ_i so that this expression is Max

$$\frac{\partial G}{\partial \mu_i} = \frac{x+\mu_i' N}{\mu_i} - \frac{y+N(1-\mu_i')}{1-\mu_i} = 0$$; solving for μ_i and $1-\mu_i$ in terms of x, y.

.35
$$\mu_i = \frac{x+\mu_i' N}{x+y+N} \qquad (1-\mu_i) = \frac{y+N(1-\mu_i')}{x+y+N}$$

Noting that $\ln x! \approx x \ln x - x + \frac{1}{2}\ln x + \frac{1}{2}\ln 2\pi$; .30 becomes

.37
$$k \cdot \left[(x+y)\ln(x+y) - x\ln x - y\ln y + \ln(x+y+1) + \frac{1}{2}\ln\left(\frac{(x+y)2\pi}{x \cdot y}\right)\right]$$

.38
$$+ \sum_{i=1}^{k}\left[(x+\mu_i' N)\ln\frac{x+\mu_i' N}{x+y+N} + y + N(1-\mu_i')\ln\frac{y+(1-\mu_i')N}{x+y+N}\right]$$

2.7.99 SMFT STEIN

We want to find x, y such that the expression of 31.37, .38 is Max.

Let $E \equiv x+y$. and let $z_i \equiv \mu_i \cdot N$; $H_i \equiv x+z_i$; $L_i \equiv y+N-z_i$; $F = H_i+L_i = x+y+N$

For 31.37 we get

.04 $k\cdot\left[E\ln E - x\ln x - y\ln y + \ln(E+1) - \frac{1}{2}\ln\left(\frac{E\cdot 2\pi}{x\cdot y}\right)\right]$ and 31.38 becomes

.05 $+\sum_{i=1}^{k}\left(H_i\ln H_i + L_i\ln L_i - F\ln F\right)$

So we want to find x and y to Maxz. (.04+.05).

.09 The program STEIN7.Bas [illegible] would enable us to do this: The data is for N=45; k=18

From appstenori data, The values of x and y that maximize the probability of this data are $(x+y)=450$; $x=.275\cdot 450$; $y = .725*450$.

However, If we investigate behavior of .04+.05 in STEIN7.Bas, we find that There is no maximum for that expression: If we let $x = .25R$, $y = .75R$, then .04+.05 seems to approach ∞ as $R\to\infty$.

Doubling R increases (.04+.05) by about 6. so exp(.04+.05) is proportional to about R^{10}.

For fix large R, and $x = U\cdot R$, $y=(1-U)R$: There seems to be a peak for U about .25.

What this means is that μ_i is $\simeq .25$; independently of i.

For the present data, this gives not bad results. (much better than using $\mu_i = \mu_i'$), but still not correct.

2·8.99 SMFT STEIN.

Some possible objections to 31.01 – 32.40 :

1) That we are using 2+k params to describe k datapts. (i.e. the μ_i's).
2) That the method described isn't really Baysian.
3) The a priori of the x and y params. isn't clear.

~~[illegible]~~ Replys.

1): Actually we are using 2+k params to describe kn binary data bits.
2) It doesn't have to be Baysian (?) Hvr., it is in accord with the Mixed Corpus Theorem: which is Baysian.
3) We can give x and y a uniform apriori. since $\binom{x}{y} < \infty$:
It can't be done directly: Make the d.f. uniform from $\boxed{0 \text{ to } V}$
Say V = 1000: ~~[illegible]~~ If there was a peak, then as we increase V, the peak would remain about the same.
However, this does not occur, since the peak is usually at V.

The main Argument for the method is the Mixed Corpus Theorem.
I don't think the punctuation costs are relevant. — they're concerned with the relative pc. of various Models (= x, y values)

Perhaps look at the double Gaussian Version of the problem — see if it gives reasonable answers.

2.8.99: 1) BIG objection: Clearly $G \to \infty$ as $R \to \infty$: yet G can't be as large as ϕ, which is probability 1.
2) on the other hand, the D.F. P (31.11) is for proby density, not proby.
3) Or If I integrate over all R, then, using some d.f. on R so that $\int_0^\infty P(R)\,dR = 1$, I will mainly use low R.

12870

$\boxed{e^G \approx \propto R^{\frac{1}{2}} \ ?}$

4) say I know R is betw. 1 & ∞. I use $\frac{1}{R}$ has uniform d.f. betw. 0 & 1.
$R = R, R+\Delta$; $\frac{1}{R} = \frac{1}{R}$; $\frac{1}{R} - \frac{\Delta}{R^2}$ if $\frac{\Delta}{R^2} = \delta$; [illegible]
So $\boxed{R \text{ has } \frac{1}{R^2} \text{ d.f.}}$ —— ∴ e^G would still diverge!

$\frac{8!}{24^2} = 70$

$\frac{24}{4} = 6$

5) try n=1 analytically see if $e^G \propto R^{\frac{1}{2}}$:

$\ln x! \sim x\ln x + \frac{1}{2}\ln x + \frac{1}{2}\ln(2\pi)$

$\frac{2x!}{x!\,x!}$: $\ln(2x!) = \cancel{2x\ln x} + 2x\ln 2 + \frac{1}{2}\ln x + \frac{1}{2}\ln 2 + \frac{1}{2}\ln 2\pi$

$-2\ln x! = \cancel{-2x\ln x} - \ln x - \ln 2\pi$

$-\ln 2\pi$

$= 2\backslash$

$2x\ln 2 - \frac{1}{2}\ln x - \frac{1}{2}\ln\pi$
$\cancel{-\frac{1}{2}\ln 2\pi} + \cancel{\frac{1}{2}\ln 2}$
$\cancel{-\frac{1}{2}\ln x}$

o.k. $\frac{2x!}{x!\,x!} \approx \frac{4^x}{\sqrt{\pi x}}$

x = 2: $\frac{4^2}{2^3} = 6$ $\frac{16}{\sqrt{2\pi}} = 6.3$

x = 8   12870 $\left(\frac{4^8}{\sqrt{8\pi}} = 13072\right.$

$\frac{4^8}{\sqrt{8}}$

8 = x

$\frac{48}{8}$ $\frac{2^{16}}{\sqrt{2^3}} = 2^{13}$

$\frac{2x!}{x!\,x!}$ ~~$\approx 2^x$?~~ $\approx \frac{4^x}{\sqrt{x}}$

$\frac{x\ln x + \frac{1}{2}\ln x}{2x\ln 2x + \frac{1}{2}\ln 2\pi}$
$-2x\ln x - \ln x$

$2x\ln 2 + \cancel{[illegible]}$
$-\frac{1}{2}\ln x$

$1\frac{1}{2}\ln 2$ $\frac{2^{2x}}{\sqrt{x}}$ $\frac{4^x}{\sqrt{x}}$

x = 2 $\frac{16}{\sqrt{2}}$

2 899 SMFT STEIN → see 35.17 - .40

(also try $\frac{(x)!}{(x)!\,((1-x)x)!}$)

say $R \equiv 2r$.

any way: from 32.04! $E \equiv R = 2x$

$k \cdot (\frac{4^R}{\sqrt{\pi R}} (R+1) \cdot \frac{R^a}{2\pi R}$

$\frac{2r!\,(2r+1)}{r!\,r!} = \ldots = 2^{2r}$.

$\propto \frac{2^{2r}}{\sqrt{\pi}\cdot\sqrt{r}}(2r) = 2^{2r}\cdot\sqrt{r}\cdot\frac{2\sqrt{r}}{}$

$\frac{A(y+A)}{x} = A + A\cdot\frac{A}{x} = A(1+\frac{A}{x})$

.09 in 31.27 $\prod_{i=1}^{k}\left(\mu_i^{\,x+\mu_i' \cdot N}\;(1-\mu_i)^{y+N-\mu_i\cdot N}\right)$ from 31.35: $\rightarrow \left(\frac{x+\mu_i' N}{x+y+N}\right)^{x+\mu_i' N}$

$= \left(\frac{x}{x+y+N} + \frac{\mu_i' N}{x+y+N}\right)^{(\,)} = \left[\left(\frac{x}{x+y+N}\right)\left(1+\frac{\mu_i' N}{x}\right)\right]^{(\,)} = \left(\frac{x}{x+y+N}\right)^{(\,)}\cdot e^{\frac{\mu_i' N(x+\mu_i' N)}{x}} \rightarrow \left(\frac{x}{x+y+N}\right)\cdot e^{\mu_i' N(1+\frac{\mu_i'}{x})}$

.16 products of y & x factors: $\left(\frac{y}{x+y+N}\right)^{(\,)}\cdot e^{(1-\mu_i')N} = \left(\frac{x}{x+y+N}\right)^{x+\mu_i' N}\left(\frac{y}{x+y+N}\right)^{y+N-\mu_i' N}\cdot e^{N}$

.17 In R. case $x=y=r$ we get $\left(\frac{1}{2}\right)^{2r+N}\cdot e^{N}$ mult. by $\frac{4^r\,r\,2\cdot 2}{\sqrt{2\pi}} \rightarrow e^{N}\frac{\sqrt{r}\cdot 2}{\sqrt{\pi}}$

for each factor: so we get the power of .17 for the entire expression.

.19 ~~Ratio is~~ $e^{N\cdot k}\cdot\boxed{r^{\frac{k}{2}}}\cdot\left(\frac{2}{\sqrt{\pi}}\right)^{k}$ which is $\propto r^{9}$ since $k=18$. whereas observed exponent or $\frac{v6}{\ln 2} = 8.656$: so not far off!

from ~30.22 T. exponent ~ $\frac{6}{\ln 2} = 8.656$ approx holds.

This is for $U=\frac{1}{2}$; Actually I used $U = .25$ or $.265$.

.20 — NOTE by using $x = uR$, $y = (1-u)R$, I began t. harmonic ratio $\propto \sqrt{R}$ for mildly large R, → 13. when

Actually it's a prob. pc of R → see 35.22 $(\propto u < 1)$ must be $< \frac{1}{r}$, so $\left(\frac{\sqrt{r}}{r}\right)^{k} \rightarrow r^{-\frac{k}{2}}$. → Eqn $\propto$ [illegible] $\cdot\frac{\sqrt{R}}{\sqrt{u(1-u)}}$ see 35.22

so any convergent (& even non-convergent!) assignment of a prior to R will give a limit of ϕ for e^{ϕ} as $R\to\infty$. Try a prob for R of $\propto\frac{1}{R}$ then $\frac{1}{R^2}$! see if result is much different. Hm, a thing that annoys: I seemed to get a U of .25 rather than .265. for t. peak! Check this out!

Anyway, inserting a factor of R^{-1} or R^{-2} (or $R^{-1.5}$) is for d. At:

.25	29.70
.255	4.98
.260	4.5917
.265	4.617
→ .258	4.5513
.257	4.5577
.259	4.5519

.27 So: try t. output STEIN 7

Since e^{G} has a $E+1$ $(\equiv R+1)$ factor, Getting rid of it is equiv to append $\propto\frac{1}{R}$, $R>1$.

So STEIN 9.B ≡ STEIN 7.B, but G has no $\log(E+1)$ term.

I ran t. prgm: ~~It has a peak~~ its max (for $R \geq 0$) is ~ $R=0$! (hm R not so good to be ≤ 1, say because a prior diverges around R=1)

.32 ~~Stein 7~~ Stein 9: .265 450 -5246

for R = 450 best U was .259

for R = 10 best U was ~ .225

instead of $\ln\left(\frac{R+1}{R}\right)$ use $\ln\left(\frac{E+1}{\sqrt{R^2+10^2}}\right)$

This is a significant change! $\ln(E+1) - \frac{1}{2}\ln(R\text{^}2 + 10\text{^}2)$

This gave a peak for R near 10 ← using 100^2 here

Peak was at around R = 80

.22	.0087	.0187
~~.23~~	.0052	.05
.24		.0651
.225		2.9997

2994 SMRT Stein :

This is disturbing! ~~Peak~~ Peak depends critically on shape of apriori for R!

try $2^{-(\log_2 R + 2\log_2\log_2 R)} = \frac{1}{R(\log_2 R)^2}$ $2^{\log_2\log_2 x}$ $(\log_2 x)^{-2}$

This causes convergence at hy R values,

— But that's not the problem! (Note how the factor $\frac{1}{(\log_2 R)^2}$ only occurs when $\log_2 R$ is > 0, i.e. $R > 1$, — so the trouble around $R=0$ would persist.

Apriori of $\left(\frac{1}{\sqrt{R^2+A^2}}\right)$ seems to 8NE a peak at $\sim A$, ∴ critically dependent on "A" —

Using $A = 45$ gave peak in R at ~ $\boxed{30 \text{ to } 35}$ say 33 is v.g. 04

3E8 - 3E5

.17 Stein10.Bas

uses $R = 2^{.5/2}$

In Stein10.B ! ~~try~~ try + ln(E+1) → + (ln(E+1))/2. This flattens things out. Even if it worked, it would be equivalent to an apriori of $\sim \frac{1}{\sqrt{R}}$: expect trouble at low R

.19 Maybe try it in Steng: which has a loop for incrementing R..

.17 didn't seem to work; ~~[illegible]~~ e^G ↓ as R ↑, rapidly at first 2.3 /doubling of R at R≈1. Then less ↑/doubling — only .4 for R = 500 → 1000. Using u=.255 gave ≈ results. — the result seems very "flat" in u 1.4

Still! Even if in some sense there was a peak, clearly we need an apriori for R that is ~~[illegible]~~ at least as rapidly as $\frac{1}{R}$ for hy R. This simply cuts out hy R —

.22 R ; $\boxed{r \equiv R}$ $x = \overset{a}{u}R$, $y = (1-u)\overset{b}{R}$! $\ln\left(\frac{R!}{x!\,y!}\right) \approx$ $r\ln r + \frac{1}{2}\ln r$ $\left(+\frac{1}{2}\ln 2\pi\right.$ $-\frac{1}{2}\ln(2\pi)$

$- aR\ln^a r - \frac{1}{2}\ln ar$ $\quad -\frac{1}{2}\ln 2\pi$

$- bR\ln^b r - \frac{1}{2}\ln br$ $\quad \left.-\frac{1}{2}\ln 2\pi\right)$

$r\ln r - (a+b)r\ln r - ar\ln a - br\ln b$

$+ r\ln r$ cancel. $+\frac{1}{2}\ln r$

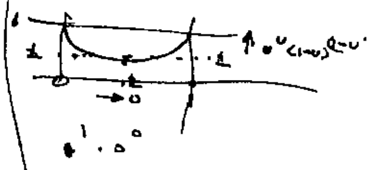
$- ar\ln r - ar\ln a$ $\quad -\frac{1}{2}\ln r - \frac{1}{2}\ln a$ $\quad -\frac{1}{2}\ln 2\pi$

$- br\ln r - br\ln b$ $\quad -\frac{1}{2}\ln r - \frac{1}{2}\ln b$

$a\ln ar = ar\ln r + ar\ln a$

$-\underbrace{(a\ln a + b\ln b)\cdot r}_{\text{(major term!)}} - \frac{1}{2}\ln r - \frac{1}{2}(\ln(a\cdot b)) - \frac{1}{2}\ln 2\pi$

.30 $\boxed{(a^a\cdot b^b)^{-r} \quad \sqrt{\frac{1}{r(a\cdot b)\cdot 2\pi}}}$ mult by $\boxed{r+1}$



If $x = y = \frac{1}{2}$ this ↑ is $\left(\frac{1}{2}^{\frac{1}{2}}\cdot\frac{1}{2}^{\frac{1}{2}}\right) = \frac{1}{2}$ This factor cancels that ~~[illegible]~~ inverse factor occurring in 34.16 : $\left(\frac{x}{x+y+N}\right)^{x+N}$ $\frac{2^{a\cdot r}}{(b^b)^r} = (a^a)^r$

and the corresponding y term away

So for large R ; $\left(R \gg 45\right)$ we put $E+1 = r \approx r$ so

$\left((r\cdot a\cdot b\cdot 2\pi)^{-\frac{1}{2}}\cdot r\right)^k = \left(\sqrt{\frac{r}{a\cdot b\cdot 2\pi}}\right)^k$ which is about e^G for "$\log_{10} R$".

$= \left(\frac{r}{a\cdot b\cdot 2\pi}\right)^{\frac{k}{2}}$ $a \equiv u;\ b \equiv (1-u)$

2.16.99 SMFT · STEIN:

Going back to the original Argument: Say $\bar{\mu}$ is the ~~[illegible]~~ "mean of means" & [illegible] $[\mu_i]$ are the true means. ~~[illegible]~~ Say σ_1^2 is the var of the "$\bar{\mu}$" d.f. & the μ_i's all have var σ_2^2.

If the d.f.s were Gaussian, Then it is cheaper to describe a set of μ_i that cluster closely about $\bar{\mu}$.

e.g. say $\bar{\mu}=0$; the set μ_i described about ϕ has p.c. $= \prod \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{\mu_i^2}{2\sigma_1^2}} = (2\pi\sigma_1)^{-k} \cdot e^{-\frac{\sum\mu_i^2}{2\sigma_1^2}}$ | $\frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{\mu_i^2}{2\sigma^2}}$

We select σ_1 so expt of e is $-\frac{k}{2}$ so $pc = (2\pi\sigma_1 e^{\frac{1}{2}})^{-k}$. So $pc \propto \sigma_1^{-k}$:

So smaller σ_1^2 is better (more likely).

On the other hand, if $[\mu_i]$ are the ~~[illegible]~~ empirical averages, we have to pay extra to move ~~[illegible]~~ the μ_i away from the μ_i, toward $\bar{\mu}$.

Using Bern derns: Given $[\mu_i]$ we want $\prod p(U, R, A_i) = \max$

Woops! Am I using the right normal. const for the Bern. μ_i d.f.? I think so: $\int_0^1 df(x)\,dx = 1$.

.17 $$\left(\beta(x,y)^{-1}\right)^k \prod_{i=1}^{k} \mu_i^x (1-\mu_i)^y = \max$$

$$k \ln\left(\frac{(x+y+1)!}{x!\,y!}\right) + x \sum \ln\mu_i + y \sum \ln(1-\mu_i) = \max$$

$$\prod_{i=1}^{n} \mu_i \equiv A \; : \; \prod_{i=1}^{n}(1-\mu_i) \equiv B$$

$A' \equiv A^{\frac{1}{k}}$ $B' \equiv B^{\frac{1}{k}}$

.20 ~~viz 3.1.30~~ viz 25.30 $$\left((a^a b^b)^{-(x+y)} \cdot \frac{x+y+1}{\sqrt{x+y}} \cdot \frac{1}{\sqrt{a\,b\cdot 2\pi}}\right)^k \cdot A^x \cdot B^y$$

looks like there is no optimum x, y!

.23 $$k\left(-\ln(a^a b^b)\cdot(x+y) + \tfrac{1}{2}\ln(x+y)\right) + x\ln A + y\ln B = \max.$$ for $x+y \gg 1$, ~~is wrong!~~ → 37.20

Wrong! → $$k\left(\tfrac{1}{2} - \ln(a^a b^b)\right)(x+y) + x\ln A + y\ln B = \max. \quad \text{or}$$

no! $$\left(\tfrac{1}{2} - \ln(a^a b^b)\right) k r + r(a\ln A + b\ln B) = \max \quad \text{— so } r \text{ is irrelevant!}$$

No prob, a (& b = 1 − a) can be found. (the maximality solved if we set $A' = \frac{A^{\frac{1}{k}}}{[illegible]}$, $B' = \frac{B^{\frac{1}{k}}}{[illegible]}$ = Geom means of μ_i's.

So this says that the optimum mean of the d.f. can be found, but there is no best Var! Which seems Crazy!

- Maybe not exactly right: Check algebra!

$$\ln(a^a b^b) = a\ln A' + b\ln B' + \tfrac{1}{2}$$

$$a^a b^b = (A')^a \cdot (B')^b \cdot e^{\frac{1}{2}} \; . \quad \Big| \; e^{\frac{1}{2}} \doteq 1.65$$

.17 ff may be on the right track!

.29 Var of $t^x(1-t)^y$: second moment − (first mom)^2

first moment = ~~$\frac{(x+y+2)!}{(x+1)!\,y!}$~~ ~~[illegible]~~

first moment, $$\frac{x+y+1!}{x!\,\cancel{y!}} \cdot \frac{x+1!\,\cancel{y!}}{x+y+2!} = \frac{x+1}{x+y+2}$$

Second moment: $$\frac{x+y+1!}{x!\,y!} \cdot \frac{x+2!\,y!}{x+y+3!} = \frac{(x+1)(x+2)}{(x+y+2)(x+y+3)}$$

$$\frac{(x+1)(x+2)}{(x+y+2)(x+y+3)} - \frac{(x+1)(x+1)}{(x+y+2)(x+y+2)} = \frac{x+1}{x+y+2}\left(\frac{x+2}{x+y+3} - \frac{x+1}{x+y+2}\right)$$

$$\frac{(x+1)(x+2)(x+y+2)^2 - (x+1)(x+1)(x+y+3)^3}{(x+y+2)(x+y+2)(x+y+3)}$$

$$\frac{x+1}{x+y+2}\left(\frac{(x+2)(x+y+2) - (x+1)(x+y+3)}{(x+y+3)(x+y+2)}\right)$$

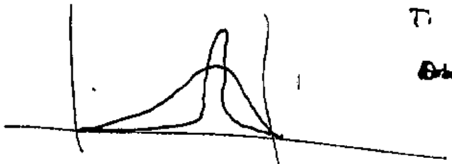
.40 $$= \frac{(x+1)(y+1)}{(x+y+2)^2(x+y+3)} = \frac{x+1}{x+y+2} \cdot \frac{y+1}{x+y+2} \cdot \frac{1}{x+y+3}$$

$$= \frac{x+1}{(x+y+2)^2(x+y+3)} \cdot \left[x^2 + xy + 2x + 2x + 2y + 4 - x^2 - x - 3x - y - 3 \atop y+1\right.$$

2.10.94 SMRT STEIN.

.01 If $x' = x+1$, $y' = y+1$ then $\frac{x'}{x'+y'} \cdot \frac{y'}{x'+y'} \cdot \frac{1}{x'+y'+1}$

.02 or $\approx \frac{u \cdot (1-u)}{r}$ which vaguely suggests that 36.20 & 35.30 may be wrong — i.e. perhaps ab should be in numerator. 35.30 should be easy to check, numerically



T. norman constant should ↑ as x & y ↑ because $u^{x}(1-u)^{y}$ will ↓.

~~Dave~~ D[redacted] have it backwards! I checked!

Norm. const $\int P(t)\,dt = 1$,

$$\frac{~~Gamma~~\ E^{E}}{x^{x}\ y^{y}} = \left(\frac{x}{E}\right)^{x}\left(\frac{y}{E}\right)^{y} = \frac{x^{x} \cdot y^{y}}{E^{x+y=E}} = \left(u^{u} \cdot (1-u)^{1-u}\right)^{R} \cdot E+1 \cdot \sqrt{\frac{E}{AB \cdot 2\pi}} = \sqrt{\frac{1}{u \cdot (1-u) \cdot E \cdot 2\pi}}$$

$$\frac{E}{A} \cdot \frac{E}{B} \cdot \frac{1}{E}$$

Normalization const $= \left(u^{u}(1-u)^{u}\right)^{-R} \cdot \left(\frac{E+1}{\sqrt{E}}\right) \sqrt{\frac{1}{u(1-u)\,2\pi}}$

$$\approx \left(u^{u}(1-u)^{1-u}\right)^{-R} \cdot \sqrt{\frac{1}{u(1-u)}} \cdot \frac{1}{\sqrt{2\pi}} \cdot \left(\sqrt{R} + \frac{1}{\sqrt{R}}\right)$$

$$\underbrace{\left(u^{u+\frac{1}{2R}}(1-u)^{1-u+\frac{1}{2R}}\right)^{-R}}_{\text{Between } \frac{1}{2} \text{ \& } 1} \cdot \left(\sqrt{R} + \frac{1}{\sqrt{R}}\right) \cdot \frac{1}{\sqrt{2\pi}}$$

So t. norm. const ↑ w. R.

.20: 36.23 36.23 is $\boxed{\text{a bug.}}$ $-\ln(a^{a}b^{b}) \cdot R + \frac{1}{2}\ln R + R(a\ln A' + b\ln B') = \max.$

.24 $R \cdot \left(a\ln A' + b\ln B' - \ln(a^{a}b^{b})\right) + \frac{1}{2}\ln R = \max$

$a\ln A' + b\ln B' - \ln(a^{a}b^{b}) \equiv \alpha$ $R\alpha + \ln R = \max$

$\alpha + \frac{1}{R} = 0$ $R = -\frac{1}{\alpha}$ $R = -\frac{1}{\alpha}$.

.26 $R_{\max} = -\left(\frac{1}{2} \frac{1}{a\ln A' + b\ln B' - \ln a^{a}b^{b}}\right) = -\frac{1}{2}\frac{1}{a\ln\frac{A'}{a} + b\ln\frac{B'}{b}}$ $A' = \left(\prod_{i=1}^{k} u_i\right)^{\frac{1}{k}}$, $B' = \left(\prod_{i=1}^{k}(1-u_i)\right)^{\frac{1}{k}}$.

$\frac{A'}{a}$ small, $\frac{B'}{b}$ small.

Not very reasonable.) $= \frac{1}{2} \frac{1}{\ln a^{a}b^{b} - a\ln A' - b\ln B'}$ $= \frac{1}{2}\left(a\ln\frac{a}{A'} + b\ln\frac{b}{B'}\right)^{-1}$ Equals A' is usually $< a$, B' " " $< b$.

So St 36.17 has a peak for $R (\equiv x+y)$

What about 26.05? $B(x,y)^{-k} \prod_{i=1}^{k} u_i^{x}(1-u_i)^{y} \cdot u_i^{u_i' \cdot N}(1-u_i)^{(1-u_i')N}$

$$\prod_{i=1}^{k} u_i^{x+u_i' N}(1-u_i)^{y+(1-u_i')N}.$$

Say we decide on a value of u ($x = uR$, $y = (1-u)R$) : u will be some kind of mean of the data $\{u_i'\}$.

Then we consider various α values $(0<\alpha<1)$: $\ni u_i = \alpha u + (1-\alpha)u_i'$ . .

.33 for each α compute ① The pc of the set of u_i's (as a funct of α) ~~[illegible]~~ w.r.t. the optimum fit of (.20 - .26)

② Compute the total pc of t. empirical data using these u_i's & u_i''s: — This, too, will be a funct of α.

① ~~[illegible]~~ should ↓ w. α ; it's max when $\alpha = 0$. (corresponds to $R = \infty$)

② should ↑ w. α ; it is max when $\alpha = 1$. (corresponds to $R = 0$)

Then presumably, t. product of ① ~~②~~ & ② should have a max value betw. 0 & 1. — Unless one or other dominates completely! — 2.11.94

2·11·99 SMFT STEIN.

Another possy: that usually U or $[U_i']$ dominates completely (?).

① & ② of 37.33 ff correspond to the 2 factor groups of 26.03.

$$\prod_{i=1}^{k} \left(\beta(x,y)^{-1} \; U_i^{x}(1-U_i)^{y}\right) \quad \text{v.s.} \quad \prod_{i=1}^{k} ~~[illegible]~~ \; U_i^{95U_i'} (1-U_i)^{95-95U_i'}$$

look at logs & compare. $U_i = \alpha U + (1-\alpha) U_i'$.

107.6

Perhaps let $x = UR$, $y = UR$; use 27.26 to est R from $[U_i']$

Go back to 37.20 : we can also find optimum a, b : The shortcut

R ~~[illegible]~~ for max, is in terms of a, b, A', B' ! We can also optimise 37.21 :

Then min indep of R : $\frac{d}{da}\left(a \ln A' + \underset{(1-a)}{b} \ln B' + a \ln a + b \ln b\right) =$

$\ln A' - \ln B' + \ln a \boxed{+1 - 1} - \ln(1-a) = 0$ $\quad +(1-a)\ln(1-a)$

$\ln \frac{A'}{B'} = \ln \frac{a}{1-a}$ $\quad \frac{A'}{B'} = \frac{a}{1-a}$. so $a = \frac{A'}{A'+B'}$, $1-a = b = \frac{B'}{A'+B'}$

$R = \frac{1}{2}\left(a \ln \frac{a}{A'} + b \ln \frac{b}{B'}\right)^{-1} = \frac{1}{2}\left(\frac{1}{A'+B'}\left(A' \ln \frac{1}{A'+B'} + B' \ln \frac{1}{A'+B'}\right)\right)^{-1}$

$= \frac{1}{2}\left(\ln\left(\frac{1}{A'+B'}\right)\right)^{-1} = -\frac{1}{2}\frac{1}{\ln(A'+B')} = \frac{1}{2}\left(\ln\frac{1}{A'+B'}\right)^{-1}$ $[A'+B' < 1]$

.20 So $a = \frac{A'}{A'+B'}$; $b = \frac{B'}{A'+B'}$; $R = \frac{-1}{2\ln(A'+B')} = \frac{-1}{\ln(A'+B')^{2}} = \frac{1}{\ln\frac{1}{(A'+B')^2}}$ $\left\{ \approx \frac{1}{2}\frac{1}{1-p-q} \;\; \text{cf } 21.13 \right.$ $\quad \approx x+y \left\{ \frac{x+y=}{\frac{1}{1-p-q}+\frac{1}{2}} \right.$

This result is almost same as 21.13

Now if we move the μ_i's toward a (& $(1-\mu_i)$'s toward b,

presumably A'+B' get closer to a+b = 1 ; ∴ $\ln(A'+B') \downarrow$ & $R = \frac{1}{2}\frac{1}{\ln(A'+B')}$ ↑.

What about $\underline{\underline{p}}$? As μ_i's & $(1-\mu_i)$'s move toward a & b, resp, I assume that $\frac{A'}{A'+B'}$ & $\frac{B'}{A'+B'}$ remain about the same, but A' & B' both ↑ so A'+B' → 1.

$$\prod_{i=1}^{k} \mu_i' \to \prod\left(\mu_i'(1-\epsilon) + \epsilon a\right) = \prod\left(\mu_i' + \epsilon(a - \mu_i)\right)$$

$$= \prod\left(\mu_i'\left(1 + \epsilon\frac{(a-\mu_i')}{\mu_i'}\right)\right) = \prod_i \mu_i' \cdot \prod_i\left(1 + \epsilon\left(\frac{a}{\mu_i'} - 1\right)\right)$$

$$\prod_i (1 + \epsilon z_i) \approx \exp(\epsilon \Sigma z_i) \;\Big|\; \exp\left(\epsilon\left(\left(\Sigma\frac{1}{\mu_i'}\right) - k\right)\right)$$

This is not small but it may become small if combined w. corresp'g "b", $(1-\mu_i')$ · terms.

SN The Approp'st of R & U only occurs once in the derv ! Check this out, hvr! I think this is correct. !

Hvr, a different error : The values of U_i chosen are given prob'y densities for them in pc ; this is of course, wrong. The U_i are continuous reals hvr. — Normally, in dealing with a seq. of continuous reals, we assume some (narrow) precision, Δ, so we mult. the density by this Δ, to obtain a prob'y. In the present case, we would have to mult each μ_i by its "width" — the width obtained in deriving its ~~[illegible]~~ (N scaled) sub corpus. This "width" is $\tilde{\propto}\left(\sqrt{U_i(1-U_i)}\right)^{-1}$

.38 It modifies the d.f. a bit. — maybe ~~[illegible]~~ ! (Hvr, I may have made that "error" — but w/ a success!)

→ 40.01

Nailboxes etc.

Kennel: 7.22 + 35

76 Arlington Watertown

5.20 · 3-7 PM.

661 2622.

$9

To Brook

No. Dist. Fireplace.

5 bags

$7.57 4

9 am 70%

To Aron ALG

Phone

Nyborn rent.

335 482

46

Compost

Processing.

11 Feb.

User CRAZHY

Password. MAEVA

Naburn.com.

Lotteries

STEIN 6 → 7

6 to 9 8

7 → 9

8 → 10

9 → 11

2.12.99 INDEX

(39)

Random Notes: ① t. aprgd should be only used once in a problem; It is then modified in subsequent problems. I seem to be using the same aprgd k times!

② T. result I got wrt. to aprgd seems smooth — tho at one pt I seemed to have gotten a "local max" in R (see 29.12R).

Stein pgms:

.07 Sam1.bas (9.17) Examines rate of convergence of certain successive approxn systems relating to the Gaussian stat'.

$$X_{n+1} = -(\mu + X_n)(\sigma^2 + X_n^2)/ks^2 .$$

Stein2.bas Sam2.bas ~21.13

computes $\rho \equiv \underset{.2569572}{\left(\prod_i A_i\right)^{\frac{1}{k}}}$; $Q = \underset{.7313933}{\prod_i (1-\mu_i^2)^{\frac{1}{4}}}$ and $\frac{1}{1-\rho q} = 85.7639$

Also computes Same for various approx. methods; $\frac{1}{2(1-\rho q)} = \underline{42.8819}$

i.e. using $\mu_i = \mu_i'$; using "S form"; (using ~~[illegible]~~ some other way, method) of [illegible]

NB — 19.05–.40 compares both A's for x & y to date: This is done again on 36.07–.23, 37.20–.38.40 : The results are almost the same, but x+y differs by about ½ (addition)

$\ln x = 1 + x - 1 - \frac{(x-1)^2}{2} \dots$

I think the earlier result may be more exact since it is based on the approxn. $\psi(x) \approx \ln(x+\frac{1}{2})$ which is a v.g. approxn. T. more recent result is based on $x! \approx x^x \cdot e^{-x} \cdot \sqrt{2\pi x}$, which isn't as close.

~~[illegible] 34.27~~
~~Stein 7 old 729.05~~

Stein10 (old) 35.17

Steing old 35.19

See (ST5.01–.20R) for more notes on this content

2.12.99 SMPT STEIN

.01:38.40 : 3 years $\binom{35.22}{35.40}$ ends up w. an expression for $\frac{(x+y)!}{x!\,y!}$: $x = ar$, $y = br$ $a+b=1$. $r >> 1$

.02 $\frac{x+y!}{x!\,y!} \doteq (a^a b^b)^{-r} \frac{1}{\sqrt{(r\cdot a\cdot b\cdot 2\pi)}}$: we mult by $(r+1)$ to get the normaliz. const

.05 $'' \cdot (r+1) \approx (a^a b^b)^{-r} \frac{\sqrt{r}}{\sqrt{2\pi\, ab}}$

.10 we mult .04 by $\left(u_i^x (1-u_i)^y\right) \prod_{i=1}^{k}\left(u_i^{45\mu_i'-\frac{1}{2}}\cdot(1-u_i)^{45-45\mu_i'-\frac{1}{2}}\right)$ (the ($-\frac{1}{2}$'s) are new w.r.t. 35.55

~~In [illegible]~~ 35.38 The "$-\frac{1}{2}$"'s might change $\bar{A}$ a bit, but wouldn't affect R (= x+y) at all.

.15 $\prod_{i=1}^{k} u_i^{x+45\mu_i'-\frac{1}{2}}\,(1-u_i)^{y+45-45\mu_i'-\frac{1}{2}}$; so $u_i' = \left(\frac{x+45\mu_i'-\frac{1}{2}}{x+y+45-1}\right)$ $1-u_i' = \frac{y+45-45\mu_i'-1}{x+y+45-1}$

$\boxed{\text{STEIN 40.BAS}}$ (slite modification of STEIN??.BAS i.e. (5.10) → also $+\frac{1}{2}\ln(E+1) \to +\ln(E+1)$ (using $U = .255$

~~Doubling R changes G by only [illegible]~~

result from R (= x+y)/= 1.4 to 1024, G was monotonic R.

for Doubling R ΔG = .9 for R = 1→2

" " " " 5.8 " R = 512→1024

" " " 7 " R = 524288→1048576

For R = 8,388E6, doubling R gives ΔG = ~ 81) Really crazy! Trouble w. basic (interpreted)? or Q basic or whatever.

For larger R, G ~~[illegible]~~'s increases oscillates, oscillates!

Get Power "Basic" Math. (This is the interpreted Q Basic that came w. DOS 6.22

It would seem possi to show that exp(G) is like $\propto R^{\alpha k}$ w. α say ½ — but > 0.

If so, that the psm is clearly wrong because exp(G) can't be > 0 — all p's are ≤ 1.

Now, the normaliz factor .02 (is always < 1 — w.h. legs) ← woops! no! its always >> 1. a or b is very close to 1 —

The factor (.15) can't be > 1 — its the product of factors, u_i, $(1-u_i)$ — all ≤ 1.

So how did I get $e^G \propto R^{\frac{k}{2}}$? — ⇒ 34.19 — this was for u = ½, but

r83 35.22 — .30 In 35.30 we get Normaliz factor $\boxed{(a^a b^b)^{-R}\cdot\frac{\sqrt{R(+1)}}{\sqrt{2\pi\, ab}}}$

For the rest of the e^G: 34.16 gets $\left(\frac{x}{x+y+N}\right)^{x+\mu_i' N}\cdot\left(\frac{y}{x+y+N}\right)^{y+N-\mu_i' N}\cdot e^N$

So 34.16 looks like $a^{aR}\cdot b^{bR}\cdot e^N = (a^a b^b)^R\cdot e^N$.

~~Multiplying the 2 factors:~~ More exactly: $a^{aR}\, b^{bR}\cdot a^{\mu_i'}$

Well 34.16 needs cleaning up!

2.12.99 SMRT STEM.

Start w. 34.09 !! Just one factor: $\prod_i \left(\frac{x+\mu_i' N}{x+y+N}\right)^{x+\mu_i' N} \Rightarrow \prod_i \left(\left(\frac{x}{x+y+N}\right)\left(1+\frac{\mu_i' N}{x}\right)\right)^{x+\mu_i' N}$

$$= \prod_i \left(\frac{x}{x+y+N}\right)^{x} \left(1+\frac{\mu_i N}{x}\right)^{x} \left(\frac{x}{x+y+N}\right)^{\mu_i' N} \cdot \left(1+\frac{\mu_i' N}{x}\right)^{\mu_i' N}$$

$\prod_i \sim \quad a^{x} \overset{''}{=} a^{aR} \quad \overset{''}{e}^{\mu_i' N} \quad a^{\mu_i' N} \quad e^{\frac{(\mu_i N)^2}{x}}$

[illegible] $\prod_{i=1}^{k} \mu_i^{x+\mu_i' N}$

corresp'g "y" products $\quad b^{bR} \quad e^{(1-\mu_i')N} \quad b^{(1-\mu_i')N} \quad e^{\frac{((1-\mu_i)N)^2}{y}}$

End product $(a^{a} b^{b})^{\underline{\underline{R \cdot k}}}$

~~which totally dominates the~~ $(a^a b^b)^{-R}$ of 40.33 (i $\frac{35:30}{\underline{35.30}}$)

No! The Norm. factor is taken to the R^{th} power!

$\prod_{i=1}^{k} \mu_i^{x}$! for $U_i = a$; becomes $(a^{x})^{k} \cdot (a^{a})^{Rk}$. – so .08 is rite!

so $\underbrace{(a^a b^b)^{Rk}}$ $\cdot\, e^{kN} \quad a^{N \cdot \Sigma \mu_i'} \quad \cdot\, b^{N - N \Sigma \mu_i'}$ $\quad e^{\frac{(\Sigma \mu_i'^2) \cdot N^2}{x}} \cdot e^{(\Sigma(1-\mu_i)^2) \cdot \frac{N^2}{y}}$

$a^{N^2 \overline{\mu_i'}} \cdot b^{N^2(1-\overline{\mu_i'})}$

$\text{[illegible]}(a^{a} b^{b})^{N^2}$

The factor ↑ cancels the corresp. [illegible] factor.

$\exp \frac{N^2}{y} \cdot \sum_{i=1}^{k} \left(\mu_i^2 \, (1-\mu_i)^2\right)$

None of the other terms are much dependent on K.

$$(a^a b^b)^{Rk + N^2} \cdot \exp\left(kN + \frac{N^2}{R}\left[\frac{\Sigma \mu_i^2}{a} + \frac{\Sigma (1-\mu_i')^2}{b}\right]\right)$$

so I still looks like $\overset{e^{a}}{}$ must be $\tilde{\propto}$ $R^{\frac{k}{2}}$ for large R – which is impossl.!

Putting in the "$-\frac{1}{2}$" term of 40.10 doesn't make much difference.

So, if I could really prove that the formula for G has to be > 0 for very large R — that would prove that I couldn't be the correct formula!

•23 Another possy! (very likely!) That the width correction factor should be like $\sqrt{\frac{U_i(1-U_i)}{R}}$

This just cancels out the $R^{\frac{k}{2}}$, but I think I [illegible] fixed it:

so in ~~[illegible]~~ $\boxed{\text{Stem 40}}$ $-\frac{1}{2} \to +\frac{1}{2}$ & divide by K!

•28 Stem 41.Bas

Dividing by $E\sqrt{E}$ elims E from $\frac{1}{2}\log(E/A/B$....

From $R = 1$ to 8.3M, G is [illegible] funct of R! Above R = M, the QBasic may not work! for R > 1000.

rather constant from $R \approx k$ to M [illegible]

Actually G may flatten out ↑ plateau

→ 2 non-monotonic glitch at 262k → 524k!

also 80k → 160k is non monoton.. Also $R = 3.8M$ – 480.8

Also local max at $R = 452k$; $G = -482$

I should be able to show that $R \approx 0$ is worse than $R = \infty$.

In fact, I know that! I.e. $R = 0$ means we use $\mu_i = \mu_i'$!

$R = \infty$ means we use $\mu_i = \overline{\mu}$, which is better.

Maybe do as double prec. in PowerBasic

2-14-99 .SMFT ST (≡ STem) $\binom{A}{B} \equiv \frac{A!}{B!(A-B)!}$ 42 $\boxed{342\frac{1}{2}}$

.0c ST:42 ! Contrast 2 ways of doing data : about common $\bar{u}$ & about individual $\bar{u}_i$; $\bar{s}$ (= u_i''s). See if both are logits or whether normal, punch is

Needed: Remember that method of doing a {Binary} Bar scap. of unknown "noise" ; P.C. was $\frac{n+1!\,m+1!}{n+m+2!}$ — Wrong! — See 43.01 — $\frac{n!\,m!}{(n+m+1)!}$

SN ; it was correct to use the "pre corpus 0,1", using pre corpus 1,0 would give a new set of codes. — must of them = old "0,1" codes.

So contrast the 2 ways of deriv. $\frac{x+y!}{x!\,y!} \approx (a^a b^b)^{-R} \cdot \frac{1}{\sqrt{2\pi a b R}}$ $a \equiv \frac{x}{x+y}$ $b \equiv \frac{y}{x+y}$ $R \equiv x+y$

.10 So the $\bar{U}$ method → $\left(\bar{U}^{\bar{U}}(1-\bar{U})^{(1-\bar{U})}\right)^{NK} * \sqrt{2\pi\, U(1-U)\, NK}$

.11 U_i' method. → $\boxed{\prod_i \left(\left(U_i'^{\,u_i'}(1-\bar{U})^{(1-\bar{U})}\right)^{N} * \sqrt{2\pi\, U_i'(1-U_i')\, N}\right)}$ $\approx \frac{x!\,y!}{x+y!} = \frac{x!\,y!}{NK!}$ Not reciprocal

~~It will be easier to do it~~ $\bar{U}$ & U_i' will ~~be~~, (clearly) be modified by

$\frac{x}{x+y} \to \frac{x+1}{x+y+2}$. ; how, omit this term, to start.

U_i' method ~~$\sum z \ln z + (1-z)\ln(1-z)$~~

$z = N U_i'$: H = z ; L = N − z ; F = N Modific. of Strain 9 (st9.bas) No :

.20 $\bar{U}$ method: $k=1$; $U = \bar{U}$: $\boxed{\sum_{\text{k}} \left(U \ln U + (1-U)\ln(1-U)\right) N K + \frac{1}{2}\ln\left(2 * \pi * U(1-U) * N * K\right)}$ ~~Met~~ Web. Mit.edu Search Soc. for refer. am. py

6.28318

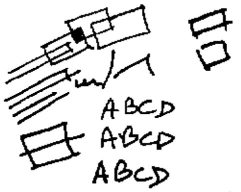
U_i' method: .20, but $U = u_i'$; $k=1$ ≤ (20) This will work for both methods!

So just do prog twice;

① $k=1$, $u = \bar{U}$

② $k=18$, $N=45$, in loop, $U = .045 * A(i,3)$

Modific. of Strain 9



ABCD
ABCD
ABCD
ABCD

abcdef

457/9 =

9.5 · 2.4 = 810

290 to 240.

ABC
DEF

.29 for u_i' method : G = −494.9771 ; for $\bar{u}$ method. : G = −464.9104

~~471.8094~~
~~23.1577~~ Δ u

$\bar{U}$ is ≈ e^{20} times as good as $[u_i']$!

So ΔG = 30 ~~[illegible]~~ ~~[illegible]~~ seems CRAZY !

$\ln \frac{10!\,40!}{50!}$ = −23.0527

Prog. got −23.06146 ~~[illegible]~~ ~~[illegible]~~

U = .2, 50

(.2, 50)

Woops! using $\bar{u} = .25$ & out G = −452.061 ~~for~~

So .25 is better than .265 !) o.k.! o.k.

.2 → −401.97) .5 is worst, 0,1 best.

u = .001 → −~~[illegible]~~ 55.54 .0001 → 1.16 0 prob → 0.

.5 → −557.87 Woops! .8 → ~~[illegible]~~ 481.9708 as it should.

$\bar{U}$ = .2653889

This seems to question the correctness of the $\frac{x+1!\cdot y+1!}{x+y+2!}$ method of coding!

try the x → x+1, y → y+1 correction.

2.14.99 SMFT : ST

.01 $\frac{x+1 \cdot y+1}{x+y+2}$ v.s. $\frac{x!\,y!}{(x+y)!}$: $\frac{(x+1)(y+1)}{(x+y+1)(x+y+2)} = \frac{x+1}{(x+y+1)}\left(\frac{y+1}{x+y+2}\right) \approx u(u-1)$

.3 × .7 = .21 which would make $\{u_i'\}$ even worse!

Actually, t. | Pc is $\frac{x!\,y!}{(x+y+1)!}$ [illegible] — which is much lower pc than $\frac{x!\,y!}{(x+y)!}$ ...factor of $\left(\frac{1}{x+y+1}\right)$: in this case $\frac{1}{45+1}$

01:010 $\quad \overset{\cdot}{\frac{1}{2}} \quad \overset{1}{\frac{1}{3}} \Big| \overset{0}{\frac{2}{4}}$

So mult by (N+1)

.09 T. differences: betw. the 2 pc's should be ≈ pc for $\bar{U}$: $\frac{1}{\sqrt{2\pi U(U-1)\,kN}}$ | $\frac{1}{kN}$

for $\Pi[u_i']\,\Pi \frac{1}{\left(\sqrt{2\pi\,(\ldots u_i(u_i-1)N}\right)}\cdot N$

omit for, checking 42.29

IN 5.4$\frac{2}{}$: I used k=88, $\bar{u}$ = .265

~~for N=278.2789 N=250, Both PS Gul's were ≈ (within ...) 00.~~ Bug in Fem.

~~N=280~~

~~So ΔN = Δ ... exp(Δ) → × 5.3~~

N	$G_{\{u_i'\}} - G_{\bar{u}}$
~~260~~	~~−3.06~~
~~250~~	~~−4.72~~
~~270~~	~~−1.39~~
~~280~~	~~+ .29~~
278.2789	+ .00097
300	109.35
50	43.3
11	22.5
4	.1528

) Δ ~ 1.66 exp(1.68) = 5.3

) Δ ~ 1.68

So $G_{u_i'}$ is always better — which seems unreasonable

~~Tryd $\left(\sqrt{R+1}\right)^{-1}$ term.~~

~~∴ Same Equal result $[U_i]$ is always ahead; ... difference → 0 as N → 0.~~

$\frac{x!\,y!}{x+y! \cdot (x+y+1)} = (2^2 b^b)^R \frac{\sqrt{2\pi ab R}}{R+1}$

N	$G_{\{u_i'\}} - G_{\bar{u}}$
100	−.16
200	−1.66
270	
	10.02
209.9486	−.056
220.	+1.590
210	−.0485

,28

.30

← $\frac{1}{R+1}$ is interim

So N ≈ 210 is the transition.

N±10 gives exp(Δ) = × 5.2 or × $\frac{1}{5.2}$

So N±10 → factor of 5 betw $\bar{U}$ & $\{u_i'\}$.

← Those are for $\bar{U}$ = .265 for $\bar{U}$=.2053889 converges ~ N≈201.4

Δ = 1.65 or .165 for ΔN = 1

exp(1.65) ~ × 5.2

⇒ : with N ≈ 200, T. observed u_i' ... equally ... (they to have been done by a single person sep. or by 15 Barnsleys!

This latest result isn't so crazy! That the transition of w/s betw $\bar{U}$ & $\{u_i'\}$ should be so sharp, is certainly not expected. But N≈200 is only ~ ½ of 450! — which is what I expect the "crossing @ x+y" to be!

For by program ~~xxxx~~ ln x! = x ln x + ½ ln (x · 2π) − x + $\frac{1}{12x}$

in $\frac{x!\,y!}{x+y!}$ the −y cancel but the $\frac{1}{12x}$ doesn't so + $\frac{1}{12}\left(\frac{1}{x} + \frac{1}{y} - \frac{1}{x+y}\right)$ or $\frac{1}{12}\left(\frac{(x+y)(x+y)}{xy \cdot x+y} - \frac{(xy)}{xy(x+y)}\right)$

$\frac{1}{x} + \frac{1}{y} - \frac{1}{x+y} = \frac{y \cdot (x+y) + x(x+y) - xy}{x \cdot y \cdot (x+y)} = \frac{(x+y)^2 - xy}{xy(x+y)} = \frac{x^2 + xy + y^2}{xy\,(x+y)}$.

$\frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}$

For x=2 ... ≈ x! gives 2.00065

(1/x + 1/y − 1/(x+y))/12

for x=y=2 this amounts to $\left(\frac{1}{2} + \frac{1}{2} - \frac{1}{4}\right)/12 = \frac{3}{4} \cdot 12 = \frac{1}{16}$

2.16.99 SMFT STEIN

.01 Actually, the $\frac{x!y!}{(x+y+1)!}$ formula for pc of Bin/Bern. string is just using the "Laplace's rule" for each bit.

The "precorpus" of 0's, 1's is just a way of making it easy to remember.

Using $\bar{u}$ alone for pc. would give $\left(\bar{u}^{\bar{u}}(1-\bar{u})^{(1-\bar{u})}\right)^{NK}$. By not needing to specify $\bar{u}$, we increase the pc of ~ $\sqrt{2\pi\,\bar{u}(1-\bar{u})}$ $\left(\frac{\sqrt{NK}}{\sqrt{NK+1}}\right) \approx \sqrt{2\pi}\cdot\sqrt{\frac{\bar{u}\cdot(1-\bar{u})\cdot}{NK}}$ $\frac{z+1}{\sqrt{z}} = \sqrt{z}+\frac{1}{\sqrt{z}}$.

I think that usually ~~$\left(\prod_{i=1}^{k} u_i^{u_i}(1-u_i)^{1-u_i}\right)^{\frac{1}{k}} < 1 \approx 1$~~, no! if $\left(\prod u_i\right)^{\frac{1}{k}} \equiv A$, $\left(\prod(1-u_i)\right)^{\frac{1}{k}} \equiv B$

Then $\frac{A}{A+B} \approx u_i$!

u_i is also $\approx A$, but not as closely as

ON 43.09 - .40! Just how did I vary N? OH! easy! It's the totally difficult to vary, There is something peculiar going on here!

.16 using the old SO(64b code, I got pc of ~ $(a^a b^b)^N\sqrt{\frac{2\pi\, ab}{N}}$ } for a N bit corpus.

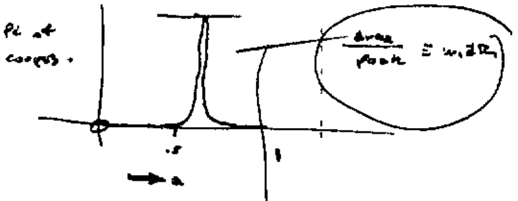
Using the statement of a (= u) followed by the corpus: I get $\sqrt{\left(\frac{ab}{N}\right)}$ for the pc of deriving "a": It's width $\sqrt{\frac{ab}{N}}$.

So $(a^a b^b)^N\cdot\sqrt{\frac{ab}{N}}$; so how come the extra $\sqrt{2\pi}$ here? ($\approx$ 2.5066)

2 × this is more reasonable & $\sqrt{2\pi} \approx 2.5$ times that is more exact

It could be that the σ of the d.d. is not closely related to its width — in fact 2σ would be closer! And it may well be that if one takes a wtd mean of all cases, one will get $\sqrt{2\pi} \approx 2.5$.

There is prob'y a very easy exact way to get the desired figure.



$\frac{\text{area}}{\text{peak}} \equiv$ width

$\left(a^{\bar{u}}(1-a)^{1-\bar{u}}\right)^N$

peak = $\left(\bar{u}^{\bar{u}}(1-\bar{u})^{1-\bar{u}}\right)^N$

area = $\frac{\left(\bar{u}^{\bar{u}}\,1-\bar{u}^{1-\bar{u}}\right)^{-N}\sqrt{N}}{\sqrt{2\pi\,\bar{u}(1-\bar{u})\cdot}}$ Approx.

so the width may, indeed be just $\sqrt{2\pi} \approx 2.5$.

$\int_0^1 x^x(1-x)^y\,dx = \frac{x!y!}{(x+y+1)!}$; $a = \frac{x}{x+y}$; $= (a^a b^b)^{-(x+y)}$

So, the moral is, we may get just about exactly the same result if we use either (SO64b) for ~~[illegible]~~ or (the best value for u, multiplied by its width times the pc of the corpus wrt. that u.) to get the pc of a Bern seq,

.30
~~.33~~

A source of Systematic Error in the Baseball data: Players w. low scores at the start of the season tend to be discouraged, so they have fewer times at bat for the rest of the season

.32 So, the present problem: According to recent analysis, for n < 210 (say N = 45) the ~~[illegible]~~ u_i are essentially useless for predn. of u_i's. $\bar{U}$ is much much better!

Hvr, in fact, the u_i's were useful. i.e. the ordering of u_i was useful for ordering out of sample u_i'. only 3 cases out of 9 of players in top 9 to start, going to bottom 9 in "out of sample".

SEE 140.05 - 140.40 (more specifically 140.32ff) for a good discussn. & understanding of this problem)

$\frac{\Gamma(x+y+1+1)}{x!\ y!}$. $x+y+1!$

2.16.99 SMPT Stam:

So maybe write review:

I nature of problem.

II Stam's soln(s)

III My solns:
ⓐ using Gaussian eff.
ⓑ " $t^x(1-t)^y$ eff. — I think way that seems to work using $\frac{1}{SSZ} = \frac{1}{SSZ_1} - \frac{1}{SSZ_2}$
ⓒ " 50/64b : sort of works but 49.32 suggests maybe not! ⓒ is a special case of
46.29 – 51.40: conv ~~only~~ R=0 & R≠0 coeffs.

IV Give Main "Math" results in easily readable form:

$\frac{x+y!}{x!y!}$; $\frac{x+y+1!}{x!\,y!}$ etc. "width" of d.f.s. Stuff about $\phi(x)$, etc. (≡ var)

Also tell what each $\boxed{\text{Stam param}}$. Maybe use WSS to change names from Stam to "ST". (The fields not in etc., now that I'm dropping off the project! (see (50.01 – .20R) for some param ...

V Tell what major diffys & conclusions are.

2.1699 SMFT STam:

46
∋ 46½

Paradox! Consider 26.03:

.02 $e^G = \prod_i \frac{x+y+1!}{x!y!} \underbrace{\mu_i^{x+\mu_i \cdot N}(1-\mu_i)^{y+N-\mu_i N}}_{\text{"}\mu_i\text{ factor"} = F_i(\mu_i)}$:

for each i, for each of the A_i factors there is a peak value of μ_i: $\frac{x+\mu_i \cdot N}{x+y+N}$: If V_i is the value at the peak, it's

.05 width is $\frac{1}{V_i}$, since $\int_0^1 F_i(\mu_i)\,d\mu_i = 1$. (The "width" being defined ∋ width peak $= \int_{-\infty}^{+\infty} f(x)dx$.)

See .14 ff for resoln. of this difty.

One immediate diffy w. this Argt., is that it is used usefully in 44.26–.33 to show the Equivalence of 2 ostensibly different ways to get the pc of a Bin. Bern. seq.

→ from 44.01 $\frac{x!y!}{x+y+1!}$ is the pc of a Bin. Bern seq. w. x "1's"; y "0's".

→ Also $= \int_0^1 p^x(1-p)^y\,dp$. — this $\int$ is the "sum" of all different possible params, p, being used to code the sequence.

.14 So perhaps, for a given $x+\mu_i \cdot N$, $y+N-\mu_i \cdot N$, in .02, the pc should be $\frac{x+\mu_i \cdot N!\; y+N-\mu_i \cdot N!}{(x+y+N+1)!} = \frac{x+A_i!\; y+B_i!}{(x+y+N+1)!}$

Instead, I used $A_i = \mu_i \cdot N$; $B_i = N - A_i$ $\boxed{\left(\frac{A+x}{x+y+N}\right)^{A+x}\left(\frac{(B+y)}{x+y+N}\right)^{B+y}}$

$\ln \frac{x+A!\; y+B!}{x+y+N!\;(x+y+N+1)} = \underline{x+A\ln x+A + y+B\ln y+B - x+y+N \ln x+y+N} + \frac{1}{2}\ln\left(\frac{2\pi(x+A)(y+B)}{(x+y+N)(x+y+N+1)^2}\right)$
$+ \frac{1}{2} \cdot \ln(\quad)$

$\ln\left(\frac{u!v!}{u+v!}\right) = u\ln u + v\ln v - (u+v)\ln u+v + \frac{1}{2}\ln\left(\frac{2\pi\, u v}{u+v}\right)$

I used instead just ————, omitting the factor $\sqrt{\frac{2\pi(x+A_i)(y+B_i)}{(x+y+N)(x+y+N+1)^2}}$ —

which amounts to an ↑ in x by ½ & y by ½ & also a factor of $\sqrt{\frac{2\pi}{x+y+1}}$

so it's like $\sqrt{\frac{2\pi \quad x \cdot y}{x+y+1}}$; which is the "width" of the distribution

which I may have should included.

So This Seems to Resolve the difty of .02–.05

So the exact expression we want is:

.28 $e^G = \prod_{i=1}^{k} \frac{x+y+1!}{x!y!} \frac{(x+A_i)!\,(y+B_i)!}{x+y+N+1!}$ Seems correct;

was that due to a width argt.? not

ST41 uses $E\ln E - x\ln x - y\ln y + \frac{1}{2}\ln\left(\frac{1}{2\pi XY}\right) + \ln(E+1)$

$E = x+y$

G = $\boxed{\begin{aligned}&E\ln E - x\ln x - y\ln y + \tfrac{1}{2}\ln\left(\tfrac{E}{x\cdot y\cdot 2\pi}\right) + \ln(E+1)\\ &\quad H_i = x+A_i;\ L_i = y+B_i\quad F = H_i+L_i\\ &+ \sum_{i=1}^{k} H_i\ln H_i + L_i\ln L_i - F\ln F - \ln(F+1)\end{aligned}}$

See 69.01 .... for a good way to look at this

ST 46 from ST41

These G values are pretty off: due to +s in H & L of ST41 This was Corrected on 47.10

I ran it w. $x = uR$, $y = (1-u)R$ $\boxed{u=.275}$

R = 100 to 9500; G was −484.056 ← min
R = 320 −483.762 max

R = 10 gave G = −489.9

G = (47.5) gave G = −484.8 (↓ by 1 from peak at R = 320.

~~.5~~ $\boxed{=.265}$ very broad min of G = −483.3907 at R = 800

R ≈ 60 ; G = −484.35

δG = −1,

u = .250 max G = −484.0606 at R = 240

closer to post hoc $\bar{u}$ / $\frac{n}{m}$ .275

(u = .270) max G = $\boxed{-483.337}$ R = 11416
11916

(G ≈ −484.5 at R = 00, but the pgm is probly not correct for R > 12000 I need double precision.

2.17.99 SMRT Stm.

$U = .225$ $G = -483.3765$ at $R = 9036$.

$U = .280$ $G = -483.44$ at $R = 13576$ but R this high is not accurate

Looks like $U = .270$ is best(w. $R = 1141$, which gives lots of wt to $\bar U$.

④ → The broadness of the max in R suggests something is amiss

$U = .2$ gives max of -488.79 at $R = 28.28$ — so the peak in U is somewhat sharp.

Check apriori again.

The peak is so broad, that its beyond the 5 or 6 digit precision I'm using in Basic.

.10 Well! Stm 41 had a "+2" term in the G (see 40.10, and 38.38

This is inappropriate in present form. (I removed it.)

.13 Now, $U = .270$ gives no finite peak, but looks like peak is $R \approx \infty$ maybe ($G = -468.7168$ at 1900 (due precision error.

.26 " gives max at $R = 565$, $G = -468.7065$

.25 " Max at $R = 238$, $G = -469.0483$

.17 .265 " " $R = 672$, $G = -468.6654$

.18 $N = 300$; $U = .265$; peak at $R = 47.5$, $G = -3087.97$ | $R = 24, 80$ for $\Delta G = 1$ so sharper peak.

$U = .25$ " " $R = 47.5$ $G = -3087.867$

$U = .26$ " " $R = 47.568$ $G = -3087.846$ $R = 24, 80$ for $\Delta G = 1$

.20 Max → $\boxed{U = .255}$ Wierd $R = 47.566$ $G = -3087.81$ $R = 24, 80$ for $\Delta G = 1$

.21 $N = 600$ $U = .255$ max at $R = 40$ (24, 67)

.22 $N = 1200$ $U = .255$ " " $R = 40$ (24, 67)

→ See 94.14 for More on larger N

That peak R should be the same, is not unreasonable; The relative wts. of $\bar u$ & u_i are being mult by 2, hvr. See 75.20-23 for More on this.

The results of .10-.17 are kind of wierd! — but less so than before! The peak in G is very mild. The apriori 96.23 is quite simple in theory. — Likely to be correct. If we had a convergent apriori for R, we could get "expected values" for the U_i estimates.

What we want is $E\left(\frac{R\bar{u} + u_i N}{R + N}\right) \doteq E\left(\frac{R\cdot\bar{u} + u_i N}{R + N}\right)$

.29 Maybe $\int_0^\infty \frac{R\cdot\bar{u} + u_i N}{R+N}\cdot p(R)\,dR$: where $p(R)$ is the product of the apriori of R and the aposteriori R as given by $\exp(G)$ [viz Bayes]

Because of the flatness of $\exp(G)$ as a function of R, the apriori we use will probably have to actually converge for large R. Using $2^\wedge(\log_2^* R)$ may not be inappropriate

$R^{-1-\epsilon}$ may be O.k. for large R, but it's unclear as to what to do near $R=0$.

2.17.99 SMFT Storm.

Notes

.01 1) Try to Get expected values of U_i's: One way: for each x, y selected, we have this d.f. for u_i: (46.02: $f(u_i)u_i^{x+A_i}(1-u_i)^{y+B_i}$)

The expected value of u_i is $\int_0^1 u_i f(u_i)\,du_i \Big/ \int_0^1 f(u_i)\,du_i$

T. second term factor is $\frac{(x+A_i)!\,(y+B_i)!}{(x+y+N+1)!}$; t. first is $\frac{(x+A_i+1)!\,(y+B_i)!}{(x+y+N+2)!}$

So ratio is $\boxed{\frac{x+A_i+1}{x+y+N+2}}$

If we let $x=\bar{u}R$, $y=(1-\bar{u})R$ then $Eu_i = \frac{R\bar{u}+A_i+1}{R+N+1}$. To get final Exp. Value, we then use 47.29.

.12 2) I could improve accuracy of t. formulae by putting in the $\boxed{\frac{1}{12x}}$ term in to t. X! approxn.

$X! \approx \sqrt{2\pi}\, x^x \cdot \sqrt{x} \cdot e^{-x+\frac{1}{12x}}$ — so $\frac{1}{12x}$ is an additive correction to ln X!

Exponents cancel in $\frac{x!\,y!}{(x+y)!}$

3) Try this "fitting" pgm using t. data for t. entire season. T. present pgm could be readily modified to deal w. players who have different CSZ's.

4) The value of u_i used was simply $\frac{\text{no. of hits}}{45}$; "The straight rule"

a better estimate would be $\frac{\text{no. of hits}+1}{45+2}$ $\boxed{\text{Laplace's rule}}$

Since t. batting averages are ≈ .25 this amounts to only 11 hits, so the correction

.21 is usually large $\textcircled{.25} \to \frac{.25\cdot 45+1}{46} = .2663$; so .25 → .2663 is a fair increase. ($\Delta = .0163$)

The season average was .27, so this would reduce error appreciably.

On t. other hand, the observed mean of .265 itself is not very accurate: its σ is $\sqrt{\frac{.25\times .75}{810}} = .0152$

Which is about same size as "Laplace rule" correction. $\textcircled{.21}$

---

5) On $\boxed{\text{Apriori for } R}$. If $f(R)$ is ∋ $\int_0^\infty f(R) = 1$, then

.25 $\int_0^1 \alpha F(\alpha R)\,dR = 1$; this $\alpha F(\alpha R)$ expands $F(R)$ in the F direction & contracts it in the R direction

Any normalization [other than $\frac{A}{R}$ which is "invariant"] can be distorted this way, — [well, almost all say $\alpha F(\alpha R) = F(R)$

one soln. is $F = \frac{1}{R}$; are there any other solns? $F(\alpha R) = \frac{1}{\alpha} F(R)$

From $F(1) = A_0$ $F(\alpha) = \frac{A_0}{\alpha}$. ← so this is t. most general soln: $F(R) = \frac{A_0}{R}$.]

Could I use $\frac{A}{R}$ for an a priori? If t. fcn to be normed was 0 for R=0 & t. function approached 0 at R=∞, t. product could "converge" — but what value of "A" to use?

T. only reason to use $\frac{A}{R}$ is that it can't be distorted by .25 — it has no inherent "size" in t. R direction — but the "A" factor gives it "size" in t. F(R) direction.

I was considering using $2^{-\lceil \log_2^* R \rceil}$ but I still have diffy .25.

6) A way to use $P(R)$ (if $\int_0^\infty P(R) \to \infty$] for apriori.

do $\lim_{R\to\infty} \frac{\int_0^R F(R)P(R)\,dR}{\int_0^R P(R)\,dR}$ ← If $f(R) = \frac{1}{\sqrt{2\pi+R^2}}$ and $\lim_{R\to\infty} F(R) = \alpha$, then this $\to \alpha$

2.18.99 SMFT Stam

.2033 × .297

7) even for N=99, it looks like R has a broad d.f. (like a factor of 2 or 3 "width".

8) What I want is a d.f. for the "next time at bat" This is a 1 param d.f. (μ_i):
I prob'ly want Expected Value of μ_i (w.r.t. the D.f. on $\bar{\mu}$ & R (!)).

9) I could use "rest of season" data to predict first 4 "at bats": the pgm Stam 46 is easily modified to deal w. difrnt ss's for player

10) On the d.f. of R: If the density of R at R=0 was ∅, I could use a uniform

107 D.f. on $\frac{1}{R}$. Hvr: $\frac{x+y+1!}{x!\,y!}\;\frac{x+A_i!\;y+B_i!}{x+y+N+1!}$: at R=0; $\frac{1!}{0!\cdot 0!}=1$; $\left(\frac{A_i!\,B_i!}{N+1!}\right)\neq 0$.

.20 At R=∞, I can compute this exactly; $\frac{(x+1)(\cdots)(x+A_i)}{(R+2)\cdots(R+A_i+1)}\;\frac{(y+1)(\cdots)(y+B_i)}{(R+A_i+2)(\cdots)(R+N+1)}$

For R=∞ this is $\boxed{a^{A_i}\cdot b^{B_i}}$ $\boxed{x\equiv aR,\ y\equiv bR,\ a+b=1 \quad a\equiv\bar{u}. \quad A_i\equiv\mu_i N,\ B_i\equiv N-A_i\ldots}$

This is wierd! $\mu_i^{x}(1-\mu_i)^{y}$ is a δ funct at $\bar{\mu}$ when R=∞, So it should have value ≈ zero in the $\mu_i^{A_i}(1-\mu_i)^{B_i}\equiv f_1(\mu_i)$ region. Hvr, apparently what happens; f_1 has enuf values in the δ funct region for the expected value of its product to be >> 0 (!)

In fact, it has the value ($a\equiv\bar{u}$) $\boxed{a^{A_i}b^{B_i}}$. If N (& A_i, B_i) are large, $a^A b^B$ will be quite small, but still ≠ 0.

.20 For R=∞; the d.f. $a^{\Sigma A_i}\,b^{\Sigma B_i}$. $=\left(a^{\bar{u}}\,b^{1-\bar{u}}\right)^{N\cdot k}$ arith mean. — for $\bar{u}$=.265, R=310. T. max of the d.f. is ≈ $\bar{u}$ w. "width" $\sqrt{\frac{2\pi\,\bar{u}(1-\bar{u})}{N}}$

At any rate $\prod_i a^{A_i} b^{B_i}$ is max when $a=\bar{A}_i$, $b=\bar{B}_i$) arith mean.

(If the players don't have = ss's, we wt. each acc. to his no. of times at bat. — so simple arith. mean)

.25 .07: (at R=0) $\frac{A_i!\,B_i!}{N+1!}$ = ~~$\left(\frac{A_i}{N}\right)^{A_i}\left(\frac{B_i}{N}\right)^{B_i}\cdot\sqrt{2\pi\frac{A_i}{N}\frac{B_i}{N}}\cdot\frac{1}{N+1}$~~ ← woops! mult by 18 (Σk) = −468.35987

.26 $= \frac{A^A\cdot B^B\cdot N^{-N}}{N+1}\cdot\sqrt{\frac{2\pi A\cdot B}{N}} = \left(\frac{A}{N}\right)^A\left(\frac{B}{N}\right)^B\cdot\frac{1}{N+1}\sqrt{\frac{2\pi A\cdot B}{N}}$

$\approx \left(\left(\frac{A}{N}\right)^{\frac{A}{N}}\left(\frac{B}{N}\right)^{\frac{B}{N}}\right)^N\cdot\sqrt{2\pi\frac{A}{N}\left(\frac{B}{N+1}\right)}\cdot\frac{1}{N+1}$

" $\left(\frac{A}{N}\right)^A\cdot\left(\frac{B}{N}\right)^B$

These 2 factors are "almost" equal when $a=\bar{u}$

So a^G is much smaller at ∅ than at ∞: By factor of ½ $\left(\frac{2\pi}{N+1}\right)^{\frac{k}{2}}\cdot\left(\frac{A_i}{N}\frac{B_i}{N}\right)$ ← $\frac{k}{2}$ power geometric mean $\sqrt{\left[(\prod\mu_i)(\prod(1-\mu_i))\right]}$

it's 1. other way around at R=∅: see 70.18-40

prob'ly wrong! see 70.18-.40

wrong! see

exponent should be 45.18 = N.k = 810

.735
.265
$(.5608947)^{45}$ = 5.0073 E-12
ln = −26.02

.38 So G as funct of R must be

I don't know if there is a max or peak here.

− 469.6654

Well! on 47.17 I got for $\bar{\mu}$=.265: R=673; G = −468.6654, which is more than R=∞ value of −468.3598

14 $\left(\left(.265^{.265}\,.735^{.735}\right)\right)^{45\times18}$ = ~~[illegible]~~ −468.35987 checks w. .26R

$\boxed{.20-.26R}$

−.35

Mult this by $\frac{300}{45}$ → −3122.399) which is much worse than the peak of −3087.97 for N=300 on 47.18

Woops! R=300 had a peak at U=.255 of −3087.81 — looks wrong, but it shows that peak is >> the value at R=∞

In the case of $\bar{\mu}$=.265, the difference betw the 2 values at ∞ could be due to error in approx of X! (48.12) (not roundoff). Using a more accurate value of $\bar{u}$ ~~[illegible]~~ could be relevant.

I want $\left(.265^{\bar{u}_{true}}\cdot .735^{1-\bar{u}_{true}}\right)^{kN}$

2 ·19·99

50

$\left(\log \sum \left(.265^{U_T} + .735^{(1-U_T)}\right)\right) * 810 =$ −468.6812 for G_{oo} R=∞, which is slightly worse than the −468.6659 that I got for U=.265, R≈672 on 47.17

Using U = .2653889 I get peak of ~ −468.671 at ~R=672

6 but ≥ 150

So precision is imperfect. → (−468.6702 at R= 800 and 1131 ! −468.6718 at R=931 ≡)

Using P. Basic 85! peak for G = $\boxed{468.66\overset{6660367}{}\ \text{at}}$ R = 800

(Set U to .2653889)

For large R ; = 45254.83 G= −468.$\overset{680845}{}$; slowly ↓

R = 4.54 G= −468.6809188

The peak at R=800 is very small w.r.t. 954

$\boxed{\text{Woops! seems to be a small peak of} - 468.67899 \text{ at } R = 11892}$

25

81

Woops! R=10⁻⁴ G= −416.3813

10⁻³ G = −445.595

I'm getting crazy stuff (as ε 2π)! R = 10⁻⁴ to R = .81 ; G ↓ from −411 to −482

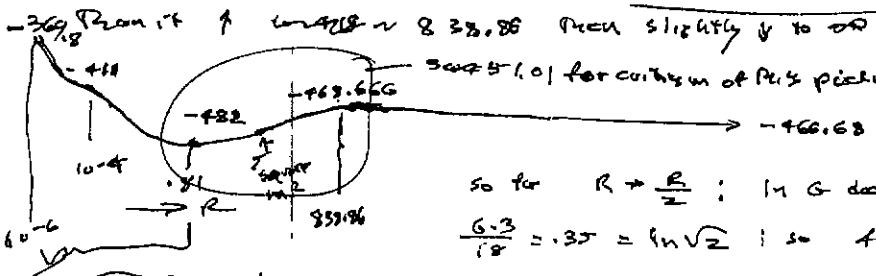
−369.18 Then it ↑ to −468 ~ 838.86 then slightly ↓ to ∞.

13

800

67

.20



So for R → $\frac{R}{2}$; ln G does − 6.3.

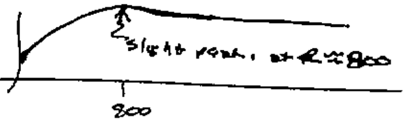
$\frac{6.3}{18}$ = .35 = ln√2 ! so for small R ; G = $\frac{1}{R^{.35}}$ ≥ R^{-9}

This section is invalid

Strange! ~~Maybe~~ AH! Could be due to error in approx of K!

Hmm. X! approxns don't work at all for R<1 ; So this pgm really can't deal w. R<1 at all.

From 49.20 ff. the plot prob'ly looks like



872

8 299 better

468.66660089

R = 792.7 ± 6.1 is a peak for G = −468.666659270322

R=4.54 presumably −468.6808918... very slightly worse.

STEIN 50.Bas!

Print G − log G → : has broad peak close to R=80 !

Print G − 2 log G → broad peak at R ≈ 43.8 — sharper than ↑ for R ÷ 2 or ×2; ΔR ~ .5.

R40 — down .36

R160 — down .26

∗ 33

The idea is that in original D.F. (ST46.B) T.P.D. became constant > 0 for large G! This is reasonable from the model. If R is big, we have δ fund at $\bar{U}$, and the value of e^G is $\prod_i \bar{U}^{A_i}(1-\bar{U})^{B_i}$ —

which is $\sum \left(\bar{U}^{\bar{A}}(1-\bar{U})^{\bar{B}}\right)^{K}$

ST (8, 2, 10) Checks X! approx by integrating

ST11 : no "Σ" correction

ST5 : compute Geom means of u_i', $(1-u_i')$

ST6 ~ to 5

ST7 ! calculates & maybe of U's avoid'n ; also get Grand mean (Aps!)

$\boxed{\bar{U} \sim .2653889}$

ST3 ; computes Geom means of u_i', $(1-u_i')$ & is arith mean.

$\frac{6.3}{18}$ = .35 = ln √2.

50.20 & 50.33 about cusp: In 50.20 we have an ↑ of ~ 13.33 in G. W. a 1000 fold change of R: $\ln 1000 = 6.9$; $\frac{13.33}{6.9} \sim 2$ so e^G ↑ by R^2 from $R = .8/8$ R = 8386. If we mult ~~[illegible]~~ subtract $2 \ln R$ from G we should have no peak) — sort flatness!

OH: T. really rapid rise of G was from $R = 1.6 \rightarrow 481.8$ to $R = 25.9 \rightarrow 471.5$ & G = 10 $R = \times 16$; $\ln 16 = 2.77$; $\frac{10}{2.77} = 3.6$: So e^G was like $R^{3.6}$

.10 So $\frac{e^G}{R^2} \rightarrow R^{1.6}$. So it would peak.

Anyway, t. moral is: That t. position of t. peak depends much on our approx. This assumes also that $R > 1$, (not unreasonable)

Also, for $\sigma^2 \propto \frac{1}{R}$, even tho there is a peak, if one takes expected value of μ_i, $R = \infty$ gets all of t. wt (I think). (using t. trick "renormalization" method)

SN: T. behavior of t. approxn used in t. Pgms (ST 46, 50 & previous) don't work at all for $R < 1$. Hvr, for $R = 0$, t. analysis of $\boxed{49.26}$ holds, & I think I can get a good approx for $0 < R < 1$.

49.20 deals w. $R = \infty$.

Anyway: Since I'm fairly sure about behavior at $R = 0$ & ∞; To criticise t.

System: Compare it w. a Gaussian System: which seems to work reasonably", Ways! : I'm not so sure it did work!

(Also ~~[illegible]~~ Consider Simpler Laplace analysis of 42.01–44.30 — can I compare w. present treatment [redacted]

Essentially it uses 2 coding methods only: $R = 0$ & $R = \infty$ & ~~[illegible]~~ wts t. 2 methods. ($\boxed{\frac{G(\infty)}{G(0)}}$):

($R = \infty$ wins so $\bar{\mu}$ gets $\simeq$ all wt, for $N < 210$ ~~[illegible]~~ $R = 0$ gets all wt for $N > 220$. There is a small transition region of comparable wts. $N = 210 \pm 10$)

Mainly, I'm interested in t. correspc. betw. $\frac{1}{R}$ & σ^2: Why does G Gauss ~~[illegible]~~ o.k., but Laplace not o.k.: Is there a problem w. apprxd [Laplace / Gaussn.] of σ^2 or σ in Gauss system? — I imagine it's 0 to ∞ (!).

.3) Another Deriv method is to use $\mu_i = c\bar{\mu} + (1-c)\mu_i'$: T. tricky pc is use of c & $\bar{\mu}$. Perhaps use uniform apriori for $c, \bar{\mu}$ over t. unit square; integrate over that square. This gives us t. pc for each data set $[\mu_i^j] = P([\mu_i^j])$ to make a predn, we compute $P([\mu_i^j]^{\alpha_j})$ } This sounds quite rigorous!

$[\mu_i^{\alpha_j}]$ is same data set $[\mu_i^j]$; but $\mu_j^j \rightarrow \frac{N\mu_j^j + 1}{N+1}$, $N_i \rightarrow N_i + 1$

2·20·99 SMFT STEIN

A Biblio review of STEIN:

1½.01 Dern of STEIN effect from Sci Amer May 77 pp119-127
I was uncertain about defns of various variables in the eqns, hvr. Later, I figured out
the defns. [Ref? 1½.01 →05; 8.01 — By 8.12 I hadn't ~~read~~/referenced papers into SA ART.
10.01 starts to analyze one of the refs (Cover, Ralph)

7.20 points out diff from my early idea of what σ² was.

By 10.20–.23 I got a reasonable defn. of σ² as used in the S.A. paper

So by 9.25 & 10.20 I think I understood what the SA results were.

3.01 is a biblio. review of recent work on "SOY" in TM & SM.
2.20 is my early Gaussian version of the ST. problem. 4.20 ff develops it.

.01 : 15.40 ; on 15.01–.40 ; Consider the expected ms error in predicting r_{ij} for both methods.

.02 Using μ', the ~~error~~ ms mean error would be $\sum_i \sum_{j=1}^{l_i} (\mu' - r_{ij})^2 \Big/ \sum_{i=1}^{k} l_i$

.03 In simple case $l_i \equiv l$ = const; $\frac{1}{l}\sum_{j=1}^{l}(\mu'-r_{ij})^2 = \frac{1}{l}\sum(\mu'^2 - 2\mu' r_{ij} + r_{ij}^2) = \mu'^2 - 2\mu'\mu_i' + \mu_i'^2 + \sigma_i'^2$

$= (\mu'-\mu_i')^2 + \sigma_i'^2$.   from

.04 → omit ~~$\sum_j(-2\mu' r_{ij} + r_{ij}^2) = -2\mu' l_i \mu_i' + \sum_j r_{ij}^2$; $\sigma_i'^2 = \frac{1}{l}\sum_j(\mu_i' - r_{ij})^2$; $\mu_i'^2 - \frac{2\mu_i'}{l_i}\sum_j r_{ij} + \frac{1}{l}\sum_j r_{ij}^2$~~ $= \frac{1}{l}\sum_j r_{ij}^2 - \mu_i'^2$

.07 $\boxed{\sigma_i'^2 = \frac{1}{l}\sum_j r_{ij}^2 - \mu_i'^2 \; : \; \frac{1}{l}\sum r_{ij}^2 = \sigma_i'^2 + \mu_i'^2}$

So ms error if $l_i = l$ = constant is from .02, .03, .04 & .07 :

$\frac{1}{k}\sum_{i=1}^{k}\left((\mu'-\mu_i')^2 + \sigma_i'^2\right)$ $\frac{1}{k}\sum_{i=1}^{k}(\mu'-\mu_i')^2 = \frac{1}{k}\sum\mu_i'^2 - \mu'^2$

$\equiv s'^2$

$s'^2 + \frac{1}{k}\sum\sigma_i'^2 = s'^2 + \overline{\sigma_i'^2}$; This is the ms error in using μ' to predict r_{ij}.

$\frac{1}{k}\sum\sigma_i'^2$ " " " " " μ_i' " " .

} leaves out ~~[illegible]~~. Also the factors $\frac{kl+1}{kl-1}$ & $\frac{l+1}{l-1}$ resp.

.19 So if l is of reasonable size, $s'^2 + \overline{\sigma_i'^2}$ is always appreciably $> \overline{\sigma_i'^2}$

↓ Have omitted the effect of non-independence : see 15.25–.29

Ent. Case of the Baseball scores: $\sigma_i'^2 = \mu_i'(1-\mu_i')$ (l ?) (o.k.) !

One way of getting it is: $\sum$ error $= (\mu-1)^2\cdot\mu l + \mu^2(1-\mu)l$ $= (\mu+1-\mu)l\cdot\mu(1-\mu) = \mu(1-\mu)\cdot l$

This is close to, but not exactly what got in 37.01–.02 : 36.44 $\boxed{\frac{x+1}{x+y+2}\cdot\frac{y+1}{x+y+2}\cdot\frac{1}{x+y+3}}$

So ms error is $\mu(1-\mu)$.

$s'^2 = \left(\frac{1}{k}\sum\mu_i'^2\right) - \mu'^2$ $(\mu' \equiv \frac{1}{k}\sum\mu_i')$

$\sigma_i'^2 = \mu_i'(1-\mu_i')$; $\overline{\sigma_i'^2} = \mu' - \frac{1}{k}\sum\mu_i'^2$, $= \overline{\mu_i'^2}$

$s'^2 + \overline{\sigma_i'^2} = \overline{\mu_i'^2} - \mu'^2 + \mu' - \overline{\mu_i'^2} = \mu' - \mu'^2$

$s'^2 + \overline{\sigma_i'^2} = \mu' - \mu'^2$

$\overline{\sigma_i'^2} = \mu' - \overline{\mu_i'^2}$

Anyway, .19 means that the μ_i''s always get $\geq \frac{1}{2}$ of the wt.! — which seems unreasonable —

also S.A. results got wt of ~~only~~ .791 for $\bar{\mu}$ & only .209 for μ_i'.

(NB) Actually, the ratio is Not just 1 to $1+\alpha$ $(0<\alpha<1)$

It's 1 to $(1+\alpha)^k$! So the μ_i' term would win completely almost always !

This is a very poor coding method.

43.01

39.20) in optimum
21.13) k for 2 batsmen