.01 Hora ran lam Moldc: D In ST50.bas, using a prior of $\frac{1}{R}$ Save pool 24 ~80! Using $\log^*_2$ and integrating to get μ_i would ~~[illegible]~~ probly push Expected value of μ_i out further.

~~Bate~~ Integrating w.r.t my version of $\log^*_2$ may be easier because it has constant values for long ~~[illegible]~~ intervals of arg. We don't have to integrate to ∞: only out to where ~~[illegible]~~ becomes constant (i.e. "G_0") we can then just mult $\int_{G_0}^{\infty} 2^{\wedge}\log^*_2 R\, dR$ by that constant:

.07 This is easily computed: its just $1-\int_0^{G_0} 2^{\wedge}\log^*_2 R\, dK$.

---

2) For x, y d.f. perhaps have apriy width of d.f. be uniform betw. $0 \,\&\, 1$

$\sqrt{\frac{2\pi\ \mu(1-\mu)}{R}}$: is uniform betw $0 \,\&\, 1$, for constant μ : $P(R)$ is $\propto R^{-\frac{1}{2}}$. There may be a $R \to R+1$ or some ~ thing; see $36.28 - \overset{x}{} 37.01$

$\therefore P(R) \propto \frac{1}{R^{\frac{3}{2}}}$.

$\frac{d}{d\sigma} = -\frac{1}{2} R^{-\frac{3}{2}}$

of course we could assume $\sigma^2 \propto \frac{1}{R}$ is uniform betw $0 \,\&\, 1$ ~~[illegible]~~ as well so $P(R) \propto \frac{1}{R^2}$

~~[illegible]~~ tried $\frac{1}{R}\cdot\left(\frac{1}{\ln R}\right)^2$ converges. ~~[illegible]~~ (This is like $2^{\wedge}\log^*_2 R$ .01–.07)

or try $\frac{1}{R}\frac{1}{\ln R}\left(\frac{1}{\ln\log_2 R}\right)^2$ which converges slower.

.19 $\int \frac{1}{R}\left(\frac{1}{\ln R}\right)^2$! P203, G&R! 2.721 (2) $\int \frac{\ln^m x\, dx}{x} = \frac{\ln^{m+1} x}{m+1}$

if $m=-2$ $\to \frac{\ln^{-1} x}{-1} = -\frac{1}{\ln x}$!) $d\frac{1}{\ln x} = -\frac{1}{(\ln x)^2}\cdot\frac{1}{x}$ fine!

~~[illegible]~~

So it converges, but $-\frac{1}{\ln x}$ doesn't work for $\int_{20}^{\infty}\frac{1}{x\ln^2 x} = \left.\frac{1}{\ln x}\right|_{20}^{\infty} = 0 + \frac{1}{\ln 20}$ ok.

So $\int_x^{\infty}\frac{1}{x\ln^2 x}dx = \frac{1}{\ln x}$, which diverges at $x=1$.

$\frac{d}{dx}(\ln\ln x)^{-1} = \frac{-1}{(\ln\ln x)^2}\cdot d\ln\ln x$ | $d\ln\ln x = \frac{1}{\ln x}\cdot\frac{1}{x}$

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

$\frac{d}{dx}(\ln(\ln\ln x))^{-1} = -\left(\frac{1}{\ln\ln\ln x}\right)^2\frac{d}{dx}\ln\ln\ln x$ | $\frac{d\ln\ln\ln x}{dx} = \frac{1}{\ln\ln x}\cdot d\ln\ln x$

So probly $\frac{d}{dx}(\ln^{(n)}x)^{-1} = -\frac{1}{x}\cdot\frac{1}{\ln x}\cdot\frac{1}{\ln\ln x}\cdots\left(\frac{1}{\ln^{(n)} x}\right)^2$

Mora relevant:
$\frac{d\ln^{(m)}x}{dx} = \left(\ln^{(m-1)}x\cdot\ln^{(m-2)}x\cdots\ln x\cdot x\right)^{-1}$

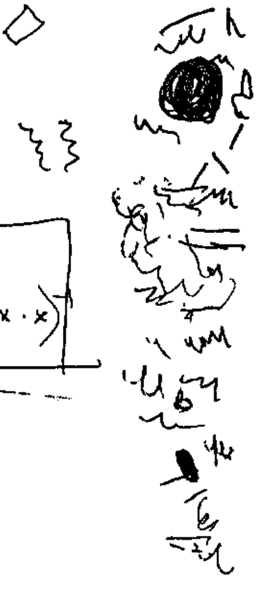
.32 But $\ln^* x \neq \left(\lim_{n\to\infty}\right)\ln^n x$. (!) ☹.

Say $G' = e^c$. I want $\int_1^{\infty} G'\, P(R)\, dr$: Say G' is constant for $R > G_0$.

I have to get G' more directly! mult by Normaliz. consts: $\prod\left(\frac{A_i!\ B_i!}{N+1!}\right)^{-1}$

No! I want Expected value of N_i.

(SN) That stuff about $\frac{1}{x}$ being the only apriori that didn't "Select" a scale in x direction: False. x^n can be expanded by α in x direction $\to \frac{x^n}{\alpha^n}$: this contracts it by α^n in $f(x)$ direction. The stuff about $\int_0^{\infty}$ is also wrong. We will usually integrate from some pt α $\int_{\alpha}^{\infty} f(x)\,dx = 1$.



22399 SMFT STEIN: $\boxed{\text{Log}_\alpha^* x}$ 58.

ON The Apripd of the integers: (a/o the apripd of the reals between 1 & ∞.)

The aprpd of the positive integers is wrt. a reference UMC, M_0. $P_M(N) = \sum 2^{-|s_i(N)|}$ $s_i(N)$ is the ith pgm that will cause M to print N & stop. — or $P_M(N) = e^{-s'(N)}$ $s'(N)$ being the shortest such pgm.

$P_1(N) \doteq 2^\wedge(-\log_2^*(N))$ is meant to be an approx to $P_M(N)$. $P_1(N)$ has the property ~~of converging~~ that $\int_0^x P_1(x)dx$ converges very slowly, but it does converge.

$\int_0^x P_M(x)dx$ converges more slowly than any recursive function. —

($P_M(x)$ is itself partial recursive). ← NO! — its incomputable — which is not the same as p.r.

$P_1^\alpha(N) \equiv \alpha^\wedge(-\log_\alpha^*(N))$: As α goes from 1 to e, $\int_1^x P_1^\alpha(x)dx$ converges more and more slowly.

.13 for $P_1^e(x)$, $\int_0^x P_1^e(x)dx$ does not converge.

To show ~~this~~ the divergence of .13:

$\ln^* x$ = ~~[illegible]~~ $\ln x + \ln\ln x + \ln\ln\ln x \cdots$; We only take positive terms.

so from $x = 1$ to e: $e^{\ln^* x}$ = x

e to e^e " = $\ln x \cdot x$

e^e to e^{e^e} " = $\ln\ln x \cdot \ln x \cdot x$

etc.

Now $\frac{d}{dx}\ln^{(m)}x = \frac{1}{\ln^{(m-1)}x \cdot \ln^{m-1}x \cdots \ln x \cdot x}$

look at 54.19-.32

So $e^{\ln^* x}$ is ~~not~~ integrable in close form from 1 to e | from e to e^e | from e^e to e^{e^e}, etc.

for each of these intervals, the integral is: $e - 0. = e$

I.e. for each interval, the $\ln x$ goes from some value to that value plus e

.213 So $\int_1^\infty e^{\ln^* x}dx = 1+1+1+1+ \cdots = \infty$.

---

for $1<\alpha<e$, the above arg. doesn't hold, & the $\int$ converges slowly.

Consider $f(x) = x^{-1-\epsilon}$ $\int_1^x f(x) = \frac{-x^{-\epsilon}}{\epsilon}$ which converges slowly for small $\epsilon>0$.

But $\frac{1}{x\ \ln x \cdot \ln x}$ converges more slowly than $\frac{1}{x}\cdot\frac{1}{x^\epsilon}$ for any $\epsilon>0$.

$\int_1^x \log_\alpha^* x\ dx$ converges very slowly for $\alpha = e-\epsilon$. (small $\epsilon>0$).

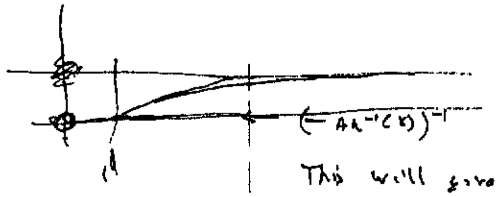
The $\log_\alpha^* x$ is not analytic (it is piecewise analytic), it is Primitive Recursive.

→ A Non-p.r. function that can converge more slowly: Consider Ackermann fnct. $Ak(x)$

$Ak^{-1}(x)$ is its inverse: it approaches ∞ as $x \to \infty$, but very slowly.

$[Ak^{-1}(x)]^{-1} \to 0$ as $x \to \infty$, but very slowly.

If we take $\frac{d}{dx}\cdot[-Ak^{-1}(x)]^{-1}$ and integrate it, it will approach some constant from below, very slowly.



unfortly $[Ak^{-1}(x)]^{-1}$ is a discontinuous function so its derivative is δ funct between long stretches of zeros

To circumvent this, we can modify $Ak(x)$ by making it piecewise linear, or by using a more complex "spline".

This will give us a function of N that converges, but very slowly —

2.23.99 SAPT STRM. $\boxed{\log^*_\alpha x}$

More slowly than any Prim. Rec. function. Th. way to prove this: If there were any prim rec. function that converged more slowly than 1. once I've shown, then we could construct a prim rec. function that ↑ more rapidly than Ak(x) : — which is impossl.

[N.B. T. way that Akerman showed that Ak(x) was not prim. rec., is that Ak(x) increased faster than any prim. rec. function]. [See Kleene's Book for discn]

Just as this ~~[illegible]~~ function based on Ak(x) ↑ faster than any prim. rec. funct — It is itself a total rec. funct. ~~[illegible]~~ $P_M(x)$ can converge even slower because it is not recursive ~~a partial rec. funct~~.

.07 I don't know if $P_M(x)$ is the most slowly converging funct possl. (in some sense). → See .26 (small "proof")

There are functions that ↑ slower than any recursive function

---

.10 Definition of "slow convergence": (lim x→∞) Say $\lim_{x\to\infty} f(x) = A$.

~~Then if f(x) converges slowly, $\frac{1}{A - f(x)}$ will → ∞, but very slowly.~~

~~By look~~ We can look at how fast $A - F(x) \to 0$ as $x \to \infty$.

We can compare 2 functions G(x) & F(x) on "how fast" t. corresponding functs → 0.

F beats G if $\exists\ x_0 \ni$ for all $x > x_0$ $A - F(x) > B - G(x)$ ← $B \equiv \lim_{x\to\infty} G(x)$.

$\left(\frac{1}{\log_\alpha^{(m)} x}\right)^{1+\epsilon}$ ~~approaches zero~~ converges more rapidly (i.e. $\int_0^\infty (\ )\,dx$) Than $\frac{1}{\log_\alpha^{(m)} x}$ $\frac{1}{(\log_\alpha^{(m+1)})^2}$ (i.e. $\left(\frac{1}{\log_\alpha^{(m)} y}\right)^\epsilon > \frac{1}{(\log_\alpha^{(m+1)})^2}$ for all values of $0<\epsilon<1$. → 57.19

.22

So 55.01 - 56.22 gives a perspective on how good Rosenau's $1^\wedge \log^*_2 x$ is as an approxn to $P_M(x)$ — with respect to behavior for large x.

$P_M(x)$ is within constant factor betw. say 2 UMCs, P_{M_1} & P_{M_2}.

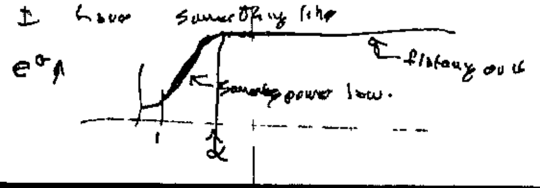
.26 For me (.07) is an open Q — as to whether ∃ a pr funct that converges more slowly than $P_M(x)$: Suppose there were such a function: T. p.r. functs are effectively enumerable. M could simulate that function, & This could probly be used to show that no p.r. funct (or any rec funct) could → its limit significantly more slowly than $P_M(x)$.

A second Q about Aprioris is The "horizontal scaling factor". I think this is much more impt than convergence rate. For all UMC's This "scaling factor" is machine dependent. For any $P(x) \ni \int_0^\infty P(x) = 1$, then $\forall \alpha$ $\int_1^\infty P(\alpha x)\,dx = 1$ also. The function $P(\alpha x)$ is stretched out in the x direction by a factor, α.

Consider $P(x) = \frac{k}{x^2+A^2}$: The value of A gives us a horizontal "scale factor"

In My present approach to the Stein problem, before I apply the apriori of R,

I have something like



The important thing is not how our apriori P(R) behaves after α (by R), but how it behaves for 1<R<α.

α is, of course, a priori unknown, so

2.23.97 SMPT 5 am $\boxed{\log_\alpha^* x}$

It would seem that ~~we are pretty likely~~ choice of apriori (≡ behaviour [illegible] on $[1, \alpha]$) will influence our result tremendously!

Well, as we ↑ amt. of data (eg. N) t. apriori should ↓ in importance, ↑ in N w.o. modific. of t. $[M_i']$ is, hvr, a picture that has to be interpreted differently from ↑ of N w. natural variation of t. $[M_i']$. [In this second case, usually R would ↓ as N ↑, but would settle down to some finite value as $N \to \infty$ — in which case t. apriori would not be important.

In present case, t. Vars ~~[illegible]~~ (σ^2, R) can be any thing betw 0 & 1. We could look at t. σ^2 of t. $\boxed{\text{previous season(s).}}$ ← use them dif. to der. a sequence of previous yrs' M_i's. A uniform d.f. in σ or σ^2 should be not bad]

Uniform in σ^2 means $P(R) \propto \frac{1}{R^2}$. uniform in σ means $\frac{d}{dr}\frac{1}{\sqrt{R}} = \frac{1}{R^{3/2}}$

Look at ΣM_i for $\frac{1}{R^2}$ & $\frac{1}{R^{3/2}}$. See what occurs as ~~[illegible]~~ $N \uparrow$ (w. same $[M_i']$ set)

Also, look at Gaussn ~~approx~~ Model. See if t. apriori is t. same. It should map very closely to t. Laplace Model (≡ Bernoulli Model).

$((2 \times 3)+2) \times 5 = 40$

Another approach: Actually generate M_i' data Monte Carlo using some known x, y or $(R, \bar{\mu})$: then using some N, generate ~~[illegible]~~ associated M_i'. See how well t. "true" $R, \bar{\mu}$ fits t. data, as compared to other $R, \bar{\mu}$.

.19 : 56.22 The integration of $\log_\alpha^* x$ is not easy at all! Only easy if $\alpha = e$! — So how did

.20 Riss do $\alpha = 2$? He tells how in his paper "A universal Prior for integers ..." 1983 (I have) - pp429-430

p.424 of that paper suggests that $\alpha \wedge \lg_\alpha^* x$ does not converge for $\alpha \geq 2$ (or that convergence is uncertain) — seems unlikely that $\lg_2^*$ would be t. "edge"!

'I don't know if ~~[illegible]~~ Riss. knows this for sure.

.23 I did have an idea that t. summation for α could be mapp'd onto t. summation for e. — That ~~for~~ each interval of x in which $\ln^* x$ ~~[illegible]~~ went from ~~[illegible]~~ an integer n, to n+1, could be mapp'd into a similar interval for $\lg_\alpha^* x$. — by a simple mult factor w/o a relative constant for both.

- Start w. $(x \ln x)^{-1}$ v.s. $(x \log_\alpha x)^{-1}$ they differ by factor $\ln \alpha$.

$(x \ln x \cdot \ln\ln x)^{-1}$ " $(x \log_\alpha x \log_\alpha \log_\alpha x)^{-1} = \left(x \frac{\ln x}{\ln \alpha} \cdot \frac{\ln}{\ln \alpha}\left(\frac{\ln x}{\ln \alpha}\right)\right)^{-1}$

Unclear as to how to integrate this → $= \left(x \frac{\ln x}{\ln x} \frac{\cdot 1}{\ln \alpha} (\ln \ln x - \ln \ln \alpha)\right)^{-1}$

— But it may be possi. to show it converges for $\alpha < e$.

.32

Say $\alpha = 1+\epsilon$. Will it converge? | as $\epsilon \to 0$ $\log_{1+\epsilon} x \to \infty$ $((1+\epsilon)^y = x$

So it seems likely ~~the~~ that | $\ln_{1+\epsilon} x \approx \frac{\ln x}{\epsilon}$ $e^{\epsilon y} = x$; $\epsilon y = \ln x$

as $\alpha \to 1$, $\log_{1+\epsilon}^* x$ will converge; | Note $\ln(1+\epsilon) \approx \epsilon$ so $\frac{\ln x}{\ln(1+\epsilon)} \approx \frac{\ln x}{\epsilon}$ → $y = \frac{\ln x}{\epsilon}$

for $\epsilon = 0$, $\log_1 K = \infty$ so I think $\log_1^* K = \infty$ i.e. $\frac{1}{\log_1^* x} = 0$ But I'm not at all sure: I'd have to calculate look into this ~~[illegible]~~ in more detail.

→ In Riss paper of .20, he shows how to integrate $2^{\wedge}(-\log_2^* K)$ ~~from~~ this, it might be possi.

.37 (in fact is likely) that one could find a way to integrate $\alpha^{\wedge}(-\log_\alpha^* K)$. → 58.07

.38 looks very simple! → Note also see 58.01 $\boxed{\text{Another approach}}$: Consider the e intervals mapping into t. α intervals: discussed on .23 – .32 Would it be possi. to show t. integrand for α is at least $(\log \alpha)^k$ times smaller than t. corresp. "e" interval (k is t. no. of log iterations for that interval) → 58.01 spec

22599 SMFT **STEIN** $\boxed{\mathrm{Log}^*_{\alpha}\, x}$

.01 Re: 57.38 : Well, not so easy! The correspg. intervals for e & α differ considerably in "x" values. In general, $\ln^* x >$ ~~[illegible]~~ $\log^*_\alpha x$ if $\alpha < e$.

Proby $\int_1^\infty \alpha \wedge (\log^*_\alpha x)\, dx$ converges for $\frac{1}{e} < \alpha < e$: i.e. the behavior of $\log_{\frac{1}{\alpha}} x$ and $\log_\alpha x$ is pretty much the same except for sign.

for $\alpha < 1$, hvr., the defn of $\log^*_\alpha x$ has to be modified — so only negative terms are included.

So we only have to ~~[illegible]~~ show the convergence for $1 < \alpha < e$.

.07: 57.37 from Riss (57.28) : Say $\lg(x) = \log_\alpha x = \frac{1}{\ln \alpha} \ln x = \beta \ln x$ $\left(\beta = \frac{1}{\ln \alpha}\right)$

say $\lg^{(k)}(x)$ is the k'th iterated lg of x.

then $\frac{d}{dx} \lg^{(k)} x = \frac{d}{dx} \beta \ln(\lg^{(k-1)} x) = \beta \frac{1}{\lg^{(k-1)} x} \frac{d}{dx}\left(\lg^{(k-1)} x\right)$

$\therefore \quad = \beta^k \frac{1}{\lg^{(k)} x \cdot \lg^{(k-1)} x \cdots x} = \beta^k \cdot \alpha \wedge -\lg^{(k+1)} x$ ← for a certain x range.

Which probly makes it easy to integrate it exactly ; So why did Riss have to use upper & lower bounds & get an approxn ? (of known smallness, hvr).

This all has to be checked out ~~[illegible]~~ carefully!

.15

While I just want to show that $\int_1^\infty \alpha \wedge -\log^*_\alpha x$ converges for $1 < \alpha < e$, getting an exact soln. can be useful.

Also, I suspect. that what Riss was after, was a prior for the integers ; So these integrals are used by him to find upper & lower bounds on the sum. Of course, for $\alpha = 2$, there is a much easier way to do it — but I think it converges more slowly than Riss' function

On the other hand a good prior for ~~x~~ $1 < x < \infty$ would be useful! — & its what I seem to need for STEIN !

(2 loop) Ak(x) ↑ faster than any (one loop) prim rec. funct. Can we devise a 3 loop funct. that ↑ significantly faster than Ak(x) ?

2.26.99 SMPT STEIN N $\mathrm{Log}^*_{ex}\,x$ <u>Review.</u> 59

I want to stop work on this problem, write a review of results obtained, discn of what seem to be main problems, & list open problems along w. possible approaches for each.

1) 2 related problems ⓐ sprd for $1, 2, \ldots \infty$
ⓑ " " $1 < x < \infty$

2) [redacted] [redacted] A/P gives (incomputable) soln: ~~[illegible]~~ call it. <u>$P_0(n)$</u> (for integers).
Main constraints on $P_0(x)$: ⓐ its a ↓ funct of x. ⓑ $\sum_{n=1}^{\infty} P_0(n) = 1$
ⓒ we'd ~~[illegible]~~ like $S(n) = \sum_{i=1}^{n} P_0(i)$ to ~~[illegible]~~ converge ~~[illegible]~~ as slowly as possl., since $\sum_{i=1}^{n} P_0(n)$ converges more slowly than any recursive function. (See 56.10 for dctn of "<u>slow</u>" convergence)

3) For any recursive ~~[illegible]~~ $P(n)$, we can always modify $P(n)$ to converge more slowly, by removing some P from small n ~~[illegible]~~ adding it onto P for larger n.

3.5) If one uses $P(x) \propto \alpha$^ $\log^*_{ex} n$, then $P(x)$ converges for $1 < \alpha < e$: As $\alpha \uparrow$ $P(x)$ converges more <u>slowly</u> — for $\alpha = e + \epsilon$ it converges <u>very</u> slowly for small ϵ — Also $\frac{1}{n^{1+\epsilon}}$ converges slowly. For discussion of integration of α^$-(\log^*_{ex} n)$ see 57.19–.37 58.02–.15

4) Rizs proposes a ~~[illegible]~~ funct for $P(y)$ that is Prim. rec.
Functs that are <u>not</u> prim rec. (e.g. Ak function & functs based on it) can ~~[illegible]~~ increase more rapidly than any prim. rec. fun., and ~~[illegible]~~ modifns of the Ak funct can be made that converge more slowly than any prim rec. fund.

5) Whether these Ak based funct is in any sense a "best" $P_0(\ )$ is unknown. Probably <u>there is no best</u> among the recursive functs

6) Can one define (in any sense) a $P(n)$ based on <u>partial recursive</u> functs?, $P_0(n)$ is <u>incomputable</u> not partial recursive.

-24 7) ~~For a discussion of the recurs~~ The ~~[illegible]~~ proof that Ak's funct ↑ more rapidly than any Prim. rec. funct was ~~[illegible]~~ what Ak used to ~~show that~~ demonstrate it as the first ~~known non-pr~~ function <u>known</u> to be <u>non-prim. rec.</u>

- That $\log^*_{ex} x$ is prim rec. follows from Peter's demo. that all prim rec. functs are expressable by simple loops & compositions of functs.
For discn. & proofs & refs for .24 ff see 5. Kleene's "Introdn to Meta Mathematics"

8) For a discussion of ~~[illegible]~~ <u>how good Rizs' function is as an approxn to $P_0(n)$</u> — $\boxed{\text{see } 55.41 - 56.22}$

9) The foregoing is all on behavior of $P_0(n)$ for <u>large n</u>. For most applicns, (as Rizs points out) the difference betw. 2^$\log_2^* n$ and $\frac{1}{~~[illegible]~~}\left(\frac{1}{\log_2 n}\right)^2$ is not very much.

·36 10) For many the impt. Q is behavior for <u>small</u> n. Any $P(n)$ can be ~~[illegible]~~ modified by something like $\frac{1}{\alpha} P(\alpha n)$ to yield a convergent $P(n)$, with signifly diffrnt ~~[illegible]~~ wt. for low n. In fact, $P_0(n)$ (the A/P function) can do this also, by changing reference machines — (for large n, the "behavior is within a constant factor").

22699 sm ET STEIN $\log^*_\alpha x$ Review.

.01 As I see it, w.r.t. applicns for STEIN: I'm mainly worried about relative wt. for small v.s. larger n.

~~[illegible] 5896234~~

N399: I'd forgotten this, but 55.23 shows that $\int_0^\infty dx\, e^{-\ln^* x} = 1+1+1\cdots = \infty$.

Lavin Mentioned this divergence at the time, I didn't know how to prove it ——

I think I had proved it in "CLEVeland" when I was analyzing $e^{-\ln^* x}$.

$\int_0^\infty dx\, e^{-\ln^{(n)} x} \approx n$ (* — $\ln^{(n)} x$ is $\ln x \ln\ln x$ n times (or less if it becomes imaginary).

2.27.99 SMFT Stein

.01 : Random Musings on STEIN problem!

.8 .2

T. final soln that ≈ STEIN got was WH. avg of $\bar{\mu}$ & μ_i' w. wts ~ .8 & .2 resp. | 4+1

1 to 4 gives rel. σ's of 1 to 2. I think 9 to 1 might have Been Better. | 180:45

$\underbrace{.265 \quad .290}_{25} \quad \underset{110}{.400}$

.156 .239 .265 .378

83 26

$\bar{\mu} = .25$ gives $\sigma_{\bar{\mu}}^2$ of $\frac{\frac{1}{4}\cdot\frac{3}{4}}{45} = 4.1\dot{6} \times 10^{-3}$

$\sqrt{\sigma_{\bar{\mu}}^2} = .0645$. so S of $\bar{\mu}$ would be ~.0322

$\frac{1}{15\times16}$

. R G=0 1

Print(G − G1)/ln 2

G1 = G

STEIN61.Bas from Stein 50 Gives slope of ln R v.s. G. (G = ln(…)).

Max slope is ~ 4.5 at R ≈ 9.5

Peak is at R ≈ 829.4 : G = −468.666685

~~we are down~~ G is down .86 from peak at R=64 $e^{.86} = 2.36$

G at 16M (700) is −468.68.08957 : down .0191 from peak.

.20 (SAD) T. apriori for abs wt. of $\bar{\mu}$ should be indep of n.

So say for n=1 we will have rel wts c & d for $\bar{\mu}$ & μ_i' resp: c+d=1.

Give c uniform apriori. So

μ_i estimate = $(c\bar{\mu} + (1-c)n \cdot \mu_i')/(c + (1-c)n)$

so c is always handicapped by factor n (n≈45), the size of each player

.24 (ΣN) No matter how much data one has, the Apriori can be so badly selected, that the prdn. results will be very bad! — Does this maxim apply in my choice of aprioris for R in the Stein problem? → (34)

A reasonable guess for the apriori of $\bar{\mu}$ would be (the σ^2 of the entire raw data set) — (mean σ^2 of the players). Why didn't I get this in the more rigorous approach to STEIN?

.31 In combining 2 p.d.s, sometimes one convolutes & sometimes one multiplies — The results are quite different (Multiplying narrows the result. Convol. broadens the result.

.33 Trouble is, w. the Bern. dist. I can't convolute since I have to stay on interval (0,1).

.34 Re: (24) To get an idea of what the actual apriori is, pretend that I have data for 00 previous yrs.

Re: (.31 − .33) Instead of convoluting the 2 Bern. p.d.f.'s, I've been multiplying them!

Multiplication gives $\frac{1}{S_{total}^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$; $e^{\frac{(x-a)^2}{2\sigma_1^2}} + \frac{(x-b)^2}{2\sigma_2^2}$ + x^2 coeffs in exponent are $\frac{x^2}{2\sigma_1} + \frac{x^2}{2\sigma_2^2}$

Con $S_{total}^2 = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}$ } This number must be smaller than either σ_1^2 or σ_2^2.

Convdn. Gives S^2 total = $\sigma_1^2 + \sigma_2^2$.

2.28.99 SMAT STEIN.

See 116.32 – 120.40 for workout eq. 62.03

Say in the Gaussian case: $\bar\mu$, s^2; μ_i, σ_i^2: we have only k data pts, rather than k·n.

T. pc. of the observed $\{\mu_i'\}$: — no constant factor

(.03) $$\prod_{i=1}^{k}\left(\frac{1}{s}\cdot e^{-\frac{1}{2}\frac{(\mu_i-\bar\mu)^2}{s^2}}\ \frac{1}{\sigma_i}e^{-\frac{1}{2}\frac{(\mu_i-\mu_i')^2}{\sigma_i^2}}\right); \rightarrow \boxed{s^{-k}\cdot\prod_{i=1}^{k}e^{-\frac12\frac{(\mu_i-\mu_i')^2}{\sigma_i^2}}}$$

(.04) $$= s^{-k}\cdot\exp-\frac12\left(\sum_{i=1}^{k}\underbrace{\frac{(\mu_i-\bar\mu)^2}{s^2}}_{\text{say sum to k}}+\frac{(\mu_i-\mu_i')^2}{\sigma_i^2}\right)$$

$$\boxed{\frac{k}{2}\cdot\ln(s^2)-\frac12\sum\frac{\mu_i-\mu_i'}{\sigma_i^2}}$$

.05 $s^2 k = \sum(\mu_i-\bar\mu)^2$ $\sum 1 = k$. $= ks^2 = \sum(\mu_i-\bar\mu)^2 = \sum(z\bar\mu+(1-z)\mu_i'-\bar\mu)^2$

$\sum((z-1)\bar\mu-z\mu_i')^2 = \sum(1-z)\bar\mu+z\mu_i')^2$

$\sum((z-1)\bar\mu+(1-z)\mu_i')^2 = (1-z)^2\sum(\bar\mu-\mu_i')^2$

say $\boxed{\mu_i = z\bar\mu+(1-z)\cdot\mu_i'}$. $\bar\mu = \frac1k\sum\mu_i'$. No d

(.07) $G = \ln(.04) = -\frac12 k\ln(s^2) - \frac12\left(\sum\frac{(\mu_i-\bar\mu)^2}{s^2}+\frac{(\mu_i-\mu_i')^2}{\sigma_i^2}\right)$

$$\frac{d}{dz} = \frac12 k\frac{1}{(s^2)}\frac{ds^2}{dz} - \frac12 k + \frac12$$

$\mu_i-\bar\mu = z\bar\mu-1\bar\mu+(1-z)\mu_i' = (1-z)(\mu_i'-\bar\mu)$

$s^2 = \frac1k\left(\sum(\mu_i'-\bar\mu)^2\right)(1-z)^2$ $\frac{\partial s^2}{\partial z} = -\frac2k\left(\sum(\mu_i'-\bar\mu)^2\right)(1-z)$

$\mu_i-\mu_i' = z\bar\mu+(1-z)\mu_i'-\mu_i' = z\bar\mu-z\mu_i' = z(\bar\mu-\mu_i')^2$

$= z(\bar\mu-\mu_i)$

$\frac{(\mu_i-\mu_i')^2}{\sigma_i^2} = z^2\frac{(\bar\mu-\mu_i')^2}{\sigma_i^2}$

$$\frac{d(.07)}{dz} = -\frac12\frac{k}{\sum(\mu_i'-\bar\mu)^2(1-z)^2}\frac{2}{k}\sum(\mu_i'-\bar\mu)^2(1-z) - \frac12 k - \frac12 z^2\overset{A}{\boxed{\sum\frac{(\bar\mu-\mu_i')^2}{\sigma_i^2}}}$$

.20 Actually, $\boxed{\bar\mu = \left(\sum\frac{\mu_i'}{\sigma_i^2}\right)\Big/\sum\frac{1}{\sigma_i^2}}$

Does $s^2 = \frac1k\sum(\mu_i'-\bar\mu)^2$? No! $s^2 = \frac1k\sum(\bar\mu-\mu_i')^2$

.03 → .06
.04 → .08
.07 → 0.11

$$\frac{d(.08)}{dz} = +\frac{1}{1-z} + \frac12 k + \frac12 z^2 A = 0$$

→ all terms but $\frac{1}{1-z}$ are > 0 so $1-z<0$ — unlikely! so z is a funct of k & A only! > !

$$\frac{1-\frac12k+2zk}{1-z} = z^2A$$

Looks like cubic eqn!

.26 from .07

$= \frac12 k\left[(\ln(1-z)^2)+\ln\frac{\sum(\mu_i'-\bar\mu)^2}{k}\right]$

$$\boxed{-\frac12 k\ln s^2} - \frac12\underbrace{\sum\frac{(\mu_i-\bar\mu)^2}{s^2}}_{=\frac{k}{2}} - \frac12\sum\frac{(\mu_i-\mu_i')^2}{\sigma_i^2}$$

$$\frac{d(.07)}{dz} = -\frac12 k\frac{1}{s^2}\frac{ds^2}{dz} - \frac{z^2}{2}\boxed{\sum\frac{(\bar\mu-\mu_i')^2}{\sigma_i^2}} \equiv A$$

$\boxed{s^2 = \frac{\sum(\mu_i'-\bar\mu)^2}{k}(1-z)^2}$

$\mu_i-\mu_i' = z\bar\mu+\mu_i'-z\mu_i'-\mu_i'$
$= z(\bar\mu-\mu_i')$

.30 from .05 $s^2 = \frac1k\sum(\mu_i-\bar\mu)^2$ $\mu_i-\bar\mu = (z\bar\mu+\mu_i'-z\mu_i')-\bar\mu$ $-\left(\frac{-\frac12k\cdot2}{1-z}\right)$

.31 $= \frac1k\left[(1-z)^2\sum(\mu_i'-\bar\mu)^2\right]$ $= (z-1)\bar\mu+(1-z)\mu_i' = (1-z)(\mu_i'-\bar\mu)$

$\frac{\partial s^2}{\partial z} = \left[\frac1k\cdot\sum(\mu_i'-\bar\mu)^2\right]\frac{d}{dz}(1-z)^2 = -\frac2k\sum(\mu_i'-\bar\mu)^2\cdot(1-z)$ $\mu_i-\bar\mu$

$$\left(-\frac{k}{2}\right)\cdot\frac{k}{(1-z)^2\left(\sum(\mu_i'-\bar\mu)^2\right)}\cdot\left(-\frac{2}{k}\sum(\mu_i'-\bar\mu)^2\right)(1-z) - \cancel{\frac{k}{2}} - \frac{z^2}{2}A$$

$\frac{k}{1-z} - \cancel{\frac{k}{2}} - \frac{z^2}{2}A = 0$ | $\frac{1}{1-z} - \cancel{\frac{k}{2}} - \frac{z^2}{2k}\cdot A = 0$ so z is a function of $\frac{A}{k}$ only.

$z^2(1-z) = \frac{2k}{A}$ $\frac{1}{1-z} = \frac{z^2A}{2k}$

.36 $\frac{k}{1-z} - zA = 0$. $\boxed{z(1-z) = \frac{k}{A}}$ Look at $\frac{A}{k} = \frac1k\sum_{i=1}^{k}\frac{(\mu_i'-\bar\mu)^2}{\sigma_i^2}$

$$\frac1k\sum\left(\frac{\mu_i'^2}{\sigma_i^2}-\frac{2\bar\mu\mu_i'}{\sigma_i^2}+\frac{\bar\mu^2}{\sigma_i^2}\right) = \frac1k\left(\sum\frac{\mu_i'^2}{\sigma_i^2}-\frac1k 2\bar\mu\sum\frac{\mu_i'}{\sigma_i^2}+\frac{\bar\mu^2}{k}\cdot\sum\frac{1}{\sigma_i^2}\right)$$

$$\frac1k\sum\frac{\mu_i'^2}{\sigma_i^2} - \frac{2\bar\mu}{k}\underbrace{\sum\frac{\mu_i'}{\sigma_i^2}}_{\bar\mu\cdot\sum\frac{1}{\sigma_i^2}} + \frac1k\bar\mu^2\left(\sum\frac{1}{\sigma_i^2}\right) = \boxed{\frac1k\left(\sum\left(\frac{\mu_i'^2}{\sigma_i^2}\right) - \bar\mu^2\cdot\sum\frac{1}{\sigma_i^2}\right) = \frac{A}{k}}$$

.01 look at $\frac{A}{K}$ = 62.36 R; Say σ_i^2 all = 1. $\frac{A}{K}$ will almost always be > 1. So $\frac{K}{A} < 1$

if $\bar{\mu} = 0$

$z(1-z) = \frac{K}{A} = \alpha$ ~~[illegible]~~ $z - z^2 = \alpha$ $z^2 - z + \alpha = 0$

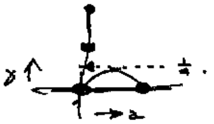
$z = \frac{1 \pm \overset{\downarrow ?}{\sqrt{1-4\alpha}}}{2}$ ax^2+bx+c; $x = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$

$\mu_i^* = z\bar{\mu} + (1-z)\mu_i'$

[illegible] $x = \frac{1}{2}$; $\alpha = \frac{1}{4}$ is a soln.

So soln is real only if $\alpha < \frac{1}{4}$.

For $0 < z < 1$; $z(1-z)$ ~~[illegible]~~ has max of $\frac{1}{4}$ when $z = .5$; it only goes small as 0.



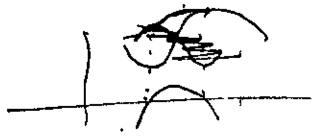
So $\frac{A}{K}$ has to be > 4 for real soln. to z!

and if z is a soln, then 1-z is a soln! ~~[illegible]~~ !

So if z is a good wt. for $\bar{\mu}$ then 1-z is also a good wt. for $\bar{\mu}$)!

Well, while there may be 2 solns, 1 can be max, the other min!

~~[illegible]~~ when z = .5, it's neither max nor min. but an inflection pt;



This is definitely not behaving in a Reasonable Manner!

.20

Going thru it again!

$\frac{d}{dz}$ $S^2 = \beta \cdot (1-z)^2$ (62.26R)

from 62.07 : ~~[illegible]~~

$\boxed{\mu_i' - \bar{\mu} = z(\bar{\mu} - \mu_i')}$

↓

$\boxed{z^2 \sum \frac{(\mu_i' - \bar{\mu})^2}{\sigma_i^2}} = \boxed{z^2 A}$

.07 : $= -\frac{1}{2}k\ln\beta - k\ln(1-z) - \frac{1}{2}k - \frac{1}{2}\left(\frac{\sum(\mu_i - \bar{\mu})^2}{\sigma_i^2}\right)$

$= -\frac{1}{2}k\ln\beta - k\ln(1-z)$ ~~$-\frac{1}{2}k$~~ $- \frac{1}{2}z^2 A$

$\frac{d}{dz}$ = $\frac{+K}{1-z} - zA = 0$

In baseball case: $\sigma_i^2 = \frac{\mu_i'(1-\mu_i')}{N}$. $A = \sum \frac{(\bar{\mu} - \mu_i')^2}{\sigma_i^2} = \sum \frac{(\bar{\mu} - \mu_i')^2 N}{\mu_i'(1-\mu_i')}$.

$z = \mu_i'$ $\bar{\mu} = .265$

ST63 $\boxed{A = A + ((\bar{\mu} - z)\wedge 2)/z/(1-z)}$

A = A * N / K A = 45/18 = $\frac{A}{K}$ = 1.09104 : so no real soln! — Really Crazy!

← from ST 4

22.899 SMFT ST

Try a pgm for 62.03 → .07 ! G as a function of ~~z~~ a.

.03 → $s^{-k} \cdot \prod_{i=1}^{k} e^{-\frac{1}{2}\frac{(\mu_i'-\mu_i)^2}{\sigma_i^2}}$ | $z \equiv \mu_i'$.

.07 → $-\frac{k}{2}\ln(s^2) - \frac{1}{2}\sum\frac{\overset{zz}{(\mu_i'-\mu_i)^2}}{\sigma_i^2}$ $\equiv$ SM

~~$s^2 =$~~ ~~[illegible]~~ ~~$\sum\mu_i'$~~

$\bar{\mu} = \frac{1}{k}\sum\mu_i'$; ① $s^2 = \frac{1}{k}\sum(\mu_i')^2 - \bar{\mu}^2$ $\equiv$ SS ! just use $\bar{\mu} = .265$; ~~[illegible]~~

→ use $s^2 = \frac{1}{k}\sum(\mu_i'-\bar{\mu})^2$ in pgm. — so s^2 is never <0 ; this it will be ϕ if A=1

will be ϕ if A=1

② (zz) $\mu_i' = a\bar{\mu} + (1-a)\mu_i'$

(SGS) ③ $\sigma_i^2 = \frac{\mu_i'(1-\mu_i')}{N}$ ← SGS

$k = 18$; $N = 45$

ST 64.805 I got Max at $a=0$, It ↓ monotonically to −9.22 at a=1. — The peak at a=0 is symmetric ~~for~~ about a=0 so a=−1 gives same G as a=+1. G = .59

Wrong formula for $s^2 \equiv SS$.

a=0 means wt=1 ~~[illegible]~~ for $\mu_i' \equiv z$

$U \equiv \bar{\mu} = .265$

$(z = \mu_i')$: $zz = a \cdot U + (1-a)z$

$SGS = z*(1-z)/N$ $(= \sigma_i^2)$

Summing using $\sum$ (SS = SS + zz*zz) Then SS = SS/K − UU } gave peak at (a) slightly >1. ~~[illegible]~~ a = +1.011

Apparently ~~[illegible]~~ ln s^2 ~~term~~ is quite small compared to other term.

I had factor of $\frac{k}{2}$ inverted! Change $\frac{2}{k}$ to $\frac{k}{2}$ | Now monotonic ↑ from a=0 to 1

This time the peak is at a=1.01 maybe.

There may be a singularity near a=1 (→ for a = 1.01 or 1.001 it says "illegal funct. call"

.991 → 107.49 ; 1.019 illegal.

.999 → 128.197 ; 1.049 ok.

1 → 174.05

O.K.! $s^2 = \frac{1}{k}\sum(\mu_i)^2 - \bar{\mu}^2$ can be <0

So ln s^2 is troublesome!

T. formula for s^2 is wrong!

$\bar{\mu}_i = \bar{\mu}_i' = \bar{\mu}$

Trouble may be due to my not using ~~exactly~~ right $\bar{\mu}$.

Using ~~[illegible]~~ more exact $\bar{\mu}$ → U = .265 3889

I got from .99 to .998, w. illegal .999

a=1 was legal at G= 179.14 a very sharp peak.

G ~ 121 for ~~[illegible]~~ A = .99 or 1.01

The actual peak was A<1. (very slightly less).

A = 1 means all wt. on U.

A = 1.000001 gives Illegal Funct call.

A=1 doesn't

A = 1.049 doesn't:

A = 1.049 is illegal.

A = 1.045 is legal G = 51.66

1.0448 illegal.

1.0449 illegal

.0445 ill

1.04499 legal G = 173.149

1.0449 ? illegal

~~.95 illegal~~

Well, it seems clear that there is a ~~[illegible]~~ singularity at ~~A~~ a=1. This corresponds to (1−a)=0 which is all wt at $\mu_i = \bar{\mu}$. At (1−a)=0, $s^2 = \frac{\sum(\mu_i'-\bar{\mu})^2}{k}(1-a)$ (62.26) → 0. So s^{-k} → ∞.

Its a simple pole; I don't know if ln (x) has a simple pole at x=0.

So This looks like same problem w. ~~[illegible]~~ Binomial distribution. as $R (= x+y) \to \infty$. $R = \infty$ corresponds to $(1-a) = 0$.

A possl. error in t. [redacted] Gaussn stuff here: I assumed $\bar{\mu}$ was known $\bar{\mu} = \frac{1}{k} \sum \mu_i$.

Try U	Peak
w .265	1 → 179.14
.24	1 → 155.122
.23	1 → 153
.22	trouble at A=1. A=.995 → 70.79. ← —
.2	1 → 158
.1	1 → 125
.26	165
.27	trouble at A=.95
.28	" " " "
.30	" " " "

That s^2 is neg means it is being calculated incorrectly.

s^2 should be $\sum (\bar{\mu} - \mu_i)^2$

SS = SS + (zz − u)^2 z

SS = SS/k

In this case, 1 is clearly a pole, no matter what $\bar{U}$ is.

For U = .4 there is a ~~pole~~ sing at A = 0

" U = .5 Things seem to change discontinuously, but still, 1 Peak, sing at a = 1.

I think what happens is that t. Gaussian ~~term~~ factor becomes imp.

Re: t. calcn. of a for max pc: (62.56 – 63.20 [redacted]) Usually ~~[illegible]~~ $\frac{A}{K}$ is slightly > 1 so $\frac{k}{A}$ is slightly < 1 but $\gg \frac{1}{4}$, so there ~~[illegible]~~ are no real peaks.

Looking at eq. 62.03 — It really looks like there has to be a ~~[illegible]~~ pole at ~~[illegible]~~ $(1-a) = 0$.

Perhaps look at it from pt. of view of coding: ~~Say I simply have a set μ_i to code directly~~

As s^2 gets small, it should become very diffk to code data distant from t. mean. From t. pt. of view of coding, I don't see how one could ~~[illegible]~~ have a pc = ∞! ☺.

~~[illegible]~~ Using eq 62.03 as a basis for coding: For each value of a, we have a code for t. corpus.

Exactly how does one use eq 62.03 to code a corpus) — ask same q. for t. divergent Binomial d.f. code.

→ [Well, t. $\frac{1}{s}$ pole is a familiar δ-function! The peak is ~~[illegible]~~ high, but Also narrow. T. product of width & ht = 1. + so it's width is irrelevant]

In 62.03, we use an initial s, $\bar{\mu}$ & a to derv. t. corpus; we then integrate over all s, $\bar{\mu}$, a to get total pc.

Also, note in 62.03 t. factors $\underbrace{\prod_i e^{-\frac{1}{2}\frac{(\mu_i - \bar{\mu})^2}{s^2}}}$: It has a "super zero" that dominates any pole! While $\frac{1}{s^k}$ → ↑ has a peak at $s^2 \to \frac{1}{k} \sum (\mu_i - \bar{\mu})^2$, it is a very narrow peak if t. corresp. s^2 is small.

zero ↙ ~~essential~~ singularity.

8.1.99 SMFT STEIN

Re: 62.03!

$\underline{S}$, $\underline{\bar{\mu}}$ & $\underline{a}$ are indep. der. params of the data set. ($\in$ $[r_{ij}]$)

$\mu_i = a\bar{v} + (1-a)\mu_i'$.

No! Sits and its param. This is merely the position of $\underline{a}$ peak.

$$\boxed{s^2 = \frac{1}{k}\sum(\mu_i-\bar{\mu})^2}$$

$$\boxed{S^2 = \frac{(1-a)^2}{k}\sum(\mu_i'-\bar{\mu})^2}$$

$\mu_i - \bar{\mu} = (1-a)(\mu_i'-\bar{\mu})$ (62.30)

Here: $\sum(\mu_i-\bar{\mu})^2 = (1-a)^2\sum(\mu_i'-\bar{v})^2$ — (in numerator of exponent)

$\mu_i-\mu_i' = a(\bar{v}-\mu_i')$ so $\sum(\mu_i-\mu_i')^2 = a^2\sum(\mu_i'-\bar{v})^2$ $\frac{1}{\sigma_i^2} = \frac{N}{\mu_i'(1-\mu_i')}$?

So 62.03 $= \frac{1}{s^k}\cdot\exp-\frac{1}{2}\left(\sum(\mu_i'-\bar{v})^2\right)\left(\frac{(1-a)^2}{s^2}+Na^2\sum\frac{1}{\mu_i'(1-\mu_i')}\right)$

62.03: $\frac{1}{s^k}\prod_i\frac{1}{\sigma_i^2}\exp-\frac{1}{2}\left(\boxed{\frac{\sum(\mu_i-\bar{\mu})^2}{s^2}}\overset{(1-a)^2\sum(\mu_i'-\bar{\mu})^2}{}\;\boxed{\frac{\sum(\mu_i-\mu_i')^2}{\sigma_i^2}}\overset{a^2\sum(\mu_i'-\bar{\mu})^2}{}\right) \rightarrow a^2\sum\left[\frac{(\mu_i'-\bar{\mu})^2 N}{\mu_i'(1-\mu_i')}\right]$?

$\left(\prod_i\frac{N}{\mu_i'(1-\mu_i')}\right)^{\frac{1}{2}}$ $\exp-\frac{1}{2}\cdot(\sum(\mu_i'-\bar{\mu})^2\cdot($

(I Think $\sigma_i^2 = \frac{\mu_i'(1-\mu_i')}{N}$ is o.k.! The σ_i^2 are supposed to be "Given".)

62.03 $s^{-k}\;\underbrace{N^{\frac{k}{2}}\cdot\prod\frac{1}{\mu_i'(1-\mu_i')}}_{\text{constants.}}\;\exp\left[-\frac{1}{2}\left(\frac{(\sum(\mu_i'-\bar{\mu})^2)\cdot(1-a)^2}{s^2}+N\sum\left(\frac{(\mu_i'-\bar{\mu})^2}{\mu_i'(1-\mu_i')}\right)\cdot a^2\right)\right]$

Simple expression in terms of 2nd moment of $\{\mu_i'\}$ & $\bar{v}$ — (Must mean)

Similar remarks apply to This

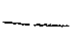
What it will look like: $\sim s^{-k}\cdot\text{const}\cdot\exp-\frac{1}{2}\left[\left(A^2+(\bar{\mu}-v_1)^2\right)\frac{(1-a)^2}{s^2}+N\left(B^2+(\bar{\mu}-v_2)^2\right)a^2\right]$

I'm not so sure of this structure!

taking ln's simplifies. $-\frac{k}{2}\ln(s^2)$

for each s^2 value we can ~~[illegible]~~ perhaps solve for $\bar{\mu}$ & a to minimize the exponentential part. So the exponential part is a funct of s^2. Then we minz $\frac{k}{2}\ln(s^2)+f(s)$.

or just do a 3 dim optzn "by hand".



On the other hand, it's easy to maxz w.r.t. s^2, first which simplifies things a lot!

We then ~~[illegible]~~ have a ~~[illegible]~~ function of $\bar{\mu}$ & a to maxz! The max is at $a=1$ (Max=∞)

To get actual useful output, integrate over all $s,\bar{v},a$. Perhaps have range 0,1 for each — so $\int$ over the unit cube to get total pc. of code.

One trial: fix $\bar{\mu}$ around .265, Then, for each a value, integrate over all s. Maybe there will be a peak in a.

For any value of $a=1$, the pole in S ~~[illegible]~~ at $s=0$ will be removed by the $e^{-\frac{M}{s^2}}$ factor.

While the pgm ST64.Bas could be easily modified to do this integration! I don't like ST64.B. or ~~[illegible]~~ Romberg's pgm. Much neater!

3-1-99 SMFT STEIN

.01 K=18, N=45

```
U (≡ Ū) = .265      (= Ū)
A = (set by hand)      | A=2
For A = .1 to 1, step .1 (maybe modify)
  For s = .1 to 1 step .1
                                  UU ≡ $U_i$ ! $u_i = a\bar{U} + (1-a)U_i'$
    For J=1 to k
      Z = A(j,3)                  | Z ≡ $U_i'$.
      UU = A*U + (1-A)*Z          | UU ≡ $U_i$
      S1 = $\Sigma(u_i-\bar{u})^2$     S1 = S1/s^2. after J loop
      S1 = S1 + (UU - U)^2
      S2 = $\Sigma((u_i-u_i')^2/\sigma_i^2) = \Sigma (u_i-u_i')^2/u_i'/(1-u_i')$ ! S2 = S2*N. after j loop.
    Next J
  After J loop.
  print a, s, s^(-k/2) * exp(-1/2 (S1 + S2))
```

.16

ST67

Same as .01-.18

K=18, N=45

U = .269

A = .1

$$S1 = \left(\Sigma (u_i-\bar{u})^2\right)/s^2$$

$$U_i = A\bar{U} + (1-A).u_i'$$

$$S2 = \Sigma\left(\frac{(u_i'-\bar{u})^2 N}{u_i'(1-u_i')}\right)$$

```
30 For JJ = 1 to 20 : s = JJ/20
   S1=0 : S2=0
   For J = 1 to k
     Z = A(J,3)/1000
     UU = A*U + (1-A)*Z                UU = $u_i$ ; U = $\bar{u}$   Z = $u_i'$
     S1 = S1 + (UU-U)^2
     S2 = S2 + ((UU-Z)^2)/Z/(1-Z)
   Next J
   S1 = S1/(s^2)  ←
   S2 = S2*N
   print A, s, s^(-k/2) * exp(-(S1+S2)/2).
   NEXT JJ
   ~~Input A ; Go to 30~~
```

For A = .1 ; sharp peak at s ≈ .09 peak ; e^G = 3 million at s = .075 ; ≈ 1.4×10^{-9} at s = .025

The Increase in e^G as S→0 is rapid, then after it gets to its peak, it drops down very rapidly.

A	peak
.2	~~.095~~ .08
.3	.07
.4	.06

as A increases toward 1, the wt. of S becomes closer & closer to S=0.

A	peak
.5	.05
.6	.04
.7	~~(A=.025) (S=.025 e^G = 5.5×10^9)~~ .03
.8	.02
.9	.01
.95	<.005

Ran O.K. first time (Eguess!)

3.1.99 SMRT STEIN

Try integrating wrt S, ~~[illegible]~~ $\int_0^1$ for each A value.

ST68.BAS

so INT=0

for JJ=1 to 100 : s= JJ/100

~~LPRINT~~ INT=INT+ S^(-k(2))x . . . .

Next JJ

Print INT/100.

Input A.

"INT" is special word! Use INTT

So it looks like almost all wt is at A≈1, S≈0. U=.2653889

I could try .25

	INTT U=.2653889		U=.25		U=.28		Double prec. U=.28
.1	2.18M		1.786M		1.79M	x3.4	6.1M
.2	4.16M		3.43M		3.3	x3.6	11.9M
.3	7.41M		6.17M		5.57	x3.93	21.9M
.4	12.77M		10.28M		8.86 M		
.5	22.65M		19.47 M		14.1 M		
.6	45.94M		40.3 M		25.3 M		
.7	124.75M		112.7M		59.3 M		
.8	764.20M		724.0M		310 M	x8.39	2.678E10
.9	58458.59M	5.845E10	55900.M	5.55E10	1.916E10		
.95		4.98E11		5.67E11 greater!	1.688E11		

$AB\,c\,d\int_{-\infty}^{100} e^{-\frac{x^2}{2}}dx$

Seems to be big difference betw. singles double precision! I wonder where it occurs!

Maybe error in computing exp(x)?

I do do 100 additions, so this could lose precision a lot.

Anyway, the results are not encouraging! It says that the $\bar{U}$ is always best estimate!

Try N = 450. U=.28

A	INTT × 10^6	try U=.265
.1	~~[illegible]~~ 2.66	15.6 M
.2	.43	5.9
.3	.0125	.7
.4	6.7E-5	.02
:		
.8	2.2 E-20	
.9	1.5 E-29.	~~4~~ E-11
.95		6 E-16
.99		

N = 200 U=.265 units of M. Add!

N = 150

A	N = 200 U=.265	N = 150
.01	13.7	13.7
.1	22	23
	24	32
	17	32
.4	8	25
.5		16
.6	~~5~~.05	9
.7	.122	6
.8	.046	5
.9	.0087	38
.95		27.9
.99		9

M

strange!

STRANGE

See
116.32 –
120.40
for Mmp (default) work on arc.
63.03 ff

Maybe this is correct! — i.e. for N as small as 45, all wt is in $\bar{U}$.

I did get results like this by ~~[illegible]~~ comparing P cost of Bernoulli codes

① using one $\bar{U}$ for entire corpus v.s. using k U_i's for k corpus: as N passed thru 210

There was rapid change in betw. which method had most Pc, & one would quickly dominate the other. 93.01–.30 deals w. this result.

3/2/99 SMFT STEIN

Some Note on previous pgm. (ST 97) ← did I mean ST67?

1) Try 1000 pts of integration a/o ($\pm$EXT) extended precision.

2) On t. General problem of correctness of concepts usd:

~~Initially~~. Say F_p is t. D.f. for ~~[illegible]~~ each player about his ~~[illegible]~~ mean, μ_i.

F_u is t. D.f. for means, μ_i

F_F is t. D.f. for for t. entire ~~corpus~~ Corpus.

If we convolute

If we convolute F_u, F_p ; $\sigma_F^2 = \sigma_u^2 + \sigma_p^2$. } ← These are easily shown for Gauss D.f.'s.

.10 " " multiply F_u, F_p : $\frac{1}{\sigma_F^2} = \frac{1}{\sigma_u^2} + \frac{1}{\sigma_p^2}$ } (o'

.11 Using Bin D.f.s instead of Gaussian, we don't have a convolution but we have something like t.

.12 $$\int_0^1 \left(F_u(\mu_i) \underbrace{\left(\mu_i^{\mu_i' \cdot N} \cdot (1-\mu_i)^{(1-\mu_i')N}\right)}_{F(\mu_i', \mu_i)}\right) d\mu_i = F_F(\mu_i').$$ ← Normalizn consts. should be inserted. This is an integration over

all μ_i that could give rise to t. observed $\mu_i' \cdot N$ data pts.

.14 To get t. final D.f. we mult all these together, ~~[illegible]~~ $\prod_i F(N\mu_i')$ ← Giving a (Geom mean) of all these "Pseudo convolutions" ← i^{th} power of

T. reason $F(\,)$ is not a convolution, is that for each μ_i' point, $F(\mu_i', \mu_i)$ ~~[illegible]~~ has a somewhat difnt form! This ~~[illegible]~~ makes it possl. for all of these "pseudo ~~convolutions~~"

to keep on t. interval (0,1).

I believe that I did implement .12 - .14 in my ~~[illegible]~~ last best, version of t. Bin.d.f. omial

i.e. 46.28 and ST46

In 46.28, t. norml proby of a hit on t. next "at bat" by t. i^{th} player is just $\frac{X+A_i+1}{X+Y+N+1+1}$

~~[illegible]~~ One way to look at it: That rel proby of hit v.s. out is $\frac{X+A_i+1}{Y+B_i+1}$

normd proby of hit is $\frac{X+A_i+1}{X+A_i+1+Y+B_i+1} = \frac{X+A_i+1}{X+Y+N+2}$ $X = \bar{\mu} R, \; Y = (1-\bar{\mu})R$

To get a prediction, we integrate this ↗ w. wt. of corresponding values of (X,Y) (or $\bar{\mu}$, R)

To get an idea of what t. apripd of ~~[illegible]~~ R, $\bar{\mu}$ looks like: One way would be to

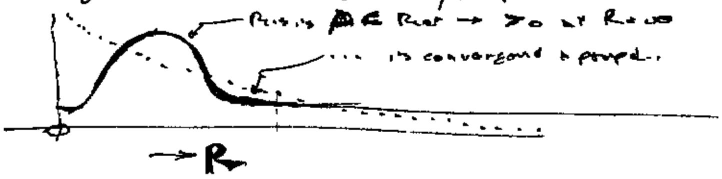
ABCD EFG

consider many previous yrs: try to get apripd ~~induced~~ on X Y, induced by that (large) data set.

.31 My Guess is that for a finite data set, ~~[illegible]~~ $\bar{\mu}$ ~~[illegible]~~ has a reasonable D.f. betw 0 & 1, but R will have a pc > 0 for R=0 and (unfortly) a pc > 0 for R → ∞!

Not so bad! Alth t. df for large ssz will be ~~[illegible]~~ # > 0 ~~[illegible]~~ for R=∞, it will be very small. This will not, hvr, be a true apripd, since $\int_0^\infty (\,) = \infty$.

.34 We have to use an axillary apripd that does have $\int_0^\infty (\,) = 1$. Multiplying by t. D.f. of (31) will give us a d.f. that is fairly indep of t. behavior of this (.34) apriop. at ∞.



T. product of t. 2 D.f. of course, converges.

3.2.99 SMFT STEIN c.p.? R?

70

I had a discussn of c.p. of ≈ ~~46~~ 46.28 at R=0 & at R=∞ : 49.07 ff

$a=\bar{u}, b=1-\bar{u}$: for R=∞ ~~$~~ 49.07 ⇒ $a^{A_i} b^{B_i} = \left(a^{u_i'} b^{1-u_i'}\right)^N$

For t. entire set this becomes: $\left(\prod_i a^{u_i'} b^{1-u_i'}\right)^{kN}$. T. max value this can have is $\frac{1}{2}$ — No! Is can be large: vpb!

(when $u_i' = a$ & $b = 1-u_i'$). So t. value at R=∞ will be $\boxed{\leq 2^{-kN}}$ ln = $\frac{kN}{\ln 2} = \frac{18 \cdot 45}{\ln 2} = \frac{810}{.7} = -1168.6$

I didn't get this small at ∞! ~~for~~ G(R=∞) < −1168.6 : I got for R=∞ −468.68 on 61½

from 49.20 ;, T. value when $a=\bar{u}$, is $\left(\bar{u}^{\bar{u}}(1-\bar{u})^{(1-\bar{u})}\right)^{kN}$ = .56 (!) | .2653889

~~(1−.56)×810 = −469.65~~

does it fit exactly for u=.2653889 : $810 \times \ln\left(u^u (1-u)^{1-u}\right) = -468.680\boxed{9}$

on 61½ I got for R=16.37 : −468.680$\boxed{8}$957 very close!

O.k. now while ~~(.56)~~ $(.56)^{810}$ seems very small, in t. pgm STG1.B, It was t. largest value of G in t. range R=1, ∞

Actual max value at R=∞ is $\left(u^u(1-u)^{1-u}\right)^{kN}$ this is max at ½ (= ½ → 1) Max at u=0 or

I probly left out some Normzn factors:

.17 from 46.02 they are $\prod\left(\frac{N+1!}{A_i!\,B_i!}\right)$ ln = ~~N~~(N ln N) − ~~A~~ $A_i \ln A_i - B_i \ln B_i + \ln(N+1)$ $-\frac{1}{2}\ln\left(\frac{N}{A_i B_i \cdot 2\pi}\right)$

.18 Whether G is small at R=∞ is unclear : Depends on ~~whether~~ Whether G has hy peak(s) before R=∞

→ T. main Q now is whether t. discn. of 69.31−40 is Basically correct.

← Get size of this in terms of u_i', N :

Try $\frac{N!}{(pN)!\,(qN)!}$ ~~(...)~~ $N\ln N - pN\ln pN - qN\ln qN$ ~~N ln N~~ $- pN\ln p$ ~~− pN ln N~~ $- qN\ln q$ ~~− qN ln N~~ = $-N(p\ln p + q\ln q)$

$+\frac{1}{2}\ln\left(\frac{N \cdot 2\pi}{pN\,qN \cdot 2\pi}\right)$ $= -\frac{1}{2}\ln(pqN2\pi)$

p+q=1 & sub. known

$\approx \frac{\left(p^p q^q\right)^{-N}}{\sqrt{2\pi Npq}}$ Mult by N+1 = $\left(p^p q^q\right)^{-N} \cdot \underbrace{\sqrt{\frac{N}{2\pi pq}}}_{(\text{width})^{-1} = \text{Height}} \left(1+\frac{1}{N}\right)$

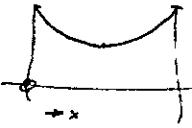
.23 $$\frac{N+1!}{A_i!\,B_i!} \approx \left(u_i'^{u_i'}(1-u_i')^{(1-u_i')}\right)^{-N} \cdot \sqrt{\frac{N}{2\pi u_i'(1-u_i')}} \cdot \left(1+\frac{1}{N}\right)$$

This factor would exactly cancel out the $\left(\bar{u}^{\bar{u}}(1-\bar{u})^{(1-\bar{u})}\right)^N$ factor at ∞.

Hvr, at ∞ ~~e~~ e^G will be >> 1 if these factors cancel! because of t. $\left(\sqrt{\frac{N}{2\pi pq}}\right)^{k}$ factor.

Hvr, t. 2 factors don't cancel. One is usually > t. other. i.e. the product of t. $\left(u_i'^{u_i'}(1-u_i')^{1-u_i'}\right)$ factors,

is usually > t. N power of $\bar{u}^{\bar{u}}(1-\bar{u})^{1-\bar{u}}$ because of a (Convexity arg't) ← Take the curve, then convexity is in a linear environment.



.27 $x^x(1-x)^{(1-x)}$

.28

So for lrg Nk, the $\prod_i \sqrt{\frac{N}{2\pi u_i'(1-u_i')}}$ factor is $\approx \left(\frac{N}{2\pi \bar{u}(1-\bar{u})}\right)^{\frac{k}{2}}$

and t. other factor is something like $\left(\alpha^{2N}\right)^{\frac{k}{2}}$ (α < 1)

So it depends ~~on other~~ on rel size of N and α^{2N}.

If t. u_i' have a broad d.f. then α is small. If u_i' are close together α is closer to 1.

In fact we can compute an approxn to α, if we know $\bar{u}$ and the dispersion (σ^2) of $[u_i']$ about $\bar{u}$.

Re: Behavior for R=0! See 49.26! The Normzn factor for $v^{A_i}(1-v)^{B_i}$ cancels out t. — & it should since t. factors are x, y'=1.

Value at R=0 exactly : so value is exactly 1. — in contrast to value at R=∞ which is usually

.34 << 1 (WRT R=0) Also, t. concavity arg't enables us to tell how narrow t. D.F. is in t. $\bar{u}$ direction,

α has a broad peak at $\bar{u}$ = arith mean of all u_i'; but this peak is much sharpened by t. power kN (.28)

For N=95, k=18, I'm not sure G(R=∞) < G(R=0) : for very large Nk it would seem very likely to be much smaller at R=∞. Depends on α & k.

$\alpha \equiv \left(\prod_{i=1}^{k}\left(u_i'^{u_i'}(1-u_i')^{(1-u_i')}\right)\right)^{\frac{1}{k}} / \bar{u}^{\bar{u}}(1-\bar{u})^{(1-\bar{u})}$.

3394 SMFT Stem.

for R=1 G << at R=∞ for present run. (ST46)

See if this is because ~~[illegible]~~ $\alpha^{2N} > N$. (70.28) : or is G at R=0 much different from for R=1?

eq 46.28

~~[illegible]~~ $\frac{(\frac{1}{2}+A)!(\frac{1}{2}+B)!}{(N+1+1)}$ How much different from $\frac{A!B!}{N+1}$?

$\frac{1!}{\frac{1}{2}!\,\frac{1}{2}!}$ how much different from 1.?

$(x+\frac{1}{2})! \approx x!\sqrt{x+\frac{3}{4}}$

$= \frac{A!B!\sqrt{(A+\frac{3}{4})}\sqrt{B+\frac{3}{4}}}{N+1!\,(N+2)}$

So its not very much different.

.07

$\frac{1}{2}! = \frac{1}{2}\sqrt{\pi} = .88622$

$(\frac{1}{2}!)^2 = \frac{1}{4}\pi = .7853$

$(\frac{1}{4}\pi)^{-1} = 1.273$

This can be done better! $(x+a)! \approx x!(x+b)^c$ is not bad approx. See 72.23 for better approx for G(1)

ST71 Compute ~~[illegible]~~ ln α, α: ~~[illegible]~~ $\underbrace{\frac{1}{k}\sum \mu_i'\ln\mu_i' + (1-\mu_i')\ln(1-\mu_i')}_{\equiv A}$ ~~[illegible]~~ $-\bar{u}\ln u - (1-\bar{u})\ln u$. | is $ss \overset{?}{=} e^A$?

Since α should be > 1! α = ~~[illegible]~~ 1.7411818 ?

ss = exp( ) = .97523271823 — seems way too large!

$\bar{u}^{u}(1-\bar{u})^{(1-\bar{u})}$ = .5606724 (for $\bar{u}$ = .2653889.)

$\frac{ss}{(\ )}$ = ~~1.7411814~~ = α ---

.7032421

$\alpha \equiv \frac{ss}{\beta}$: $\beta = F_1(\bar{\mu}) \doteq \bar{\mu}^{\bar{\mu}}(1-\bar{\mu})^{(1-\bar{\mu})}$

ss = .5672587943

β = .5606724

α = 1.011747218

say $F_1(x) = x^x(1-x)^{1-x}$; $F_1(\bar{u}) = .5606724$ $F_1(.346) = .51$

.17 error in run! → $\frac{\overset{\exp(A)}{.5672587943}}{\beta}$

$\alpha = 1.011747218$ $\frac{1}{\alpha-1}$ = ~~[illegible]~~ 85.12653

This begins to seem much more reasonable!

Looking at 70.28 $\alpha^{2N} \approx e^{90\times.01174}$ = 2.86076 ~~[illegible] ≈ e with s's << N (=85)~~

from 7.17ff So: the normaliz constant of $\prod_i \frac{N+1!}{A_i!\,B_i!} \approx \prod_{i=1}^{k} \underbrace{(f_1(\mu_i))^{-N}}_{=\beta^k} \sqrt{\frac{N}{2\pi\mu_i(1-\mu_i)}}\,(1+\frac{1}{N})$

$\approx \beta^k \cdot \left(\frac{N}{2\pi\bar{\mu}(1-\bar{\mu})}\right)^{\frac{k}{2}} \cdot e^{\frac{k}{N}}$ $\approx \sqrt{\left(\frac{N}{2\pi\bar{\mu}(1-\bar{\mu})}\right)}^{k} \cdot e^{\frac{k}{N}}$.

$\frac{k}{N} = \frac{18}{45} = \frac{2}{5} = .4$

$= \left(\beta \cdot \sqrt{\frac{N}{2\pi\bar{\mu}(1-\bar{\mu})}} \cdot e^{\frac{1}{N}}\right)^k$ = +32.83027062

6.06 ln .1944776 .0673 6.06 6.196

← ~~ln = 1.8239 × 18 = 32.8367~~ should be difference between G at R=1 & R=∞.

= 3.51473 → 1.2589642 ln( ) → × 18 = 22.625356

Running ST46! Got a local max of -468.6718 at R = 1024. This is for V = .2653889

-468.67341 " " when I took π out to 6 more places to 3.141592654 not = !

OK! that's ok. on 50.20 I got -468.666 at R = 838.96

ΔG bet R=1 & R= 4.2M is only $\underset{R=+1}{\overset{G}{-482.1129}} - (\underset{R=4.19M}{\overset{G}{-468.6808}}) = \boxed{13.43}$

Also, if the normaliz factor is only $(3.5)^4$ then the Value of G should be ~~[illegible]~~ normaliz constant, ~~[illegible]~~ i.e. $e^{22.6}$ not e^{-482}. Wooops! [illegible] I left out a -N exponent here

$(.5672587)^{-810}$

.37 = +459.221528 in log domain. adding 32.83022 was $e^{+492.051786}$: Mult this normaliz by

Should grow +1 (i.e. zero in log domain, not ~ +10 !

3399 SMFT STem.

72

T. approxn of $\frac{N+1!}{A!B!}$ of 70.18ff may be wrong. ⊃70.23

N=10 ~~A=5~~, B=4

Check 70.23! Say N=10, A=5 B=4. $\frac{11!}{6!\,4!}$ = ~~[illegible]~~ 2310

$(.6^{.6}\,.4^{.4})^{-10}\sqrt{\frac{10}{2\pi\cdot .4\cdot .6}}\cdot(1.1)$

~~2.2946016~~

837.724 × 2.83267 ~~[illegible]~~ = 2373. Not far off!

$$\frac{N!}{A!B!} \to \frac{N^N}{A^A B^B} = \frac{N^A\cdot N^B}{A^A\ B^B} = \left(\frac{A}{N}\right)^{-A}\cdot\left(\frac{B}{N}\right)^{-B} \to \left(\left(\frac{A}{N}\right)^{\frac{A}{N}}\left(\frac{B}{N}\right)^{\frac{B}{N}}\right)^{-N}\cdot\sqrt{\frac{1}{\frac{A}{N}\cdot\frac{B}{N}\cdot N\cdot 2\pi}}$$

$$\frac{N^N}{A^A B^B}\cdot\sqrt{\frac{2\pi N\cdot\ \cdot N}{2\pi A\cdot 2\pi B\cdot N}}$$

$N+1 = N(1+\frac{1}{N})$

$a = \frac{A}{N}$, $b = \frac{B}{N}$

$$\boxed{\frac{N+1!}{A!B!} = (a^a b^b)^{-N}\left(\sqrt{\frac{N}{a\cdot b\cdot 2\pi}}\right)\left(1+\frac{1}{N}\right)}$$

O.K.

.1906536 $\frac{18}{45} = \frac{2}{5} = .4$

$$\prod_{i=1}^{k}\frac{N+1!}{\mu_i N!\,(1-\mu_i)N!} = \underbrace{\prod_{i=1}^{k} f_1(\mu_i)^{-N}}\left(\frac{N}{\mu_i(1-\mu_i)\cdot 2\pi}\right)\left(1+\frac{1}{N}\right)$$

$$= \beta^{-Nk}\left(\frac{N}{\bar\mu(1-\bar\mu)\cdot 2\pi}\right)^{\frac{k}{2}}\cdot e^{\frac{k}{N}}.$$

~~lo~~ lnc: 459.23366 | 37.565 32.634741 + .4) = = 492.268 in log domain: = 71.38 ??

I'm off by a factor of $(3.09)^9 \sim \sqrt{\pi}^k$

Well ~~[illegible]~~ R=1 ≡ R=0. On 71.07 I got an additional factor $\frac{N+2}{\sqrt{(A+\frac{1}{2})(B+\frac{1}{2})}} \approx \frac{N}{N\sqrt{\mu_i(1-\mu_i)}}$

$\sim\frac{1}{\sqrt{\bar\mu(1-\bar\mu)}} = 2.24$ {nothing there!} $\frac{(A+\bar\mu)(B+1-\bar\mu)}{N+1+1!}$

.23

neither is not good for ε = ½ ≈ .5

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

How bad is $x!(x+\epsilon)$? No for $\epsilon=0$!

$\boxed{(x+\epsilon)! \sim x!\cdot(x+1)^{\epsilon}}$ this is close for $\epsilon = \frac{1}{2}$ (not great approxn. works for $\epsilon=0$ & $\epsilon=1$.)

or $\sim x!(x+\epsilon)^{\epsilon}$

$\frac{1}{e^1}\cdot\sqrt{2\pi} = .92$

Also Stirling is not so good for x=1.

$\frac{N+1+1!}{A!\cdot B!\cdot A^{\bar\mu}B^{1-\bar\mu}}$ $\frac{N+2}{A^{\bar\mu}B^{1-\bar\mu}} = \frac{N+2}{N(a^{\bar\mu}b^{1-\bar\mu})} \approx \frac{1}{\beta}\left(1+\frac{2}{N}\right)$

— is it in the other direction}

$\beta = .56725$ $\beta^{-1} = 1.7629 \approx \sqrt{\pi}$ — so that's about the size.

~~This effect makes Norm. Const. larger so [illegible] G(1) larger.~~

Well, in ST46 I calculated G(R=1). It larger by a factor of ~~[illegible]~~ π^9 than G(R=0)

So its G(R=0) ~~[illegible]~~ should be ~ −452.11 minus 10 or N−492.11 which is what I want.

So it may be O.K. : (But check again!)

Then the Q is: Is 69.31 — 40 correct?

Another check should be a "by hand" calcn. of G(R=1)/G(R=80) — hence i.e. I'd like an expression for it! So far, 'my "hand calcns" of G at R=1 is seem OK.

Actual an expression for G(∞)/G(0) is of more interest! I think (46.28) ≟ 7201 are on this (45.00,25) 70.17ff deals w. Normzn Factor.

(69.31–40) So I really want at least G(0), G(∞), the normz constant (~~[illegible]~~ recall G(0)=1)

Also position, ht., & width of peak. — How they vary w. N, k, dispersion of the "μ_i" & of the "true μ_i". What does G(R) look like for N~500 and k→∞.? This is $R^{\frac{k}{2}}$ ... If I did it for some dispersion of the μ_i, it would be more like the ... of R.

7=11
16 to 5 7=
15 285 48 = 1.46 ℓℓd.

So G as funct of R w. ~~[illegible]~~ $\overset{U}{\text{for}}$ Max value ~~[illegible]~~ value $\bar{U}$:

1. $G(\phi) = 1$;

$A_i \equiv NM_i'$; $B_i = N(1-M_i')$. N = 45, k = 18, $\bar{\mu}$ = .2653889.

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

.03 2. Norman Const: from B.CC, 70.17

$$A_n \equiv \prod_{i=1}^{k} \frac{N+1!}{A_i!\, B_i!} \approx \prod_{i}^{k}\left[f_i(M_i')^{-N} \cdot \left(\frac{N}{2\pi M_i'(1-M_i')}\right)(1+\tfrac{1}{N})\right] \quad (70.23)$$

$$\approx \beta^{-Nk}\left[\prod_{i=1}^{k}\left(\frac{N}{M_i(1-M_i)}\right)\right] e^{\frac{k}{N}}$$

unnecy to use approxs ~~[illegible]~~ $\prod_i \frac{1}{M_i'(1-M_i')}$ is available.

$$\approx \beta^{-Nk}\left(\frac{N}{\bar{\mu}(1-\bar{\mu})2\pi}\right)^{\frac{k}{2}} \cdot e^{\frac{k}{N}}$$

: β = .5672587943 (ST71, Bas, 71.17)

$\bar{\mu}$ = .2653889

$$\approx \left(\beta^{-N}\left(\frac{N}{\bar{\mu}(1-\bar{\mu})2\pi}\right)^{\frac{1}{2}} \cdot e^{\frac{1}{N}}\right)^{k}$$

$\boxed{\frac{N}{\bar{\mu}(1-\bar{\mu})2\pi}} = \frac{45}{1.224954977} = 36.73604$

$\boxed{\sqrt{}} = 6.0610267$

Ln(↑) $\approx \left(-25.51224 + 1.801879 + .0263158\right)\cdot n$

Ln( ) $\approx -(23.6671756)\cdot k$

.14 3) $G(\infty) = A_n \cdot \underbrace{(F_1(\bar{U}))}_{f}{}^{k\cdot N} = A_n\left(\bar{\mu}^{\bar{\mu}}(1-\bar{\mu})^{(1-\bar{\mu})}\right)^{N\cdot k}.$

$F_1(\bar{U})$ = .5606724

$\alpha \equiv \frac{\beta}{F_1(\bar{U})}$ = 1.0117 4218 1.0117 4218

$$= \underbrace{\left(\frac{\overbrace{.5606724}}{\beta}\right)^{Nk}}\left(\left(\frac{N}{\bar{\mu}(1-\bar{\mu})2\pi}\right)^{\frac{1}{2}} \cdot e^{\frac{1}{N}}\right)^{k}$$

$$\left(\left(\frac{1}{\overset{\alpha}{1.01174218}}\right)^{N}\left(\frac{N}{\bar{\mu}(\bar{\mu}-1)2\cdot\pi}\right)^{\frac{1}{2}} \cdot e^{\frac{1}{N}}\right)^{k} ;$$

So want to know how ~~α^{N} compares with~~ $\sqrt{\alpha}^{N}$ compares with

α^{N} compares with $\left(\left(\frac{N}{\bar{\mu}(1-\bar{\mu})2\pi}\right)^{\frac{1}{2}} \cdot e^{\frac{1}{N}}\right)$: The second factor is $\sqrt{}$ = 6.0610267 · $e^{\frac{1}{45}}$ = 6.1972238

for what N ~~[illegible]~~ is G(∞) = 1 ?

$\alpha^{N} = \left(N^{\frac{1}{2}} / \bar{\mu}(1-\bar{\mu})\cdot 2\pi\right) e^{\frac{1}{N}}$

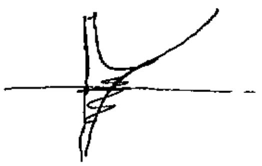
$(\alpha^{2})^{N} = N / \sqrt{\mu(1-\mu)\cdot 2\pi} \cdot e^{\frac{2}{N}}$

Try $A\,\gamma^{N} \equiv N$ $e^{\delta\cdot N + a} = N$. $\delta N + a = \ln N$

$\delta N \equiv N'$ $N = \frac{N'}{\delta}$. $e^{N' + a} = \frac{N'}{\delta}$ $N' + a = \ln N' + \ln\delta$

$N' = \ln N' + \ln\delta - a$ $N' = \ln N' + B$ $X = \ln x + Z$ $e^{x} = x \cdot e^{Z}$

$e^{Z} = \frac{e^{x}}{x}$; so I want the inverse of $\frac{e^{x}}{x}$



Surprized that it is 2 valued!

$\alpha^{2} \equiv e^{D}$ so $\underbrace{D = 2\ln\alpha}_{.02334755}$ $e^{DN} = \frac{N\cdot D}{D\sqrt{\mu(1-\mu)\cdot 2\pi}}$ | $X = DN$: $e^{X} = \frac{X}{D\sqrt{\mu(1-\mu)2\pi}}$

$\underline{D \cdot 1.10677 6819}$ = .025840528

$e^{x} = \frac{x}{.025840528} = x \cdot \underbrace{38.6585}_{Z}$

$e^{x} = x \cdot Z$ (Z ≈ 38.7)

$x = \frac{e^{x}}{Z}$; 5, 3.835 : 028 : .026578 .02653653 54 395 $x = \boxed{.02653 5399}$

is one soln. $\boxed{N = \frac{x}{D} = 1.13}$; This is not the soln. Dream.

$X_{n+1} = \ln x_n + \underset{3.6553}{\ln Z}$ 5 5.26 5.317 267 85 85 66 84 2014

$X \approx 5.3249684$ $\boxed{N = \frac{x}{D} = 228.24}$

N = 228.24 is ≈ the N at which

crossover pt. for the 2 coding methods

$\underline{(43.11-.20\ \text{derbs}\ +302.\text{Bas})}$

64 ← 43.30 (ST 42.Bas)

I then I constructed P.C. of $\frac{N k\bar{\mu}!\ Nk(1-\bar{\mu})!}{Nk+1!}$

with $\prod_{i}\left(\frac{NM_i'!\ N(1-M_i')!}{N+1!}\right)$ for k = 18 & various N values.

3699 SMRT STEIN

Look at ~~#~~ ST42.B23 More carefully: That 228 ≈ 201 is prob'y no coincidence! —

~~E~~ See what ~~ST42.6~~ does:

.03 Basic Eq: $\boxed{\frac{x!\,y!}{x+y\,!} \approx (a^a\, b^b)^R \cdot \sqrt{2\pi\, a b R} \qquad a = \frac{x}{R} \quad b = \frac{y}{R} \quad : R \equiv a+b.}$

check $\frac{4!\,5!}{10!} \approx (.4^{.4}\; .6^{.6})^{10} \cdot \sqrt{2\pi \cdot 4 \times .6} = 4.63813\text{E-}3$

$\parallel$ | ~~1.1958E-3~~ | 1.19 4396E-3

4.762E-3 .736 ok.

.693

.51066

so $G'' = (F_1(\bar{u}))^{NK} \cdot \sqrt{2\pi\, \bar{u}(1-\bar{u}) \cdot N \cdot K} \;/(NK+1)$

(.10) A_n $G'' = \prod_i \left[(F(u_i'))^N \sqrt{2\pi\, u_i'(1-u_i') \cdot N} \;/(N+1)\right] \equiv$ (Norman constant of 73.03)$^{-1}$

.11 Hvr, Value of $G(\infty)$ is $A_n \cdot (F_1(\bar{u}))^{NK}$ or $\frac{F_1(\bar{u})^{NK}}{(.10)}$

In ~~[illegible]~~ ST42B23 We find N ∋ $\frac{F_1(\bar{u})^{NK}}{.10} \cdot \frac{\sqrt{2\pi\,\bar{u}(1-\bar{u}) \cdot N \cdot K}}{NK+1} = 1.$

In 73.14–.40, we find N ∋ (.11) = 1

That the 2 results should be close, means that the relative values of the 2 expressions change very rapidly w. N.

.20 Anyway as $N \gg 228$, $G(\infty)$ gets smaller as $\sqrt{N} \cdot \alpha^{-N}$ $\alpha = 1.0117428$

each ↑ of N by 85 ~~w~~ make $G(\infty)$ smaller $\approx \sqrt{N}\; e^{\frac{-N}{85.16}}$

.23 by ~ × e^{-1}. (since $\sqrt{N}$ doesn't ~~[illegible]~~ ↑ much.

T. discuss of .20–.23 are for the effect of N only: Actually the effect is $\left(\sqrt{N}\; e^{\frac{-N}{85.16}}\right)^K$

So for N > 238 (say N=300), ~~#~~ increases in K become very important in ↓ $G(R=\infty)$.

Next, get peak of $G(R)$ curve & perhaps its height at peak

.27 $\left(\begin{matrix}46.28\\ \text{So, unnormed!}\end{matrix}\right) \left(\frac{x+y+1\,!}{x!\,y!}\right)^N \cdot \prod \frac{x+A_i!\; y+B_i!}{x+y+N+1\,!}$ · use $x = \bar{u}R \quad y = (1-\bar{u})R$

from .03: $(R+1) \cdot F_1(\bar{u})^{-R} \cdot (2\pi\,\bar{u}(1-\bar{u})R)^{-\frac{1}{2}} \cdot \underbrace{\prod_{i=1}^{K} \frac{(R\bar{u} + N u_i')!\; R(1-\bar{u}) + N(1-u_i')\,!}{R+N+1\,!}}_{\equiv\; \overline{\prod}}$

$\overline{\prod} =$ [illegible] $\prod_i \left[F_1(a_i)(2\pi\, a_i b_i (R+N+1))^{\frac{1}{2}}\right]$ $a_i \equiv \frac{R\bar{u} + N u_i'}{R+N+1}$; $b_i \equiv 1-a_i$.

.27 ~~th~~ doesn't look promising!

Think about Apriori for R: Obtained by considering Many Previous years' data.

Hvr, note that Unless $N \gtrsim 238$, $G(\infty)$ will be >1

Try d.f. ~~of [illegible]~~ G(R) for N→∞! ~~I [illegible]~~ I did ≈ this at . ST46.8: See results 47.6–.20

Shape of d.f. becomes ~~[illegible]~~ indep of N as N→∞.

46.28 and 74.03: $\frac{x+y+1}{x!\,y!} = (F_1(\bar{u}))^{-R}\,(2\pi\,\bar{u}(1-\bar{u})R)^{-\frac{1}{2}} \cdot (R+1)$

.37 $\frac{x+Nu_i'!\;\; y+N(1-u_i')!}{R+N+1\,!} = (F_1(a_i'))^{R+N+1} \cdot \sqrt{\frac{2\pi\, a_i b_i}{R+N}}$ ~~[illegible]~~ $\frac{1}{1+\frac{1}{R+N}}$ ~~[illegible]~~

.38 So it looks ~~[illegible]~~ like as a funct of R $\underbrace{\left(\frac{\prod F_1(a_i)}{F_1(\bar{u})^K}\right)^R}_{\equiv\, \alpha^K} \cdot \frac{\sqrt{R}}{(\sqrt{R+N})^K}$

.37 is being cancelled by the Norman constant 73.03

For any large (but finite) N, in 74.37 For R<<N, we have $\prod F_1(a_i)^{R+N+1}$ ← may cancel w. numerators.

• Then $\prod_i F_1(a_i)$ first is > $F_1(\bar{\mu})^k$ and causes G(R) to ↑ w. R.

as R >> N; This factor becomes $\prod F_1(\bar{\mu}) = F_1(\bar{\mu})^k$ and so G(R) starts to ↓.

Unfortly ~~this~~ in 47.10–.20, This transition didn't occur when R became > N,

but at a fixed pt, of $\overset{R}{}\sim 40$. for N = ~~6~~00 ~~thru~~ 1200

74.37 after multg Normaz const. (73.03):

$$\underbrace{F_1(\mu_i)^{-N}\cdot\sqrt{N}\cdot F_1(\mu_i)^{R+N+1}}\cdot\frac{1}{\sqrt{R+N}}$$

$$=\prod_i\Big[F_1(\mu_i')^{R}\sqrt{\frac{R}{R+N}}\Big]\cdot F_1(\bar{\mu})^{-R}\cdot\sqrt{R}$$

$$\underset{>1}{(\alpha)}^{kR}\left(\sqrt{\frac{R}{R+N}}\right)^{k}$$ which → ∞ as R → ∞. ∴ wrong!

Look at 46.28! $\frac{x+y+1}{x!\,y!}\rightarrow F_1(\bar{\mu})^{-R}\cdot\sqrt{R}$

$\frac{(x+A_0)!\,(y+B_0)!}{(x+y+N+1)!}\rightarrow F_1(\bar{\mu})^{R+N}\cdot\frac{1}{\sqrt{R+N(\;)}}$ } Product is indep of R at R→∞, ~~&~~ its → 1.

.20 Look at 74.38! $\left(\frac{\prod_{i=1}^{k}F_1(\mu_i')}{F_1(\bar{\mu})^{k}}\right)^{R}\cdot\left(\sqrt{\frac{R}{R+N}}\right)^{k}$

T. first factor becomes close to 1 as R gets > N. so ↑ slows down.
i.e. the μ_i''s → $\bar{\mu}$.

T. factor $\left(\frac{R}{R+N}\right)^{k}$ decreases more slowly as R gets > N. : It could be that these

.23 effects result in a peak that's indep of N for large enuf N.

I could just write a pgm to do it! — its a bit more complex hvr. 1.6

Instead of R+N its $\frac{R+N}{\mu_i'(1-\mu_i')}$ These "μ_i" are really $\frac{N\mu_i'+R\bar{\mu}}{N+R}$

This xfm should be about what ST 46.R does for large N.

Study t. 47.10–.20 d.f. See fits ~ lognormal. It seems ≈ lognormal.

Do a run from JJ=0 to 22 R = 20 * 2 ^ (JJ /10) $2^{.1} = 1.07$.

— Trouble is t. value of $\bar{R}$ that ends up w. will probly be much dependent on t. params I use on t. apriori. — but take a look) T. fact that its broad & lognormal, biases it much up by ~~R~~ size also t. D.F. in 47.10–.17 (for N=45) definitely pushes R higher.

SN I'm mainly concernd w. how narrow t. R d.f. is ? NO!

SN The width of the x, y d.f. can't be infinitely narrow since we don't have enuf ssz for $\bar{\mu}$. The ssz for $\bar{\mu}$ is 810.

If one had only 1 "at Bat" for each of k players, T. best est. would be
$\mu_i = \bar{\mu}$ ($\bar{\mu}$ = average for these k data pts). Say ssz → 2 (2k) ← what then?

→ So a poss'l way to get a reasonable apriori; Use large N; assume a log normal d.f., Then ~~get expected value of~~ mult ~~through~~ Then use this a prio. w. t. real data i get E & K and t. ~~expected value for each~~ probty that each player will hit at next bat.

This "aprio." is a way to simulate data from previous yrs.

— Trouble is, its not a real apriori.

3799 SMRT STEIN

(Well, it is an a priori in the sense that I more or less remember previous yrs being pretty much like this once

A ≈ "True" ~~a priori~~ a priori would use ~~[illegible]~~ large no. of previous yrs, plus some a priori.

The weirdness is that as we ↑ amt. of previous yrs' data, the import of this ⟶ a priori ↓ into nothingness (or at least it should!).

(5N) A weirdness! 47.18 – .20: Seems to be a peak at $U = .255$! ? !

Check this at double precision, w. good value for π.

{ The result of 23.14 – .40 and 74.20: That N has to be > 228 before we can get much ↓ of σ with R — No matter how large K is! } Very unreasonable

50 0278

0292

100 0224

What value of R gave best predns for rest of years? Use MS error & pc error

See how broad this R is. Wrt. pc distribution for various k; what is the meaning of the "width"! Is a factor of e, say in pc, the width? — How is the "width" [illegible] "usable"? It can have a ~~[illegible]~~ use in parallel codes:

so a ↓ of pc by e would (

18
54

2161

300 2156

400 2192

5
36

313

(3899) A New approaches:

① for conceptual debugging: write MCarlo pgm to produce data sets $[u_i]$ as a funct of $\bar{u}$ R, N, k. Should be easy to do: I can then use this ~~[illegible]~~ generator to test various ideas I have on statistics of the set $[u_i]$

② Compute σ^2 theoretical of $[U_i]$ as funct of $\bar{u}$ R, N, K [I think it's mediocre]

.25 }
.25 } ?

It is $\sigma^2_{\bar{\mu}, R} + \frac{1}{k}\sum_{i=1}^{k}\sigma_i^2$: The second term is $\underset{\text{normaz.}}{A_n} \int_0^0 u^x(1-u)^y \frac{u(1-u)}{N}\,du = \frac{x+1!\,y+1!}{R+3!\,N} \cdot \frac{\overbrace{R+1!}^{\equiv A_n}}{x!\,y!}$

$$\boxed{= \frac{(x+1)(y+1)}{(R+2)(R+3)N}} \approx \frac{\bar{U}(1-\bar{U})}{N}. \qquad \sigma^2_{\bar{\mu}, R} \approx \frac{\bar{U}(1-\bar{U})}{R}$$

~~[illegible]~~ LATER: 5.5.99: I don't understand this ↑ argt! — I got the eqn of a fin. dif. later; maybe same ~~[illegible]~~ but much more complicated derivation.

See (44.15–.29) for ~~[illegible]~~ an derivation from STEIN2.BAS

Get errors in various models.

Also get P, Q

$\frac{1}{1-p-q}$

So ~~total~~ $\sigma^2_{\text{total}} = \bar{U}(1-\bar{U})\left(\frac{1}{N} + \frac{1}{R}\right)$.

ST76.BAS

So got σ^2_{Tot}; ~~[illegible]~~ $NN = \frac{\bar{U}(1-\bar{U})}{\sigma^2_{Tot}}$; ~~[illegible]~~ = equivt R_T : $\frac{N \cdot RN}{N - RN} \approx R$. $\overset{=}{\cancel{[illegible]}} NN$

I may have done ≈ this. $\overset{5.8}{4.583889\text{E-3}}$; $\overset{NN}{\cancel{[illegible]}\ 42.53105}$ $\overset{R}{\cancel{[illegible]}\ 225.183}$!

$$\frac{1}{NN} = \frac{1}{N} + \frac{1}{R}$$

42 45

I got ~~5~~775.183 NN = 42.53.

It may be that my approxs. $\frac{U(1-U)}{N}$ etc. for Vave are inaccurate for coarse ~~[illegible]~~ dpc. (.256) is exact.

ST76A.BAS ww

(24)

~~A~~ Note: ~~[illegible]~~ Gave of ST2 R(75) is not the best one to use!

SS ~~[illegible]~~ R=200

Strange! When I did R(≡ wax) = 10⁶, I got $\overset{\Sigma}{\text{ms. error of}}$ .0242535322 ″ .0242535322

The ST paper got .075374, ~~see~~ My SIG x 18 = .0825!

For R = ~~75~~ error 13% same as R = ∞: Error = .02~~4~~ 28

.37

So [illegible] ST paper used ~~a~~ ~~R~~ 313. ; R = 152 is best. R = ∞ is not bad.

Serr = .0216 Serr = .02126 Serr = .02425

For R = 275 Serr = .02278

21	.021611
200	.02127277
190	.02126926
150	.02127 7
175	.021286
192.	.02126912 ← or (28)

SS1: err in all

SS2: err in (A7) Stein

SS3: error in my approxn.

~~[illegible]~~ SS1 = error π_i as calculated ~~[illegible]~~ V

SS2 error w/ ~~[illegible]~~ Stein

SS3 error ~~[illegible]~~ calc. of V

SS4 used → $\bar{U}$

3899 SMFT Stein

In ST76.Bas (& 76.25R), I found a $R, \bar{\mu}$ s.t. expected μ & σ^2 of the $[\mu_i]$ set produced, was that observed in $[\mu_i^o]$. This may make the $R, \bar{\mu}$ most likely generators of $\Sigma[\mu_i^o]$, but that's not usually what I'm after. I'd like more, the d.f. of $R, \bar{\mu}$. That I got a peak in ST76.B. (I guess a peak) is weird! It certainly was nowhere near the peak in ST46B!

ST46.B finds the D.F. of (μ, R) that would produce the observed $\Sigma[\mu_i^o]$;
Here, I want the D.F. of (μ, R) that produces a $\Sigma[\mu_i^o]$ set of the observed σ^2, only.
ST46.B is more like ALP. ST76.B is more like conventional Statistics.
~~For (example), there is more certainty of the σ^2~~ in ST76. So I think to do it write in ST76.B, / the knowledge of / the would have to be included.

ST 77A.Bas

Anyway, the next pgm is one that generates data sets for arby $\bar{\mu}, R, N, k$.

```
DEFDBL A-Z
Dim A(~~[illegible]~~, 3)
R=  u=  k=  N= .
For JJ=1 to k
X=0
For J=1 to R
If RND ~~[illegible]~~ < M Then X=X+1     0<Rnd<1
Next J
~~A(JJ, 3) = X/R~~
~~A next [illegible]~~
                    Y=0
                    For L=1 to N :
                    If RND < X then y=y+1
                    Next L
                A(JJ, 3) = Y/L
    Next JJ
```

This pgm. will at least check my Algebra!

ST77B.Bas Modification of ST76.Bas Instead of 76.25, averaging over all possible $\bar{\mu}$ getting

(75.25R) Use mean $\frac{\overline{\mu_i^o(1-\mu_i^o)}}{N}$; so $\frac{\overline{\mu_i^o} - \overline{(\mu_i^o)^2}}{N}$ instead of $\frac{\bar{\mu}(1-\bar{\mu})}{N}$.

Do I know $\bar{\mu}$ consistent w. ssz of 775? Barely?

The difference of $\frac{\overline{(\mu_i^o)^2} - \bar{\mu}^2}{R}$ would seem to be quite large!

Now I want $\sigma^2 = \frac{\bar{U}(1-\bar{U})}{R} + \frac{\frac{SS}{k} - U^2}{N}$; AA = SIG ~~[illegible]~~

$R(R \equiv \frac{\bar{U}(1-\bar{U})}{R}$ $(U - SS/k)/N$

$= SIG - U/N + SS/k/N$

$AA = \overline{SS/k - U^2} - U/N + SS/k/N.$

$\frac{\bar{U}(1-\bar{U})}{R} = AA = \left(\sigma^2 - \frac{\frac{SS}{k} - U^2}{N}\right)$: $R = \frac{U(1-U)}{AA}$

I got R = 551.69 (instead of 775) : Some improvement!

R	Σerr²	
775	.02273	
552	.02357	
192	.02126	Bush
313	.02216	Stein

Using ST77A, I can see if This pgm or ST76B.
② Gives ~~better~~ R on the average
⑥ which one has more variance in R.

$$\frac{1}{k}\left(\Sigma(\mu_i^o - \bar{U}\right) = \frac{\bar{U}(1-\bar{U})}{R} + \frac{\frac{1}{k}\Sigma \mu_i^o(1-\mu_i^o)}{N}$$

$$\frac{\bar{U}(1-\bar{U})}{R} = \underbrace{\frac{1}{k}\Sigma\mu_i^{o2} - \bar{U}^2}_{SIG} - \left(\frac{\bar{U} - \frac{1}{k}\Sigma\mu_i^{o2}}{N}\right)$$

$- \bar{U}/N +$

✓

3999 SMFT **STEIN** $\boxed{\text{INDEX}}$ to pgms! ((.28 – .40)R) ST50.Bx thr ST72B.bas
11 pgms → $\boxed{\text{go for next index}}$ $\boxed{\text{→ Also 80}}$

Re: Apriori of R: Its not possl to have R<0 in some cases; (Re In baseball it seems unlikely!). Anyway, This suggests we ~~would~~ may want $\frac{1}{R}$ to be smooth ~~to~~ ~~to~~ ~~be~~ ~~say~~ around zero. So try apriori of $\frac{1}{R^2}$ for $1<R<\infty$

In such a case, ~~the~~ t. first moment of t. product d.f. would be ∞.

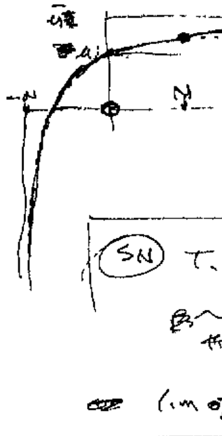
AbCdefg

[But Arr, I'm not interested in this: I want t. ~~exp~~ ratio of t. expected values of t pc. for all R, all $\bar{\mu}$, all $[\mu_i]$, all $[\mu_i]$ w. μ_i' omitted $\left|\begin{matrix} N=N \\ \mu_i'=\mu_i' \end{matrix}\right)$ v.s. $\begin{matrix} N=N+1 \\ \mu_i'=\mu_i'+1 \end{matrix}$;

In ~ST46.B, I get t. D.f. for R from ~1 to +∞: Actually, t. values could be anywhere from −∞ to +∞. Large neg values (like −1000) means $\frac{R\bar{\mu}+\mu_i'N}{R+N} \approx \bar{\mu}-\frac{N}{R}\mu_i'$. Actually, I don't see How this could occur. (I'm t. "true" R should always be >0.) Tho an "apparent" large negative R could be t. result of taking ~~diff~~ t. reciprocal of t. difference betw. 2 large, noisy quantitys.

$$\equiv\left(\bar{\mu}+\frac{\mu_i'N}{R}\right)\Big/\left(1+\frac{N}{R}\right)$$

$$\approx \bar{\mu}+\left(\mu_i'-\bar{\mu}\right)\frac{N}{R}-\left(\frac{N}{R}\right)^2\mu_i'$$



This suggests that t. selection soln. converges More rapidly w. R→∞, than expected —

It may be possl to use $\frac{1}{R}$ as t. density (No!) Multiplying by $\frac{1}{R}$ → $\frac{1}{R^2}$ which does converge.

.20 (SN) T. way $\frac{1}{R}$ is used as density: $\lim_{R\to\infty}\int_1^R \frac{1}{R}\cdot f(R)\,dR \Big/ \int_1^z \frac{1}{R}\,dR$

~~So f(∞) must → 0.~~ ~~then~~ $\lim_{R\to\infty} f(R)$ must → 0 If ~~f(R→∞)~~ if $\lim_{R\to\infty} f(R)$ is some constant $\alpha\neq 0$:

~~so~~ lim of .20 → α (Woops!) ~~wont~~ ~~→ 0~~ if α=0, t. limit is φ. doesn't work!

try apriori $\frac{1}{R}\cdot\frac{1}{(\ln R)^2}$ ← diverges at R=1. Try range [2, ∞]

ABCDEFGH
ijklmnopqrst

$\boxed{\int_\alpha^\infty \frac{1}{R}\frac{1}{\ln R^2}dR = \frac{1}{\ln\alpha}}$: $\frac{d}{dR}\frac{1}{(\ln R)} = -\frac{1}{(\ln R)^2}\cdot\frac{1}{R}$

So ln α is t. normaliz. constant. T. larger α is, t. less it is sensitive to α.) ?

Anyway, try it for various α values. Start w. α ~ 2.

.28 Well, O.K.: Try integration of ST46. pgm; from α to β: first Mult by $\frac{R\bar{U}+\mu_iN}{R+N}$ for $\mu_i' = .2$ say.

Also use Normzn factors: $\frac{1}{R}\frac{1}{(\ln R)^2}$ and $\frac{1}{R^2}$ >

No! [& simultaneously integrate t. normzn factors to get normzn const. (tho in both cases, t. ∫ is computable, so I can check.)

$$\int \frac{1}{R}\frac{1}{(\ln R)^2} = \frac{-1}{\ln R} \quad ; \quad \int\frac{1}{R^2} = -\frac{1}{R}$$

If ∫ takes too long, approximate $\frac{G}{R^2}$ by a log normal d.f. as on 50¢.1

$e^G\cdot\frac{1}{x^2}$ or $e^G\frac{1}{x}\frac{1}{(\ln x)^2}$ are wtg factors that have to be integrated to get t. normzn factor for

e.g. $\int_\alpha^\beta e^G\cdot\frac{1}{x^2}\cdot\frac{R\bar{U}+\mu_iN}{R+N}\,dR$.

Use α = 10 or 20; use β = 1000 or 2000
But Note G is constant for R ≥ 200 → $\boxed{80.01}$

STEIN50.Bas
Gives $\frac{e^G}{R^2}$ or G − 2 ln R

ST61 finds out slope of ~~a~~ e^G on log log paper.

ST63: finds ~~S~~ $\frac{1}{18}\sum_{i=1}^{18}\frac{(\mu_i'-\bar{\mu})^2 95}{\mu_i'(1-\mu_i')}$ = mean ratio of σ's
Got 1.09104

Gaussian D.f.'s used.

ST64. This is for 62.03–.07
Pc of corpus as a funct of values of $\bar{\mu}$, μ_i' for 0<A<1
$A\bar{\mu}+(1-A)\mu_i'$

ST67 I don't get this!
ST68 Integrates ST67 for a range of S values from 0 to 1

? ST76A Small Modifn of ST76 (Done after ST72B)
ST77A MonteCarlo Data Generator
ST77B Modifn of ST76

ST71 computes $\left(\prod_i \mu_i(1-\mu_i)\right)^{\frac{1}{n}}$ and $\bar{U}^{\bar{U}}(1-\bar{U})^{(1-\bar{U})}$ & gets ratio

ST76: latest very approx way to get R.

3.10.99 SMPT **STEIN**

Thots on APRIPD

I) In Most cases One can chose an apripd quite easily, and th. amt. of damage done by non-optimum choices minimal.

— — In others, it is not so easy, & t. choice can strongly influence t. final apsip. Suggests Suggests

Hvr. no matter how much data there is, its always possl. to pick an apriy So bad, that the apsip is terribly wrong!

If t. apripd is Universal, hvr, using enuf data will eventually patch things up.

— This is not necly true of non-universal apripds.

III In t. case of STEIN, t. apriy seems very impt, critical, & non-obvious:

T. apripd is for t. d.f. of t. μ_i. Up to now, I've been assuming that it is of t. form $\mu_i^{R\bar u}(1-\mu_i)^{R(1-\bar u)}$, w. some ~~—~~ d.f. for $\bar u$ & R, ~~=~~ & that ~~[illegible]~~ t. d.f. for $R, \bar u$ is of t. form $A(R)\cdot B(\bar u)$. (i.e. ~~#~~ R, u indep.)

In fact, our apri info is that: μ_i is usually sharply limited to $.1 < u_i < .4$.

The mean of t. d.f. is ~~mm~~ more sharply limited to .2 to .3. ok.

R must be large enuf so $\sigma^2 = \frac{\bar u(1-\bar u)}{R}$ wouldn't let μ_i get out of t. $(.1, .4)$ range.

.17 So $\sigma < .1$ $\sigma^2 < .01$ $\frac{.2}{R} < .01$ $(\cong .25(1-.25))$ $R > \frac{.2}{.01} = 20$; so $R \geq 20$:

.18 Also R can't be too large & since $\bar u$ is uncertain to $\frac{\bar u(1-\bar u)}{KN}$ ← = total ssz.

.19 So probably R can't be $> KN$ (= 810 in this case)

.20 Using the limits of 20 & 810 (on R), perhaps I could do a uniform d.f. on R & get reasonable results

3199 SMFT STEIN

INDEX (also see 78 & 90)

.01 (spec 78.40) Hvr, $\frac{RD+\mu_i N}{R+N} \cdot \left(\frac{1}{R^2} \text{ or } \frac{1}{R}\cdot\frac{1}{(\ln R)^2}\right)$ may not be easily integrated.

Anyway, see how long it takes to do integration: I really don't know how often I'll need to do it! Also try 79.20 (using $\int_{20}^{810}$ and a uniform d.f.)

The new encouragement, is that the thing I have to integrate, looks like $\frac{1}{R}$ for large ~~E~~ R (i.e. R >> N). This does reduce importance of apriori dist by R. — but see what I get! Also, if I only need consider R > 20 (79.11–17) Then this can also help a lot!

If it looks like its working, try some Mt. Carlo generated data sets & see how well it works! (I may want to speed up the function of ~~E~~ ST46.BAS: see if double prec is necc; Many places I can speed things up.

Also, upper bnd for R can be K·N (= 810 in present case) (79.18–19).

Anyway, if this works, write a long review of stuff like 79.01 ff & try to see how I could have shortened the long discovery process!

Also: ~~E~~ Try to get some shorter approxns. to the optimum results.

one: $\sigma_T^2 - \overline{\sigma_i^2} = \left(\frac{1}{k}\Sigma\mu_i^2\right) - \bar{\mu}^2 \; \underline{-\frac{1}{k}\Sigma\,\mu_i(1-\mu_i)} = -\bar{\mu} + \frac{1}{k}\Sigma\mu_i^2$

$= 2\frac{1}{k}\Sigma\mu_i^2 - \bar{\mu}^2 - \bar{\mu}$. from modifn of ST 27B!

$2 \quad -.185789$

$\left(\frac{1}{k}\Sigma\mu_i'^2 - \bar{v}^2\right) - \frac{1}{k}\left(\Sigma\,\mu_i'(1-\mu_i')\right)$ — should be divided by N

$\underbrace{-\frac{1}{k}\Sigma\mu_i'}_{\bar{v}} + \frac{1}{k}\Sigma\mu_i'^2 = \frac{2}{k}\Sigma\mu_i'^2 - \bar{v}^2 - \bar{v}$

Modfn of ST27B.BAS 20 $\boxed{SIG - v(1-v)/(k)/N}$: No (SIG = 4.5839 E-3) (= σ_T^2) (error this is SIG!)

$\frac{1}{\sigma_T^2} = 218.$ So I think this is "it" "!

→ NO!

SIG − v(1−v)/N

I got 2.51512 E−4 = σ_T^2

$\frac{1}{\sigma_T^2}$ = 3975.9 (400)

From 76.37: R = 192 has least apsi of error = .02126

Stein gave R = 313 = .0216

So this is closer to Minimum.

~~I could try this w. my Mt Carlo data.~~

So t. R = 218 ~~was ≈ $\frac{1}{\sigma_T^2}$!~~

$\left(\frac{\bar{\mu}(1-\bar{\mu})}{N}\right) = 4.3323918$ E−3

$(\quad)^{-1} = 230.81938$

$\boxed{\begin{matrix}4.583904\text{E-}3\\ SIG =\end{matrix}} \quad \overbrace{\frac{\bar{v}(1-\bar{v})}{R}}^{3.2076514} + \overbrace{\frac{1}{k}\Sigma\frac{v_i'(1-v_i')}{N}}^{.0013762526}$ from ST.B so $\underline{R} = 311.75458 * \frac{\bar{v}(1-\bar{v})}{}$ = 60.760968

.342, .174 / .1949

SIG = $\frac{1}{k}\Sigma(\mu_i - \bar{v})^2$ = observed var. (perhaps should be ~~[illegible]~~ by (k−2)?

.074017 "Sig" using $\frac{1}{k-2}$ → 5.1568875 E−3 → R = 242.67

Perhaps SIG = $\frac{\frac{1}{k}\Sigma v_i(1-v_i')}{R} + \frac{\frac{1}{k}\Sigma\,u_i'(1-u_i')}{N} = \frac{1}{k}\Sigma\mu_i'(1-\mu_i')\left(\frac{1}{R}\;\frac{1}{N}\right)$

gives $\frac{1}{R} = \boxed{\frac{\overset{.0740158}{SIG}}{\frac{1}{k}\Sigma\mu_i(1-\mu_i')}} - \frac{1}{N}$ = .0517935 : R = 19.3 not so good.

Needed for ~~cont~~ constants:

$\bar{\mu}$, ~~E~~ $\frac{1}{k}\Sigma\mu_i$

(Fewer sums of μ_i', $1-\mu_i'$)

α, β, etc.

Step 1 9.07 } 39.07

2 21.13

Calculates P, Q

Step 7 {29.14 / 32.05 maybe / 34.27 random

9 34.32

10 35.17

5 27.4 more [illegible]

27.01

25.01 !

Step 1/20 on convergence of approxn.

2 1/31 data compute P, Q

3 1/13 So i... etc Get 1/(1−P−Q) P, Q,

4 2/4

5 2/5

6 2/5

7 2/12

8 2/07

9 2/14

10 2/09

11 2/10

40 2/12

41 2/12

42 2/16

46 2/14

So 2/20 : 2/30

3.11.99 SMPT STEIN:

Some Constants of int:

$\bar{u} = .2653889$

$\bar{u}(1-\bar{u}) =$ [illegible] .1949

$\frac{\bar{u}(1-\bar{u})}{N} =$ [illegible] 4.3329318 E-3 (recip = 230.81938

.07 $\frac{1}{k}\sum u_i'(1-u_i') = \bar{u} - \frac{1}{k}\sum u_i'^2 = (.0619314)$? Enowpet .19037374 $= (.27395547)^2$

$\frac{1}{k}\sum u_i'^2 = .07501516$ — (?? $\frac{1}{k}\sum u_i' = \bar{u} =$ ~~.2653889~~

$\frac{\bar{u} - \frac{1}{k}\sum u_i'^2}{45} = .0013762526$ | reciprocal = 726.61077

.5 $\boxed{\text{ST81.Bas}}$: used to find values of these constants.

.13 $S'G = \left(\frac{1}{k}\sum u_i^2\right) - \bar{u}^2 = \frac{1}{k}\sum(u_i' - \bar{u}) =$ 4.583904 E-3 (recip = 218.15485) for.

$\frac{1}{k}\sum u_i'^2 = .07501516$

.17 $\boxed{\text{ST81A.Bas}}$: Interation of STEM46 w. $\left(\frac{1}{R^2}, \frac{1}{R(\ln R)^2}, \frac{1}{R}\right)$ ← Apriori $\left(\frac{R\bar{u} + N\overset{uu}{u_i'}}{R+N}\right)$

Range R = 10 to 800 uu = .2153889 ($= \bar{u} - .05$)

Apriori: $\frac{1}{R^2}$ first. R = 10 to 800

I Got S1 = 4.58449 E-206 S2/S1 = .5135676 ! unlikely Pr's is never near .5.

S2 = 2.354447 E-206

try from 700 to 800. Got S2/S1 = .307 again impossl.

S1, S2, ( ) .3605

Print R, S1, S2

Print S1, S2, S2/S1

.28 for $\int_{10}^{800}$: I Got .265 u, .248447 = α, .215 uu

$\alpha = \frac{R \cdot u + N \cdot uu}{R + N}$ (45) 26 21 25

$\alpha R - uR = N \cdot uu - \alpha N$

$\to \alpha R + \alpha N = Ru + N \cdot uu$

$(\alpha - u)R = N \cdot uu - \alpha u$

$R = \frac{N \cdot uu - \alpha N}{\alpha - u}$

$= N \frac{uu - \alpha}{\alpha - u}$

45. × .215 − .248447 .2653889

I Got R = 89: Not so good!

Integrate from 20 to 500: R = 93.11

20 to 2000 R = 99

R = 20 to 800, $\frac{1}{R(\ln R)^2}$! R = 137.03,

R = 20 to 2000 R = 167.7

R = 20 to 800 : $\frac{1}{R}$ R = 181.9

R = 20 to 2000 $\frac{1}{R}$ R = 249.1

R = 20 to 810 $\frac{1}{R(\ln R)^2}$ R = 137.47

R = 20 to 5000 " R = 193

39 R = 20 to 10,000 " R = 210

R 20 100,000 " R = 252

(R = 92 is best post hoc soln. (76.37))

R = 20 to 810 : $\frac{1}{R(\ln R)^{1.5}}$: R = 147

SS3 error R 137 .02537

192 .02126 (Best)

313 .0216 Stein

Try new, uu = .4 R = 137.47

Result same as uu = .215 to 14 digits precision!

This indicates that it's a linear effect.

3 1199 START STEIN

Since the results are indep of UU, let UU=0.

$\boxed{R \cdot U / (R+N)}$ So final P.O. is S1, S2, S2/S1, N*S2/S1 /(S2/S1+U) = R value

A short cut would be to find the peak of (G−2ln R) − or ... /(R ln²R) then find its second deriv. int. the domain.

.04 Actually we can simplify it further by letting U=1 in the update of S2 and in the final P.O. So S1 = S1+(GD); $\boxed{S2 = FGD*R/(R+N)}$ Result RR = N*S2/S1/(S2/S1−1)

So we want the expected value of $\boxed{\frac{R}{R+N}}$; this ends up as $\frac{S2}{S1}$ = α ; and $\boxed{RR = \frac{N\alpha}{\alpha-1} = \frac{N\alpha}{1-\alpha}}$

Something wrong: $\frac{R}{R+N}$ is always <1 so α < 1, so RR would always be <0.)

I'm getting neg. R but right size!

$\alpha = \frac{R}{R+N}$: αR+αN = R ; αR − R = −αN ; R(α−1) = −αN ; $R = \frac{-\alpha N}{\alpha-1}$

I could see if I could simulate the STEIN result using the Sm−1 x term. Also look at their methods in which σ_i^2 are not all the same!

Note! I really can't tell if my method is better than Stein's, from just this one case. — So Get Stein result for my M. Carlo data & compare w. my results.

Note on Integration in 81A.Bas!

Changing integration step from ~~...~~ ΔR=1 to ΔR=10 ($\int_{20}^{810} dR$) takes RR from 137.47 (ΔR=1) to 135.57: the amt. of the M.S. error is negligible.

ΔR=2! 137.236 ; | ΔR=20 133.7 | ΔR=40; 131.6 | ΔR=100: 124.97

ΔR=.5 ! 137.600

Try $\int_{20}^{1M}$ w. ΔR=10 $\int_{0}^{100k}$ was only 252 : still better than Stein.

(SN) Expect mean of ~~data~~ the set ≠ mean of test set! A better test set would have same mean:

expected error in $\bar{u}^2$ $u = \frac{.265(1-.265)}{45}$ $\frac{.194775}{5/0}$ = ~~...~~ ; √ = .0657 ; √ = .0155

change in Σ err² ≈ 18 × err² = .0043! not much!

ST82 Try it anyway: From ST77B — A(5,2) take out A(X,5) s+2. } This finds $\bar{u}$ of test data.

S5 = S5 + V*VV

S6 = S6 + VV

Print "Dat=" S5/S6 : $\bar{U}$ = .2653|85888 ; U_r = .2744|888

Δ = .0091 Σ = the u of test v.s. trg. data.

ST82A old $\bar{u}$: .2653889 ; Σ err²

.0212692	R=192
.0229743	192
.0228869	175
.022869	R=192
.029144	225
.024059	100
.022986	250
.023164	300

new $\bar{U}$ = " +.0091

R=100 200 R=300

Stein error: .021611 better than Stein by .0051)

" : .02239178 ! better than Stein by .0091!

Nature.com search indexes
CAZHV
MAENA
Fab(issue

CAZHV
MAENA
MAEVA

FSMKX
Stanley
Mkt index
Inv. Grade Bonds
(FBNDX)
(Fabl) Nature Letters.
ISSUE INDEX by date
Saved and Alerts
Nature.com
661
9650
D. Sackel
@ Bowery

31299 SMRT STEIN

83

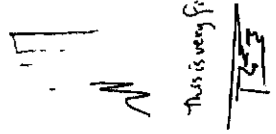
So, I ↑ $\bar{U}$ of data by .0091, so it was ≡ $\bar{\mu}$ of test set.

I also ↑ $\bar{U}$ of stein by .0091.

T. result was err of .021611 → .02239178 (↑ of error in Stein)

But err2 in using $(R\cdot\bar{U}+N\cdot Z)/(R+N)$ as approxm to test data

Seems to have min error of ≈ .022869 at R=192 ! Seems unlikely; its > Stein error!



Seems that moving the mean to the Test mean would ↓ ms error! but it increases it.

This could have to do w. unequal weightings in finding mean

King Shildus

When I use wtd error (as time of But), error for R=192 is < that of STEIN.

R	err2	~~err≤stein~~
182	6.10	
192	6.0932	6.2466 : corresponds to → so R=∞ (using $\bar{U}$ as μ_i) is better than STEIN in this case.
212	5.97788	After I used wtd mean Σ err2.
225	5.9406	
250	5.8872	
300	5.8257	
400	5.745122	
450	5.79536	
600	5.815	
425	5.794460	→ Max pc of ops data is for R≈410 (see STEIN4.BAS)
2000	5.958	
10 K	6.04458	
137	6.468	worse than Stein.
160	6.29345	— stein = 6.2466
∞	6.069885	← U used as approxm of μ_i.

– 10 ST82A output.

.18 ≈ min — 425

– 20 → ∞

ST83 Modify ST 81A so $\bar{U}$ = U + DEL, Z = Z + Del (Del = .0091)

~~Woops! ST81A looks very wrong! forgot to divide by 1000~~

R = 20 to 840 using $\frac{1}{R(\ln R)^2}$ for a priori of R;

I got R = 14.68 instead of 137.47

ST83A Modification of ST83: uses (as likelyhood score: $\sum \nu\nu \times (p_i \ln p_i' + q_i \ln p_i')$

This is (I think) more stringent than Σ err2.

Error! Should not be mod. of ST83; But Modific of ST 82A.Bas,

Score for STEIN = −3907.959 — So ≈ RR=160

RR	Score
160	−3907.977
192	−3907.478
150	−3908.200
225	07.579
300	6.875
450	6.771
500	6.778
250	6.872
1000	6.962
425	6.771
400	6.774
475	6.773
→ 437.5	−390 6.770 max

RR	Score	
137	−3908.563	$\sum_{20}^{800} \frac{1}{R(\ln R)^2}$
181.9	7.607	″ $\frac{1}{R}$
10 K	−3907.373	

3(389 SCH FT STEIN!

Couldn't find 85.
It may not exist!

.01 I've been doing correction in Test Data backwards! Correction should be −.0091 not +.0091!

~~[illegible]~~ Not Quite! But ST82A; W, t. test data, should not be modified.

It is .0091 too big. We only ↑ the data by .0091. (also $\bar{u}$).

ST84 So! fixed! ST84

Single Const for RR = 425, is 5.794468 | w.o. t. .0091 correction converts 5.794460 (Not much difference!)

try RR=212 : 5.977882 Same as 83.10!,
RR=60 : 6.069885 " " 83.20!

$(.0091)^2 * 18 = \sim 10^{-4} * 18 =$ .0018 (it should show up!) No!

Not "18"; ~ 200 times 18. — much more saving.

Trying Del = .009 → .
= .01 6.075
.00 6.620495 | woops! , ≠ 83.20

RR = ∞ → sterm = 61.0925 10 times as large!

O.K. 83.10 & .20 were w.t. .0091 correction ; but I also increased test data: which I shouldn't have done!

Try double prec. original for integration in 81A.Bas.! See if it makes any difference.

It was double prec for R=20 to 810 $\frac{1}{R(\ln R)^2}$; it gave R = 137.476836055008

for single prec I got "division by 0", It may be that 10^{-206} is too small for single prec. $10^{-206} = e^{-206 \cdot \ln 10}$ ln 10 ≈ 2.3 460

I did G = G + 460 : Then ran worked ok., I got R = 137.4819,

which is quite enuf precision! I left in the (G = G + 460 instruction) it does no harm; — Results are same for single prec.

This takes .1 on double prec. (Using step .1 for integration, increased R by 137.4768 → 137.695 (double prec.)
← << .1 sec. on single prec.

3/4/99 SMRT!

This is a replacement for a lost or non-ex, but ~~for~~ Page 85!

85

I noted on 81.38 R that the result was indep of UU. Also, (82.04) its indep of "U" as "U" ~~it~~ appears in the integrand.

.07 Since $S = \int_\alpha^\beta e^G \frac{u_i N + \bar{U} R}{N+R}\, dR$: The integral is a linear function of u_i & $\bar{U}$, $= A' u_i' + B' \bar{U}$

~~also~~ let $A u_i' + B \bar{U}$ be normed result. $= \alpha$ (81.28) $= \frac{N u_i' + R \bar{U}}{R+N}$ $(A+B=1)$

$\frac{N}{R+N} = A$ so $R = \frac{N}{A} - N = \frac{N-NA}{A} = N\left(\frac{1-A}{A}\right) = N\left(\frac{1}{A} - 1\right)$

So $\int e^G \frac{1}{N+R}\, dR$ has enuf info to compute R.

On e^G: using an integration step of 1, we can express e^G exactly as the product and quotient of factorials: There are just N·K factors that don't cancel & updating involves maybe 3K factors. ~~That's much better~~ — But I should check on this!

I did use this exact form to get e^G for $R = \infty$ (& maybe for $R = 0$).

31399 SMFT STEIN:
31399

.01 Well, I have some understanding of t. Baysian approach! T. apriod ~~[illegible]~~ I'm using does appear to effect t. result a lot — I.e. t. R value I get depends much on t. apriod of R:
Hvr, t. estimate of μ_i doesn't depend so much on t. apriod.

My impression is that if $l > \sim 200$, t k will give a sharp apsip, relatively indep of t. apriod.

.06 Hvr. some Q's: ① for a given initial R that generates (M. Carlo) t. data, how accurate is my R estimate & how accurate is my μ_i estimate? ——

.08 Are t. estimates unbiased? What is actual uncertainty in μ_i, [illegible]?

.10 How do t. answers to .06 – .08 compare w. those for t. various "STEIN.
.11 Solns"?

$\longrightarrow$ (88.01)

.12 Going back to 1½.01 ~~#~~ – 2.46 & 10.01ff:
Reading the Efron – Morris ~~[illegible]~~ JASA ~~[illegible]~~ 1975 paper:
They give us Stein's estimator:

$$\mu_i + \left(1 - \frac{(k-2)}{S}\right)(\overset{\overset{\mu_i'}{\downarrow}}{X_i} - \mu_i) = \frac{k-2}{S}\mu_i + \left(1 - \frac{k-2}{S}\right)\mu_i' .$$

" t. $[\mu_i]$ are any initial guess of t $[\theta_i]$.
What t. $[\mu_i]$ are on p 311 eq.(1.4) is unclear: The X_i are of unity variance.

[illegible] $S \equiv \sum (X_i - \mu_i)^2$:

Tentatively set $\mu_i \equiv \bar{\mu} = \frac{1}{k} \sum X_i$

so ~~[illegible]~~ $\frac{S}{k} =$ ms ~~[illegible]~~ deviation of t. X_i from their mean.

$$k - \frac{(k-2)k}{(k-2) + \underbrace{\sum(\theta_i - \bar{\mu})^2}_{\text{use } k\ ([illegible]!)}}$$

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

$$k - \left(\frac{k-2}{1 - \frac{1}{k-2}}\right)$$

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

$$\approx k - 2 + 1 \sim k - 1$$

$$k - (k-1) = 1$$

On p 312 eqs. 1.7, & 1.8: They put estimator

$\left(\mu_i \cdot \frac{1}{\tau^2} + X_i\right) / \left(\frac{1}{\tau^2} + 1\right)$: i.e. rel wts of $\frac{1}{\tau^2}$ & 1 for the 2 ~~[illegible]~~ estimators μ_i & X_i: T. estimator μ_i has var. τ^2 (= its wt.); t. estimator X_i has var 1; (its wt.)

From ~~eq~~ 1.9 $S \approx (1+\tau^2)\overset{\frown}{\chi^2_k}$ = chi sq. dist'n k deg. freedom.

$\frac{k-2}{S} \approx \frac{1}{1+\tau^2}$ so $\frac{S}{k+2} \approx 1+\tau^2$ ∴ $S = (1+\tau^2)(k+2)$ so $E(\chi^2_k) \approx k+2$?

$E(\chi^2_k) = k+2$? $E\left(\frac{k-2}{S}\right) = \frac{1}{1+\tau^2}$ so $E\left(\frac{1}{S}\right) = \frac{k-2}{1+\tau^2}$.

$\frac{1}{S} = \frac{1}{1+\tau^2}(\chi_k^2)^{-1}$ $E(\chi^2_k)^{-1} = ?$

$1 + \tau^2 = \frac{S}{k-2}$ ~~[illegible]~~ τ^2 will be usually << 1 so $\tau^2 = \frac{S}{k-2} - 1$ will be a very noisy estimator of τ^2: so they propose using $\tau^2 = 0$ if $\frac{S}{k-2} > 1$.

Condon & Odishaw p 1-161 (12.241) give $\kappa_n = 2^{n-1}(n-1)!\,k$ as the nth cumulant. T. first & second cumulants are μ & σ of t. dist.

I need t. ~~[illegible]~~ $\underset{E}{\text{mean}}$ (first moment) so its $\kappa_1 = 2^0\, 0!\, k = k$, not $k-2$.
Hvr. I think we need t. [illegible] of $\frac{1}{S}$ not $E(S)$: $E\left(\frac{1}{S}\right)$ i.e.
Hvr: $\frac{1}{1+\tau^2} \approx \frac{1}{S} \cdot \chi^2_k$ E of both sides gives $\frac{1}{1+\tau^2} = \frac{1}{S}(k)$ as before; (not $(k-2)$).

3 13 99 SMRT stein;

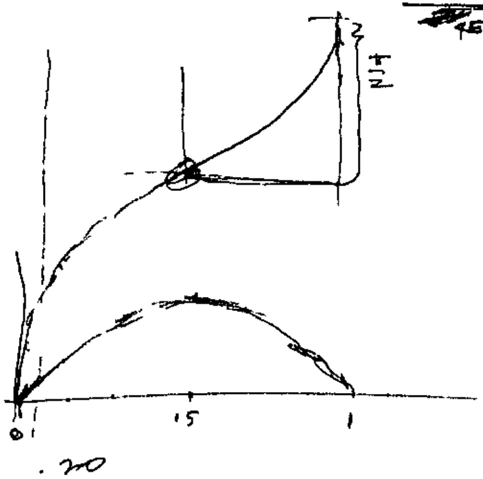
But should we use $E\left(\frac{1}{\chi^2_k}\right)$? This is confusing!

Anyway; Stein used $\frac{1}{1+\pi^2} = .791$: $\pi^2 = .26422$; $\frac{1}{\pi^2} = 3.7846$

$\frac{45}{\pi^2 \cdot}$ = 170.31 $/u(1-u) = .19477$; = 874.4. $\frac{45}{.2} = 225 \approx \frac{1}{6_i^2}$

They do $u_i' \to 45^{\frac{1}{2}} \cdot \sin^{-1}(2u_i - 1)$ to get all vars ≈ 1,

as opposed to $\frac{u_i'(1-u_i)}{45}$ $\sin^{-1}(2u_i - 1)$ v.s. $\sin^{-1}($



$\int_0^x x(1-x) = \int x - x^2 = \frac{x^2}{2} - \frac{x^3}{3}$

$\boxed{x^{uN}(1-x)^{(1-u)N} \cdot \sin^{-1}(2x-1).}$ we could just fiddle this to test for $\approx$ constancy of σ^2 i.e. indep of u.

Perhaps instead of $\sin^{-1}$, use $\frac{1}{x(1-x)}$.

Value of $x^{uN-1}(1-x)^{(1-u)N-1}$ is

$\approx \frac{(uN-1)((1-u)N-1)}{(N-2)^2}$ If N is large, this is much dependent on u.

3.14.99 SMAT STEIN

.01: 86.11; Some points that I'd like to work on;

1) Get t. R d.f. for t. apsi data: It should ~~[illegible]~~ have a peakish form;
The ~~[illegible]~~ Apsp R d.f. would be ~ to t. D.f. obtained from previous Year(s).
We would have to include a t.(remote) Apriori factor like $\frac{1}{R(\ln R)^2}$ or $\frac{1}{R^2}$ ~~[illegible]~~

2) Using t. E & M method, ~~[illegible]~~ exactly, see how well it does on t. average for Mt. Carlo data, Using several R values. Compare w. results for 81A.Bas.
Use both M.S. error & ~~[illegible]~~ apsi of data. as criteria.

3) I think t. E&M method is ~~[illegible]~~ M_i estimate $= \bar{V}\cdot\frac{k-3}{\Sigma(M_i'-\bar{M}_i)^2} + (1- \quad)M_i'$.

This assumes t. M_i' all have unity variance. If they all have same variance σ^2,

We use $\frac{\sigma^2(n-3)}{\Sigma(M_i'-\bar{M})^2}$ rather than $\frac{k-3}{\Sigma(M_i'-\bar{M})^2}$

The xfm. $X_i = n^{\frac{1}{2}}\cdot\sin^{-1}(2Y_i-1)$ is used, so t. data gets ~~[illegible]~~ more close to 1.

---

§3 of t. E&M paper deals w. case in which ~~[illegible]~~ σ_i^2 are not all t. same.

I really need a good approxt. soln. to this case. Perhaps compare it to t. pure Bayesian case. They use Gauss. d.f.s, tho ~~[illegible]~~ t. appropriate d.f. is t. Binomial d.f.
Unfortly, t. data they give is not raw ~~[illegible]~~ frequencies, so it looks like a Gaussian analysis
Mite be best

.22 (3A) Try 81A.Bas w. various α, β $\int_\alpha^\beta$:

α : 1, 10, 20, 30
β : 810, ∞

Apriori
$\frac{1}{R^2}$
$\frac{1}{R(\ln R)^2}$.

$\ln(1+\epsilon)^2$
$\sim \epsilon^2$
$\frac{1}{\epsilon^2}$.

1) $\frac{1}{R(\ln R)^2}$: ~~[illegible]~~

β	α	R
810	20	137.48
	10	134.36
	5	134.055
	2	134.057
	1	impossl. because $\ln R = 0$ $\frac{1}{(\ln R)^2} = \infty$
	1.1	1 ~~[illegible]~~ 133.856 (step .1 in integration
	1.01	132.605 step .01 ~~[illegible]~~ in integration.

So it looks like α is not critical.

something wrong!

$\frac{1}{(\ln R)^2}$ is not integrable near R = 1!

$\frac{1}{R(\ln R)^2}$

100k		2	~~141.07~~	
		2	246.3191	(step 1) 246.1302 for step 10 ~~[illegible]~~
200k		2	255.6	
	10k	2	204.9425	step 10
500k		2	297.7	step 10
	k	2	141.310	
750k		2	~~41.446~~	
1000k		2	1328.7	step 10
850k		2	~~998.8~~	

.30

does this converge?

$\int\frac{1}{R(\ln R)^2} = -\frac{1}{\ln R}$

$\frac{1}{R^2}$ apriori, all, step 10

810	2	88.6
10k	2	97.23
100k	2	97.96

$\frac{A}{\ln 10k} - \frac{A}{\ln 100k}$ $\frac{45}{1.5}$ =
$\frac{A}{\ln 10k} - \frac{A}{\ln 10k + \ln 10}$

$\frac{1}{R}$ all step 10

100k	2	567.6119
100k	1	567.611

~~RRR~~ 2 $\frac{1}{R}$

100k	10	567.65
810	10	181.56
810	20	182.029

$\frac{1}{R(\ln R)^2}$ from 1.1 to 2 step .1
→ 1.2

1.01 to 2 step .01
→ ~~1.024~~

1.001 to 2 step .001
1.0097

3-14.99 start STEIN

There does seem to be something wrong in $\int_{1+\epsilon}^{2} \frac{1}{R(\ln R)^2}$: it should ↑ as ε↓ but in fact, it → 1 from above! [illegible]

No! What I'm getting is not the ∫ but the ratio of $\frac{\int_{1+\epsilon}^{2} \frac{\text{[illegible]}}{R(\ln R)^2}}{\int_{1+\epsilon}^{2} \frac{1}{R(\ln R)^2}}$

as ε → 0 this ratio approaches the value of the integrand for R = 1. — which may be ≃ 1.

Should make ~~for~~ more diff for $\int_{1+\epsilon}^{\text{large}}$

$\int_{2}^{\alpha} \frac{A}{R(\ln R)} dx$ ~~[illegible]~~ α 63.9 ~~4~~ 44.4 1082.41

α	S	$\frac{1}{\ln \alpha}$	Δ	S	Δ	S1
k	141.07	.1447		.201209		
10k	204.96	.1096	.0361	.2175785	.0163	
100k	246.3	.0866	.0220	.2243677		
kk	1328.7	.0724	.0142	.2566954		

O.K.! R is not the S at all! Also, note ~~convergence~~ divergence of $\frac{1}{R(\ln R)^2}$ near R = 1 (((88.30 → 40)R)). The integrals are S1 & S2 & their ratios are S1/S2

Also, it appears that the integrand, e^G approaches $\bar{U}$ as R → ∞ : for R = 10^6 its ~~[illegible]~~ .265377. ; in such case RR → ∞ as R → ∞.

~~[illegible]~~! S1 = ∫ $e^G/(R(\ln R^2))$

.24 S2 = ∫ $e^G/R(\ln R^2) \cdot \frac{RU}{R+N}$ ~~[illegible]~~ my ratio of integrands → U as R → ∞. ∴ RR → ∞ as R → ∞ as — BAD!

It was S1 ~~[illegible]~~ e^G [illegible] by R.

So S1 will be closely related to $\frac{1}{\ln R}$ for large R — But it ~~[illegible]~~ doesn't seem so!

For M to M+10 S1 = 1.46265 - 11) ~~[illegible]~~ [illegible] — [illegible] by $\frac{1}{R(\ln R)^2}$.

10G 10G+M S1 = 2.14 E -9)

At (~~single~~ double prec. step 10) R = 100k got RR = 2~~4~~5.52 (246.13 single prec)

R = M " $\boxed{RR = \text{~~22~~}.866}$ (1328.7 single prec.

Well, if I have to add 10G items, single prec. is ~~[illegible]~~ only ~~6~~ or ~~7~~ decimals, so we could run into trouble. But for ~~the~~ R = 10^6; ~~at~~ step 10; only 10^5 items! Still, it was the difference between two close quantities, I'm not sure I really understand what went wrong! — also from .24, RR → ∞ as ~~[illegible]~~ β ($\int_2^\beta$) → ∞.

~~[illegible]~~ Think this out more carefully)

$\int_2^{10M}$ step 100 : ⇒ RR = 399 !

try more precision

$\int_2^{100G}$ step 1000 → RR = 1082. (maybe → ∞ ? as ~~[illegible]~~ β → ∞.)

e^G is constant for all values of R > 2.

$\int_{1000}^{\beta} \frac{1}{R(\ln^2 R} \left(\frac{RU}{R+N}\right) \int_{1000}^{t} \frac{1}{R \ln^2 R}$ I think it shouldn't → U as β → ∞. 1 ~~[illegible]~~ $\frac{N}{R}$

it should differ by like $\int \frac{1}{R \ln^2 R} \cdot \frac{N}{R+N}$ U ∫ $\frac{N}{R+N}$

=

So! Compute RR for β = ∞, ≐ exactly. ≚ compute $\int^{\infty} e^G\left(U - \frac{VR}{R+N}\right) \int \frac{1}{R \ln^2 R} dr$

3.15.97 SMFT STEIN

from 78 | INDEX → 165

Also from 80

This is a check integration of $\boxed{\text{ST81A}}$ — it checks.

$\boxed{\text{ST90}}$: This obtains ratios:

$A = \int_2^{1k}$; $B = \int_2^{10k}$, $C = \int_2^{100k}$ $D = \int_2^{M}$;

$BA = B - A$; $CB = C - B$; $DC = D - C$!

$\frac{CB}{BA}$ should = .6 ; $\frac{DC}{BA}$ should = .9 (if G is constant for $R \geq 1000$

I got .595 & .396 resp., using Double prec. in all calcs. So O.K.

A worrysome effect; 89.24; $\frac{S2}{S1}$ seems empirically to → 0 as $R \to \infty$!

My integrations for large β may be way off !

Step 100 is o.k. for much of the integrand is constant but for the $\frac{1}{R \ln^2(R)}$ it is not.

I'd like $\int_\alpha^\beta \frac{1}{R(\ln R)^2} \cdot \frac{1}{R+N} \, dR$. $\int_\alpha^\infty$

$R = e^x$

$dR = e^x dx$

$\int \frac{e^x}{x^2} \frac{1}{e^x+N} dR$

$\int \frac{dx}{x^2(e^x+N)}$ $\frac{ds}{(s-N)\ln^2(s-N)} \cdot \frac{1}{s}$

$d\frac{1}{\ln x} = = \frac{1}{\ln x} \cdot \frac{1}{x}$ ∴ int by parts:

$d\, u \cdot v = u\,dv + v\,du$

$u \cdot v = \int u\,dv + \int v\,du$ ∴ $\int v\,du = u \cdot v - \int u\,dv$.

$U = \frac{1}{\ln R}$; $V = \frac{1}{R+N}$ $\int v\,du$ is desired: $u \cdot v - \int \frac{1}{\ln R} \, d\left(\frac{1}{R+N}\right)$

$d\frac{1}{R+N} = -\frac{1}{(R+N)^2}$ so $\int \frac{1}{\ln R \cdot (R+N)^2}$ $\int \frac{dx}{(x+a)^2 \ln x}$

$\int_{s^a} \frac{1}{\ln(s+N)}$ G&R P552 4.281.2

?

$\int^\infty \frac{dx}{x^2(\ln p - \ln x)} = \frac{1}{p} \, li(p)$

$\int \frac{dx}{x^2 \ln(x+A)}$

$x + A = e^y$

$dx = e^y \, dy$

$\int \frac{dx}{(x+N)^2 \ln x}$

$x = e^y$

$dx = e^y dy$

$x = e^y$

$\frac{1}{R+N} = \frac{1}{R}\left(\frac{1}{1+\frac{N}{R}}\right) = \frac{1}{R}\left(1 - \frac{N}{R} + \left(\frac{N}{R}\right)^2 - \frac{N}{R}^3 \ldots\right)$

$\int \frac{1}{R^2(\ln^2 R)}\left(1 + \frac{N}{R}\right.$ so $\int \frac{1}{R^n \ln^2 R}$ = G&R reduces

thing to $\int \frac{dR}{R^n \ln R}$ & 2.4 2.724.1 ibid 2.724.2 $\int \frac{x^n}{\ln x} = li(x^{n+1})$

li(x so that may solve it ! I'd have to get a (program formula) for li(x).

Maple. Actually, Prob. best to Get Maple to do t. $\int$. !

ST81 (from ST77A
Uses Monte Carlo generation of data

ST81A. Main pgm for integrating e^G w. $(R(\ln R)^2)^{-1}$ & $\frac{U \cdot R}{R+N}$.

ST82 (from ST77B)
Is used to find wtd. mean of eps1 data

ST82A (from ST76)
Used to calculate R based on Gaussian D.F. (sent to STEIN)

ST83 (from ST81A) ↓ production pgm.
Moves $\bar{U}$ by .0091 in TNG Data.

ST83A
from ST76 → ST82A
uses U → U+.0091
Also uses P cost core for 2 ways to do pred.

ST84 | from ST82A
Correction of ST82A
See 84.01

ST90: Checks on integration of ST81A

3.16.99 SMRT STEIN

Actually, the soln. isn't so diff!! $\frac{1}{R+N} = \frac{1}{R}\left(\frac{1}{1+\frac{N}{R}}\right) = \frac{1}{R}\left(1-\frac{N}{R}+\frac{N^2}{R^2}-\left(\frac{N}{R}\right)^3\cdots\right)$

we want $\int \frac{1}{R(\ln R)^2}\frac{dr}{R+N} := \int\frac{1}{(\ln R)^2}\frac{1}{R^2}\left(1-\frac{N}{R}+\left(\frac{N}{R}\right)^2\cdots\right)$

$\not\equiv$ $\int\frac{R^n}{(\ln R)^2}$ reduces to $\int\frac{R^n}{\ln R}$ by int by parts. (P204 G&R)

$\int\frac{R^n}{\ln R} = li(R^{n+1})$: $li(x)$: from B.S. P510 & 504! $= U(1,1,-\ln R^{n+1})$

$$U(1,1,z) = \frac{\pi}{\sin\pi\cdot b}\left(\frac{e^z}{\Gamma(1+1-b)\,\Gamma(b)} - \frac{e^z}{\Gamma(a)\Gamma(2-b)}\right) M(1,1,z) = e^z$$

$a\to1,\ b\to1$ $= \frac{\pi}{\sin(\pi b)}\ e^z$ This is not clear!

$li(x) = Ei(\ln x)$: $\boxed{Ei(x) = \gamma + \ln x + \sum_{n=1}^{\infty}\frac{x^n}{n\,n!}}$ so this converges rapidly.

$= Ei(n+1\cdot\ln R)$

$e^x = \sum_{n=0}^{\infty}\frac{x^n}{n!} = 1 + \sum_{n=1}\frac{x^{n-1}}{n!}$ | $\frac{d}{dx}Ei(x) = \frac{1}{x} + \sum_{n=1}^{\infty}\frac{x^{n-1}}{n!}$

$\frac{e^x-1}{x} = \sum_{n=1}^{\infty}\frac{x^{n-1}}{n!}$ so $\int\left(\frac{e^x}{x}-\frac{1}{x}\right) = \sum\frac{x^n}{n\,n!} = Ei(x) - \ln x - \gamma$

Does I. $Ei(x)$ power series converge rapidly for large neg x?

I'd like something that converges faster. $Ei(x)$ may be $\approx a + bx$ for $x > .6$!

From pp 242-243 $x\,e^{-x}\,Ei(x)\to1$ as $x\to\infty$ for x as smallest z, this

expression is only 1.34 from fig 5.1 P 228 $Ei(x)$ really looks linear for $x > .81$,

1.65 ..
1.90
3.55

So $Ei(x) \doteq \frac{e^x}{x}$

Also $Ei(x)\,x\,e^{-x} + x\,e^{x}\,E_1(x) \doteq 2.1$ which is good, because there are more formulas for $E_1(x)$

not so good approx!

But also $Ei(x) \equiv \int_{-\infty}^{x}\frac{e^x}{x}dx$. $li(x) = \int_0^x\frac{dt}{\ln(t)}$ | $\int\frac{x^n dx}{\ln x} = li(x^{n+1})$

$\int\frac{1}{}$

Actually, what I need is $\frac{\frac{s_2}{s_1}}{1-\frac{s_2}{s_1}} = \frac{s_2}{s_1-s_2}$

$s_1 - s_2$ is $\int_\beta^\infty\frac{1}{R+N}$: By making $\beta \gg R$ this is $\approx \int_\beta^\infty\frac{1}{R}$

So $\int\frac{1}{R^2(\ln R)} =$ G&R P204 (n=-2, m=2) $= \frac{R^{-1}}{\ln R} + \frac{-1}{1}\int\frac{1}{R^2\ln R} = \frac{1}{R\ln R} - \int\frac{1}{R^2\ln R}$

$= \frac{1}{R\ln R} - li\left(\frac{1}{R}\right)$

$n=-2,\ m=2$ $\frac{x^{-1}}{1\cdot(\ln x)^1} + \frac{-1}{1}\int\frac{x^{-2}}{(\ln x)}$ | $= \frac{1}{R\ln R} - \int\frac{1}{R^2\ln R}$

$\int\frac{x^{-2}}{\ln x} = li(x^{-1})$ so I need $li(x^{-1})$ for large x; say $x = 1000$

$li(x) = Ei(\ln x)$ for $x > 1$.

$li(x^{-1}) = Ei(-\ln x)$

$li\left(\frac{1}{R}\right) = \boxed{Ei(-\ln R)}$

From Complex Vars by Spiegel: P275 $Ei(z) = \int_z^\infty\frac{e^{-t}}{t}dx$ actually use finite k max.

$\sim e^{-z}\sum_{k=0}^{\infty}\frac{(-1)^k k!}{z^{k+1}}$ is asymptotic expansion. Sometimes error is < first term discarded.

If $R = 1000$ $\ln R \approx 7$ $Ei(-7)$

$$\sim\frac{1}{1000}\underbrace{\boxed{\Sigma\ \frac{1}{7} + \frac{2}{7^2} + \frac{6}{7^3}}}_{.2011} + \frac{6}{7^3}\cdot\frac{7}{7}$$

$= .0175$

$1.83\,E-4$

1.84

3.16.99 SMFT STEIN

What I need is $\int_{\beta}^{\infty} \left(\frac{1}{R(\ln R)^2} \cdot \frac{1}{R+N} \right) dR$.

say $\beta = e^7$

well ok. $Ei(-x) = -E_1(x)$: from Burose Stands.

$E_1(7) = 1.1548\text{E}-4$ $\quad$ ($e^7 = 1096.63 \approx 1000$).

$\frac{e^7}{45} = 24.36 \; (\frac{1}{x}) = .0410$

use $\int_{\beta}^{\infty} \frac{1}{R^2 \ln^2 R}$: for $\beta = e^7$; result is at most 4.1% hy.

$= \int \frac{1}{R^2 \ln^2 R} = \frac{1}{R \ln R} - \int \frac{1}{R^2 \ln R}$

$\int \frac{1}{R^2 \ln R} = li(\frac{1}{R}) = Ei(-\ln R) = -E_1(\ln R)$

$\int \frac{1}{R^2 \ln^2 R} = \frac{1}{R \ln R} + E_1(\ln R)$

Trouble is, $\int_{\beta}^{\infty} \frac{1}{R^2 \ln R}$ is > 0 for $\beta > 1$

Actually, I can est $\int_{\beta}^{\infty} \frac{1}{R(R+N)\ln^2 R}$ w. known accuracy!

$\sum_{i=0}^{\infty} f(i) < \int_0^{\infty} f(x)dx < \sum_{i=1}^{\infty} f(i)$.

Make step size so integrand ↓ 1% in each step.

So each time R doubles, step size is ×4.

w. R = 1000; if I integrate to R = 100k

the remaining sum will be < 1%. $\quad \int \frac{1}{R^2} = -\frac{1}{R}$

for R = 1000 I can use step 10,

.67 26
.4 15
.26 10
.2 .16

I use steps of $\boxed{ST83}$:

IGD = Exp(G)/R/LL

S1 = S1 + IGD

S2 = S2 + IGD * R * l /(R+N) $\quad$ "l" replaced by "1"

print $\quad N * S2/S1 / (1-S2/S1)$ $\quad$ This is same as $\frac{N * S2}{S1 - S2}$

we can just comput S3 = S1 − S2 = w. S3 = S3 + N/R+N.

so initialize S3 = 0;

print N * S2/S3.

Integrate from 1000 to 100k step 10. First go it for ST83.bas;

from ST50 The strange thing is the ↑ from β = 100k to β = kk! I would expect < 10% changes in S1 − S2!

$\boxed{ST93}$

1k	148.5437
10k	217.6987
100k	262.1534 ~16"
M	292.0644

What has happened! S3 converged at R = 10k "+"?

But S2 continued to grow. (S2 = S1 $\textcircled{+}$ S3.) $\quad$ Should be minus

S1 is over normen. So easily integrated to ∞

$\int \frac{1}{R \ln^2 R} = -\frac{1}{\ln R}$.

.38 $\quad$ So value of S1 at R = Million is 3.0296488E−208 (did I use step = 1000?)

S1 did go down as $\frac{1}{\ln R}$. $\quad$ So S1 for R = ∞ = 0 as far from S1: R = M

as S1, R = M is from S1; R = K $\quad \frac{1}{\ln K} : \frac{1}{\ln KK}$

→ $\frac{1}{2}$ $\quad$ 0

$E_1(z) = \int_z^{\infty} \frac{e^{-t}}{t} dt$

$E_1(z) = \int_1^{\infty} \frac{e^{-zt}}{t} dt = \int_z^{\infty} \frac{e^{-x}}{x} \frac{dy}{z}$

$zt = x$

$z\,dt = dx \qquad dt = \frac{dy}{z}$

So $E_1(z) = E_1(z)$

$Ei(x) = -\int_{-x}^{\infty} \frac{e^{-t}}{t} dt$. $\quad$ so $Ei(-x) = -E_1(x)$

$Ei(y) = -E_1(-x)$?

$\frac{1}{e^{10}} \left(\frac{1}{10} + \frac{2}{1000} + \frac{6}{10k} + \frac{24}{100k} + \right.$

2 3 4 5

$\frac{1}{7} + \frac{2}{7^2} + \frac{6}{7^3} + \frac{24}{7^4} +$

$E_1(7) =$ 1.1548 E−4

3.16.99 SMRY STEIN

So: T. result appears to be $\boxed{R = 532.097}$ for Integration from R=20 to ∞.

for Integral for 20 to 1k, 10k, 100k, M, & S followed the $\frac{1}{\ln R}$ law,

So extrapolating to ∞, I got $S1_{\infty}$ for $\int_{20}^{\infty}$.

I had $S3_{\infty} (\equiv S1 - S2)$ for $\int_{20}^{\infty}$; $S2 = \boxed{S1+S3;}$ — S2 should be $S2 = S1 - S3$

I needed $A = N * S2/S1 / 1 - S2/S1)$ (using v=1 in the last 3 insts of ST83/ST93)

$= N \frac{S2}{S1 - S2} = N * \frac{S1+S3}{S3} = N \left(\frac{S1_{\infty}}{S3_{\infty}} + \right)$ ok.

$S1_{\infty}$ was 4.378191 E-208
$S3_{\infty}$ was .4044475 E-208] N=95 so this gave $\boxed{532.097}$ for R of $\int_{20}^{\infty}$.

For other calcns: Get S3 for, say $\int_{1}^{10k \text{ or } 100k}$.

Get S1 for R = 10k and 100k: from this calculate value for R = ∞.

$\frac{1}{3}$ $\frac{1}{4}$

.14 $S1_{10k} + 4(S1_{100k} - S1_{10k})$ $S1_{\infty} = S1_{1k} + 4 S1_{10k} - 4 S1_{1k} = 4 S1_{10k} - 3 S1_{1k} = 4.382679$ E-208

2D − A These results shouldn't disagree! $2 S1_{M} - S1_{K} = 4.37191$ E-208

1k 10k 100k M
$\frac{1}{3}$ $\frac{1}{4}$ $\frac{1}{5}$ $\frac{1}{6}$
A B C D

5C − 4B

to get $\int_{10k}^{\infty}$ it would be easiest to get e^{G}, = constant, then use $\int_{\alpha}^{\infty} \frac{1}{R(\ln R)^2} = \frac{1}{\ln R}$

$5 S1_{100k} - 4 S1_{10k} = 4.3765$ E-208

So the results are cons:

Some Morals:

1) Computing error in the S. in designing the calcn. so that S2−S1 (=S3) was calculated, was not difflt to do, but it took a long time to realize this!

2) The only ∫ I had to do analytically was $\int \frac{dR}{R(\ln R)^2} = \frac{1}{\ln R}$

Next Project! Get apsid of R. using suitably wtd data.

Use total data (i.e. include first 95 "at bats".

Start w. new total data!

ST93 from STEIN50 A(,2) is test data $\boxed{A(,5) \ 5352 \quad U \text{ changed to } U + .0091}$ ← Modification of STEIN50

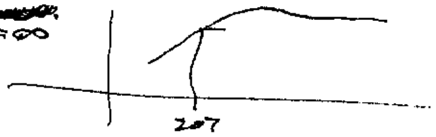
Prelimy result: Is this ≥ before! Newtom: Down $\frac{1}{e}$ from max at a slight peak at ~ R ~ 700 Down maybe $e^{-.2}$ from R=∞

Down $\frac{1}{e}$ from "Peak" at R ~ 207

I think it's necy to mult by $\frac{1}{R^2}$ or $\frac{1}{R(\ln R)^2}$.

Using $\frac{1}{R^2}$; a slight peak at R ~ 207 G = −3918.36 down $\frac{1}{e}$ at ~R=88, 586

"Peak" is not quite symmetric in log domain.



3.16.99 SMRT STEIN

(.01) Running ST93 but ommiting $N = A(J,5)$ [so $N=45$] : keeping $U = U + .0091$. 36

peak in $\frac{1}{R^2}$ output at $\sim R \overset{\sim}{=} 103$ 24

~~→~~ In G output ; down ~~▲~~ 1 from $R=1658$, at $\sim R = 207$ 30

Comparing 1. 2 D.F.'s : ① Both are down $\Delta G = 1$ for $\sim R = 207$; | wrt $R = 1658$ 13
| $\approx R = \infty$. 10

.07 ② In ~~R~~ $N=45$; from $R=.8$ to $R = 1658$; $\Delta \overset{G}{~~G~~} = 20$ 7

" $N = A(J,5)$ " " $\Delta \overset{G}{~~G~~} = 25$

.09 ③ In ~~N~~ $N = 45$: $\frac{1}{R^2}$ peak is at $\sim R = 103$ ~~[illegible]~~ $\overset{G}{}$ down from peak by 1 at $R \approx 36, 586$

In $N = A(J,5)$ " " ... $\sim R = 207$ $R = 80, 586$

Well, ~~[illegible]~~ (.01 ff) isn't just right! ~~●~~ It uses $A(I,\underline{\underline{2}})$ data w. $N=45$ — which is not realistic. — But anyway, using $A(I,2)$ data w. $N = 450$ for all ~~●~~ Players 39.90

$\underline{N = 490}$ $R = 207$ up $\Delta G = .6$. from $R = 1658$ 36.20

\underline{did} Give a peak (w.o. $\frac{1}{R^2}$) at $R = 207$; $\Delta G = 6$ at $R = 1658$! 3.70

.14 $N = 369$ $R \approx 207$ $\Delta G = 3.7$ up from $R = 1655$

Try ~~N~~ $N = 300$: peak at $R = 293$; only $\Delta G = 2$ up from ~~R~~ $R = 1658$.

$N = .250$ " " $R \approx 293$ " $\Delta G = 1$ " " $R = 1658$ 7.25

$N = 200$ " " $\sim R \approx 500$ " $\Delta G = .2$ " " $R = 1658$ 6.71

.54

Turns out $N = A(J,5)$ \underline{\underline{has}} a \underline{\underline{peak}} at 1658, Its down only ~~[illegible]~~ $\overset{\Delta G =}{.35}$ at $R = 6635$

. Mean Values $A(J,5) = \boxed{369.38}$ surprising like $\bar{N}$: so I'd \underline{expect} a peak!

See 47.10 for work on Large N

(.22) Well, \underline{I got} a peak in G ~~[illegible]~~ using ~~[illegible]~~ $A(J,5)$ at $R = 586$! ΔG down only .25 at $R = 1658$

ΔG .54 " $R = 6635$

But note (.14) ~~[illegible]~~ ($\bar{R} = 369$) ; using ~~N~~ $= 369$ for all \underline{players} Got a \underline{much sharper peak} at $R = 207$ at a ~~[illegible]~~ much lower R !!

$\Delta G = -3.7$ at $R = 1659$!!

It may be that \underline{large} N_i ($\equiv A(I,5)$) is very impt for peaks & small N_i is of no effect at all. So if we have a distribution ~~of~~ of N_i about $\bar{N} = 369$.

→ This is \underline{wrong}: $\overset{N=}{}A(J,5)$ has peak at \underline{\underline{586}} : down only $\frac{\Delta G = -.2}{~~[illegible]~~}$ at $R = 1658$, i.e. line (.22)

For ~~A~~ $A(J,5)$ value of G at $R = 6TM$ (essentially ∞) was -3907.43580 : down $\sim .7$ only, true

6.247 peak at ~~[illegible]~~ $R \sim 572$

Try \underline{Mult by $A(J,5)$ by 10}! ~~[illegible]~~ Max at $R = 128$, nearby about ~~[illegible]~~ $\Delta G = 15$ from $R = \infty$

Peak was down by ΔG ~~[illegible]~~ $= 1$ w. ~~[illegible]~~ \underline{$R = 64, 512$}. (Sharper peak) 74.35

~~Was~~ After Mult by $\frac{1}{R^2}$, peak was at about same position. (maybe R max $\div$ ~~1~~.4 or 2) 59.48

Contrast w. old \underline{STEIN 90}: ($N = 45$, $Z = A(J,3)/1000$ ... (Try. data)

Peak ~~at~~ $R \sim \underline{1200}$; About $\Delta G = .004$ above $\overset{R = 8,200.}{}$ & $R = \infty$ } very broad "peak"

Down $\Delta G = -1$ from peak at & $R \sim$ ~~[illegible]~~ 64 } very gradual to ∞. .845

ΔG betw. ~~[illegible]~~ peak ($\approx R = \infty$) & $R = .8$: \underline{13.5} .849

82.3

For $\overset{N=}{A(\cancel{I},\cancel{5})}$ $A(J,5)$ (.07 data) ΔG betw $R = .8$ to $\approx R = \infty$ was 25 (About \underline{twice} as steep) $- 468.8985$

13.5

~~For Down $\Delta G = 1$ from peak at R, at~~ See $\binom{.07}{.09}$ for data on $N = A(J,5)$

Uniform df. of $\frac{1}{R}$ from 0 to 1 would mean $P_C \propto \frac{1}{R^2}$ from 1 to ∞. Did I ever do $\int_1^\infty$ for $P_C = \frac{1}{R^2}$?

3 1999 SMFT STEIN

What I've forgotten, In ST 4, I found best pc score predns. using wts of $.105 \overset{.105}{\approx} + \overset{.895}{~~.885~~}\,.275$ $\approx \bar{u}$ of test data

probly rel wts of 1 & 9 → .45, .405 so R ≈ 405.

~~Bomes~~. ~~Ts~~ R = 429 gave min error² on 83.18

I think STEIN was equivalent to R ≈ 109.

Using $\int_{20}^{\infty} \frac{~~de~~ e^{a}}{R(1+R)^2}$ I got R = 532.097 | probly $\int_0^{\infty} \frac{e^{a}}{R(1+R)^2}$ would be $\overset{\text{should}}{\text{the same}}$.

$\int_0^{\infty} \frac{e^{c}}{R(1+R)^{~~2~~}}$ has $\overset{\text{little}}{~~[illegible]~~}$ a priori info. Using closer to $\ln^{th} R$ would probly not affect result much, but I'm not sure.

So Reg ↑ would seem like an O.K. soln.

.09 Def .10 Re: Using ~~a~~ "test data" as Tng data. (the N_i are not constant), → error see 96.20

In ST 81A, the final result "$\tilde{R}$": $\tilde{R} = \frac{N_i S_2}{S_1 - S_2}$: $= N_i \frac{\boxed{S_1+S_3}}{S_3} = N_i\left(1 + \frac{S_1}{S_3}\right)$.

.11 $\boxed{S_3 = \int (A) \frac{dR}{R+N_i}}$ .11R : $S_1 = \int (A)\, dR$ If we do .09

.12 If $\tilde{R}$ was same for all i (which is desirable). $\overset{\text{is same for all } i}{\boxed{\tilde{R} = N_i\left(\textcircled{0} + \frac{S_1}{S_3}\right)}}$ ← from ST81A : $S_3 = \frac{S_1}{\frac{\tilde{R}}{N_i} - 1}$

~~Sum~~ $\tilde{R} = N_i + \frac{S_1 N_i}{S_3}$; ~~$N_i = \tilde{R} - \frac{S_1}{S_3}$~~ Because of .11R: ∴ S_3 is a ↓ funct of N_i which is

~~contrary to this~~ ~~are~~ contrary to ↑ in which S_3 is an ↑ function of N_i

so $\tilde{R}$ must be a function of ~~N_i~~. Should be $\textcircled{-1}$ see 96.20

$\frac{S_1}{S_3}$ ↓ , $-\frac{S_1}{S_3}$ ↑

$\frac{\tilde{R}}{N_i}$ ↓ ; $\frac{\tilde{R}}{R_i} - 1$ ↓ $\left(\frac{\tilde{R}}{R_i} - 1\right)$ ↑ ← (if $\frac{\tilde{R}}{R_i} > 1$) | $\frac{S_1}{\frac{\tilde{R}}{N_i} - 1}$ ↑ | So S_3

$\tilde{R} = \frac{\frac{N S_2}{S_1}}{1 - \frac{S_2}{S_1}}$

.20 $\boxed{\frac{\tilde{R}}{N_i} = \textcircled{0\,1+} \frac{S_1}{S_3}}$ Should be −1

So N_i & S_3 cannot move in same direction ~~as~~ if $\tilde{R}$ is indep of N_i.

Hvr. from (.11R) S_3 is a ↓ funct of N_i.

$\frac{\tilde{R}}{N_i} = \frac{S_2}{S_3} = \frac{S_1+S_3}{S_3} = 1 + \frac{S_1}{S_3}$

Well! If $\frac{\tilde{R}}{N_i} > 1$ then N_i & S_3 move in same direction. If $\frac{\tilde{R}}{R_i} < 1$ then R_i & S_3 move in opposite directions. $\frac{\tilde{R}}{N_i}$ can't be < 1 because $\frac{S_1}{S_3}$ is always ~~[illegible]~~ > 1

Actually, $\frac{\tilde{R}}{N_i}$ is ~~the~~ equiv to assign wts to $\bar{u}$ & u_i. $\frac{\tilde{R}}{N_i} = 1 + \frac{S_1}{S_3} = \frac{S_1+S_3}{S_3} = \frac{S_2}{S_3}$

So the relative wts, $\frac{\tilde{R}}{N_i}$ ~~as~~ is $\textcircled{1} + \frac{S_2}{S_3}$ which is always > 1 ! & even when N_i is large!

No $\frac{S_2}{S_3}$ can be betw ~~[illegible]~~ 1 & ∞ : so $\frac{\tilde{R}}{N_i}$ can be betw. 0 & ∞ as desired.

This is the same unreasonable result that I got [redacted] before, but w. a simpler ??? calcn. (Couldn't find this calcn. hvr.) &: 2 to 3 #60 6-8 ~~[illegible]~~ years ago.

It was : Something like getting rel. wt. of 2 methods of ~~[illegible]~~ predicting whether players would get a hit at next "at bat". Hvr, ~~a~~ in that ~~[illegible]~~ analysis, I was always u_i that got ~~[illegible]~~ > ½ of wt. There I considered ~~[illegible]~~ the ~~[illegible]~~ rel. errors of the 2 methods!

$= \frac{1}{N}\bar{u} - \frac{1}{N}\overline{u_i^2}$

$\frac{1}{k} ~~[illegible]~~ \sum_{i=1}^{k} ~~[illegible]~~ \frac{u_i(1-u_i)}{N_i}$ ~~[illegible]~~ mean error, using each u_i for predn.

$\frac{1}{k}\sum \frac{u_i}{N} - \frac{1}{k}\sum \frac{u_i^2}{N}$.

$= ~~[illegible]~~$

~~[illegible]~~ $(u_i - \bar{u})^2$ mean error using $\bar{u}$ for predn.

$= ~~[illegible]~~ \frac{1}{k}\sum u_i^{2} - \frac{1}{k}\bar{u}^2$

$= \overline{u_i^2} - \frac{1}{k}\bar{u}^2$

While I did try this ↑, That's not the one I'm thinking of.

1

.011 on the other hand, ~~[illegible]~~ $(\overset{95.20L}{[illegible]})$ isn't so reasonable. $(\frac{\tilde{R}}{N} = 1 + \frac{S_1}{S_3})$

It should be possib. to get any $\frac{\tilde{R}}{N}$ ratio w. suitable data, & ~~[illegible]~~ a (95.20L) means this impossl.

S_1 is normaliz. const. S_2 is $\int (f) \frac{dR}{N+R}$ (integral)

with suitable shape of F, $(\int F dR = 1)$ ~~[illegible]~~ $\frac{S3}{S1}$ can have any value betw. $\frac{1}{N}$ & 0.

O.K. if ~~[illegible]~~ so 1+ ~~[illegible]~~ $(\frac{S_2}{S_1})^{-1} \Rightarrow 1+N$ to ∞. So it can't even get near 1!

This still seems wrong!

$\int f \; (\frac{R\bar{U}+NU_i'}{R+N}) dR$ ~~[illegible]~~ $R_0 \bar{U} + N_0$ say $f(R) = \delta(R-R_0)$

say! $\bar{U}=1$; $U_i'=0$ using $f=\delta(R-R_0)$: $S = \frac{R}{R+N}$: can have range from ~~1~~ to 0. ; If it's 1, $R=\infty$; if it's 0, $R=0$.

I think it was the setting of $\bar{U}=1$, $U_i'=0$ that caused the trouble.

$S1 = \int f \; dR$; $S2 = \int f \cdot \frac{R\bar{U}+NU_i'}{R+N}$; | in ST81A: $S1 = \int f$: $S_2 = \int f \frac{R}{R+N}$
$S3 = \int f \frac{N}{R+N}$

It could be that to simplify I introduced in the final 3 lines ($U0=0$, $U=1$) will not work when we have different N_i values.

$\frac{R}{R+N} = \frac{S_2}{S1} = \frac{S_1}{S_2} = 1+\frac{N}{R}$: $\frac{S1-S_2}{S_2} = \frac{N}{R}$; $S1-S_2 = S3$; $S1 = S_2+S3$ | $\frac{S_3}{S_2} = \frac{N}{R}$:

(.20) in 95.10 ; $S_2=S1+S3$ This is wrong! $S2=S1-S3$. $S2 = S1-S3$

$R = N\frac{S2}{S3}$

So $\frac{\tilde{R}}{N_i} = \frac{S_1}{S_3} - 1$ S_3 is betw 0 & S_1 so $\frac{S_1}{S_3}$ is betw. 1 & ∞

$= N(\frac{S_1-S_3}{S_3})$

$\frac{S1}{S3}$ is betw 0 & ∞ ; as desired.

$= N(\frac{S1}{S3}-1)$

~~S_3 is [illegible]~~ $S_3 = \int f \frac{N}{R+N} dR$: S_3 is an ↑ funct of N $\frac{S_1}{S_3}$ is a ↓ funct of N.

$\frac{\tilde{R}}{N_2}$ is a ↓ funct of N. — so it may be O.K.!

. ST96 ~~[illegible]~~ First ~~[illegible]~~ modify $\overset{ST83}{~~[illegible]~~}$ so it can compute $\int_1^\infty (\;) \frac{1}{(R\ln R)^2}$:

First I find out how far out to go so that ~~[illegible]~~ G is constant := RR say RR ~ 1000.

Get ~~[illegible]~~ value of $e^G(RR)$. $\equiv A_0$

Compute $\int_{~~[illegible]~~ 1}^{1000RR} (\;) \frac{1}{(R\ln R)^2} \cdot \frac{N}{R+N} = \boxed{S3}$ (Using step 1, ~~[illegible]~~ the code of in ST93)

Comput $\int_1^{~~[illegible]~~ RR} (\quad) = S1$

' $S1 = S1 +$ ~~[illegible]~~ $A_0/\ln RR$: ~~[illegible]~~

Then compute $\tilde{R}$ as before

.34 To find RR: ~~[illegible]~~ find G(R=1000): Then ~~E~~ R=2000, 3000 ect until ~~[illegible]~~ G(~~RR~~ R) ~~[illegible]~~ G(~~RR~~ R−1000)+.01 | If ... Then RR=R.

Use large SRTN to find G(R). ~~[illegible]~~ first use it to find RR, Have it print RR and e^G ~~[illegible]~~ Then do usual integration.

```
RR=1
 Do until G<G1+.01 ~~[illegible]~~        G1 = -~~[illegible]~~100000
 ~~[illegible]~~
 Go sub 100
 ~~[illegible]~~
 Loop
 Print RR: END          => (97.11)
```

RR=RR+1000
R

```
RR=1, G1=-100000
10 Gosub100
  If G<G1+.01 Then print RR
   RR=RR+1000 ~~: G1=G~~
   Goto
```

3 20 99 SMFT : STEIN!

.01 ⇒ Well, I think this is it wrt t. optimum Soln. via Bayes:

Getting an optimum R value is not really t. rite way, recall N_i are not t. same.

At any rate, t. theoretically correct way to do Bayes, is to get t. total a priori. (integrating ~~[illegible]~~ wrt R)
using uniform D.F.'s for $\bar{U}$ & t. U_i's.

We then get t. relative probys for a 0 or a 1 being t. next data pt. for t. i^{th} Player.

So, say we have $[N_i, U_i]$ for t. corpus; we want t. rel. probys ($\equiv \frac{C_2}{C_1}$)

for ① t. $\overset{\text{same}}{/}$ corpus, but ~~[illegible]~~ for one of t. players, $N_j \to N_j+1$, $U_j \to \frac{U_j \cdot N_j+1}{N_j+1}$ — making a "hit"

② " " " " $N_j \to N_j+1$, $U_j' = \frac{U_j \cdot N_j}{N_j+1}$ — making an "out".

This is also t. ideal method to use if all of t. N_i are $\overset{\text{initially}}{/}$ t. same,

---

So :1) Make avgln of eG a srtn:

.10 2) Get RR viz ~~[illegible]~~ 96.34

.11 : 96.40 STEIN(cont) STG6 (Cont) So, [redacted] after we reduce RR, Get GG=G.

~~After R integration~~

Actually, In ·ST81A, we integrate ~~[illegible]~~ $S2 = \int e^G \frac{1}{R(\ln R)^2} \cdot \frac{R}{R+N}$

$S3 = \int e^G \frac{1}{R(\ln R)^2} \cdot \frac{N}{R+N}$.

In t. context of varying N_i's; this "N" is meaningless.

All I really need is $S1 = \int e^G \cdot \frac{dR}{R(\ln R)^2}$.

.001	2001
.0001	2001
.00001	2001

Another point in t. RR determination, G ↓ as it ^ becomes constant).

G > G1 −.001

— .01 G
— .0097 G1 a larger
.0003 ~~G1~~

G $\gtrless$ G1 ~~.00~~

RR, G, G1

I Got RR (≅ 2000) from t. preliminary pgm.

Then $S1 = \int_2^{2000} = 1.941698E-208$

$e^{GG}/\ln R = 2.451589 - 208$ ← integral for R > 2000

} So most of S for S1 occurs ~~[illegible]~~ for R > 2000 !

BAT = 0 or 1

If $J = \overset{CC}{\blacksquare}$ then N = 46 : $A(J,3) = (A(J,3) * 45 + \overset{BAT}{\#})/46$

CC = (1 thru k) ; BAT = 0 or 1.

CC
17 316 178 244 408
A(3)

CC = 10 → 273
Hit 11919
31688
16 → 273
Same as ~~R=20~~ CC = 10

Putting in that "IF" line changes RR from 24 to 235K if Hit=0
" " 236K " Hit=1) not reasonable.

For Hit=0 ; CC=1 ; RR ~~= 373~~ 200K Now I get ~~[illegible]~~ 288

$\frac{400 \times 45 + \cancel{1}}{46}$ should be > 400.

So for Player #1 I got ~~.28~~ 288 batting avg.

Stein Got 290 seems unreasonable! $\bar{U} = 265$!

Try CC = 18 .265

hit 1.2565
3.1008
.28836

0 → 32.004 A
1000 → 11530 B

$\frac{B}{A+B} = \left(\frac{A}{B}+1\right)^{-1}$

3 21 99 SMRT Stein:

	CC	Raw Avg	My value	Stein
265 135	1	400	288	290
	10	244	273) ?	259
	11	222		254
265 104	18	156	265	239
				254

using Dal = .0091

avg values

Avg w. Dal=0 changed things a lot!
(Back)

280 367 265

263 200 265
262 .125
255 70 265
265

using Dal=0

120	15
19	2 ← ?
99	10

70 → 264!

1 413.04) So to modified Bar avgs. check.
391.3

50470
18030
263 !

247 → 11 : $\frac{12}{46}$, $\frac{11}{45}$

The cc = 10 result seems nutty

Check perm w. cc=0 ; Dal=0 against ST83 (92.28 ff)
Check C1 Also check total integral against those cases where I did it better! 93.14
I think I did $\int_{20}^{R}$

93.14! running ST93 for Dal=.0091 !! R=1k → S1 = 1.681106 E-208 × 3 = } → $4S1_{10k} - 3S1_{1k}$ =
10k S1 2.35645 " " × 4 = } 4.3825105E-208
= $C1_\infty$

This agrees w. 93.14 Now Go to ST96 C=0 Dal=.0091

R = 5000

- ×3 $S1_{5k}$ = 3.502295 -205 ; $S1_{\infty}$ = 6.85996 6.5596 8598
+ ×4 $S1_{50k}$ = 4.2151036 3

Try S_{∞} using different G-G1 limit.
→ 6.3535 3734 -205
≠

Using |G-G1|<.0001 RR= 16000 , $S1_\infty$ = 6.85085
very close.

Using |G-G1|<.01 R = 2000 $S1_\infty$ = 6.8771 -205

~~" " R = 2000 gives $S1_\infty$ =~~ .

|G-G1|<.00001 R = 492 $S1_\infty$ = 6.848 1657 so its rather stable.

The disparity is appreciable! I checked Arithmetic.

6.85 vs. ~ 7.7% error
6.35

- ×3 S1 = 15k $S1_{15k}$ = 3.885 2584
· ×4 S1 = 150 $S1_{150k}$ = 4.457768) → $C1_\infty$ = 6.17520

$\frac{1}{\quad}$ 5k to 50k. No! I have to do 1k & 10k. see (93.14 R) ff

5k, 50k works directly.

$\frac{1}{\sqrt{5k}}$ $\frac{1}{\sqrt{50k}}$.

-4 × $C1_{10k}$ 3.7546 -205
5 × $C1_{100k}$ 4.373 -205) → 6.848 E-205 as it should be E-205

So it checks out fine. C1 may be ok.

Using the on 92.28, I Got $C1_M$ = 3.029 ... E-208 A factor of 100 smaller.

Maybe I did stop 1000! No: running ST83.Bas! R = 20 to 10k I got S1 = 2.356 E-208

work ST83: $\int_{1000}^{1040}$: S1 = 4.28091 -211 Dal = .0091

ST96 $\int_{1000}^{1010}$: S1 = 6.6119 E-208 ~~Dal=.0091~~ Dal=0

3.21.99 bMFT STEIN ~~[illegible]~~ 0 to 10 of 11

99

Differences Between ST96 & ST83.

1) ST96 doesn't do Del in the So Sub loop!

Still, it shouldn't make this much difference!

	R	$e^{G}/R(\ln R)^2$
ST96	1000	6.049682−209 = S to " $\int_{1000}^{1000}$ "
	1010	5.9723049−209
	2000	2.487 −209
	5000	7.8844 −210
ST83	1000	6.046958−209 Del=0

So Del = .0091 means a by difference! $\int_{1000}^{1010}$ are same for ST83, ST96; [illegible] Del=0 makes [illegible].

$H = R\bar{u} + z \quad L = R(1-\bar{u}) + N - z \quad F = R+N$

$G = \text{const} + H\ln H + L\ln L - \boxed{F\ln F}$ ← indep of z

$+\frac{1}{2}\ln(H\cdot L\cdot 2\pi/F)$

$(H+1)\ln(H+1) + (L-1)\ln(L-1)$

$H\ln(H+1) + L\ln(L-1) + \ln(H+1) - \ln(L-1)$

$2\pi\sqrt{(H+1)(L-1)}\ \frac{H+1}{L-1} \rightarrow \sqrt{\frac{(H+1)^3}{L-1}}\ ,\ (H+1)^H\cdot(L-1)^L$ terms ratios =

$\sqrt{\frac{H+1^3}{H-1}\frac{L-1}{(L+1)^3}} \qquad \left(\frac{H+1}{H-1}\right)^H\cdot\left(\frac{L-1}{L+1}\right)^L$

$\sqrt{\frac{H+1\ (L-1)}{(H-1)(L+1)}} \qquad \frac{(H+1)(H-1)}{(L-1)(L+1)} \qquad \left(\frac{H+1}{H-1}\right)^H\left(\frac{L-1}{L+1}\right)^L$

$\frac{(H+1)^{H+1\frac{1}{2}}}{(H-1)^{-H+\frac{1}{2}}} \quad \frac{(L+1)^{-L-1\frac{1}{2}}}{(L-1)^{L-\frac{1}{2}}} \qquad \frac{H^2}{L^2}\cdot e^{\frac{1}{H}-\frac{1}{L}}$

$\frac{e^{1+\frac{1.5}{H}}}{e^{1-\frac{.5}{H}}} \quad \frac{e^{-1-\frac{1.5}{L}}}{e^{-1+\frac{.5}{L}}} \quad = \quad \frac{H^2\ e^{\frac{1}{H}}}{L^2\ e^{\frac{1}{L}}}$

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

$\frac{1}{H}+\frac{1}{H^2}$

$= \frac{2}{1-\frac{1}{H}} - 1 \ ,\ 2\ \ 1+\frac{2}{H}+\frac{2}{H^2}$

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

$-\frac{1}{L}+\frac{1}{L^3}\ \ldots$

$= 1-\frac{2}{L}+\frac{2}{L^3}-\frac{2}{L^4}\ \ldots$

So $H \rightarrow H+1$ v.s. $H \rightarrow H-1$ produces ratio of the second terms of

$\frac{H^2}{L^2}\frac{e^{\frac{1}{H}}}{e^{\frac{1}{L}}}$ for large R. But then $\frac{H^2}{L^2} \rightarrow 1$ for large R,

$\sim \frac{H^2 + H}{L^2 + L}$ So the change in E1 occurs at small values of R — say $R \sim N$.

4 x 95 = $\boxed{150}$

400 − 265 / 5

265 − 239 = 26

.28 I fixed ST83 on cc=1; Got 279,509 — Same as below. So 1. cc effect looks symmetric about

		raw
cc=11	→ 262.125	222
=12	→ same.	
16	→ 260	200
17	→ 257.75	178
18	→ 255	156
8	→ 266.54158	267
7	→ 268.756	289
5	→ 270.9706	311
10 3	→ 275.3548	356 — 26

100

1000 — $\boxed{4.2}$ forward

ΔStan. 35 $\boxed{3.8}$ ← year

A−Ū		A	265 Stan.	B	B−Ū	Ratio
134.6	1	~~[illegible]~~ 400	280	279.509	~~[illegible]~~ 14.1	9.54
90.6	3	~~99.6~~ 356	281	275.3548	9.96	9.096
45.6	5	~~[illegible]~~ 311	273	270.97	5.61	8.128
23.611	7	289	268	268.756	3.367	7.012
1.611	8	267	264	266.54	1.111	$\boxed{1.45}$!
−21.389	10	244	259	263	−2.3889	8.9
−43.4	12	222	254	262.125	−3.2639	$\boxed{13.3}$?
−65.4	16	~~−56.4~~ 200	249	260	−5.3889	12.11
−87.4	17	178	244	257.75	−7.64	11.44
−109.4	18	~~−400~~ 156	239	255	~~[illegible]~~ −10.4	10.52

3.21.99 SMRT STEIN

[illegible] is conceivable that it is [illegible]; Write pgm to calculate values exactly automatically!

Then get ~~[illegible]~~ Σ err².

I could put values into ~~the~~ Data table. I only need to calculate 2 more values.

— But this hand calcn. is not very accurate!

A correction! When we have CC>0, $\boxed{U \to (U*810 + HIT/1000)/811}$ This make some difference. ✓

Δ to my predn.		old predn.	Δ to exp data	new ratio
1:	278.7560	279.504	13,367	10.06957
3				
5				
7				
.575812 8	265.964 ✓	267.264	267	1.6112	2.8011

Recover N17 18 Jg

Processed, answd within 2 days.

In the case of CC=1 (May 4?) my predn of 279.504 → 278.756 dropd ~ .75

So maybe "[illegible] R" ↓. for small $M_i' - \bar{u}$ the effect was perhaps more reasonable?

$\boxed{ST100}$! Same as ST96 but modify to do both Hit=P & 1000.

A * 45/46 ; (A*46 + 1000)/46 = A + 1000/46

261.195
282.934
~~[illegible]~~
267

for CC=8, seems to get same answer, but when I ↑ A(8,3) why does CI ↓? Well, because an Hit is less likely than an "out".

1.06079
1.0835
1.1258

$1/\left(\frac{H0}{H1}+1\right) = H1/(H0+H1)$ $A \to \frac{A*45}{46}$

$A = (A - 1000/46) * 46/45$

$= A*46/45 - 1000/45$

BA = final result. : so I want. $R_{o.} = (BA - [illegible]1000*UU) / (A(66,3) - 1000*UU)$ ↓ div by

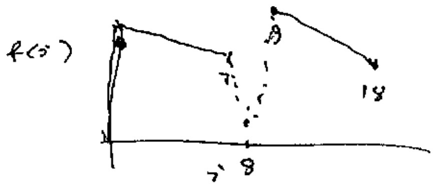
Well! The output of ST100 is Disturbing!

$f(i) \equiv \text{Avae} \frac{M_i' - \bar{U}}{U_i - \bar{U}}$ M_i' = prediction.

f(i) ≈ 9.5 for i=1 then → 8 at i = 7; drops to 2.7 for i=8

then up to 11.4 at i=9; drifts ~~[illegible]~~ monotonicly down to 9.8 at i=18.

.32



~~[illegible]~~ Also, as EPS (|G-G| < eps determines RR) goes from .001 to .01 the f(i) values change by ×1.06 to ×1.125

for EPS=.0001 values are (relatively) close to those for EPS=.001

3 22 99 SM FT STEIN.

This "ruffness" of f(z) (of 100.32) is very disturbing!

Possl. causes:

1) I should also consider the width of the distn. in the Ū direction, and ~~m~~ multiply the S by it.
2) Hardware bug(s)
3) S.W. Bug(s)
4) Some other thing!

(NB) → 114.12ff, ~~E~~ ST114.bas tests ST values ~~[illegible]~~ & gets even more "Ruffness".

→ At anyrate, what I could do is get Gsrc for Apcip — both. My ~~#~~ stuff & Stein's - Ruff tho it may be, it may have good Gsrc. Actually, it looks like equivt. R is very large, so it probly does have a good Gsrc.

Modifn. of ST100:

After BA is computed.

use ~~BA/1000~~ $BA = BA/1000$: $S2 = \prod BA^{A5 \cdot A2/1000} (1-BA)^{A5 - A5 \cdot A2/1000}$

~~S2~~ Σ ~~[illegible]~~ $A5 \cdot A2/1000 \ln BA + \ldots$

Also do it for S3 A4 (stein) instead of BA

prob'ly would be good idea almost to divide all data by 1000.

A(5) ss2
A(2) test data.

S2 = BA (my predn)
S3 = Stein! } log demarn:

$S2 = -3907.93876$ ~~S2~~ BA
$S3 = -3909.7935$ ST

So BA is better than Stein by a ~~log likelyhoods~~ log likelyhood of 1.85479 → $e^{1.85479} = \times 6.4$

Not very much better.

Try S4 = using $x = UU \cdot .9 + .1 * A(cc,3)/1000$

S4 got -3907.95274 only .014 better in log likelyhood = × 1.014 ☺

There is a no optimum ~~radius — it is close~~ wt. for UU, its close to .9 hvr.

I can speed up the function of S3 & S4 by running R from 2 to 2 $\frac{B}{N}$

Woops! I got a Gsrc for BA of -3911.47! Most of the equivt ~~B~~ R values were > 10 (except cc = 1,2,3, which were ~12)!

Using UU as estimator gives -3908.837, Better than Stein, but worse than BA.

~~It may be~~ It seems that this log likelyhood is not very sensitive!

The above results were for EPS = .01; for EPS = .001,

S2 = -3907.8531 : slightly better than EPS = .01

S3 =

EPS = .0001

S2 = -3907.84029 very slight further improvement,

using EPS = .01 and UU = UU + .0091, I got -3907.86 — which is better than corresp. EPS ~~[illegible]~~ = .01 value for UU = .265289...

(Not surprising, hvr, tho the BA values it obtained were a bit weird!

ST101 : This is an attempt to clean up ST100 by dividing all data by 1000;

~~May not work~~ It hangs up around last line of SRTIV100 !

(with EPS = .01 and UU = UU + .0091)

I got S2 = -3907.86 BA
S3 = -3909.7935 ST
S4 = -3906.765 ← (Best) Because of +.0091

Well! its getting old ~~[illegible]~~ answers!

3 23 99 SMPT STEIN

Well, I don't quite know what ~~to~~ make of it! ~~The~~ log likelyhood score seems insensitive.

Try Σ err² score.

BA S2 = .0219944 worst

→ ST S3 = .021611 Best — almost identical. So ST did better ⇒ Σ err² score!

.9,.1 mix S4 = .02194 next

BA
ST
.9,.1 mix.

The differences do seem small! — Not enough to distinguish between methods!

I don't know if this is true of the ln likelyhood score, perhaps also true!

I could easily set it up to print out all 6 scores: I need to initialize S5, S6, S7.

I inserted a wt. of A(55,5) for t. ~~Now~~ Σ err² score!

BA 6.2505 Best — Δ=.66 — Stein 6.915 — .9 & .1 wts 6.2570

5.8306 — Δ=.43 — 6.246 — U Best 5.7948 should be v.f. if model is exactly right.

If I changed data by A(55,2) → A(55,2) − .0091 so means of test & tuning data were same.

Using U as ~~se~~ as estimate. S2 w. wts. & $\boxed{\Delta \to .0091}$

S2 BA
S3 ST
S4 .9 + .1

Σ err² was 29.98 (~30) for

No: s

estimate score.
BA −3846.65
ST −3847.816
.9+.1 −3846.559 ← Best
M_2 −3869.68 — very bad
$\bar{U}$ −3847.23 better than ST!

— But this is using correct $\bar{U}$.

Log like score: −3909.6096 better than ST?

−3909.60969

Stein gets −3847.8 under these circumstances

BA gets −3876.6

~~M2 Best~~

.9M + .1Mi gets −3846.559 best

3.23.99 SMRT Stein

O.k. Recreating f. STEIN estimator

$x_i = (N)^{\frac{1}{2}} \sin^{-1}(2u_i - 1)$; $\mu_i = \bar{x} \frac{k-3}{V} + (1 - \frac{k-3}{V}) \cdot x_i$! $\Sigma = (x_i - \bar{x})^2$

$= \Sigma x_i^2 - \frac{(\Sigma x_i)^2}{k}$.

ST 103

```
S=0 , SS=0
for J=1 to k : X= 2*A(J,3) - 1
~~A(J,6)=~~ Y = ~~[illegible]~~ N^.5 * ~~[illegible]~~ ATN ( X*(~~[illegible]~~1-X^2)^-.5 )   | sin^-1 X = tan^-1 X/√(1-X^2)
      A(J,6) = Y                              ( X*(1-X^2)^-.5 )           ATN
  S = S+Y
  SS = SS+Y*Y
Next J
for J=1 to k     S = S/k ,  V= SS - S*S/k ,  W = (k-3)/V
  ~~S(J)=~~ A(J,7)= ~~S*(k-3)/V + (1-(k-3)/V) * A(J,3).~~
Next J.            \ S * W + (1-W)·A(J,3)
Print W., S
```

$\frac{W}{1}$ $\frac{S}{\sqrt{N}}$ $\tan^{-1} -.204124 = -.20135$

$\tan^{-1} -.2013$ = ~~[illegible]~~ $\sin -.2 = -.203$

$-.1987$ $-.2013$

$u_i = \left(\sin\left(\frac{X_i}{N^{\frac{1}{2}}}\right) + 1\right)/2.$

$\sin\theta = -.2$, $\cos\theta = \sqrt{1-.2^2}$

$\tan\theta = \frac{\sin}{\cos} = \frac{-.2}{\sqrt{1-.2^2}} = -.204124$; $\theta = .20135$

E and M used integer values for no. of hits per player. This accounts for fig 1 col(4) — using simple 3 signif. digs for Y_i doesn't give their x_i —

on the other hand, ~~[illegible]~~ taking the mean of these values gets $\boxed{-3.465}$! V, W, S [illegible]

they give $\boxed{-3.275}$! — probably ~~[illegible]~~ typo

So far, I got $W = \frac{k-3}{V}$ right. $= .791$ So I got these values "correctly" — just was with their values of X_i.

~~[illegible]~~

$\boxed{\frac{R + 45}{W + (1-W)}}$ ~~[illegible]~~ $\frac{W}{1-W} \cdot 45$ R : 45

The R value equivalent to W = .791 is

$\frac{W \cdot 45}{1-W} = 170.3$

$\frac{R}{45} = \frac{W}{1-W}$.

for factor (k-3) "R" = 170.3

k-2 → = 248.0

k-1 → 389.76

k-0 → 841.36

2.6234

So this integer / does seem to be important. — well, for small k. [0,1,2 ...]

3.24.99 SMRT Stein

108
1.16 to 3.23

So the wt on $\bar{u}$ is $\approx \frac{\overline{\sigma_i^2}}{\Sigma^2}$ Σ^2 is the var of the total ~~Bs~~ d.f., $\overline{\sigma_i^2}$ is the mean σ^2 of the ~~u~~ i's,

If R is large, then the ratio is close to 1, and most wt is on $\bar{u}$.

14 96d
28
23
65 d 104pp
1.6pp/d.

$\frac{2}{3} = \frac{10}{45}$

$\frac{1}{45} = \approx .02$

.03 (Try $\frac{1}{R^2}$ for apriori.)

I think these are the first sig. figures — the whole value is −3907.9387

S.2 S3 S4

.9 to 1 ...

BA	Stein		
−7.9387	−9.7935	−7.9527	LL Gore

.07 −10.3587 using $\frac{1}{R}^2$ apriori, a bit worse than Stein

I could try $R^{-1+\epsilon}$.

−6.834 using $\frac{1}{R^{1.1}}$! better than $\frac{1}{R \ln 2R}$

−6.63447 using $\frac{1}{R^{1.01}}$

−3906.~~8~~3...

−39(0)6.615 using $\frac{1}{R}$ ~~eps=.01~~ Best.

−3~~9~~934.42 $\frac{1}{R}$ ~~eps=.001~~ ?

Very bad!

.14 −3907.9667 $\frac{1}{2R}$ eps = .1 ~~Probably eps=.01~~

But note that these are all not what they seem because they use the $\frac{e^{\epsilon}}{\ln RR}$ addition to the S, which is ~~[illegible]~~ not appropriate.

These are ~~[illegible]~~ ← All using EPS=.01

← EPS=.001

I think I did a little of this because of the Stein way, but I'm not sure.

It may be based on χ^2 d.f., which is d.f. for squares of vars. this χ^2 is easily genzd to non-equal vars, hvr — E&M may not know about it. I think they were trying to estimate the ~~Ese~~ ~~[illegible]~~ Var of the u_i generator. ([illegible]) They got a ~~d.f. for~~ d.f. (perhaps) for σ^2 of gen $= \frac{1}{R}$ so, this is like using $\frac{1}{R^2}$ a priori. — Actually I don't know what prodns $\frac{1}{R^2}$ would give (.07) is doesn't integrate from RR to ∞ properly — it would be easy to do, hvr. ~~[illegible]~~ .. $\int \frac{1}{R^2} dr = \frac{1}{R}$ just use ~~$\frac{1}{R}$~~ for the S instead of $\frac{1}{\ln R}$.

I got −3910.905 even worse than .07! — but ≈ Stein!

If σ has a uniform d.f. what is d.f. of σ^2. ? of $\frac{1}{\sigma^2}$?

(SN) Perhaps the "irregularity" of the BA results has to do with the ~~[illegible]~~ spacing of the u_i! That the estimates of u_i depends much on u_{i-1} & u_{i+1} as well as u_i (unlikely!)

	← no. ~~[illegible]~~ of hits
1	18
2	17
3	16
4	15
5	14
6	14
7	13
8	12
9	11
10	11
11	10
12	10
13	10
14	10
15	10
16	9
17	8
18	7

These are a way to get more exact values of A(J, 3) --- These were used by E&M

3.24.99 SMPT **STEIN** log

More Robson my "BA" soln ① Perhaps cc=8 got unusual results because it's RR value was significantly diff'nt from other cc's: An argt against this! — that the effect should be less prominent as EPS becomes smaller ($\equiv$ RR ↑) —— This doesn't occur.

.04 ② E&M: They assume a Gaussn for the μ_i's w. unknown mean & varc: then the observed μ_i' are of the "known" varcs.

My (old) analysis of the χ^2 situation: That if a var x was Gaussn, about zero, one knows the d.f. for x, one also knows the d.f. for x^2. If one has a bunch of Gaussn. vars. of various σ^2, the d.f.s of the sum of their squares is the convolution of the individual x^2 d.f.s. Also, the varcs of these d.f.s add.

- If the D.f.s are all about zero, one knows the D.f. of their sum.

Now, if one choses the means of those random vars, as in .04 (using a Gauss d.f.) & the varcs of the random vars are all the same, then the final d.f. of the vars is a Gaussn. Say the d.f. of μ's is $N(\bar{\mu}, \overset{\text{mean, varc.}}{\tau^2})$ & the individual vars are all $N(0, \sigma^2)$. Then the final d.f. is $N(\bar{\mu}, \tau^2+\sigma^2)$. Now Given a sample D.F. and known σ^2, to find $\bar{\mu}$ & $\hat{\tau}^2$. The mean of the sample D.f. can be used for $\bar{\mu}$.

I should think that a good estimator for $s^2+\sigma^2$ would be "approx" $(\tau^2+\sigma^2)^{u} * \frac{k}{k-1}$.

"$(\tau^2+\sigma^2)$" est $\equiv \dddot{V}$ used by E&M. I should think that $V-\sigma^2$ would be a good estimate of τ^2.

(In E&M, $\sigma^2 = 1$ so $\frac{V}{k-1}$ would be the estimator!! Instead they use $\frac{k-3}{V}$

Also in eq. 1.10 $S = \sum(x_j - \mu_j)^2$ is "approx τ^2+1". & they use $\frac{S}{k-2}$ to estimate τ^2+1

estimate of μ_i in 1.8 is $\mu_i + (x_i - \mu_i') - \frac{x_i-\mu_i'}{1+\tau^2} = x_i - \frac{x_i-\mu_i'}{1+\tau^2}$

.21 $= x_i\left(1 - \frac{1}{1+\tau^2}\right) + \frac{\mu_i}{1+\tau^2} = x_i \frac{\tau^2}{1+\tau^2} + \mu_i \cdot \frac{1}{1+\tau^2} = \frac{x_i}{1+\frac{1}{\tau^2}} + \frac{\mu_i \cdot \frac{1}{\tau^2}}{1+\frac{1}{\tau^2}}$

We use (.21 R) because $\frac{1}{\tau^2}$ & 1 are relative wts.

Analogously, in the baseball case, I minimize

$$\tau^2 = \frac{\sum(\mu_i' - \bar{\mu})^2}{K-3} - \frac{\sum \mu_i'(1-\mu_i')}{K\ N} \quad \Big| \quad \sigma_i^2 = \frac{\mu_i'(1-\mu_i')}{N}$$

w. the $\underset{\text{estimate}}{\mu_i} = \left(\frac{\bar{\mu}}{\tau^2} + \frac{\mu_i'\ N}{\mu_i'(1-\mu_i')}\right) \Big/ \frac{1}{\tau^2} + \frac{N}{\mu_i'(1-\mu_i')}$ → 106.01

.27

.28 $\boxed{\text{A POSSL TRICK}}$! I'd like to integrate the entire D.f. wrt $\bar{\mu}$: It may be possible to do this analytically — then do Num. integ. wrt R. See $\boxed{106.25 \text{ for}}$ eq.

$$\rightarrow \prod_i \left(\frac{R+1\,!}{\boxed{\bar{\mu}R\,!\,(1-\bar{\mu})R\,!}} \; \frac{\boxed{(\bar{\mu}R+A_i)!\,((1-\bar{\mu})R+B_i)}}{R+N+1\,!} \right) : \frac{R+1!}{R+N+1!} \int_{\bar{\mu}=0}^{\bar{\mu}=1} \prod_i \frac{\bar{\mu}R+A_i!}{\bar{\mu}R\,!} \, \frac{((1-\bar{\mu})R+B_i)!}{(1-\bar{\mu})R}$$

No obvious way to do! So it width in $\bar{\mu}$ direction varys much w. R.

Q: check again on whether this factor really deserves the $(\ )^k$ weight! — I prob'ly does — i.e. the dep'nd of R ($\bar{\mu}$) that goes outside the $\prod_i$ sign.

3.26.97 SMHT Stu ii:

101105.27

ST106 In/ ST106 I tried $\left(A(J,\tfrac{4}{3}) \to 1 - A(J,\tfrac{4}{3}) \right)$

Then BA = 1 − BA. , EPS = .01 Also U = 1 − .2653889,

There is no ST105!

ST106 is the simple modification of ST101

[illegible]

[illegible] Then if we do BA = 1 − BA, the output (for EPS = .01) is ~~the~~ same as ST105 to 13 decimal places) (for the GORC).

.05 Thinking about its form, changing μ_j^i to $1-\mu_j^i$ leaves the integrand in ST101 invariant.

The integrand consists of pairs of factors: one involves μ_j^i, the other ~~[illegible]~~ factor is the same except $\mu_j^i \to 1-\mu_j^i$. So, it is true.

Is it true of STEW's method? − (unlikely!)

Try it on ST103 w. this xform: Z word, seems to work exactly!!

$N^{-\frac{1}{2}} \cdot \sin^{-1}(2x-1)$; $x \to 1-x$ $(2(1-x)-1)$ $2-2x-1 = (1-2x)$

So we simply get $\sin^{-1}(1-2x) = -\sin^{-1}(2x-1)$ (This just makes the data negative, but otherwise the probs are the same. — So it should give exactly same results — as it does.

How the xform $N^{-\frac{1}{2}} \sin^{-1}(2x-1)$ works: (I think) : since the Sd of x is $N^{-\frac{1}{2}}\sqrt{x(1-x)}$

we want $f(x)$ ∋ its derivative will be $\propto \frac{1}{\sqrt{x(1-x)}}$; we note that $\to 2x-1$ xforms $\sqrt{1-x^2}$

into $2\sqrt{1-x^2}$, $\int$ of $\sqrt{1-x^2}$ & $\int \frac{1}{\sqrt{1-x^2}}$ are well known — one of them may be $\sin^{-1}$.

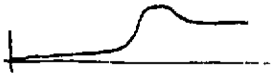
$\int = \sqrt{1-x^2}$

$\frac{d}{dx} \sin^{-1} x = \frac{1}{\sqrt{1-x^2}}$ $\int \sqrt{1-x^2} \sim \frac{1}{2}\left(x\sqrt{1-x^2} + \sin^{-1} x \right)$

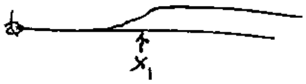
How $\int \frac{1}{\sqrt{1-}}$

$$\boxed{\frac{d}{dx} \sin^{-1}\overset{2x-1}{(2x-1)} = \frac{2}{\sqrt{1-(2x-1)^2}} = \frac{2}{\sqrt{-4x^2+4x}} = \frac{1}{\sqrt{x(1-x)}}}$$

Getting back to the aprop'd Q! ~~In~~ In general, as ssz of data ↑, ~~the~~ its r.d. → $(P_1(x))^{ssz}$.
$P_0(x)$ = aprop'b is, indep of ssz. If $P_1(x)$ has a peak > all other peaks, ~~[illegible]~~ at $x = x_0$
$(P_1(x))^{``\infty"} = \delta(x-x_0)$: $\int P_1(x)^{\infty} \cdot P_0(x)$ is essentially indep of $P_0(x)$. — it is $P_0(x_0)$ — which cancels out after normzn.

If $P_1(x)$ is like  Then ssz must be very large before we get a δ funct approxn.

If $P_1(x)$ is  then for ssz = ∞ we may get  exactly

or, conceivably, a  funct in which x_1 & the width of the step depends on ssz. } Much more common.

We are, hvr, interested in the ratio of $\int P_1(x) P_0(x)dx : \int P_2(x) P_0(x)dx$

.01 One sees ~~[illegible]~~ earlier, simpler approach to STEIN: Assume that each μ_i' is of known varc, $\sigma_i'^2$, but unknown mean, μ_i. The $[\mu_i]$ are generated by a norm d.f., of unknown $\bar{\mu}$, Σ^2.

Each datapt, μ_i' has its ~~[illegible]~~ own d.f., of mean $\bar{\mu}+\mu_i$ & ~~[illegible]~~ varc $= \Sigma^2+\sigma_i'^2$.

~~[illegible]~~ Each $(\mu_i')^2$ also has a d.f. of mean & varc, that's a function of $\bar{\mu}$, Σ^2, μ_i, σ_i^2. (Is its mean $\Sigma^2+\sigma_i^2+(\bar{\mu}+\mu_i)^2$? — or is $\Sigma^2=(\bar{\mu}+\mu_i)^2$? — I think $\Sigma^2+\sigma^2$ is more likely → same)

Anyway, if we know the μ & σ^2 of a norm d.f. we know the μ & σ^2 of its square, x^2.

.08 Assume the mean of sq. of (a norm d.f. about zero, w. varc $= \Sigma^2$) is say Σ^2.

Then convolute ~~[illegible]~~ Woops! Actually, I'd have to ~~[illegible]~~ find the d.f. directly —

But Anyway, given the mean's & varcs of the d.f.s for the squares of the μ_i''s, the d.f. for the sum of the squares is known. From the observed mean & ~~[illegible]~~ varc of the d.f. of this sum, one might be able to ~~[illegible]~~ estimate $\bar{\mu}$ & Σ^2 (No $\bar{\mu}$ is pretty well estimated as $\overline{\mu_i'}$ — so we only need to find Σ^2.)

In obtaining the varc. one can do the $k \to k-1$ or $\to k-2$ — if I find its rationale.

I do want to see what kind of answers I get using Mt. Carlo & various D.F.'s for K. — for BA & ST.

WRT a single datum μ_i': Its expected square is:

$\bar{\mu}^2+\Sigma^2+\sigma_i^2$: We then find $\Sigma(\mu_i')^2$ we expect it to be $\Sigma\left(\bar{\mu}^2+\Sigma^2+\sigma_i^2\right)$

$\Sigma\mu_i'^2$ " = " $k\left(\bar{\mu}^2+\Sigma^2\right)+\Sigma\sigma_i^2$. so $\Sigma^2 = \frac{\Sigma\mu_i'^2-\Sigma\sigma_i^2}{k} - \bar{\mu}^2$

.25 $$\Sigma^2 = \frac{\Sigma\mu_i'^2-\frac{\Sigma\mu_i'(1-\mu_i')}{N}}{k} - \bar{\mu}^2 = \underbrace{\frac{\Sigma\mu_i'^2}{k}-\bar{\mu}^2}_{\alpha} - \underbrace{\left(\frac{\Sigma\mu_i'(1-\mu_i')}{k\cdot N}\right)}_{\beta/N}$$

$$\frac{(1+\epsilon)^2+(1-\epsilon)^2}{2} \qquad \frac{2+2\epsilon^2}{2},\ 1+\epsilon^2.$$

$$= \frac{\Sigma\mu_i'^2}{k} - \frac{\Sigma\mu_i'^{\bar{\mu}}}{k\cdot N} + \frac{\Sigma\mu_i'^2}{kN} - \bar{\mu}^2$$

$$= \frac{\Sigma\mu_i'^2}{k}\left(1+\frac{1}{N}\right) - \frac{\bar{\mu}}{N} - \bar{\mu}^2$$

.30 $$= \overbrace{\frac{\Sigma\mu_i'^2}{k}-\bar{\mu}^2}^{\alpha} + \overbrace{\frac{\frac{1}{k}\Sigma\mu_i'^2-\bar{\mu}}{N}}^{\beta/N}$$

$$\frac{\Sigma\mu_i'^2}{k}\cdot\frac{1}{N} - \overset{\bar{\mu}}{\left(\frac{\Sigma\mu_i'}{k}\right)}\cdot\frac{1}{N} = \left(\frac{\Sigma\mu_i'^2}{k}-\bar{\mu}\right)\cdot\frac{1}{N}$$

.30 4.583904E-3 $= \alpha =$ "sig" in ST81 (81.13)

.32 $\boxed{\text{Mean } \sigma_i^2 = \frac{1}{k}\Sigma\left(\mu_i'(1-\mu_i')\right) = \cancel{.0619314}\ (81.07)} = \bar{\mu} - \frac{1}{k}\Sigma\mu_i'^2$ β

No! More like: ~~[illegible]~~ .1903737E

σ^2 sig

sig $-\bar{\sigma}^2$

$\bar{\mu}$ − sig

: $\frac{\beta}{N}$ = mean σ_i^2

.33 $\Sigma^2 = \boxed{\alpha-\frac{\beta}{N}}$ from (.25)

so $\boxed{\Sigma^2 = \left(\text{varc of total raw corpus}\right) - \text{mean } \sigma_i^2 \text{ of the } [\mu_i'] \text{ set}}$

$\beta/45 = 4.23052716$

$= .003207650$; $\frac{1}{S^2} = 311.7546$ / if I use $\alpha \to \alpha\frac{k}{k-1}$; $\frac{1}{S^2} = 287.58$ / for $\alpha\to\frac{k}{k-2}$: $\frac{1}{S^2} \cong 263$

~~[illegible]~~ We also need corrections relating ~~[illegible]~~ raw to ~~[illegible]~~ actual varcs.

So varc of another conf' will be at least ~~[illegible]~~ $\frac{k}{1-k}$ times as big as α.

(I have $\tau^2+(\text{mean}\sigma_i^2)=\alpha$)

~~[illegible]~~ E & M had $\sigma_i^2 \equiv 1$, so ~~[illegible]~~ $\tau^2 = \Sigma^2 = \alpha - \frac{1}{N}$ and $\boxed{1+\tau^2=\alpha}$

E & M used $\frac{k}{k-2}$ or $\frac{k}{k-3}$ for correction factor. I don't know why!

I could test it w. Mt. Carlo data.

$2\frac{42}{N}$ way to ... $\frac{1}{N}$ $\frac{k}{N}$

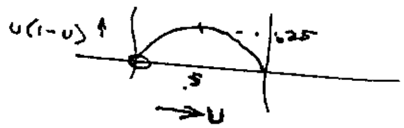
;01 estimate of $\mu_i = \left(\frac{\mu_i' \cdot N}{\mu_i'(1-\mu_i')} + \frac{\bar{\mu}}{\tau^2}\right) \Big/ \left(\frac{N}{\mu_i'(1-\mu_i')} + \frac{1}{\tau^2}\right) = \left(\tau^2 \mu_i' + \frac{\mu_i'(1-\mu_i')\cdot\bar{\mu}}{N}\right) \Big/ \left(\tau^2 + \frac{\mu_i'(1-\mu_i')}{N}\right)$

$= \frac{\mu_i'}{1+\frac{\mu_i'(1-\mu_i')}{N\tau^2}} + \frac{\bar{\mu}}{1+\frac{N\tau^2}{\mu_i'(1-\mu_i')}}$

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

So calculate $\frac{\mu_i'(1-\mu_i')}{N\tau^2}$ for each μ_i' ~~[illegible]~~

So how much does $\mu_i' - \bar{u}$ shrink estimate μ_i' ?



~~[illegible]~~ $\mu_i = A\,\mu_i' + B\,\bar{U}$: $\mu_i - \bar{U} = A\,\mu_i' + (B-1)\cdot\bar{U}$
$= A(U_i' - \bar{U})$

A+B=1
So B-1=-1

$\frac{\mu_i - \bar{U}}{U_i' - \bar{U}} = A = \frac{1}{1+\frac{\mu_i'(1-\mu_i')}{N\tau^2}}$ Which is a rather smooth funct of μ_i.

$\frac{\mu_i'(1-\mu_i')}{N\tau^2} \sim 10.$

I got $\frac{1}{\tau^2} = 311.7$ [287 for $\alpha \to \alpha \frac{k}{k-1}$ (or 242 for $\alpha \to \alpha\frac{k}{k-3}$ (k=18).

for $\mu_i' = .2$ to .4

[redacted] $\frac{311.7}{45} = 6.926$; $\left(~~[illegible]~~ \times .13166 + 1\right) = \cancel{2}1.91196$ ~~[illegible]~~

$\mu_i'(1-\mu_i') = .2 \times .8$ to .4·.6
16 to 24 so × 1.5

.156 → $\boxed{.131664}$ | .4 = .29
So factor of ~ 2

~~[illegible]~~ ×.24 +1 = 2.66

So the ~~[illegible]~~ ranges ~~[illegible]~~ from ~~[illegible]~~ 1 to 1.59 ; It is monotonic from from .156 to .90

My results are not monotonic in that region & I got a dip at μ_i = .267

I should compare this simple result w. STEIN (using the SMT xfund).

These results do seem ~~[illegible]~~ a lot different from ~~[illegible]~~ STEIN's from ~~[illegible]~~ BA.

$\frac{45}{.2}$ = 225 so $\frac{1}{\tau^2 \sigma^2}$ even as long as 311 doesn't give $\bar{u}$ ~~[illegible]~~ ~~nearly~~ nearly as much wt. as STEIN or BA.

I ~~found~~ found a post hoc ratio of ~ 1 to 9 (or 10?) to be optimum.

Perhaps it would be good to check on α & β values!

$\boxed{\Sigma^2 = \alpha - \frac{\beta}{N} = \alpha - \frac{\beta}{45}}$ 108.33

Error! → α = 4.582564 E-3 or. e.g.

ST109 (from ST81 see 108.30, 32 for defn of α, β)

→ $\beta/45$ = 4.2305 2716

(k): $\left(\alpha - \frac{\beta}{45}\right)^{-1}$ = 2829.83 ← $\frac{1}{\tau^2}$

(k-1): $\left(\frac{\alpha\cdot 18}{17} - \frac{\beta}{45}\right)^{-1}$ = 1605.08 → 7.13 → 8.13

(k-2): $\left(\frac{\alpha\cdot 18}{16} - \frac{\beta}{45}\right)^{-1}$ = 1079.48 → 4.79 → 5.79

(k-3): $\left(\frac{\alpha\cdot 18}{15} - \frac{\beta}{45}\right)^{-1}$ = 787.30 3.5 → 4.5

↑ $\frac{1}{\tau^2}$

~~[illegible]~~, multiply $\frac{\mu_i'(1-\mu_i')}{N}$ to get ~~[illegible]~~ ratio

→ 12.5 +1 = 13.5 ~ $\frac{.2}{45} = \frac{1}{225}$

Since we want ~ $1+\frac{\mu_i'(1-\mu_i')}{N\tau^2} \approx 9$

$\alpha \to \alpha\frac{18}{17}$ is not bad.

Woops! This is all wrong: ST81 makes up random data! R was set at 100

So true $\frac{1}{\tau^2}$ was $\frac{100}{\bar{U}(1-\bar{U})} = \boxed{512.93}$ so $\alpha \to \frac{\alpha\cdot 18}{15}$ would have been closest.

Woops on second thot, the pgm. reads actual Baseball data, so it's OK. (The random part is not used unless we "comment out" the "READ" statement)

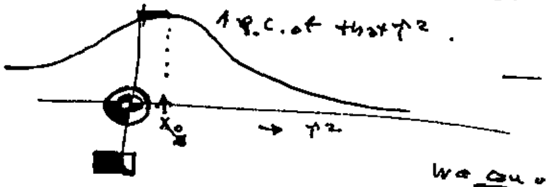
Even so, I ran ~~[illegible]~~ ~ 8 times! Got
$\frac{1}{\tau^2}$ = 467, 629, 1274, 339, 363, 356, -16080, 624. : not unreasonable.

These were all for $\alpha \to \alpha\frac{18}{15}$ denom. ($\alpha \to \alpha$ would have more negative values)

3.29.99 SMFT STEIN. 110

Well: The ~~τ^2 value I got is at~~ $\overset{x_0=}{\frac{\alpha \cdot 18}{15, 16 \text{ or } 17}} - \frac{\beta}{45}$ is only a guess of τ^2.

It does, however, give us a d.f. for τ^2. If $x_0 < 0$, the D.f. for τ^2 is close to ϕ.



— we have to mult this d.f. by the aprio of τ^2, which cuts off the neg. part. — If the d.f. is Gaussian, (~~[illegible]~~)!

We can use uniform aprio. for $\tau^2 > 0$, tho we may want to cut off τ at say .2 (since the Batting Avgs are like .2 to .4.) : we may want ~~[illegible]~~ $E(\tau^2)$.

The above analysis ~~[illegible]~~ is of uncertain value, unless τ^2 is close to ϕ or τ^2 has a narrow enuf peak ~~[illegible]~~ for that peak to be used as an estimate of τ^2.

A more exact method would be much less "ad." — i.e. ~~[illegible]~~ we would use all possib values of τ^2.

~~[illegible]~~ [redacted] Probably the best way would be closer to the "BA" method, in which ~~I get total~~ I compute ~~total~~ aprop for ~~[illegible]~~ corpus w. & w.o. a "north[illegible]".

~~$E(\tau^2)$ is [illegible]~~ Most likely τ^2 is most ad.

$E\tau^2$ is [illegible].

$E(\mu_i) = 109{,}01$: ~~the~~ it is of form $E\left(\frac{\tau^2 \mu_i + B_1}{\tau^2 + B_1}\right)$

i.e. $\mu_i^? E\left(\frac{\tau^2 + A_1}{\tau^2 + B_1}\right)$ A_1, B_1 are constants — equivalently $E\left(\frac{1 + \frac{A_1}{\tau^2}}{1 + \frac{B_1}{\tau^2}}\right)$

.21 ~~[illegible]~~ One computes it from $\boxed{E\left(\frac{1}{\tau^2 + B_1}\right).}$ and $E\left(\frac{\tau^2 + B_1}{\tau^2 + B_1}\right) = E(1)$

dz

to" dos floppy
dz" dos w floppy
~~[illegible]~~
c4" BWSE

So we just shift $\frac{1}{\tau^2}$ in the ~~[illegible]~~ direction by B_1

$$\int_0^\infty e^{-\frac{(\tau^2 - x_0)^2}{2(s^2)}} \cdot \tau^2 \cdot \frac{1}{\tau^2 + B_1} \, d\tau$$ which is .21, amounts to $$\int_0^\infty e^{-\frac{(\tau^2 - x_0)^2}{2 s^2}} \cdot \frac{1}{\tau^2 + B_1}$$

equiv to $\int_{x_0}^\infty e^{-\frac{x^2}{2(s^2)}} \frac{dx}{x + C_1}$ x_0 & C_1 are [redacted] 2 params. (or perhaps $\frac{x_0}{s}$ & $\frac{C_1}{s}$)

.25

3.30.99 ① AH! ~~[illegible]~~ Consider M.Carlo data: As $N \to \infty$, the d.f. for R should become quite sharp! Essentially we have these ~~[illegible]~~ "exact" values for the $[\mu_i]$ set and we want their ave. If the Eqns do not (analytically or "experimentally" (≡ M.Carlo simulation)) show this, then there is something wrong w. the eqns.

② Note also, that while I have tried largish N for the Baseball data sets I've not tried large k for the Mt.Carlo set. Larger k can have a somewhat different effect on the R d.f. than large N (?).

So; do a theoretical analysis of $N \to \infty$ :: see what the d.f. looks like for literal $R, \bar{\mu}$. — The d.f. will have 2 params: Then for finite N, try to match the $R, \bar{\mu}$ for that D.f. (usually) the $\bar{\mu}$ will have a narrow d.f. so its more like 1 param or by the d.f.

The factor of most interest in 46.28 is $\boxed{\prod_{i=1}^{k} \frac{(x+A_i)!(y+B_i)!}{(x+y+N+1)!}}$

The factor $\left(\frac{x+y+1!}{x!y!}\right)^k$ is relatively simple, in terms of $\bar{\mu}$ & R.

.02 $\frac{A!\,B!}{A+B)} \approx$ [illegible] $\left(\mu^{\mu}(1-\mu)^{1-\mu}\right)^{N}$: $N \equiv A+B$; $\mu = \frac{A}{N} = \frac{A}{A+B}$.

$\frac{68}{26}$
42

.03 (SN) Consider $u=1$ [for finite N we expect ~~R~~ the R d.f. to be mainly at ∞.

There are factors : $\approx \frac{x+y!}{x!\,y!}$ & ↑ so for $\bar{R}=\infty$: $\left(u^u(1-u)^{1-u}\right)^{R} \cdot \left(u^u(1-u)^{1-u}\right)^{N}$

which → ~~∞~~ exponentially as $\underbrace{\left(u^u(1-u)^{(1-u)}\right)}_{\text{}}^{R-N}$ Presumably ~~...~~ < 1.

Here, the exponent is $R-N$: for $R<N$, the ~~...~~ value is ~~...~~ small.

3978400

Or just try fitting $\bar{\pi}$, R, and the data set $[\mu_i]$ — assuming all the μ_i were exactly

Correction (i.e. $N=\infty$)

.10 $pc = \prod_{i=1}^{n} u_i^{\mu R}\,(1-v_i)^{(1-\mu)R}$ = [illegible] $A \equiv \prod v_i$; $B \equiv \prod(1-v_i)$ | $A<1$ $B<1$,

$pc \cong \left(A^u B^{1-u}\right)^{R}$ Perhaps put in Norm. z. constant for x, y!

$= \left(A'^u B'^{(1-u)}\right)^{nR}$ $A' \equiv A^{\frac{1}{n}}$, $B' \equiv B^{\frac{1}{n}}$ 1495

$\frac{x+y+1}{x!\,y!}$! ≈ $x+1 \cdot \left(u^u(1-u)^{(1-u)}\right)^{-R}$ $u > A$ $(1-u) > B$ $\frac{A+B}{2}$ $(AB)^{\frac{1}{2}}$

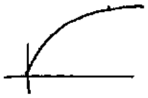
$\sqrt{\frac{R}{R\cdot u\cdot R(1-u)}}$

so

$\propto \sqrt{R}$ $\sqrt{R} \cdot (1+\epsilon)^{-R}$

R is for U

So $\frac{1}{n}\sum \ln A_i <$ ~~...~~ $\ln \frac{1}{n}\sum A_i$

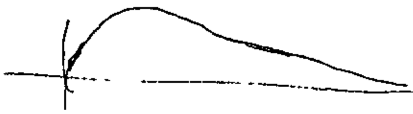


log of mean > mean of log.

.20 So, say $A' = A^{\frac{1}{n}}$, $B' = B^{\frac{1}{n}}$! $1-\epsilon = \boxed{\left(\frac{A'}{U}\right)^{u}\left(\frac{B'}{1-U}\right)^{1-u}}$ } Perhaps wrong! See 112.13

.21 $pc \propto$ ~~...~~ $(1-\epsilon)^{R} \cdot R^{\frac{1}{2}}$

so ϵ determines the peak & width of the d.f.



$pc \sim e^{-\epsilon R} R^{\frac{1}{2}}$; ~~...~~ . ln : [redacted]

$\ln pc = -\epsilon R + \frac{1}{2}\ln R$: ~~...~~ $\frac{d}{dR} = -\epsilon + \frac{1}{2}\frac{1}{R} = 0$

So $R = \frac{1}{2\epsilon}$ We can prob'ly estimate ϵ from the var. of the data set $[\mu_i]$.

We could see if it is close to ~~...~~ R that we'd get from the known σ^2 of $[\mu_i]$.

Prob'ly easier: Just check by finding $(2\epsilon)^{-1}$ of the Baseball data

Then compare it w. σ^2 of $[\mu_i]$ data set.

$\int_0^\infty t^z e^{-t}\,dt = z!$

.32 $\boxed{\text{correction}}$ σ^2 of data set should be $= \frac{u(1-u)}{R}$, $= u(1-u)\cdot 2\epsilon$

.33 → The peak is very broad, here: $\boxed{e^{-R}\cdot\sqrt{R}}$: what is width?

All wrong! See 112.01

The first 3 moments are $\frac{1}{2}!$ $1\frac{1}{2}!$ $2\frac{1}{2}!$

$\mu =$ [illegible] $\frac{m_1}{m_0} = \frac{3}{2}$ $\Gamma(1\frac{1}{2})$ ~~...~~ $\Gamma 1\frac{1}{2}$ $= \sqrt{\pi}$, $\frac{1}{2}\sqrt{\pi}$

$\sqrt{\pi}$ $\frac{1}{2}\sqrt{\pi}$

$\sigma^2 = \frac{m_2}{m_0} - \left(\frac{m_1}{m_0}\right)^2 = \frac{3}{2}\frac{5}{2} - \left(\frac{3}{2}\right)^2 = \frac{3}{2}$ ~~...~~

~~...~~ ∴ $\mu = \frac{3}{2}$; but peak is $\frac{1}{2}$!

$\sigma^2 = \frac{3}{2}$

~~$\sqrt{\frac{3}{2}}$~~

$\frac{1}{2}! = m_0 = \frac{1}{2}\sqrt{\pi}$

$1\frac{1}{2}! = m_1 = \frac{1}{2}\sqrt{\pi}\cdot\frac{3}{2}$

$2\frac{1}{2}! = m_2 = \frac{1}{2}\sqrt{\pi}\cdot\frac{3}{2}\cdot\frac{5}{2}$

3 31 99 SMRT Stem

.07 ~~Since $\int_0^\infty x^n e^{-x} dx = n!$~~ ~~$m_0 = -\frac{1}{2}!$~~ ~~$m_1 = \frac{1}{2}! = (-\frac{1}{2}!) \times \frac{1}{2}$~~ ~~$m_2 = \frac{1}{2}! \times \frac{1}{2} \times \frac{3}{2}$~~

~~$\mu = \frac{m_1}{m_0} = \frac{1}{2}$~~

~~$\sigma^2 = \frac{m_2}{m_0} - \mu^2 = \frac{1}{2} \cdot \frac{3}{2} - (\frac{1}{2})^2 = \frac{3}{4} - \frac{1}{4} = \frac{1}{2}$~~

Wrong!

111.33 is correct

Yes, $\int_0^\infty x^n e^{-x} dx = n!$ but

No! 111.33 is right:

f. funct is $e^{-R} \cdot R^{\frac{1}{2}}$

$m_0 = \frac{1}{2}!$

$m_1 = 1\frac{1}{2}! = m_0 \cdot \frac{3}{2}$

$m_2 = 2\frac{1}{2}! = m_0 \cdot \frac{3}{2} \cdot \frac{5}{2}$

$\mu = \frac{m_1}{m_0} = \frac{3}{2}$

$\sigma^2 = \frac{m_2}{m_0} - \mu^2 = \frac{3 \cdot 5}{2 \cdot 2} - \frac{3 \cdot 3}{2 \cdot 2} = \frac{6}{4} = \frac{3}{2}$

So $\mu = \frac{3}{2}$; $\sigma^2 = \frac{3}{2}$, as 111.33 : hvr., peak is $\frac{1}{2}$!

The peak of $x^{\frac{1}{2}} e^{-x}$ = peak of $\frac{1}{2} \ln x - x$: $\frac{1}{2x} = 1$! $x = \frac{1}{2}$

So, no matter how large k is we don't have ~~much~~ any more certainty in σ^2!! This sounds crazy!

~~: contrast with Gaussian fitting~~

.13 OH! I may see it! 111.20 is wrong. $1 - \epsilon = \left(\left(\frac{A'}{U}\right) \cdot \left(\frac{B'}{1-U}\right)\right)^k$ $\left(= \frac{A}{U^k} \cdot \frac{B}{(1-U)^k}\right)$

So in 111.28 : $\rho c = (1-\epsilon)^{Rk} \cdot R^{\frac{1}{2}} = e^{-\epsilon R k} R^{\frac{1}{2}}$: ϵ depends on var. of $[\mu_i']$ dist.

$\ln \rho c = -\epsilon k \cdot R + \frac{1}{2} \ln R$! $\epsilon k = \frac{1}{2R}$: $\boxed{R_0 = \frac{1}{2\epsilon k}}$ is peak.

.18 $\mu = 3R_0$ $\sigma^2 = 3R_0$ also.

$\rho c = \left(\frac{A'}{U} \cdot \frac{B'}{1-U}\right)^{kR} \cdot \sqrt{R}$

.18 is unreasonable! as $3R_0$ becomes large,

Its $\frac{\sigma}{\mu}$ is $\frac{\sqrt{\mu}}{\mu} = \frac{1}{\sqrt{\mu}}$ becomes small

– in which case we expect μ & peak to be close.

They are not!

This is also wrong! $\rho c = (1-\epsilon)^{kR} R^{\frac{1}{2}}$.

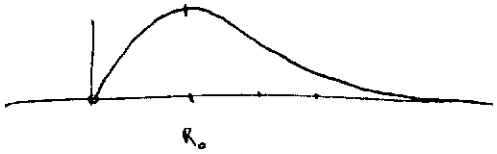
$\left(\frac{A'}{U} \quad \frac{B'}{1-U}\right)$ is expected to be close to 1

& ~~and~~ ≐ indep of R.

So ↳ = $1 - \epsilon$ from .32 ~~[illegible]~~ we expect.

var of $[\mu_i'] = \frac{U(1-U)}{R}$

But do we use the peak or 3 times the peak, for R ?



ST 112 :

```
S0=0; S1=0, S2=0
for X = 1 to ~~[illegible]~~ step .001
                  └same to 80
  Y= Exp(-x) * sqr(X)
  S0=S0+Y
  S1=S1+Y*X
  S2=S2+Y*X*X
Next U= S1/S2
Print U, S2/S0 - U*U
      1.5        1.5
```

I got

~~U= 2.224758~~

~~[illegible]~~

~~1.336485~~

limit 80 ; step .001

~~$\int e^{-x} \sqrt{x}$~~

U = 1.5

σ^2 = 1.5

peak is 1.5 = as calculated

* xy $(e^{-x} \cdot x^{\frac{3}{2}})$ has a peak at

$x = 1.5 = \frac{3}{2}$

$(e^{-x} x^{\frac{5}{2}})$ has peak at $\frac{5}{2}$.

as calculated. $(e^{-x} x^n)$ has peak at $x = n$.

$-x + n \ln x = 0$

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

X	.25	.5	.75	1	1.5	2	2.5	3	4	5	6	7
Y	.4	.42	.41	.367	.27	.19	.13	.086	.036	.015	.006	.0024

3.31.99 SMR Stein.

Going over 111.10ff more carefully

$$pc = \prod_{i=1}^{k} \frac{x+y+1!}{x!\,y!} u_i^{x} (1-u_i)^{y}$$

$$(R+1)\sqrt{\frac{R}{R\cdot u\cdot R(1-u)}}\left(u^{u}\,1-u^{1-u}\right)^{-R}\cdot u_i^{uR}\,(1-u_i)^{(1-u)R} \propto (\sqrt{R})^{k}\prod_i\left(\left(\frac{u_i}{u}\right)^{u}\cdot(1-u_i)^{1-u}\right)^{R}$$

for $R>>1$. ~~$p_c \propto R^{\frac{1}{2}}$~~.

So I forgot to put $R^{\frac{1}{2}}$ inside the $\prod_{i=1}^{n}$ so it's $R^{\frac{k}{2}}\cdot\left(\left(\frac{A'}{u}\right)^{u}\cdot\left(\frac{B'}{u}\right)^{1-u}\right)^{kR}$.

I.e. $\left(R^{\frac{1}{2}}\;(1-\epsilon)^{R}\right)^{k} \approx R^{\frac{1}{2}}e^{-\epsilon R}$ $\qquad (1-\epsilon) = \left(\frac{A'}{u}\right)^{u}\cdot\left(\frac{B'}{u}\right)^{1-u}$

Looks more reasonable: $R^{\frac{1}{2}}e^{-\epsilon R}$ has a peak at $R_0 = \frac{1}{2\epsilon}$; $\frac{\text{first moment}}{\text{zeroth moment}} = 3R_0 = \mu$

for $k=1$ the ~~width = second moment~~ σ = about μ is ~~[illegible]~~ $\sqrt{\frac{3}{2}}\cdot R_0$

However, as $k \uparrow$, the peak stays at $R_0 = \frac{1}{2\epsilon}$, but the d.f. gets narrower and narrower

so the μ moves toward the peak & σ gets small.

This is the kind of behavior I expected!

$$\int_0^{\infty} R^{n}e^{-\alpha R}\,dR = \int\frac{x^{n}}{\alpha^{n}}e^{-x}\frac{dx}{\alpha} = \frac{1}{\alpha^{n+1}}\int_0^{\infty}x^{n}e^{-x}dx = \frac{x!}{\alpha^{n+1}}$$

$\alpha R = x \qquad R = \frac{x}{\alpha} \qquad dR = \frac{dx}{\alpha}$

$R^{\frac{k}{2}}e^{-\epsilon kR}$; $\quad n = \frac{k}{2},\ \frac{k}{2}+1,\ \frac{k}{2}+2$; $\alpha = \epsilon k$

AH: ϵk is an imp't param. Indep of k, the peak is at $R_0 = \frac{1}{2\epsilon}$

$\ln R^{\frac{k}{2}}e^{-k\epsilon R}$

$\frac{k}{2}\ln R - k\epsilon R$

$\frac{k}{2R} = k\epsilon$

$R_0 = \frac{1}{2\epsilon}$

$m_0 = \frac{\frac{k}{2}!}{\epsilon k^{\frac{k}{2}+1}}$ $\quad\Big|\quad$ $m_1 = \frac{\frac{k}{2}!\cdot\frac{k}{2}+1}{\epsilon k^{\frac{k}{2}+1+1}} = \frac{m_0\;\frac{k}{2}+1}{\epsilon k}$ $\quad\Big|\quad$ $m_2 = \frac{m_0\cdot\frac{k}{2}+1\cdot\frac{k}{2}+2}{(\epsilon k)^2}$

$\mu = \frac{\frac{k}{2}+1}{k\epsilon}$ for $k>>1$; this is $\approx \frac{1}{2\epsilon}$

$\frac{m_2}{m_0} = \frac{(\frac{k}{2}+1)(\frac{k}{2}+2)}{k^2\epsilon^2} \approx \frac{1}{(2\epsilon)^2}$ for $k>>1$.

but $\frac{m_2}{m_0}-\mu^2$ (calculate done more accurately)

$= \frac{k}{2} = k'$ $\quad \mu = \frac{k'}{k\epsilon}$ $\quad \frac{m_2}{m_0} = \frac{k'+1\cdot k'+2}{k^2\epsilon^2}$; $\sigma^2 = \frac{m_2}{m_0}-\mu^2 = \frac{1}{k^2\epsilon^2}\left((k'+1)(k'+2)-k'^2\right)$

$\frac{3k'+2}{k^2\epsilon^2} = \frac{\frac{3k}{2}+2}{k^2\epsilon^2} \approx \boxed{\frac{3}{2k\epsilon^2}}$ for $k>>1$ $= \frac{3\cdot 2}{k}\cdot\frac{1}{4(k\epsilon)^2}$ $\quad \sigma \approx \sqrt{\frac{6}{k}}\cdot\boxed{\frac{1}{2n\epsilon}}$ $= R_0$

So the width of the peak is $\approx\sqrt{\frac{6}{k}}$ times the R value of the peak.

Fine! — But k has to ↑ before we get narrow peak. — (even if $N\to\infty$), we need larger k for narrow peak.

.32

Now consider $k\to\infty$, but N is finite: The simple approach of 108.01 — 110.25 would probably give quick results, but not nearly easy to apply to the Model of $\boxed{46.28}$

The worry is — why 46.28 doesn't have a peak at all.

→ But I can easily check my old pgms to see what happens w. $k=18$ & $N\to\infty$.

on 47.10 — .22 I got N up to 1200. The peak stabilized at $R = 40$, but the width of the peak stayed constant as $N\uparrow$ — as expected. I'd need to ↑ k in order to narrow the peak.

.40 I'd prob'ly need Monte Carlo data to study it empirically. → 116.09 spec

33199 SMET STEIN.

A disquieting thing about 47.~~[illegible]~~^{18} – .20! t. peak was at ~~[illegible]~~ U = .255 rather than .265) R = 300.
The difference in G was quite small, tho: ΔG = .16 betw U=.255 & .265 ; ~ 16%.

Some discn. of this is on 75.20 ff

N=∞ gave e^G = 0 at R=0 & ∞. (I think) ⟶ $e^G = (R^{\frac{1}{2}} e^{-\epsilon R})^k$

So maybe its finite N that gives $e^G > 0$ & ~~[illegible]~~ $\overset{R=}{0}$ a/o $\overset{R=}{\infty}$.

Tho it doesn't seem very reasonable!

I could write up a clear summary for ~~a~~ Alex (& myself!) See if he can find any bugs in reasoning.

101
103 (ST)
1~~0~~6
109
Final Dobs ST SS
2 3 4 5
1 > Acol,

.12 I was concerned w.t lack of "smoothness" in ST100 output!

ST114

I.e. $\dfrac{A(cc,3) - 1000*UU}{\text{~~Bar~~ BA ~~[illegible]~~} - 1000*UU}$

T. results were not a smooth funct of A(cc,3) − 1000*UU. Which I found disturbing. See 100½.3; last column.

I ran "ST100.bas" & modified it — This is ST114.bas, for output to be

$\dfrac{A(cc,3) - 1000*UU}{A(cc,4) - 1000*UU}$ ~~[illegible]~~ = ST

See 114½ output. last column: It looks as bad as, maybe worse! This is surprizing, because The ST pgm is fairly simple, & one would expect smooth output.

On t. other hand, the ST pgm, ~~shrinks~~ modifys t. μ_i''s by shrinking their xfms ($N^{\frac{1}{2}}\sin^{-1}(2x-1)$) toward t. mean of t. xfms, — then inverts t. xform those "shrunken" values. In general, t. ~~[illegible]~~ mean of xfms, is not same as xfm of t. means — so this could give much error near t. mean, (≈ A(cc,3) ≈ 1000*UU).

Well! (.12) relieves ~~[illegible]~~ some of my anxiety about my "BA" pgm! —

T. other disquietude is on how t. results seem to be much dependent on just what apriori one uses. I could see how ~~[illegible]~~^{2} Gave ~~[illegible]~~ (say MS or Log ratio) of ~~[illegible]~~ close to preds, changed w. change of apriori.

→ Did I do this already? $\boxed{104.03–.18 \text{ looks like that}}$ → But they have to be done more carefully! ← + $\frac{e^G}{\ln R}$ correction term was inappropriate.

A possl. reason why I've been having trouble, $\tau^2 + \overline{\sigma_i^2} = \Sigma^2$ ← ~~total~~ var of total $\{\mu_i\}$

Since N is "small", $\Sigma^2 \approx \overline{\sigma_i^2}$ so lots of error in τ^2: & observed $\Sigma^2 - \overline{\sigma_i^2}$ can be < 0

In such a cases perhaps apriori info is very impt. [tho in t. present cases, when we are uncertain about τ^2 & we know its ≥ 0, then, error in τ^2 isn't so impt. in assigning wts to $\bar{\mu}$ & μ_i' (in estimating μ_i). In t. present case of $\Sigma^2 - \sigma_i^2 < 0$ we ~~can~~ let τ^2 be ∅ w.o much trouble.

[(NB) final result! Once τ is known, It's easy to get t. D.f. for each μ_i. — This is done in a Bayesian way. If μ_i' is observed, it is t. result of a choice of μ_i (w. proby d.f. μ_i: N($\bar{\mu}$, τ)), followed by a choice of μ_i' w. proby μ_i': N(μ_i, σ_i^2)) [≡ μ_i: N(μ_i', σ_i^2)]. T. product of these 2 Gaussians is t. d.f. of μ_i. Its mean is t. wtd mean of $\bar{\mu}$ & μ_i'; wts are $\frac{1}{\tau^2}$ & $\frac{1}{\sigma_i^2}$, resp.

$\boxed{45}$

.01 Notes for a Review: [redacted] Things to include: $\boxed{78, 80, 90, 115, \underline{129},}$ → 126, 138.04, 143

1) Previous reviews, 52.01 (Mainly early stuff)

2) List of pgms — (their "history") — & what they do! Which ones have h.c. output.
(Note pgm. for Generation of Mt Carlo Corpus: $\boxed{ST81}$ ST109

3) List of various approaches to ST: Bibliog. of each & which pgms dealt w. each — for both Generation & Testing. (see 8)

4) What are main Conclusions? Best ways to Approximate → ⑥

5) What are Main unsolved ("open") problems? — Uncertaintys. →

6) Related to $\binom{(3)}{(4)}$: Describe in a clear way, the main approaches used:
e.g. the Methods ① 108.01–114.40, ② 46.02; 46.28ff ③ 121.13–122.40

.11 7) Genesis of STEIN: (E&M have anyone this in their bibliogy): A Big Genesis is to Main Analysis → Show copy.
Another (More extreme) is in "MIXD Cases Them".

8) Compare Empirically, using Mt Carlo data, the ~~various~~ various Approaches [ST81 is Mt Carlo tester]

Indices to pgms:
78
80 — introductory pgm
90 — & ST90

The "ruffness" of estimating sequences.
The "k−3" effect
Whether to $\sqrt{N}^{\frac{1}{2}}\sin^{-1}(2x-1)$ xform
Why Logit.
How the approxd seems to influence results excessively.

107 ~~[illegible]~~ discn of Shape of eqs: curves & effect of k; & approxn; Also 86.01

47.10–.22 Shape of d.f. for large R See 94.14

D.f. at R=0, ∞:
49.20: R=∞
50.20 discovery that e^G is flat (=0) for R ≥ 2 continued.

$\boxed{54–60}$ ~~[illegible]~~ on the approxn
~~[illegible]~~ ~~[illegible]~~ $\alpha \log_{\alpha}^{k} x$ as approxn for X.

My impression of ~~[illegible]~~ main methods:

1) 46.02 – 46.28 The Binomial d.f., w. ~~[illegible]~~ (then finally, integration over R, ST81A.Bas) then Use of integrals to estimate [Mi] (← ST101).

2) ~~Extens. of the~~ Reproduction of E&M's work of computing ST using $\sqrt{N}\sin^{-1}(2x-1)$ xform.

3) Criticism of 1):
a) That results are "Ragged" — but Notice that "STEIN" was at least as ragged.
b) $\boxed{\text{Most Serious}}$: That results depended ~~[illegible]~~ on Approxn. for R

116.32 – 120.40: On Gaussn D.f., Results Seem ~ to those for ST100, ST101 →

(99.28: – 100.40): final results using proper Bessel ~~[illegible]~~ to calc v's: Results a bit "ragged"

ST100 & implements → and gets ~~ragged~~ "ragged" results.
ST101 (cleaner form of ST100)

ST103: ~~[illegible]~~ exact computation of "STEIN" values using $\boxed{M^{\frac{1}{2}}\sin^{-1}(2x-1)}$

104.03: Redoing of [Mi] using different approxs for R! (Hvr, serious errors in calcn.) Do this again (IMPT!).

105.04 – .27 More on (E&M STEIN)
106.01 – .40

105.28: Pgm ST101 ~~[illegible]~~ much improved if I integrated w.r.t. $\bar{R}$ also

107.01 Review of effects of "ssz" on "sharpness" of D.f. Better Approxd is applied.

108.01 – 114.46 Earlier approach using Gaussian d.f. Fairly reasonably very simple: for varying σ_i^2: Hvr., I don't know why E&M used (k−3) rather than (k−1) (or why (k−2) was used)
Approxn can be used w. Binomial d.f. — also effect of approxd of R can be studied.

62.01 – 69.10 Gaussian Model
69.11... Back to Binomial D.f.'s
74.03 Good approxn ~~[illegible]~~ $\frac{x!\,y!}{(x+y)!} = (a^2 b)^k \sqrt{2\pi a b R}$

76.25 Computation of $\underline{\underline{\mu}}$ & $\underline{\underline{\sigma}}^2$ of Bin. d.f.s (may have done this better elsewhere)

ST81A.Bas Integration of RC w.r.t. R w. ~~[illegible]~~ approxd $\frac{1}{(R)}$, $\frac{1}{R^2}$, & $\frac{1}{R(\ln R)^2}$

STEIN Approach again
86.12 – 87.20

59.22 $\int_{\alpha}^{\beta}$: lower limit α found not critical. — β is impt, hvr.

94.14: Use of test data but common (large) N for all players See 43.10

95.01 ff working on wrong integral!

97.01 – .10 I got "final" w.g. way to get 4r estimates.

→ 124.01 spec

4299 SMPT STEIN.

I want to write ~~[illegible]~~ on t. d.f. of t. squares of vars; ≥ what happens to

$\sum_{i=1}^{k} V_i^2$ v.s. $\sum (V_i - c)^2 = \sum V_i^2 \; 2c \sum V_i + kc^2$. | say we let $c = \frac{1}{k} \sum V_i$

$= \left(\sum V_i^2\right) - \frac{1}{k}\left(\sum V_i\right)^2$ These 2 ~~[illegible]~~ terms are correlated, — giving trouble!

.04

find d.f. of V^2 as a function d.f. of V.

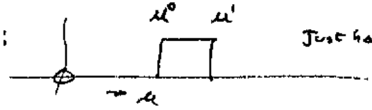
Then " " $(V-\alpha)^2$ " " " " " $V-\alpha$.

Each D.F. has a mean, ≥ t. mean of sum is sum of means,

Also [illegible]

.09: 118.40 (SN) One way to do $k \to \infty$;

119.40

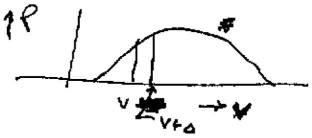


Just have a box car of uniformly dense values of μ_i' from μ^0 to μ^1.

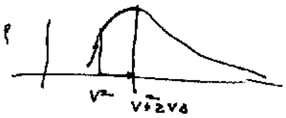
Then ~~[illegible]~~ t. density of data pts in that range; This would be w. $N=\infty$, hvr. Easy to analyze, hvr.

Now, my each of t (dense) data pts has a common σ_0^2. (or a ± σ_0). (Still unclear)

$(V^* - c)^2 = V^2 - 2cV + c^2$.



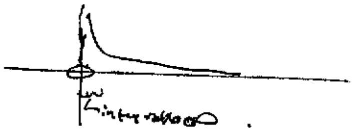
$pc = P(V) \cdot \Delta$

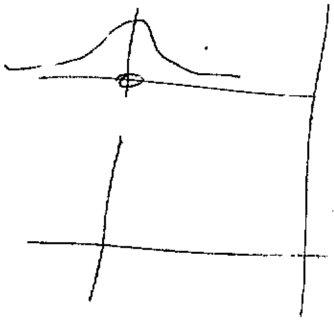


$P = P(V)$ density $= \frac{P}{2V}$.

Say $u = v^2$ density $= \frac{P(\sqrt{U})}{2\sqrt{U}}$

Say $P(V) = e^{-x^2}$; $\to \frac{1}{2\sqrt{V}} e^{-V}$





Say $P(V) = e^{-(x-c)^2}$ $\to \frac{1}{2\sqrt{u}} e^{-(\sqrt{u}-c)^2} = \frac{1}{2\sqrt{u}} e^{-u + 2\sqrt{u} \cdot c - c^2}$

This may be convertable to t. n^{th} moment of $e^{-(x-c)^2}$.

So maybe some function of erf.

So, if one has a bunch of indep vars, V_i, their squares will be indep, $(V_i - c)^2$'s will be indep.

Getting μ & σ^2 of $(V_i - c)^2$ may not be easy hvr. ; but once we got them, t. μ & σ^2 of t. sum, is easy.

.32 → Note pp 62-68: This really looks like fitting k+2 params to k data pts!

Unless we use t. "2" eq.

- But more on pp 62-68: A basic idea: We start w. unknown $\bar{v}$, s^2 : $[\mu_i]$ w. known $[\mu_i', \sigma_i^2]$.

In terms of these params, t. pc of t. corpus is:

.37 $$\prod_i \frac{1}{\sqrt{2\pi} s} e^{-\frac{(\mu_i - \bar{v})^2}{2 s^2}} \frac{1}{\sqrt{2\pi}\, \sigma_i} e^{\frac{-(v_i' - \mu_i)^2}{2\sigma_i^2}}$$. Next, we integrate each factor $\int_{-}^{+\infty} d\mu_i$.

.38 $$\int_{-\infty}^{+\infty} e^{-Ax^2 + bx + c} dx = \int_{-\infty}^{+\infty} e^{-\left(\sqrt{A}x - \frac{b}{2\sqrt{A}}\right)^2 + \frac{b^2}{4A} + c} = \sqrt{2\pi} \cdot \frac{1}{\sqrt{A}} e^{+\frac{b^2}{4A} + c} * \frac{1}{\sqrt{2}}$$

$b = 2\sqrt{A} \cdot V$ so $V = \frac{b}{2\sqrt{A}}$

$\boxed{\text{No! } \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2\alpha^2}} dx = \sqrt{2\pi}\alpha}$ so ↑

4.3.99 SMPT STEIN.

.01 $A = \frac{1}{2s^2} + \frac{1}{2\sigma_i^2}$ $b = \frac{\bar{U}}{s^2} + \frac{U_i'}{\sigma_i^2}$ $c = \frac{-\bar{U}^2}{2s^2} - \frac{U_i'^2}{2\sigma_i^2}$

.02 $\frac{-U_i^2}{2s^2} + \frac{\cancel{2}\bar{U}\,U_i}{\cancel{2}s^2} - \frac{\bar{U}^2}{2s^2} - \frac{U_i^2}{2\sigma_i^2} + \frac{\cancel{2}U_i'\cdot U_i}{\cancel{2}\sigma_i^2} - \frac{U_i'^2}{2\sigma_i^2}$ ← Exponent.

.04 $P< = \prod_i S = \prod_i \cancel{\sqrt{\pi}} \cdot \frac{1}{\sqrt{\frac{1}{\cancel{2}s^2} + \frac{1}{\cancel{2}\sigma_i^2}}} \; e^{\left(\frac{\bar{U}}{s^2} + \frac{U_i'}{\sigma_i^2}\right)^2 / \left(4\left(\frac{1}{2s^2} + \frac{1}{2\sigma_i^2}\right)\right) - \frac{\bar{U}^2}{2s^2} - \frac{U_i'^2}{2\sigma_i^2}} \; \frac{1}{\sqrt{2\pi}\,s} \cancel{\frac{1}{\sqrt{2\pi}\sigma_i}} \frac{1}{\sqrt{\sigma_i}}$

In doing parts of .04; $\prod_i s \frac{\sqrt{2^{-1}}}{\sqrt{\frac{1}{2s^2} + \frac{1}{2\sigma_i^2}}} \rightarrow \prod_i \frac{1}{s\sqrt{\frac{1}{s^2} + \frac{1}{\sigma_i^2}}} = \prod_i \frac{1}{\sqrt{1 + \frac{s^2}{\sigma_i^2}}}$;

Woops! I left out 2 "$\frac{1}{2}$" factor in $\int e^{\frac{-x^2}{s^2}}dx$. S- import

.10 $\approx e^{-s^2 \cdot \left(\sum_{i=1}^{k} \frac{1}{\sigma_i^2}\right)}$. — which ↓ rapidly as s↑:

The rest is a real mess! It looks like $e^{+\frac{N}{s^2}}$ which → ∞ superfast at s→0; → Tho maybe not too bad!

but I'm not sure. $\frac{\left(\frac{\bar{U}}{s^2} + \frac{U_i'}{\sigma_i^2}\right)^2}{2\left(\frac{1}{s^2} + \frac{1}{\sigma_i^2}\right)} - \frac{\bar{U}^2}{2s^2}$ are all the "s" containing terms in the exponent.

Since [illegible] terms are all in exponent, we want $\exp \sum_i \left[\frac{\left(\frac{\bar{U}}{s^2} + \frac{U_i'}{\sigma_i^2}\right)^2}{\cancel{2}\left(\frac{1}{s^2} + \frac{1}{\sigma_i^2}\right)} - \frac{\bar{U}^2}{\cancel{2}s^2}\right]$ Sum for all K terms!

Start by assuming $s^2 \ll \sigma_i^2$. so $\frac{\left(\bar{U} + \frac{s^2}{\sigma_i^2}\cdot U_i'\right)^2}{s^2\left(1 + \frac{s^2}{\sigma_i^2}\right)} - \frac{\bar{U}^2}{s^2} \rightarrow \frac{1}{\bar{U}s^2} \frac{\left(1 + \frac{s^2}{\sigma_i^2}\frac{U_i'}{\bar{U}}\right)^2}{1 + \frac{s^2}{\sigma_i^2}} - \frac{\bar{U}^2}{s^2}$

$\sim \frac{1}{\bar{U}s^2}\left(1 + \frac{2s^2}{\sigma_i^2}\cdot\frac{U_i'}{\bar{U}} - \frac{s^2}{\sigma_i^2}\right) - \frac{\bar{U}^2}{s^2}$ for small s this is $\left(\frac{1}{\bar{U}} - \bar{U}^2\right)\cdot\frac{1}{s^2}$.

$\frac{1}{\bar{U}}$ can be + or − & [illegible] large!

No. $\frac{\frac{\bar{U}^2}{s^4}\left(1 + \frac{U_i' s^2}{\sigma_i^2\bar{U}}\right)^2}{\frac{1}{s^2}\left(1 + \frac{s^2}{\sigma_i^2}\right)} - \frac{\bar{U}^2}{s^2} = \frac{\bar{U}^2}{s^2}\left(1 + \frac{2U_i'}{\bar{U}}\frac{s^2}{\sigma_i^2} - \frac{s^2}{\sigma_i^2}\right) - \frac{\bar{U}^2}{s^2}$

$\cancel{\frac{\bar{U}^2}{s^2}} + \frac{\bar{U}^2}{\cancel{s^2}}\frac{\cancel{s^2}}{\sigma_i^2 2}\left(\frac{2U_i'}{\bar{U}} - 1\right) - \cancel{\frac{\bar{U}^2}{s^2}}$

$= \frac{\bar{U}^2}{s^2} \cancel{2} \frac{s^2}{\sigma_i^2}\left(\frac{2U_i'}{\bar{U}} - 1\right) = \boxed{\frac{\bar{U}^2}{\sigma_i^2}\left(\frac{2U_i'}{\bar{U}} - 1\right)} = \frac{\bar{U}}{\sigma_i^2}\left(2U_i' - \bar{U}\right)$ the exponent is essentially

flat near s = 0. — otherwise, the eq. is (.10)

= [illegible] $2\bar{U}\sum\frac{U_i'}{\sigma_i^2} - \bar{U}^2\sum\frac{1}{\sigma_i^2}$ Sharp

It would be possible to compute $\prod_i$ as a function of S. If it has a peak, fine! —

Now, it may well be that the rapid ↓ of (.10) is the dominant feature ... which would be tragic!

Anyway: Best approach: Check eqns. 116.37, .56; 117.01, .02, .04: Then check

Do 117.04 for a seq. of values of s.

Another way: Use eq. 116.37: integrate over U_i for all k i's: [illegible].

Do this for many values of s. It is somewhat slow, but it may be valuable. Lots of ways to speed it up!

.37 $\boxed{A = U_i - \bar{U}; \quad B = U_i' - U_i}$ EXP (A*A* A1 + B*B* B1) $\frac{1}{\sum_{i=1} \frac{1}{s\sigma_i}}$ [illegible] step [illegible] : Then mult final sum by $\frac{1}{s\sigma_i}$

$A1 = \frac{1}{2s^2}$ (Then $B1 = \frac{1}{2\sigma_i}$ [illegible])

My guess, w. step of .005, [illegible]

0 → 1 shouldn't take long: 200 steps: × 18 = 3600 steps. per S values. So quite possible.

4 499 SMRT STEIN

DGFDBL A-Z

ST118 for ST109

```
S1=0 : N=45; K=18 ; UU=.2653859
                    Z=A(J,3)
~~For J=1 to K : A(J,6)= Z*(1-Z)/N : Next J~~ .
~~For S=~~ .001 to .4 step .01
S2=1
For J= 1 to K                      Z1=Z(1-Z)*2
  A1= 1/(S*S*2) : Z=A(J,3) : B1=N/Z1 : S1=0
    For U=0 to 1 step .005
      A = U-UU : B= Z-U
      S1=S1 + EXP(-A*A*A1-B*B*B1)
    Next U
  S2=S2*S1/S/A(J,6)
Next J
Print S, S2
Next S
```

S: 40
U: 200
i: 18

This seems to be it. Must repeated, most time consuming part.

I may need to do this in log domain.

S1*SQR(2*B1))

S2=S2*S1*SQR(2*B1)/S

$2*B1 \equiv \frac{N}{Z(1-Z)}$

Verr. check 116.37 carefully. Also 117.37-.40

I may be able to speed things up by having J, K, be integers.

.20 A way to speedup: 116.37 can be further factored: Store $\left[e^{-\frac{(u_i-\bar{u})^2}{2s^2}}\right]$ ARR1 in one 2 dim. table ($s \times i$). Store $\left[e^{-\frac{(u_i-v_i)^2}{2\sigma_i^2}}\right]$ ARR2 in another 2 dim table. (i). Then do like Cart. product. Unclear that there are any savings — since all we need at the end is 2x $s \times i$ array. No — There may be savings; v_i is actually indep of i, so it's "u". This is a 2 dim array in (u & s) (indep of i). This is a 2 dim array in (u and i) indep of s. So for various s values, one takes the dot product of a row of constant s of ARR1 w. each of the constant i rows of ARR2: This does the Integration of 116.37. We then mult. this dot product for each of the rows (for each i) to get $\prod_{i=1}^{K}$ which is our final result

.27 for that s.

The Q is: would this save any time? Computation of the integrand of 116.37 takes: 3 additions, 2 multiplies, 1 exp( ). We do this 3 dimensional times i: 1/4, s = 0/.4; u = 0/1. In .20-.27, the Arrays are free. We have to do ("n^3") operations (3f), but they are all just mult, add: no exp( ) which takes all the time. Essentially, we've made tables of all needed exp( ) values. So the method does save signif. time, but it still is "n^3" in ~~[illegible]~~ cost.

If the tables were properly constructed, going to successive entries would be incremental (very fast).

S: 40 steps, U: 200; i: 18 so 144000 steps. Could be done in < 1 sec!

I wrote ST118: S2 was monotonic ↓ from S=.001 to S=.4

Time to run ~~[illegible]~~ < 1 sec. Took .55 sec (including printing).

W.o. printing, it took .38 sec.

S	S2
.01	2.02 E+71
.02	1.6
.03	1.16
.04	.63
.05	.30
.06	.12
.07	.044
.08	.015

S	S2
.005	2.1 E+71

S = .004 or less, gives overflow!

I tried using logs in update of S2 — But no help. It may be that S1 update $Exp\left(-\frac{(U-\bar{U})^2}{2 \cdot S^2} - \cdots\right)$ was exp(− 2 very large no.) was too small — So underflow.

Perhaps something wrong! For small s, $\exp\left(\frac{-(v-\bar{v})^2}{2s^2} - \frac{(v_2'-v)^2}{2\sigma_i^2} \right)$ usually contributes nothing to $\int dv$ [illegible].... But if $(v-\bar{v})$ is small enough, $\left(\frac{v-\bar{v}}{s}\right)^2$ can be ~0. v has given steps .005.

Anyway, to compute $\exp\left(-\frac{1}{(.005)^2}\right) = \exp -200^2 = \exp -40k \sim 1E-92000$ which is likely to be underflow.

~~to~~ I could fix it ~~#~~ say X = −A*....

If X > −100~~0~~, then S1=S1+exp X.

I fixed it: W = −A*A*A1 − B*B*B1

If W > −1000 then S1=S1+EXP(W)

It took only .444 sec; < .55 sec (w.o "IF"!)
w printing

Woops!

	s	S2
	.001	2.69 E 76
	.002	8.689E 71
	.003	2.18 E 71

Maybe related to −1000 threshold!

try −100 same results!
W > −10000 still same.
W > −100k illegal funct. call.

try $\overset{v}{\text{step}}$ = .001 — This makes it very flat around s=.001

s × .001	S2 × 10^84
1	4.07
2	4.06
5	4.02
10	3.85
20	3.2
39	1.33

I set N = 100; Peak of S2 at $\boxed{s=.031}$: S2 = 1.23 E+85 to ½ values s = .051, .011

So s = .031 ± .02

N = 1000 Peak around s = .047 but narrower
N = 10000 $\boxed{s = .051}$ (<.02) maybe width same as @ ← I didn't calculate widths carefully,

It turns ~~2or~~ S2 = ∞ at s = 0! Mistake in prgm. It's not supposed to go to s = 0, since W ∞ divided by s, But at N = 1000 $\frac{\text{S2 Peak}}{\text{S2 s=.001}}$ is ~10^{25}!

For N = 100 ~~this~~ $\frac{\text{S2 Peak}}{\text{S2; s=.001}}$ is ~4.

For s = 20 flat around s = 0; Down ~ ½ at s = .031.

~~.000~~ N = 100k! Peak at $\overset{\sim}{.05} \pm \overset{\sim}{.01}$
.01

N = 100K " " .0465 ± .0075

	.040	
	.0417	1.27 1.56
Δ = .008	.046	~~[illegible]~~ 2.7913
	.0485	
	.049	~~[illegible]~~ 2.786
Δ = .01	.059	1.44
	~~[illegible]~~	
	.0105	

) peak

.0415		1.56
.0485	→	2.80
.0585		1.52

.034
~~.0585~~
.0485
.0400

peak is s = .0465
+ .016 or −.008 for ~50% down.

s = .0465: $\frac{1}{s^2}$ = 425.1125

n =	100	peak at	s = .031
for n =	~~7~~ 75	peak at	s = .023
n =	60	" "	s = .010 ~~[illegible]~~
n =	65	"	s .016
n =	63		.014
n =	62		.013
n =	61		.012

	peak at	
n = 55	~~[illegible]~~ .0009	It seems down for s < .002
n = 50	.0008	
n = 45	.0009 !	.000889 very close

Anyway, the behavior for very small s may not be correctly modeled by the prgm. I changed threshold from ~10k to −100 but the n=45 results were ~~[illegible]~~ exactly the same for all printouts. changing threshold to −10 produced some small changes.

4599 SMET STEIN

Using N = 45 : I changed ΔU from .001 to .0001 This eliminated the peak at .000$\overset{8}{5}$.

Threshold −100 → −1000 ~~speeded up the pgm~~ slowed down ~~the~~ pgm by × $\frac{2}{3}$, but left results invariant.

~~Time~~ 1.1" −100
1.48" −1000
1.04" −50 numerical results invariant.
2.7" −10000

Using −50 thresh & ΔU = .0001
N = 100 peak at [redacted] ~.0315 as before.
75 " " .025 " "
60 " " .010 same w. ΔU = .00001 but peak is small, 7.346 E122 for S=0; 7.40 E122 at "peak"
55 " " <.001 if any Hvr, if I assume S < .001 it begins to ↑ as S ↓ around S = .00005 then gets very large for smaller values; this could be due to ΔU being too small.

N=55 for ΔU = .00001, there was no "peak" other than S=.00001 the smallest S ~~[illegible]~~ calculated, & it was rather flat for larger values.

So it is likely that there is no peak for N=55 & less.

7.34618 E+122
7.40

So, my impression is that for N=60 peak is very small & near $\overset{S\approx}{.01}$ — ~~[illegible]~~ & flat for S < .01.
~~&~~ for N ≤ ~~[illegible]~~ 55 there is no peak other than S=0 & its very flat.

General remarks: These results seem similar to those using 46.28 — which uses Bin D.F. & integration over U_z's. Hvr, in the present case, I do the integration Numerically ~~[illegible]~~, & in 46.28 I did it analytically. ~~If I did the present one analytically~~ If I programed the present model analytically it would be much faster. The ~~analysis is~~ needed "analyses" is 116.37, .38, 117.01, .02, .04; I'm uncertain of the analyses, but ~~[illegible]~~ the ~~[illegible]~~ its yields could be checked against the Num. integration of ST118.

This analytical pgm for Gaussian d.f's could be then used in a Bayesian way to get BA values, by ~~[illegible]~~ integrating S; just as in ST100 ST 10[illegible].

Note on pp 62–67: 62.03 eq. & 116.37 eq. are the same, but in 62.03 ff I tried to fit the model $U_z = a\bar{U} + (1-a)U_z'$ (on 62.20 I used a different data of $\bar{U}$, hvr, ~~but~~, on second thought — (~~that~~ it may not have been so wise (— I may have dropped that data later, hvr.))

4599 SM FT STEIN

120½ ~~[illegible]~~

On the (Binomial) d.f. version: If we decide on an R, μ value, This is equivt. to an apriori | $A \equiv \mu R$, $B \equiv (1-\mu)R$

$\mu^A (1-\mu)^B$ & Using the Laplace's "Rule" derivation, its equivt. Then, for a $\mu_i N$,

of $\mu_i N + A$ hits and $(1-\mu_i)N + B$ outs so that $= \frac{\mu_i N + A}{N + R}$: This is obtainable by getting

products from $\underline{\mu^A (1-\mu)^B \cdot \mu^{\mu_i N} (1-\mu)^{(1-\mu_i)N}} = \mu^{A+\mu_i N} (1-\mu)^{B+(1-\mu_i)N}$.

→ which, using Lap's rule gives $\mu_2 = \frac{A + \mu_i N + 1}{R + N + 2}$ (I think). (See 48.01)

.08 This suggests, that if $P(R)$ is the D.F. for R — obtained via ST46, say.

.09 then μ_i should be $\int_0^\infty P(R) \frac{\overset{\equiv A}{\mu R} + \mu_i N + 1}{R + N + 2} dR$. — which is ≈ 81.17 ff, 85.03,

Around ST96 then finally ST100 & ST101 I switched over to a very general (correct) way of doing Bayesian Predn. — However, I think the results should be the same as .09 — it may be possible to show that those 2 methods are mathematically equivt.

I had been using .09 because it was intuitively reasonable — but probably its exactly right. (It may only be a v.g. approxn, however.... This would have to be worked out in detail.)

.01 on the Gaussian Approach of 108.01 – 114.40: Attempts at a clear, exact discussn.

Suppose the data set $[\mu_i']$ was generated by first chosing μ_i, from $P_1(\mu_i)$, a d.f. of known mean, $\bar{U}$, but unknown varc, τ^2. μ_i' is then generated using a (say Gaussn) d.f. of mean U_i & known varc, σ_i^2.

We want to make measurements on the dataset, that will enable us to estimate τ^2.

First say we know τ^2 and the $[\sigma_i^2]$, but we didn't know the $[\mu_i']$

What would be the expected Varc of the $[\mu_i']$ set?

.13 $\boxed{SN}$ Actually, in the Baseball case, if we know $\bar{u}$ and τ^2 (or more exactly, if we know $P_1(U_i)$, the final/ expected d.f. of the $[\mu_i']$ is completely determined:

Monte Carlowise: Get $P_1(\cdot)$ to pick a U_i: Then using N trials, w. probly U_i for a hit, Generate a Monte Carlo U_i'. We simply repeat this k times to obtain a sample. Each time, we can measure the Varc. of the final d.f.

.20 (Ms deviation from its mean). $\boxed{\text{For each value of } \tau^2 \text{ we will get a d.f. of Varcs}}$ → See 122.32 for simple exact way! ↓ Not really! 122.32 gets the d.f. of obtaining the set $[U_i']$ not just its observd. Varc about $\bar{U}$.

We then look at the varc of the true sample $[U_i']$: from .20, using Bayes, we

.23 can get the d.f. of τ^2.

From τ^2, we then know $P_1(U_i)$ and for each i,

.25 the resultant d.f. of U_i = $P_1(U_i) \cdot U_i^{\mu_i' \cdot R} (1-U_i)^{(1-\mu_i')R}$.

Since we don't know τ^2 exactly, but only know a d.f. for it, the d.f. for U_i

.27 becomes $$\int d\tau^2 \left(D(\tau^2) \cdot P_1(\tau^2, U_i) \cdot U_i^{U_i' \cdot R} (1-U_i)^{(1-U_i')R} \right)$$

Here $D(\tau^2)$ is the density of τ^2: obtained from $\boxed{.20 - .23}$

Actually, what we want is the probly of a hit at player i's next "at Bat"

the expected value of U_i in .27 may be about it. (put in the U_i one on the $\int$ sign & integrate wrt U_i as well as τ^2. If $P_1(\tau^2, U_i)$ is of the form $U_i^A (1-U_i)^B$ (which is reasonable to do) then the $\int$ w.r.t. U_i is easy.

we have $\int d\tau^2 / D(\tau^2) \left(\frac{(x + \mu_i' \cdot R)!\,(y + (1-\mu_i') \cdot R)!}{(R + N + 1)!} \right)$

$\tau^2 = \frac{\bar{U}(1-\bar{U})}{R}$; $x \equiv \bar{U}R$, $y \equiv (1-\bar{U})R$. so $R = \frac{\bar{U}(1-\bar{U})}{\tau^2}$

Is there any way to do .20 – .23 analytically? (not M.C. ones).

Even if the Ideas at .13 ff could be practically implemented by M.C. or analytical calcns, I think the apript for τ^2 would still be very impt.

So, one (Tentative, tho strong) conclusion: That for data sets in which τ^2 is small, the apripd of τ^2 will be very impt. — This will be true whether one uses Gaussian, Binomial d.f. or any other d.f. If one uses E&M's method, one ends up taking t. difference between 2 / noisy $^{\text{larger}}$ values. T. uncertainty of this difference is <u>very large</u>.

~~[illegible]~~ Which t. approx seems critical in determining τ^2 when τ^2 is <u>small</u>, what is <u>not</u> clear is how ~~[illegible]~~ much MS error or log likelihood are affected by ~~[illegible]~~ t. exact form of approx of τ^2 (or of R).

If then $\tilde{R}$ is uniform betw. 0 & 1, $\frac{1}{\tilde{R}}$ has density

$\tilde{R}$; $\tilde{R}+\delta$ $\quad \frac{1}{\tilde{R}}$, $\frac{1}{\tilde{R}} - \frac{\delta}{\tilde{R}^2}$ so density $\propto \frac{1}{\tilde{R}^2}$

If R^2 is uniform betw. 0 & 1 ;

σ uniform betw 0, 1 σ & σ+δ

$\frac{1}{\sigma^2} \to \frac{1}{\sigma^2} - \frac{2\delta}{\sigma^3}$; $\frac{1}{\sigma^2}$ is like $\frac{2}{\sigma^3}$.

.12 A Quick review E. M's method: By using $x \to N^{\frac{1}{2}} \sin^{-1}(2x-1)$

They make σ_i^2 constant (indep of i) : For ~~[illegible]~~ known $\bar{u}$ & τ^2, the d.f. of expected u_i ~~[illegible]~~ is t. convolution of 2 d.f.'s ; i.e. the d.f. $P_1(u_i)$ and t. d.f. of center u_i & vare $= \sigma^2 (= \sigma_i^2)$. So t. vare of t. conv. is just $\tau^2 + \sigma^2$. We then measure vare of $\{u_i\}$ & subtract σ^2, to get $\hat{\tau}^2$.

If the σ_i^2 are <u>not</u> all t. same, then each x_i has its own d.f. :

Its mean is $\bar{u}$ its vare is $\tau^2 + \sigma_i^2$. ~~[illegible]~~ ~~[illegible]~~ t. expected value (mean) of $(u_i - \bar{u})^2$ is $\tau^2 + \sigma_i^2$ (w.r.t. central <u>second moment</u>).

The expected value of $\sum_{i=1}^{k} (u_i - \bar{u})^2$ is $\sum(\tau^2 + \sigma_i^2) = k\tau^2 + \sum \sigma_i^2$.
so we can ~~seek~~ easily solve this for τ^2, since $\sum(u_i - \bar{u})^2$ & $\sum \sigma_i^2$ are known

.24 → There is, hm. t. k-1, k-2 or k-3 correction for $\sum(u_i - \bar{u})^2$.

.12-.24 seems like a very quick and dirty way to understand t. Stem Approach.
One <u>could</u> determine whether k-1, k-2, or k-3 is best, by M Carlo simulation

---

— Hm, an apparent impt error in .12-.24! That σ_i^2 & u_i are <u>functionally related</u> (i.e. "<u>highly</u> correlated"). <u>121.13</u> ff deals very explicitly with this correlation.

.32 : 121.20 Given $\bar{u}$, R & and N, t. <u>exact</u> d.f. in .20 <u>is</u> $u_i^{X+Nu_i}(1-u_i)^{Y+N(1-u_i)}$] ← More likely

or is it $u_i^{X+N\bar{u}}(1-u_i)^{Y+N(1-\bar{u})} = \left(u_i^{\bar{u}}(1-u_i)^{(1-\bar{u})}\right)^{R+N}$.]

.34 .32 ▷ $u_i^X(1-u_i)^Y \cdot u_i^{u_iN}(1-u_i)^{(1-u_i)N} = \left(u_i^{\bar{u}}(1-u_i)^{1-\bar{u}}\right)^R \cdot \left(u_i^{u_i}(1-u_i)^{(1-u_i)}\right)^N$

I don't know if it's easy to find first or second moments of .32 (= .34)

T. Q. is can one do $\int_0^1 du \left(u^u(1-u)^{(1-u)}\right)^N \cdot u^X(1-u)^Y$? for any fixed N, X, Y

<u>Well, it's easy to do Numerically!</u> So that solves 121.20 & 121.<u>23</u>

.01 → Well, I'm beginning to understand t. whole mess! — How t. difnt approaches relate to one another.

(SN) first a ~~brief~~ excursion: if U_i is t. probly of hit, what is t. probly of A_i hits in N times at bat? Is it $P = U_i^{A_i}(1-U_i)^{N-A_i}$ or P mult by ~~[illegible]~~ no. of difnt orderings?

i.e. $\frac{A_i!\, N-A_i!}{N!} = \frac{A_i!}{N(N-1)\cdots(N-A_i+1)}$ (T. expression ~~[illegible]~~ has to be ~~[illegible]~~ invariant $A_i \to N-A_i$)

A good way to think about it is Monte Carlo simulation.

Note $(U_i+(1-U_i))^N = \sum_{j=0}^{N}\left(U_i^{\,j}(1-U_i)^{N-j},\ \frac{j!\,(N-j)!}{N!}\right) = 1$

So sum of all probys = 1. This amounts to a sizable correction in 46,28

i.e. $\frac{A_i!\,B_j!}{N!}$, but since it's indpt of x, y ($R, \bar{U}$) it's not impt. in those calcns.

— Here, it almost cancels out t. normzn factor $\frac{N+1!}{A_i!\,B_j!}$; so we end up w. $(N+1)^K$ after normzn.

→ obtaining first, PC of corpus, given $R, \bar{U}$. Second, Integrating wrt. R & apriori for R, to get Apri PC of corpus. — Then getting relative proby of various continuns. of t. corpus.

Both $\underbrace{121.13 - 122.40}_{\beta}$ & $\underbrace{46.25\text{ ff}}_{\alpha}$ are very similar

In α, we start w. given $\hat{r}^2, \bar{U}$ or $R, \bar{U}$, or x, y: and we find t. probly that t. set $[U_i']$ was generated.

In β we look at $Z = \sum(U_i'-\bar{U})^2$: ~~[illegible]~~ ; starting w. t. same $\hat{r}^2, \bar{U}$ (at all) of α, we ask, what is t. probly ~~that a data set of~~ of generating in K trials, a data set ~~[illegible]~~ $[V_i]$ with $\sum(V_i-\bar{U})^2 = Z$. (we don't want a probly, but a density).

In ~~122~~ 123 .24, we pick ~~[illegible]~~ t. "most likely" value of $\hat{r}^2$; for small "True" $\hat{r}^2$, this val can often be <0. So, as in β, ~~[illegible]~~ if we want a more precise answer, we consider not t. only ~~[illegible]~~ most likely value of $\hat{r}^2$, but t. d.f. of $\hat{r}^2$. If it were a sharp narrow d.f. "most likely" would be fine, but it is not a narrow d.f., so we do have to consider t. d.f., & t. apriori, & t. apriori does seem to be impt.

SN I'm not really sure about. T. apriori is impt for "exact" values of R, say, but In t. large R space, that I was working w., large ~~[illegible]~~ changes in R produce very small changes in Gore (MS. ~~[illegible]~~ (or log likelyhood)) So it may be that T. apriori really isn't of much impor. w.r.t. Gore.

4.799 JMFT STEIN

.ii 115.40 Outline of Review : See ⌈ Also see 45.01 for early review outline
115.01 for many Impt ideas.

111.03 – 113.32
on N=∞.

1) Dcrb. STEIN problem & its causes. (115.11)

2) Dcrb. various solns Ⓢ $\boxed{E = M}$ 108.01 – ~~[illegible]~~ 110.20 : This is o.k.

E.g. $\tau^2 \approx s^2 = \underbrace{\frac{\sum \mu_i'^2}{n} - \bar{\mu}^2}_{\alpha} - \frac{1}{N} \cdot \underbrace{\frac{1}{k} \sum \mu_i'(1-\mu_i')}_{\beta}$: $\tau^2 = \alpha - \frac{\beta}{N}$.

But 121.13 – 122.40 gives a better understanding of the process.

Start of review.

→ 126.01

(SN) on Binomial D.F. Define $B(A,B,X) \equiv X^A(1-X)^B$. $B_0(A,B) = \frac{A!\,B!}{(A+B+1)!}$ = ~~fundamental of~~ zeroth moment $B(A,B,X)$

B_1, B_2 are first & second moments $\mu = \frac{B_1}{B_0}$; $\sigma^2 = \frac{B_2}{B_0} - \mu^2$

How, we can prob. get moments of x^2!

Say $P(x)$ is prob that x is betw. x & $x+a$.

~~Then~~ If x is betw x & $x+a$; X^2 is betw. x^2 & x^2+2xa. So Prob x^2 is betw x^2 & x^2+a is $\frac{P(x)}{2x}$

~~So $P_r(v)$ is~~ Say $V = x^2$; then $P_r(v)$ is = prob v is betw v & $v+\delta$ is $\frac{P(x)}{2x} = \frac{P\sqrt{v}}{2\sqrt{v}}$.

Which is t. density func. for x^2.

• 10 Say $P(x) = B(A,B,X)$! $= x^A(1-x)^B$! $P_r(v) = \sqrt{v}^A(1-\sqrt{v})^B \cdot \frac{1}{2\sqrt{v}}$.

$= \frac{1}{2} V^{A-\frac{1}{2}}(1-V^{\frac{1}{2}})^B$: ~~to~~ to get moments, we want $\int_0^1 V^n \cdot \frac{1}{2} V^{A-\frac{1}{2}}(1-V^{\frac{1}{2}})^B\,dv$.

~~$\int_0^1 V^2(1-x^{\frac{1}{2}})^B dx$~~. let $V = x^2$ $dv = 2x\,dx$; $\int_0^1 X^{2n}\frac{X^A}{\cancel{2X}}(1-X)\,\cancel{2X}dx = B_{2n}(A,B)$ (!!)

This seems too simple! If $P(x)$ were some other D.F. say, from $0<X<\infty$ or from $0<X<\infty$, the eq. would be as simple?

On second thot, it seems reasonable. When we want t. first moment of x^2, we want t. expected value of x^2.

Try it out numerically to test t. idea.

If true, it would make it easy to get t. μ & σ^2 of x^2 for $P(x) = B(A,B,x)$.

{ To check, use large A, B; t. first moment, ~~should of~~ for x^2 should be t. square of μ.

$\int_0^1 x(1-x) = \int x - x^2 = \frac{x^2}{2} - \frac{x^3}{3}\Big|_0^1 = \frac{1}{2} - \frac{1}{3} = \frac{1}{6} = \frac{1!\cdot 1!}{3!} = \frac{1}{6}$. checks.

$$\frac{\int X^2\, X^A(1-X)^B dx}{\int \;''\;\;''\; dx} = \frac{\frac{A+2!\,\cancel{B!}}{A+B+1+2!}}{\frac{A!\,\cancel{B!}}{A+B+1!}} \quad \frac{(A+1)(A+2)}{(A+B+2)(A+B+3)}$$ = first moment, $\approx \left(\frac{A}{A+B}\right)^2 = \mu^2$ as expected.

Second moment wrt $x^2 - \mu^2$ $$\frac{(A+1)(A+2)(A+3)(A+4)}{(A+B+2)(A+B+3)(A+B+4)(A+B+5)} - \left(\frac{(A+1)(A+2)}{(A+B+2)(A+B+2)}\right)^2$$

$$\sigma^2 = \frac{A+1}{A+B+2}\,\frac{A+2}{A+B+3}\underbrace{\left(\frac{(A+3)(A+4)}{(A+B+4)(A+B+5)} - \frac{(A+1)(A+2)}{(A+B+2)(A+B+3)}\right)}_{=\alpha}$$

$C = A+B$

$$\alpha = \frac{(A+3)(A+4)}{(C+4)(C+5)} - \frac{(A+1)(A+2)}{(C+2)(C+3)}$$

$$= \frac{(A+3)(A+4)(C+2)(C+3) - (A+1)(A+2)(C+4)(C+5)}{(C+2)(C+3)(C+4)(C+5)}$$

see $\boxed{144.15-.29}$ for simpler derivation & result & discussion of result.

$\boxed{\sigma^2 = \frac{X+1}{R+2}\cdot\frac{Y+1}{R+2}\cdot\frac{1}{R+3}}$. for $\boxed{X>-1,\ Y>-1}$ works for $R \geq 0$ (I checked it out).

$X = \mu R$; $Y = (1-\mu)(R)$; $X+Y = R$

This is o.k., but .10 ff are working on a diff. problem; i.e., t. second moment of σ^2.

.01 :124.40 A General devn. of t problem: In Sci Am May 77, pp 119 ff: Efron & Morris's array adrb. problem: K(=18) baseball players have all been at bat only N (=45) times this yr. Player$_i$ has Batting ave thus far of μ_i'. We want proby that player i will have a hit, next time at bat. If $N \gg 1$ then μ_i' is a very good estimate. If $N \doteq 1$ or some small no., then $\bar{\mu} \equiv \frac{1}{k} \sum \mu_i'$ is [redacted] a much better estimate. The problem, then, is how much wt. to $\bar{\mu}$ & μ_i' for various values of N.

In .01 we are using data from "related sources" to estimate each players ave. In this case it seems clear that this is legit. — but more generally, we can have other sources of Data (e.g. football players, or data on Monthly percipitation) — how should we decide how much wt. to use?

<u>Generalizations</u> of Stein problem: ① Meta Analysis: How to use data from "<u>related</u>" studies ② "T. Mixed Corpus Prbm" also is a (somewhat) general formulation and solution of t. problem. T. biblogy in t. S. Am art. gives a ref. to a paper by Ef & Mor. on "Genzns of Stein".

I will discuss a few mild Genzns of STEIN:

The "true" v_i were <u>created</u> by some unknown p.D. $P_0(v_i)$. Then from each v_i, a v_i' was obtained by a <u>known</u> p.d. $P^*(x_i)$, (e.g. $N(v_i, \frac{v_i(1-v_i)}{N})$)

In t. (a <u>most</u> problems / problem of .01,) $P_0(v_i) = \underset{\uparrow \text{normalizing const.}}{C}(v_i^{X}(1-v_i)^{Y})$ is a reasonable form for P_0. $0 \le v_i \le 1$; Using a Gaussn. $P_0(\;)$ gives trouble - because Gauss is to $\pm\infty$.

Here set $X = \bar{v}R$; $Y = (1-\bar{v})R$: $\bar{v}$ is t. mean of $P_0(\;)$ and R is like its bsz.

The σ^2 of t. d.f. is $\simeq \frac{\bar{v}(1-\bar{v})}{R}$.

In t. Baseball problem, $P^*(v_i) = C^1 \cdot (v_i^{\mu_i' N} \cdot (1-v_i)^{(1-\mu_i)N})$

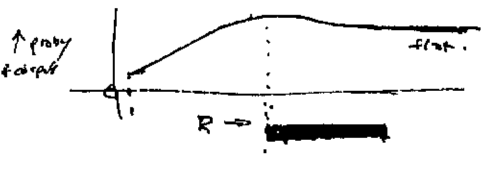
Given $\bar{v}$, R and t. data set $[v_i']$, t. probability of t. "true" averages being $[v_i]$ is

.29 $$\prod_{i=1}^{K}\left(C \cdot C_1 \; v_i^{\bar{v}R}(1-v_i)^{(1-\bar{v})R} \cdot v_i^{\mu_i' N}(1-v_i)^{(1-\mu_i')N}\right)$$

If we integrate over each v_i from 0 to 1, we obtain t. probability that t. set $[\mu_i']$ was obtained from t. D.R. $\bar{v}$, R. — and from a Bayesian View. T. resultant D.R. is a D.R. for $\bar{v}$ and R. See <u>46.28</u> for the resultant integrated form of (.29)

It was easy to integrate the v_i, since $\int_0^1 x^D (1-x)^E dx = \frac{D!E!}{(D+E+1)!}$.

We have written ST46 to compute this integrated expression, 46.28. The D.R. is rather sharp in $\bar{v}$, but <u>not at all sharp</u> in R, for the data given by E&M. for $\bar{v} = \frac{1}{k}\sum \mu_i'$, which is ≈ the peak.



[redacted]

The <u>peak</u> is very slightly > $P(R=\infty)$

The "peak" is at $R = 80$ (perhaps)

4.999 SMFT, STEIN.

.01 If the R d.f. were a sharp narrow peak, we could use it to get good U_j values:

The d.f. for U_j would be 126.29: It has the form $U_j^b (1-U_j)^E$: so we can easily

.03 find its mean (or peak) and variance (see 125.01-.40 for how to do this)

For the data of E&M, since there is no peak, we may want to use the expected value of R. Since R ranges to $+\infty$, we need an a priori for R. — such as $\frac{1}{R^2}$ or $\frac{1}{R(\ln R)^2}$. Using such a priors, we can get E(R), but this is not a rigorous argument — (we would, presumably, use that R as in .01-.03, to make predns of U_j).

A better way: For each R (assume we know $\bar{U}$), we can find an expected

(.12) value of U_j: it will be $\approx \left(\frac{R \cdot \bar{U} + N u_j'}{R+N}\right)$:

to get E U_j over all R, we integrate say $\int_1^\infty (46.38) \cdot (.12) \cdot \left(\frac{1}{R^2} \text{ or } \frac{1}{R(\ln R)^2}\right) dR$

.14 This can be done numerically.

A better way (perhaps equivalent), is to use the general "exact" Bayesian approach. Given any data set $[U_j']$ we can assume the $\bar{U}$, R model of P_0,

.17 and, by integrating over all U_j & R we obtain the a prp. of that $[U_j']$ set.

What is the relative proby that the j^{th} player will have a hit rather than an out on his next trial? Say $P([U_j'])$ is the proby we obtain in .17 for the data set $[U_j']$

Let S_j^0 be the data set $[U_j']$ but with $N \to N+1$ for player j and $U_j' \to \frac{U_j' \cdot N}{N+1}$ (an "out")

" S_j^1 " " " " " " " $U_j' \to \frac{U_j' \cdot N+1}{N+1}$ (a "hit")

.23 Then the rel probys of "hit" vs. "out" on next trial are $\frac{P(S_j^1)}{P(S_j^0)}$ from this ratio, we can

.24 readily determine U_j. I suspect this result will be about the same as that obtained in (.14)

The results of .23-.24 are obtained in ST100.BAS and ST101.BAS.

The foregoing analysis is not the way E&M do it.

The "general" idea of E&M is this:

we have $P_0(u_j)$ — It is of known mean $\left[\bar{U} = \frac{1}{k}\sum_{j=1}^{k} u_j'\right.$ but unknown variance, τ^2.

.32 If we know τ^2 & $\bar{U}$, and N we could, obtain the expected variance of the resultant data set $[U_j']$. (As a fnct of τ^2 & $\bar{U}$ & N) Then, by looking at the actual varc. of $[U_j']$, we can solve for τ^2. (since we know N & assume $\bar{U}$)

$\boxed{\text{Approximately}}$: Varc of $[U_j'] = \tau^2 + \frac{1}{k}\sum \sigma_j^2$. (Actually, exact for $\approx$ any d.f.s for "τ^2" & "σ_j^2") See 146.01-.10 for analysis

.37 so $\tau^2 = \underbrace{\left[\frac{1}{k}\sum (U_j')^2 - \left(\frac{1}{k}\sum U_j'\right)^2\right]}_{\text{Varc of } [U_j']} - \frac{1}{k}\sum \underbrace{\frac{U_j'(1-U_j')}{N}}_{\approx \sigma_j^2}$ these should be $\frac{U_j(1-U_j)}{N}$ . . . but of course, the U_j are unknown

(130.01)

See 130.01 for better way to do this!

$\boxed{\text{SN}}$ There would seem to be a systematic error in using $\frac{U_j'(1-U_j')}{N}$ instead of $\frac{U_j(1-U_j)}{N}$ — since all U_j' move toward $\bar{U}$

4.9.99 SMA+ STEIN

.01: A more exact treatment of 127.32: Given x, y (equiv. to R, $\bar{U}$ or T, $\bar{U}$).
The d.f. of any particular U_i' is "like" 46.28:

.03
$$\frac{x+y+1\,!}{x!\,y!}\qquad\frac{x+N\cdot U_i'\,!\ \ y+N(1-U_i')\,!}{x+y+N+1\,!}$$

We want the d.f. of .03 about $\bar{U}$ — or more precisely the second moment of the .03 d.f. about $\bar{U}$.

.07 $\boxed{SW}$ Perhaps express .03 as a Gaussian or better as a Binomial d.f. in U_i' — to make it easy to get the moment → 130.01

41099 SMFT STEIN:

Discn. w. Alex: 1) He didn't know what to do in a Bern seq. (by radix) when freq. of a symbol was zero.

I told him about Laplace's rule: Radix r: Same as using Sh rules, but pre counts " $a_1, a_2, a_3 \cdots a_r$.

. eg Also it was derived from uniform a priori: $x_1^{N_1} x_2^{N_2} \cdots x_r^{N_r}$ = Pcr. $\sum_{i=1}^{r} x_i = 1$

This gives pc of $x_1 \cdots x_r$ in view of data ~~x~~ $N_1, N_2 \cdots N_r$. The expected value of the ~~x~~ $\bar{x}$ vector is obtained by usual computation of a prps. $\frac{N_1!\, N_2! \cdots N_r!}{((\Sigma N_i)+r)!}$

How to get SSS (.03) w. the constraints, was ~~x~~ something I solved long ago — & don't remember how.

I think it involves ~~x~~ multiple S's in which upper ~~x~~ integration in each is a linear funct of some of the x_i. — That the formula is obtained recursively.

He was concerned w. Stein problems in which one wanted to "pool" data, but didn't know how "related" the potential "poolees" were — & just how to express & use the "closeness".

Perhaps Get more info on just what his problem is (in Speech recognition), & perhaps work on it. This seems like a very general problem probly quite impt.

?rr

→↓

4.10.99 SMRT STEIN

did I mean $\underline{\mu_i}$?

.01 ! $\overset{128.07}{127.40}$ A more exact way to do it than 127.37 !

Given $\bar{u}$, R, N, The d.f. for u_i' (if u_i' ~~[illegible]~~ were not known) would be (from 46.28)

.03 $$\frac{x+y+1!}{x!\,y!}\;\frac{\boxed{(x+u_i'N)!\,(y+(1-u_i')N)!}}{x+y+N+1!}$$: where only the red part varies w. u_i' :

So we want its D.f. We can probably get its mean & varc & approximate it as a Gauss or Binomial d.f.

We then want the D.f. of $u_i'-\bar{u}$, and the second moment (σ^2) about that $\bar{u}$.

~~[illegible]~~ say this second moment is $F(R,\bar{u},N)$.

It is the Expected value of $(u_i'-u)^2$.

If we make k choices of u_i' & add ~~[illegible]~~ their $(u_i'-u)^2$'s we Expect to get

.11 $k\cdot F(R,\bar{u},N)$. We then ~~[illegible]~~ write

$$\underbrace{\sum(u_i'-\bar{u})^2}_{\text{observed}} = k\,F(R,\bar{u},N)$$ and solve for R in terms of $\bar{u}$, N, $[u_i']$.

.13 Often, this eq. will have no usable soln. (i.e. ~~[illegible]~~ $R<0$, $\left(\frac{1}{\tau^2}<0,\ \tau^2<0\right)$

In such a case, $\tau^2=0$ ($R=\infty$) is not a bad approxn.

To do it more correctly, in (.03) if we get the $\underline{\underline{\text{varc}}}$ of the second moment of the red expression. This gives us $\sigma^2(R,\bar{u},N)$, so we have a d.f. for R. When ~~[illegible]~~ $\sum(u_i'-\bar{u})^2$ is known, this gives (by Bayes) a d.f. for R (& i.e. of $\underline{\underline{\tau^2}}$), This deals w. the "R<0" diffy of (.13)

In General! for expressions like (.03) We want the $\underline{\underline{\text{mean}}}$ = (first moment / zeroth moment)

Also σ^2 = Second moment/zeroth moment $-(\text{mean})^2$.

If we get the d.f. for $(u')^2$, then I think its mean is this → σ^2 (I'm not sure).

Say H(V) is the d.f. for ~~[illegible]~~ V ($= (u')^2$), & its mean is M_V,

Then if we take K samples $(u_i'-\bar{u})^2$, we expect their d.f. to be $H(\frac{V}{K})$ & its mean to be K Mv. We also ~~we~~ may be able to compute the varc of MV.

4.11.99 START STEIN:

.01 On the $\boxed{\text{Variance of the Variance}}$

Consider $e^{-\frac{x^2}{2}}$: Its mean is 0, varc = 1.

Say $X^2 = V$; its d.f. is $\propto \frac{1}{2x} e^{-\frac{x^2}{2}} \propto \frac{1}{2\sqrt{v}} e^{-\frac{v}{2}}$; $= \boxed{f_1(v)}$

Re. $f_1(v)$: M_0 is $\int_0^\infty \frac{1}{2\sqrt{v}} e^{-\frac{v}{2}} dv = \frac{1}{\sqrt{2}} \int \sqrt{\frac{2}{v}}\, e^{-\frac{v}{2}} dv = \frac{2}{\sqrt{2}} \left(-\frac{1}{2}\right)!$

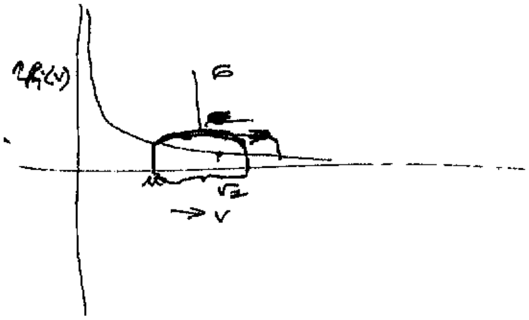
$$M_1 = \int_0^\infty \sqrt{v}\, e^{-\frac{v}{2}} dv = \sqrt{2} \int \sqrt{\frac{v}{2}}\, e^{-\frac{v}{2}} du = 2\sqrt{2}\left(\frac{1}{2}\right)! = 2\sqrt{2}\left(-\frac{1}{2}!\right)\cdot\frac{1}{2}$$

$$M_2 = \int_0^\infty v^{\frac{3}{2}} e^{-\frac{v}{2}} dv = 2^{\frac{3}{2}} \int \left(\frac{v}{2}\right)^{\frac{3}{2}} e^{-\frac{v}{2}} = 2\, 2^{\frac{3}{2}} \left(\frac{3}{2}\right)! = 2\cdot 2\cdot\sqrt{2}\cdot\left(-\frac{1}{2}!\right)\cdot\frac{1}{2}\cdot\frac{3}{2}$$

$$\frac{M_1}{M_0} = \frac{2\sqrt{2}\cdot\frac{1}{2}}{\frac{2}{\sqrt{2}}} = \frac{2\cdot\sqrt{2}\cdot\sqrt{2}}{2\cdot 2} = 1$$

$$\frac{M_2}{M_0} = \frac{\sqrt{2}\,\cancel{2}\cdot 2\cdot\sqrt{2}}{\cancel{2}}\cdot\frac{1}{2}\cdot\frac{3}{2} = 3$$

So Varc of $f_1(V)$ is $3 - 1^2 = 2$; $\sigma = \sqrt{2}$.
mean of $f_1(V)$ is 1



This is a very simple idea! Whenever we have a var. that is obtained as the sum of a bunch of random variables, we know the mean is sum of means & varc is sum of varcs:
This gives us the d.f. for the varc. of the sum variable — how uncertain of it we are.

.20: on A more exact analysis of the Gaussian case: $P_0(u) \propto e^{-\frac{u^2}{2\tau^2}}$; mean = 0.

The u obtained is then subject to a d.f. $e^{-\frac{(u-u_i)^2}{2\sigma_i^2}}$

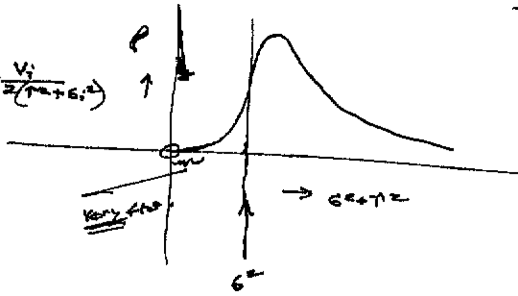
The resultant d.f. is $\propto \frac{1}{\sqrt{\tau^2+\sigma_i^2}} e^{-\frac{u_i^2}{2(\tau^2+\sigma_i^2)}}$; The varc is $\tau^2+\sigma_i^2$ because the new d.f. is the convoln of the 2 d.f.s of varc τ^2 & σ_i^2 resp.

~~Consider the case where $\sigma_i^2 = \sigma^2$ a constant~~

.26 The d.f. of $V_i \equiv u_i^2$ is $\frac{1}{2u_i(\sigma^2+\tau^2)} e^{-\frac{u_i^2}{2(\tau^2+\sigma_i^2)}} = \frac{1}{2\sqrt{v_i(\sigma^2+\tau^2)}} e^{-\frac{v_i}{2(\tau^2+\sigma_i^2)}}$

To study this, let all $\sigma_i^2 = \sigma^2$ = same constant:

.27 d.f. $\propto \frac{1}{\sqrt{v}\cdot\sqrt{\sigma^2+\tau^2}}\, e^{-\frac{v}{2(\sigma^2+\tau^2)}} \equiv f_1(v)$



But ~~[illegible]~~ first take $\sum_{i=1}^{k} (u_i - 0)^2 \equiv \Sigma^2$.

If we take k'th convolution of $f_1(V)$, we will get the d.f. for the sum of squares of k, u_i's.

The mean will be $k \times$ (mean of $f_1(v)$) (which is prob'ly $\sigma^2+\tau^2$)). This convolution will be much like $f_1(v)$ (.26 R), but the peak will be narrower relative to the mean (width $\propto \overset{\sigma}{\frac{1}{\sqrt{k}}} \cdot$ mean))

I'm not sure about how rapidly this convolution $\to 0$ at $\tau^2 = \infty$, however.

Consider the convoln of $\underset{\text{normalization const.}}{\frac{1}{2a\sqrt{x^2+a^2}}}$ w. itself. !

In the convoln, for large x, we can approximate the convoln by integrating the $\int_{x-a}^{x+a} \frac{1}{\sqrt{x^2+a^2}} dx$

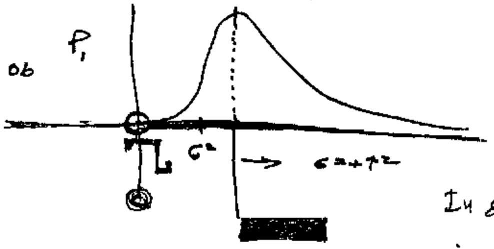
$$\approx \int_{x-a}^{x+a} \frac{1}{2a} \frac{dx}{x} = \frac{1}{2a} \ln\left(\frac{x+a}{x-a}\right) \approx \frac{\ln}{2a}\left(1 + \frac{2a}{x}\right) \approx \frac{2a}{2ax} = \frac{1}{x}$$

So for that kind of func, convolving seems to leave high x values invariant

Anyway, lets assume that we convolute, k $F_1(v)$ d.f.s. at 131.27 K:

The peak gets narrower (relative to mean), but t. behavior at hy τ^2 remains as $\frac{1}{\sigma^2+\tau^2}$.

.06



If k is small, the peak can be at $< \sigma^2$
but for large k, its very likely that peak is $> \sigma^2$
(unless τ^2 is, in fact, very small.

See doc: NN! 4.12.99, ①
www.zzvh.com

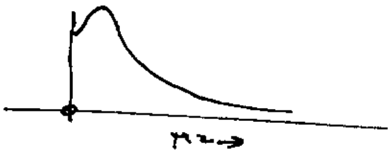
In general, t. peak will tend to be as $\sigma^2+\tau^2$, but
If k is small and τ^2 is small, it can occur at $< \sigma^2$

.09 [redacted] Th. general form of t d.f. for $\Sigma^2 \equiv \sum_{i=1}^{k}(\mu_i - \overset{\circ}{\bar{\bar{\mu}}}) = \boxed{V}$

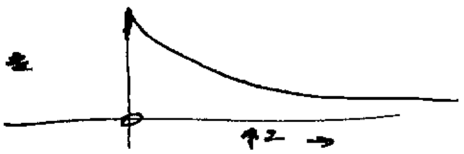
.10 131.27 $f_1(V) \sim \frac{1}{\sqrt{V}\sqrt{\sigma^2+\tau^2}} \cdot e^{-\frac{V}{2(\sigma^2+\tau^2)}}$

We know $\underline{V} = \Sigma^2$, so this gives us a d.f. for $\sigma^2+\tau^2$. (by Bayes).

If $\sigma^2 <$ t. peak of t. curve (see (.06)) we get ① if $\sigma^2 >$ t. t. peak we get ②



①



②

.21 ~~Using Bayes~~. This d.f. for τ^2 ; $\left(\mu_i' = \frac{\frac{\bar{\mu}}{\tau^2} + \frac{\mu_i^f}{\sigma_i^2}}{\frac{1}{\tau^2}+\frac{1}{\sigma_i^2}}\right)$ eq.

We can integrate τ^2 from 0 to ∞ t. product of [(.10) t. pd. for τ^2]; [t. a priori pd for τ^2] [(.21)]

(Maybe $\approx$ ~~$\frac{1}{\tau^2+\sigma^2}$ $\frac{1}{\tau^2+\sigma^2}$ $\frac{1}{\ln\tau^2+\sigma^2}$~~ $\frac{1}{(\tau^2+\sigma^2)\ln^2(\tau^2+\sigma^2)}$

.26 or $\frac{1}{(\tau^2+\sigma^2+\sigma^2)}\frac{1}{\ln^2(\tau^2+\sigma^2+\sigma^2)}$ moves the graph over by σ^2.

This integral would give us t. correct value of μ_i' — considering all possl.
.28 values of τ^2.

.29 [This eqn. ~~was~~ would be ~~at~~ T. same as ST100.Bas, ST101.Bas if t. statistic Σ^2 (.09) has as much info as $[\mu_i^f]$ w.r.t. t. estimation of μ_i'.

T. usual method ~~of ST/regression~~ ST101/[illegible] finds t. $\boxed{\text{d.f. for } R \ (= \frac{1}{\tau^2})}$ by ~~[illegible]~~ computing t. probty that each τ^2 produced known $[\mu_i^f]$.

The Gaussian approach 131.20 – 132.28 finds t. d.f. of τ^2 by computing t. probty that each τ^2 produced t. observed $\sum_{i=1}^{k}(\mu_i^f - \bar{U})^2 \equiv \Sigma^2$.

Both methods could use (MAW) .21 – .26 to compute each μ_i'.

However, in ST101.bas what we actually do is more general: We assign an a priori probty to t. data set $[\mu_i^f]$ — or any conceivable data set. We then get t. rel. probys of various possl. countings of t. corpus (= data set) by computing t. rel. probtys of t. assoc. corpi.

7.14.99 SORT STEIN

.01 This latter method is more general, & works for all problems.

Both methods ~~[illegible]~~ seem to point ~~up~~ out t. impt. of t. a priori — at hy values: In one case hy $R(=\frac{1}{1+a})$ in t. other case hy τ^2. In both cases no problem for small R or τ^2 ! — which seems crazy! It may be that in both cases, t. dependence on behavior for hy R or hy τ^2 produces ~~[illegible]~~ negligible changes in t. Gwc. ~~[illegible]~~ upon This would mean for/ rel. wts near 0 or near 1, t. exact values are not very impt.

Hvr., in General, I'm not easy w. t. state of affairs.

Another thing I'm uneasy about is G. & M's using ~~[illegible]~~ $\frac{S^2}{k-3}$ instead of $\frac{S^2}{k-1}$ to estimate V. I could use t. Mt. carlo. data generator to find, empirically, which gives best perfn ——→ 136.09

.17

Write detailed outline of talk on STEIN for Oxbridge:
Then ~~[illegible]~~ fill in outline, w. actual text; & also refs to these Notes.

3 impt probs
1) Mt carlo.
2) $\frac{101}{100}$. Bas
3) Prob to do $\sin^{-1}(2x-1)$ xfm & get E&M results exactly.
4) Prob to do ~~Gaussn~~ Gaussn case: But for $P_0(\cdot)$ Gaussn(σ^2) to $P_g(\cdot)$.
$\Sigma^2 = \left(\frac{\sum_{i=1}^{k} \sigma_i^2(1-U_i)}{N}\right)$
Did I compare this w. E&M & BA?

AH! ① Re: the $\sin^{-1}(2x-1)$ xform! While t. σ_i^2 is correlated w. μ_i, that's not impt.: t. only thing of impt. is whether t. different μ_i's are correlated or whether their σ_i^2 are correlated. — They are not — which is all we need to know.

② In ~~St~~ $\frac{101}{100}$. Bas, we integrate 46.25 in R, w. constant $\bar{U}$. Actually, tho the peak may be $\bar{U}$, the width may vary w. R & t. width may vary w. t. corpus — (so to compare ~~[illegible]~~ ~~[illegible]~~ t. a priori (≡ integral) of 2 models of t. corpus, one must see if t. widths in t. $\bar{U}$ directions are different).

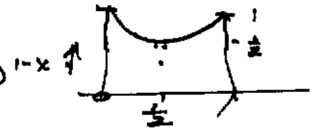
To get this width as a function of R may not be so diff/t!

I may have come close to doing this calcn in 49.07–.30; $\boxed{70.01-76.40}$

70.01–.40 is particularly relevant t. concavity argt. at 70.34

It may involve something like $\left(\frac{\prod_{i=1}^{k} \left(\mu_i'^{\mu_i'}(1-\mu_i')^{1-\mu_i'}\right)}{\left(\bar{u}^{\bar{u}}(1-\bar{u})^{(1-\bar{u})}\right)^{k}}\right)^{N}$ see 70.23

The numerator is always > (or < ?) t. denominator — by some simple convexity argt. ~~[illegible]~~. Whether it went or its reciprocal, is unclear!



at any rate, just varying $\bar{U}$ would make t. peak at $\bar{U}=\frac{1}{2}$! — So look into this more carefully!

4.16.99 SMFT STEIN 137

Not ~~in~~ in Stein ~~file but SMFT (99)~~

.01 One application of STEIN effect: Chosing Mutual funds. → Hw, see (99) 3.10 very ~~impt~~ idea

Say we list all Fid. funds that we mite consider getting into.

Or, use large list of No load funds ~~in which~~ — we use an on-line broker & switch over month or every 6 months.

Say we examine recent ~~recent~~ gains of funds: ~~wrt~~ w.r.t. $\bar{\mu}$ set to zero;

expected yield is [redacted] $\frac{[redacted]\frac{\mu_i}{\sigma_i^2}}{\frac{1}{\tau^2}+\frac{1}{\sigma_i^2}} = \frac{\tau^2\mu_i}{\tau^2+\sigma_i^2} = \mu_i\cdot\frac{1}{1+\frac{\sigma_i^2}{\tau^2}} = \frac{\mu_i}{1+\frac{\sigma_i^2}{\tau^2}}$

Say we use μ_i & σ_i^2 for exponential, last 5 yrs; 6 month data pts.

Each 6 mo. we pick a fund w. best $\frac{\mu_i}{1+\frac{\sigma^2}{\tau^2}}$ for t. "last 5 yrs".

μ_i is 6 mo. yield.

For ≈ "FT" type scheme: ① For each driven, find driver that correlates best w. it — for predn. Using Stein idea, compute for each driven i, α_i ∋

.13 $\mu_i + C_i\cdot\alpha_i\cdot p_i^r$ ~~is good~~ best possl. predictor of p_i^h:

~~Meaning~~ p_i^h is [redacted] Δ price of driven. (wrt its mean, μ_i).

p_i^r " [redacted] " " " driver. (w.r.t its mean)

C_i is coeff of corr betw. p_i^h & p_i^r.

α_i is obtained from Stein analysis.

} what is μ_i? — how do we get it? Maybe mean ~~over~~ over many (driver, driven) pairs.

Compare w. normal linear regression: we want $\beta \ni \beta(x-\bar{x}) \approx (y-\bar{y})$

$S \equiv \sum(y-\bar{y}-\beta(x-\bar{x}))^2 = \min$; $\frac{\partial S}{\partial \beta} = 2\sum(x-\bar{x})(y-\bar{y}-\beta(x-\bar{x})) = 0$.

$\sum(x-\bar{x})(y-\bar{y}) = \beta\sum(x-\bar{x})(x-\bar{x})$ so $\beta = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} \doteq$ correl. coef.

Say $\bar{x}=\bar{y}=0$; to simplify.

From linear regression it would appear that α_i (in .13) = 1.

Hw, if ssz (≡ k) is small (ε large) t. past efficacy of C_i is not as important as ~~knowldg~~ knowledge of t. ~~mean~~ mean (β). The $\frac{k+1}{k-1}\sigma^2$ effect is of some help here —

— but is it t. **whole story**? I think not! This ∃ effect tells us expected error in future,

.32 but it doesn't tell us to multiply estimate by α_i w. $\boxed{\alpha_i<1}$. !! → 137.01 136.01

Say t. corr coef $\beta \equiv \beta_\tau$ was obtained w. [redacted] window, τ. If τ is small (say 2), the error in β due to small τ is quite large. Consider R windows: divide corpus into sections of length τ; get β_τ for each section. We can then use STEIN Co-estimating for ~~estimating~~ estimating β_τ. Since we are mixing data, this amounts to using larger τ.

.36 Perhaps more relevant: Due to fact that seq. is slowly changing its params, we are limited in how large τ can be. Say τ can be as large as τ_0, but no larger. Then each consecutive section of t. corpus, of length τ_0, corresponds to a different "Baseball player" (in t. "STEIN" analysis).

4-1999 SMFT STEIN

135

628 9792
3:00 p

.01 In SMFT, τ_0 was ~ 300 days max — τ_0's was for t. entire ~~set~~ of ~~[illegible]~~ dd pairs we used. — Hvr., each dd could have its own ~~[illegible]~~ τ_0.

In ~~looking for corr. coeffs~~ trying to evaluate corr. coeffs: Sup we use ~~[illegible]~~ x windo, τ. Each day we get a new value of β. — so we get from this a σ_β^2 (σ_β^2) per day, which will be a best expected error in β. [t. corr coef. is closely analogous to Batting Average: ~~it~~ Beta gives t. p.d. for t. next data pt.] In Baseball, it's prob'y of ~~[illegible]~~ out v.s. hit. In FT, it's prob'y distribn of ~~[illegible]~~ Δp_i^k.

.11 In 134.36 ff: we could assume a Gaussian D.F. for t. corr coefs of each τ_0 sized set (subcorpus). In ideal form, these corr coefs would be statistically indep. In most time series, they would be sequentially correlated. Using x windo might be one way of doing it →

Long ago, I was disturbed by t. following analysis: I had a random var, x_i w. mean value ε (small but > 0). If I used x window to smooth x_i to obtain estimate of x_{i+1}, the s.d. ~~error~~ of my estimate was > ε, so I'd do better by chosing zero every time. (I didn't know what ε was). So ~~zero~~ estimate of zero, had less rms. error than t. "x windowed" x_i. I was really puzzled by this!

→ Actually in 134.36 & ff (135.11), one could deal w. this linearly by assuming t. signal had a slowly & a rapidly changing component. So perhaps a "Wiener filter" would be t. best way to deal w. this.

Or, in t. spirit of neural nets, find 2 or 3 ~~[illegible]~~ x window τ_i's & assoc. wts., & make predn. a sum of them. T. reason it's like ANN is that t. predn is t. sum of 3 filter outputs, & each filter has 2 params — one is size in t direction & other is amplitude in Price direction. This can be much better than simply adjusting amplitudes of several "pre-picked" wting functions ~~[illegible]~~ This is ~~[illegible]~~ an impt. idea from A. Barron's 1993 paper

Looks Good!

Hvr. it is a N.L. curve fitting problem, & prob'y best be solved using some Standard N.L. curve fitting pgm. (optim).

$\boxed{42199}$ As I see it, t. main idea is to get good predns for all stocks (funds). — we then pick t. best one, & we can estimate yield using t. ideas of "SOY" analysis.

.01: (Spec 134.32): Another way to look at this! Consider t. situation of 134.13 — .32 ; (Mean=0)

.02 Say we do linear regression for some τ/R window: we observe $\sigma_{ob}^2 \to \sigma^2_{emp.} = \sigma^2_{ob} \cdot \frac{\tau+1}{\tau-1}$.

If σ_{ob}^2 is comparable to $\frac{1}{\bar{x}^2}$ ← window τ, then we have 2 [redacted] codes ~~of~~ for t. data that are of comparable length. The pc/symbol of t. 2 codes are $\boxed{\frac{1}{\sigma_{emp}}}$ & $\boxed{\frac{1}{\sqrt{\bar{x}^2}}}$.

.08 — Hvr, t. initial pc's of t. codes are different. The "bet on zero (= $\bar{u}$)" code has hy ~~..~~ apriori. → .36

.09: 133.17 **SN** Re: E's M's $\frac{S}{k-3}$ My strong impression is that this should be $\frac{S}{k-1}$.

Say we have a Gaussian process w. var. τ^2+1, & mean = μ. We do Mt Carlo runs & observe in i run, mean μ_1 & ms deviation from that mean of σ_1^2.

An unbiased estimate of τ^2+1, is then $\sigma_1^2 \cdot \frac{k}{k-1}$, not $\sigma_1^2 \frac{k}{k-3}$.

A new, simple way of getting d.f. for sum of squares ~~of~~ of Gaussian variables! Consider first known mean = 0. We have a k dimensional ~~...~~ spherical Gaussian d.f.: say t. d.f. is $e^{-\frac{x^2}{2}} \to e^{-\frac{\Sigma x_i^2}{2}}$.

Then the r-th moment of t. d.f. is $(R \equiv \sqrt{\sum_{i=1}^{k} x_i^2})$ $\int_0^\infty R^r e^{-\frac{R^2}{2}} \cdot f(R)\, dR$

f(R) is the "area" of a k dimensional hsphere of radius R. it is $\propto R^{k-1}$ and so ~~E~~ Γ functions. — so we end up w. integrals of form $\int_0^\infty R^\alpha e^{-\frac{R^2}{2}} dR$ which are ~~...~~ well known.

.24 If t. mean is ~~not~~ zero, But ~~is~~ is unknown: we can get t. d.f. of $\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{k}$ which is t. d.f. of Σx_i^2 ~~...~~ minus $\frac{1}{k}$ × d.f. of $(\Sigma x_i)^2$.

d.f. of $(\Sigma x_i)^2$ is not so easy: first get d.f. of Σx_i.

or, we know t. d.f. of $\bar{x}$; its a Gaussian d.f. about t. mean w. var = $\frac{1}{k}$ × var of x_i: Then, given t. d.f. of $\bar{x}$, we can get t. d.f. of $(\bar{x})^2$;

.29 it is $\frac{1}{2\bar{x}} P(\bar{x})$. $= \frac{1}{2\sqrt{\bar{x}^2}} P(\sqrt{\bar{x}^2})$ → also note 139.20

So .24 – .29 would seem to be necessary soln. to t. problem of whether to use $\frac{S}{k-1}$ or $\frac{S}{k-3}$ or whatever. We could then confirm our soln. by Mt. Carlo simulation,

So this may more or less, "settle" STEIN for now.

Applying t. ~~...~~ Stein ideas to ~~...~~ SM in an optimum manner is another Question, hvr.

.36 (.08) A nice way to study this: say $y_i = a x_i + N_i$ N_i = Noise of $\sigma^2 = N$, & mean of N_i is ϕ.

Using finite τ, ~~comp~~ estimate ⓐ from t. corpus. We then use .01 – .08 to get rel. wts. of 2 predn. methods: ① using ϕ as predn. ② using $a_{obsvd} x_i$ as predn. We know expected errors for each. It would seem that as N gets larger t. predn ① gets more wt. — it should get all wt. as $N \to \infty$. From this model, I should be able to → 138.01

4 22 99 SMFT

.01 **NB** A very impt. Aspect of t. E&M approach to Stein: We have a model for t. corpus (or, more exactly, a model for an ensemble). We mite ask, "What is t. d.f. of t. params of t. model, in view of t. corpus?" Then we integrate ~~[illegible]~~ over those params to get t. apriori pd of t. corpus.

Another way (t. one E&M uses). We make a certain measurement of t. corpus. What is t. d.f. of one (or more) params of t. model ~~[illegible]~~ in view of

.02 this measurement? — E&M narrow it further by asking (as "statisticians") "What single value of t. params fits t. measurement best?"

This last is often much easier to compute ~~[illegible]~~ than t. earlier Q's of .01–.02;

— Tho often it seems doesn't give exactly rite answers, i.e. it sweeps impt. Q's a bit "aprioris under t. rug".

Anyway, E&M's method is one kind of "ruff & dirty" soln. to an induction problem.

→ In general, how do we go about inventing ~~[illegible]~~ approxns of this sort?

Some Common heuristic tricks: Instead of a d.f. for a bunch of params, use the peak, or if t. peak is too broad expected values.

[These all assume that t. underlying apriori pd is ~~[illegible]~~ "uniform"]

The idea of picking a (single) (or stat) measurement of t. corpus & getting t. "best fit." or d.f. of model params ~~from this,~~ in view of this/(a set) measurements(s), is perhaps a common trick — tho statisticians prob do it w.o. ~~[illegible]~~ realizing they are doing this heuristic Approxn.

There is also t. idea of t. "Sufficient statistic".

How to discover such measurements? Well, consider a single param of t. model; to get its pd. or "best value": Consider extreme values of that param. How would they affect t. way t. corpus looked? From this one ~~[illegible]~~ could get ideas of what measurements (s) to make on t. corpus, that would be most sensitive to this param. (or sensitive at all!)

Another trick: to find how t. measured param is a funct of τ^2 (say), first: Consider extreme values of τ^2 (say 0 & ∞): This ~~will~~ may give ideas on ~~[illegible]~~ t. form of functional relation (we may need some intermediate values)

Anyway, using Mt carlo simulation, for various τ^2 values get t. value of t. measured param. Do a curve fit on this data. Or, for various τ^2, get a Mt carlo d.f. of t. values of t. ~~pa~~ measured param. By curve fitting on this last data, one can get (by Bayes) t. d.f. of τ^2 as a funct of t. measured param.

This can be useful in A.I., if one doesn't have a good intuitive model of t. data.

4-2299 SMFT & STEIN (Rev)

.01: 136.40: Estimate ms error of any proposed "a", & any "mix" of methods ① & ②
— Any such mix is equiv. to reducing "a" by a certain factor — so I can find error of the mixture, if I know error assoc w. any "a"..

.04 **REVIEW** see 115 for list of reviews, also 143

I will actually write a "paper" on this for my own remembrance, since its an impt. idea: Outline:

1) What the Stein Problem is as in the Sci American Art. May (1977).
Desrb in terms of Baseball: Give intuitive picture of $s^2=1$ v.s. $s^2=\infty$.

2) Generalizations of the ~~St~~ Problem (why its an impt. problem) ⓐ ~~B~~ Paper by E & M on Genzns.
ⓑ Genzn to Bin. & Gauss d.f.'s ⓒ Various kinds of auxilliary data: (c1) Metaanalysis,
(c2) The Mixt Comp Thm.

3) I see 4 versions of the Problem: ⓐ Gaussian v.s. Binomial d.f.s;
ⓑ Exact v.s. approximate solns. ("Exact" means ~~[illegible]~~ Complete Bayesian soln. including aprioris. — Approxns: use mainly using $\Sigma^2 = \Sigma(\mu_i - \overline{\mu_i})^2$ as a "sufficient statistic"

4) Various Math'l methods, ~~[illegible]~~ identities, inequalities, equations, that are helpful in this area. Formulas for moments & μ & σ² of Bin. D.f. (144.15 ff)
How to get d.f. for x^2 from d.f. for x. 125.01 ff is one treatment of this
Derivative of ~~[illegible]~~ $\ln\Gamma(z)$ is $\psi(z)$ p 258 (B of standards):
$\psi(x)$ is very close to $\ln(x-\frac{1}{2})$ for $x \geq 2$.

$\int_0^1 x^{\alpha}\cdot y^{\beta}\,dy = \frac{x!\,y!}{(x+y+1)!}$ (p 258: B. of Stand)

$\frac{x!\,y!}{x+y!} \approx (a^a b^b)^R \sqrt{2\pi abR}$ 74.03; $a = \frac{x}{R}$; $b=\frac{y}{R}$; $R = x+y$.

$\ln\left(\qquad\right) \approx x\ln x + y\ln y - R\ln R + \frac{1}{2}\ln\left(\frac{2\pi\, xy}{R}\right)$

$x! \approx \left(\frac{x}{e}\right)^x \sqrt{2\pi x}\cdot \underline{\underline{e^{\frac{1}{12x}}}}$ is a very good trim on the Stirling Approx.
for x=2 it gets $2! \approx 2.00065$. It ~~[illegible]~~ gets much better as x↑.
Variance of Variance: 131.01.
Varc of Binomial d.f. 76.25 (approx); 144.15–.25 (exact)

4.30.99 SMFT STEIN: 139

There is a (ALP) model of the ~~[illegible]~~ (Gaussian) problem; Say $\bar{\mu}$ is known to be ϕ.
We have 2 methods of prediction; predn is ϕ, which has msErr = τ^2.
We have μ_i as predn, which has msErr = σ_i^2: ~~[illegible]~~
To get the pdf of the more reading, we mult. the 2 d.f.'s together, to get a Gaussian d.f.
of mean $\left(\frac{\phi}{\tau^2}+\frac{\mu_i}{\sigma_i^2}\right)/\left(\frac{1}{\tau^2}+\frac{1}{\sigma_i^2}\right)$.

Hvr, we can't find τ^2 so easily! Say we are using linear predn,
so $y_i \cong \alpha x_i$. (mean=0) We calculate past error in predn $\overset{A\ ???}{\sum}(y_i-\alpha x_i)^2$ w.
$B \equiv \sum y_i^2$. A is always < B.

Hvr, say we use a window w. T as smoothing time; we compute α on basis of
last T data pts & make predn. It has ms error of A: this can be > ms value of y.
By "Akaike methods" we can find expected value of A by looking at past only.
Use a $\frac{T+1}{T-1}$ factor. This may be ok.

.20 SN My ~~[illegible]~~ "proof" of the Akaike correction factor doesn't ~~[illegible]~~ give reasonable
answers for $k>1$ (No. of params). — But the [illegible] this N dimensional
Gaussian d.f. proof of mine — It doesn't seem to be "singular" when
$N < k$. I just get the factor $\frac{N+1}{N-1}$ .... but check this out!

Consider just error in ~~[illegible]~~ $= \sum_{i=1}^{k} \sigma_i^2$, — I.e. expected value of free variance.
This should simplify it a lot!
This is ~~[illegible]~~ to problem of $\boxed{136.24-.29}$ — which I think is rather simple, ~~[illegible]~~ if
one only wants means (or "unbiased estimates")

The expected error in mean (if all n dimensions have err = σ^2) is Σ^2 err =
$n \cdot k \cdot \sigma^2$ — $\boxed{\text{So maybe it works out after all.}}$
its $nk\sigma^2$ because its equiv. to taking $n \cdot k$ samples of each var. σ^2 ~~[illegible]~~

Well, Turns out I was doing this wrong! The method I used to get Akaike factor
for $k=1$ doesn't work (directly) for $k>1$. Say we have k variables x_i of ~~[illegible]~~ $\sigma_i^2=\sigma^2$ (constant)
the means are all ϕ, but we don't know that. The only remaining error in the mean vars,
is the mean of the means: which is essentially a single param, — so it acts ≅ like "k = 1"
So we don't get factor of $\frac{N+k}{N-k}$: (~~[illegible]~~ more like ~~[illegible]~~ $\frac{Nk+1}{Nk-1}$)

Closer to correct: $(y_i, x_{ij})\binom{i=1/N}{j=1/k}$ are vars of mean ϕ & Gaussian.
We want to ~~[illegible]~~ predict y by $y_i = \sum_{j=1}^{k} x_{ij}a_j$
We pick $\vec{a}=\vec{a}_0$ that gives min ms error. The observed error = $\sum_i (y_i-\vec{x}_i\cdot\vec{a}_0)^2 = \sum_0$

$[y_i, x_{i,j}]$
$i=1/N$
$j=1/k$.

4.30.99 SMPT STEIN

Hvr, the true Σ^2 ($\equiv \Sigma^2$ of Generator of data) is larger, since $\hat{a}_0$ is incorrect.

If $\vec{a}$ is the true value, then $(\hat{a}_0-\vec{a})^2$ tells (proportionally) how much must be added to Σ_0^2 to get Σ^2: expected error in $\hat{a}_0$ $(\approx|\hat{a}_0-\vec{a}|^2)$ can be found by looking at the second derivative of the total squared error in predn.

.05 One impt idea is that STEIN can be considered to be a special case of "Wtd mean of various coding Methods". One method I did try was perhaps ST42.bas! What I'm thinking of:

Value of Bin. P.f. 38.28

Method 1: Descr. the sequence of N·K 0's & 1's as a simple Bern seq.

.11 Its p.c. is $\frac{n!\,m!}{(n+m+1)!} = e^{G_1}$ n = total no. of outs, m = " " " hits. $n+m = KN$; $m = \sum_{i=1}^{K} M_i\cdot N$

Another way to descr. data set (Mixed composition) is

.15 $\prod_{i=1}^{K}\frac{n_i!\,m_i!}{(n_i+m_i+1)!} = e^{G_2}$ $m_i = M_i\cdot N$ = no. of hits by player$_i$, n_i = no. of outs by player$_i$, $n_i = N - m_i$.

To do predns, we use wtd. mean of the 2 descr. methods. (wts $= e^{G_{1,2}}$)

43.09–.30 are some values of G_1-G_2 (ln. of ratio of wts) for various values of N.

~~The~~ $G_1=G_2$ at $N\approx 200$, but I'm not sure of how much deviation from 200 N would need to get $G_1-G_2 = 1$ or 2 or 3: 43.28 says N±10 gives factor of 5 betw wts. i.e. $\frac{1}{e^5}$ v.s. e^5.

$N \to N+10$ gives $G \to G + e^{1.67}$; so $\Delta N = 1 \to \times 1.167$.

So this method would seem to use one method or the other; except for a narrow range of N values about $N\approx 100$.

.22
$$\ln\left(\frac{n!\,m!}{(n+m)!}\right) = n\ln n + m\ln m - (n+m)\ln(n+m) + \frac{1}{2}\ln\left(\frac{2\pi\, n\cdot m}{m+n}\right)$$
$$n\ln\frac{n}{n+m} + m\ln\frac{m}{n+m} \qquad + \frac{1}{2}\ln\left(2\pi\cdot\frac{m}{m+n}\,\frac{m}{m+n}\cdot m+n\right)$$
$$NM\ln M + (1-M)N\ln(1-M) \qquad + \frac{1}{2}\ln(2\pi M\cdot 1-M\cdot N)$$
$$= N(M\ln M + (1-M)\ln 1-M)$$

So it looks like the formulas of ST42.bas are correct.

.30

.32 At any Rate, this result seems to be much different from the results obtained using

.33 the 2nd "cascaded" difs w. τ^2 & $[G_i^2]$ ~~etc~~ (using either Binomial &/o Gaussn difs).

.34 WHY IS THIS? Well, .33 assumes some coherence in time seqs of the players — not nearly $\tau^2=0$ but not $\tau^2=\infty$ either.

.32 assumes there are only 2 ways of Generation: $\tau^2=0$ or $\tau^2\to\infty$.

.37 $\tau^2=0$ corresponds to (.11); $\tau^2=\infty$ ($\approx$ corresponds) to (.15)

.38 .33 assumes a continuum of possl. τ^2 values. → see 145.33 Probjection

NB hvr. .32 doesn't actually assume $\tau^2=\infty$, but it does give (perhaps)

.40 equal likelihood to all M_i values: More exactly, If we are considering

5.4.99 SMRT STEIN

Bin dists: .32 could assume a uniform (betw 0 & 1) apriori of μ_i.

.02 This is equiv to $\mu^{\phi}(1-\mu)^{\phi}=1$; i.e. $R=0$, in eq. 46.28, $x=\bar{\mu}\cdot\phi=0$; $y=(1-\bar{\mu})\cdot\theta=0$

Guess it just rite!

Hvr, ~~the eqs~~ for eq. 140.11 I have no immediate way of getting it out of 46.28.

As $R\to\infty$; $x=\mu R$, $y=(1-\mu)R$ both $\to\infty$ & the individual μ_i (A_i ...) become unimpt. — but not ~~this~~ converge strongly to a $PC=0$ as $R\to\infty$.

.08 The only way to get it from 46.28 would be Let ~~R N~~ $N\to\phi$, but Let R be the original N. so we get $\frac{x!\,y!}{(x+y+1)!} = \frac{N\mu!\ N(1-\mu)!}{N+1\,!}$

.10 Then $N\to\phi$ of .08 is of unclear interpretation.

Anyway, The discussn of 140.05 — 141.10 does explain away an impt. difty in t. ~~whole~~ whole analysis: i.e. why ~~are~~ were 140.32 & 140.33 so diffnt? 140.34 ff is a reasonable explain. These 2 methods ~~were~~ really assume difnt things about t. corpus! The model of 140.33 seems to incorporate more reasonable assumptions about t. data ~~set~~ source. — i.e. that the μ_i are bunched together "some what": 140.32 has only 2 models: No bunching (uniform apriori of all μ_i betw 0 & 1) & all μ_i are t. same $\bar{\mu}$, w. $\bar{\mu}$ having uniform apriori betw. ϕ & 1.

140.33 is a more general model, & ~~interpolates~~ considers a large no. of possl. models that 141.32 does not consider. 140.33 does, hvr, consider one of t. models of 140.32 (i.e. $R=0$) .02 — I'm not sure about t. other model, .09 – .11 suggests that the ~~Among~~ $N=R$ model ~~can~~ isn't accessable to 140.33. I'm not so sure! If we include t. proper normzn. constant in 141.33, we may be able to get t. desired D.F. Its probly Nec'y to put in all of t. normzn constants. — & ~~if~~ we may get that 140.32 is 140.33 w. $R=0$ & $R=\infty$. (i.e. 140.32 considers only 2 of t. models of 140.33.).

As such, it is certainly a quick, rough & dirty way to solve t. Stein problem: It usually gives wts ≈ 0, or 1 to t. 2 models. i.e. uses ~~the~~ either t. ensemble average or t. individual averages. A narrow range of N values where t. wts are not very close to ϕ & 1.

→ So essentially, it tells us whether to use μ_i or $\bar{\mu}$ as an estimate of μ_i

.36 More generally, for each item we can assign wts to ~~the~~ μ_i & $\bar{\mu}$. Say t. wt. on $\bar{\mu}$ is a; on μ_i is $(1-a)$: a is uniform apriori from 0 to 1. We consider all these models as parallel codes for t. corpus.

There is some difty in doing these "a" codes: A first attempt!

5.5.99 SMRT STEIN. 142

.01 Code the ~~[illegible]~~ jth bits of all k players at the same time: The pcost of the jth bit of the lth player is obtained by looking at all data before the jth bit, using wt. (1−a) for ~~[illegible]~~ l's previous bits,

.03 & wts. a for previous bits of all other players. (we can have a 0,1 precorpus for each player)

An easy (tho ~~[illegible]~~ inexact) way is to compare $\prod_{i=1}^{k}\left(\mu_i'^{\,\mu_i'}\ 1-\mu_i'^{\,1-\mu_i'}\right)^N$ w. $\left(\bar{\mu}^{\bar{U}}\ 1-\bar{U}^{\,1-\bar{U}}\right)^{kN}$.

.07 value = $\left(\prod_{i=1}^{k}\frac{\mu_i'^{\,\mu_i'}}{\bar{U}^{\bar{U}}}\cdot\frac{1-\mu_i'^{\,1-\mu_i'}}{1-\bar{U}^{\,1-\bar{U}}}\right)^N$. ~~[illegible]~~ Whether this ratio is

> or < 1 is indep of N. — which is quite different from 140.32 !

The main difference is the $\frac{1}{2}\ln(\quad)$ terms in 140.28–.30.

.07 only considers the N (u ln u + (1−u) ln(1−u)) part of 140.30.

.18 T. difference is $\sum_{i=1}^{k}\frac{1}{2}\ln\left(2\pi\,\mu_i'(1-\mu_i')\cdot N\right) - \frac{1}{2}\ln\left(2\pi\,\bar{U}(1-\bar{U})\cdot N\cdot k\right)$

.19 $= \frac{1}{2}\sum_{i=1}^{k}\ln\left(\frac{U_i'}{\bar{U}}\cdot\frac{1-U_i'}{1-\bar{U}}\right)$ ~~$\to \prod_{i=1}^{k}\sqrt{\frac{U_i'}{\bar{U}}\cdot\frac{1-U_i'}{1-\bar{U}}}$~~

Which ~~simply~~ changes .07 by N → N+½ (or N−½ ?)

At any rate, ~~to see~~ whether the result is > 1 or < 1 is indip of N !

So there is an error somewhere !

Well .18 is o.k., but .19 doesn't follow from it !

subtract & add
~~Add & subtract~~ $\left(\frac{1}{2}\ln 2\pi\,\bar{U}(1-\bar{U})\cdot N\right)\cdot k$ from .19.

→ .01 −.03 sounds reasonable! it is a truly sequential code (tho in fact the jth bits of all l players are not usually simultaneous).

Anyway, I have to get STEIN to Bed !

Actually, I should be able to get ~~[illegible]~~ proper (relative) pc's for codes w. various "a" values.

.32; (5.10.99) IDEA ! One of Diffys w. the Bin d.f. approach (& Gaussn) was Apparent ~~dep~~ critical dependence of soln. on apriori of "R". Hvr, instead of R, use τ², w. value betw 0 & 1 (?) & perhaps uniform d.f. : Since τ² is for a Gaussn d.f. & not exactly appropriate, its ok. for small τ² (i.e. large R), but has to be fixed up for ~~some~~ large τ² — i.e. small R. My impression was that the ~~a~~ apriori of R for small R was not so impt., so maybe letting apriori for R to be $\frac{1}{R^2}$ (since $R \approx \frac{1}{\tau^2}$ — assuming uniform d.f. for τ² at τ² = 0) for large R;

τ ≈ $\left(U_i^{\bar{U}}(1-U_i)^{1-\bar{U}}\right)^R$ for small R. I think we get uniform d.f. for U_i, when R = 0. $\int^R \frac{dR}{R^2}$ diverges at R = 0 — So we are in trouble again ! We could just cutoff at R < 1.0 → 144.01 SPEC

eliminating 0 $\frac{1}{R^2}$ < 1 will not affect things much, since N >> 1 (∴ N ≫ R in this region)

so perhaps ~~a~~ apriori of R ∝ $\frac{1}{R^2}$ would be O.K. — tho it doesn't give good results ! $\frac{1}{R(\ln R)^2}$ was ~~much~~ much better

5.999 SMRT STEIN Rev. u

Review of Reviews! 115.01 has a list of reviews! This will tell what each is about!

45 : Outline of what a Review should contain (this was before eq. 46.28 which was main line after that)

52.01 My early work on Stein: PP 1, 2, 3, 9, 10: I was just beginning to get some ideas as to what it [illegible].

79 : Index to Pgms: ST50 thru ST77B: Gives brief descr. of each pgm.

80 : Listing of ST 1 thru ~~[illegible]~~ ST50: Dates they were ~~[illegible]~~ written

For ST 5 7, 9, 10; gives pgs that ~~[illegible]~~ described or referenced them

90 ST 81 thru ST90 ST90 Brief descrs of each pgm.

115 Brief reviews of Work from ~~[illegible]~~ 46.01 thru ~~[illegible]~~ ~~[illegible]~~ 120

Also 115 Gives outline of a Review, some impt things it should contain.

124 Outline of Review.

126 -128.07 : Actual writing of review.

138.04 : another Outline of Review! (I seem to have forgotten that I had written 126-128.07 !)

~~[illegible]~~ Pgms
* 90 > ST90
93
96
100
101
103

76.25
Var. of
Bud.f. —
But not exactly

5.10.99 SMFT STEIN

.01: (Spac 142.40) Actually t^2 can't be as large as 1, so $\frac{1}{R}$ ~~[illegible]~~ $= t^2$ can't be > 1.

So R can't be < 1. — So ~~2~~ $1 < R < \infty$ is ok, w. apriori dist $\frac{1}{R^2}$ say $\frac{1}{3R^2}$ ∑ normal. const $= \frac{1}{3}$.

I'm not sure that I did it rite when I used ~~[illegible]~~ apriori dist $\frac{1}{R^2}$:

I didn't properly integrate from RR to ∞. — so I'd have to do this again.

— Also the results I got were w.o. integrating over $\bar{u}$, as I should — it was equivalent to assuming the dist. w.r.t. $\bar{u}$ was indep of R & of $\{u_i\}$!

Actually since $t^2 = \bar{u}(1-\bar{u})$,  $\frac{1}{4}$

The max value t^2 can have is $\frac{1}{4}$. — Well, this assumes size $= 1$.

.15 for the general case ~~[illegible]~~ $\left(u^{\bar{u}}(1-u)^{1-\bar{u}}\right)^R$, there is an easy way to get σ^2 as a funct of R, $\bar{u}$, using the first & second moments of

$$m_0 = \frac{x!\,y!}{x+y+1!} \; ; \; m_1 = \frac{x+1!\,y!}{x+y+2!} \; ; \; m_2 = \frac{x+2!\,y!}{x+y+3!}$$

$$\sigma^2 = \frac{m_2}{m_0} - \left(\frac{m_1}{m_0}\right)^2 = \cancel{\frac{x+1}{x+y+2}}$$

.20
$$\frac{m_1}{m_0} = \frac{x+1}{x+y+2} \; ; \; \frac{m_2}{m_0} = \frac{(x+1)(x+2)}{(x+y+2)(x+y+3)} = \frac{m_1}{m_0}\cdot\frac{x+2}{x+y+3}$$

$$\sigma^2 = \frac{x+1}{x+y+2}\left(\frac{x+2}{x+y+3} - \frac{x+1}{x+y+2}\right) = \frac{x+1}{R+2}\left(\frac{(x+2)(R+2)-(x+1)(R+3)}{(R+2)(R+3)}\right)$$

$$\overset{=y}{R - x + 1}$$

$$\frac{x+1}{R+2}\left(\frac{\cancel{xR}+\cancel{3R}+\cancel{2x}+4-\cancel{xR}-R-\cancel{3x}-\cancel{3}}{R+2\cdot R+3}\right) = \frac{(x+1)(y+1)}{(R+2)(R+2)(R+3)} \approx \frac{u(1-u)}{R}.$$

Actually $\frac{x+1}{R+2}$ is Lap's rule for u; $\frac{y+1}{R+2}$ is Lap's rule for ~~[illegible]~~ $1-u$.

∴ I don't know how to interpret "$\frac{1}{(R+2)+1}$".

.29 So $\boxed{\sigma^2 = \frac{x+1}{R+2}\cdot\frac{y+1}{R+2}\cdot\frac{1}{R+3}}$ $x > -1$, $y > -1$ See 145.20

If a d.f. has to be between 0 & 1, perhaps the most variance can have is

.32 w. ½ funct's at 0 & ½ at 1; so $\sigma^2 =$  $\frac{1}{4}\cdot\frac{1}{2} + \frac{1}{4}\cdot\frac{1}{2} = \frac{1}{4}$.

So max value of σ^2 is $\frac{1}{4}$.

We get this ⇑ w. $R < 0$: specifically $\bar{u} = .5$, $R = -\infty$.

Let $x' = x+1$; $y' = y+1$; $R' = x'+y'$: So $\sigma^2 = \frac{x'}{R'}\cdot\frac{y'}{R'}\cdot\frac{1}{R'+1}$

$$\sigma^2 = \frac{\bar{u}R+1}{R+2}\cdot\frac{(1-\bar{u})R+1}{R+2}\cdot\frac{1}{R+2}$$

Say $\bar{u}$ is fixed .... at .5 how is σ^2 a funct of R

(Woops! If R is < 0, $\sigma^2 < 0$! See 145.20 for explain! R can't be < -2 : $u^x(1-u)^y$ would diverge at $u=0$ a/o at $u=1$

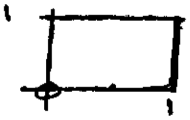
$\frac{.5R+1}{R+1}\cdot\frac{.5R+1}{R+1} \approx .5\times.5$;

The derivation of σ^2 seems to work for negative R. — Maybe not; the factorial function is weird for negative integer values. — but say $R = -5\frac{1}{2}$.

5.10.99 SMFT STEIN.

for $a = .5$ we expect σ^2 to $\to \frac{1}{4}$ as $R \to -\infty$ as (144.32 R):

for $R = 0$, the formula gives $\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3}$. seems reasonable!



& $m_0 = 1$
$m_1 = \frac{1}{2}$
$m_2 = \frac{1}{3}$

$\frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$ $\frac{3}{2}$ which is .

So what happens at $R = -2$, or $R = -3$. — we get $\sigma^2 = \infty$!

U^0 $\boxed{\text{say } u = \frac{1}{2} \quad R = -2,\ X = y = -1}$

$U^{-1}(1-U)^{-1}$; $M_0 = \frac{-1!\ -1!}{0!}$; $-1! = \infty$

$R = -2 + \epsilon$
$X = y = -1 + \frac{\epsilon}{2}$

So how do we get $\sigma^2 > \frac{1}{4}$? from 144.20 $\frac{m_1}{m_0} = \frac{x+1}{x+y+2} = \frac{\phi}{\phi}$

but say R is just close to -2. $R = -2 + \epsilon$ $\frac{m_1}{m_0} = \frac{\frac{\epsilon}{2}}{-1-1+2+\frac{\epsilon}{2}+\frac{\epsilon}{2}} = \frac{1}{2}$

$X = \frac{R}{2} = -1 + \frac{\epsilon}{2}$

So $\frac{M_1}{M_0} = \frac{1}{2}$ $\frac{m_2}{m_0} = \frac{m_1}{m_0} \cdot \frac{X+2}{R+3} = \frac{1}{2} \cdot \frac{1+\frac{\epsilon}{2}}{1+\epsilon} \to \frac{1}{2}$; $\frac{M_2}{M_0} - \left(\frac{m_1}{m_0}\right)^2 = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$.

which is unreasonable, since $\frac{1}{4}$ is the largest possible σ^2

-20 . Maybe not so unreasonable! $U^{-1}(1-U)^{-1}$ is not a proper d.f. (d.f.??) since $\int_0^{.5} = \infty$.

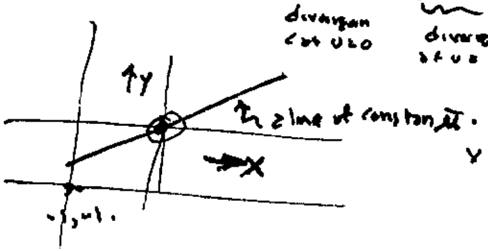
and the norm. constant could effectively put all wts. @ 0 & 1.

for ~~[illegible]~~ $R << 2$, the ~~[illegible]~~ divergence is even worse

i.e. U^{-10} diverges very badly at ~~[illegible]~~ $U \to 0$.

In general $\underset{\text{divergence at } U=0}{X > -1}$, $\underset{\text{divergence at } U=1.}{Y > -1}$. so $x+y = R > -2$: The only constraints are $X > -1$, $Y > -1$

any pt. in the x,y plane satisfying these constraints is ok.



Y can be -9, y can be +10.

33: 140.38 5.12.99 SN On second thought, it is not clear why considering all $M_i^* = \bar{U} + (1-\alpha)M_i$ models is much different from assuming wtd mean of the 2 models, $M = \bar{U}$ & [redacted] M_i & M_i.

7.14.99 STEIN:

.01 How do we get expected/observed varc, Σ^2 in terms of τ^2 & $[\sigma_i^2]$?

Not so complicated! Say we had this d.f. w. varc τ^2 (d.f. need not be Gaussian or [illegible])

we select a (Mt. Carlo) pt. from t. τ^2 d.f., then we "corrupt it" by σ_i^2.

(The σ_i^2 d.f. can be any d.f. w. finite varc). T. resultant d.f. has varc $\tau^2 + \sigma_i^2$.

We start over w. t. τ^2 & we do a new σ_j^2 & get a " " " " $\tau^2 + \sigma_j^2$

we do this for many σ_i^2's & we get many "sum" d.f.'s:

If we average over all these "sum" d.f.'s we get varc $=$ $\tau^2 + \frac{1}{k} \Sigma \sigma_i^2$.

.10 Note: σ_i^2 that we use will be $\frac{N_i \sigma_i^2 \text{ observed}}{N_i - 1}$. ($N_i$ is size for each σ_i^2).

Applicns of ST to OxG. & voice recogn.

1) 139.01 ff applicn to FT acc. Also see 132.29 ff

2) elements of corr. matrix (off diag) [illegible] corrected (134.13–.32) is unclear as to how to do it, & just what's going on!

3) Get more detail from Alex on other his problems [illegible].

4) T. idea of using last yrs' data to decide if certain params are "related" in a useful way

---

Σ sq. error w.r.t. "rest of yrs's data".

		ms error	rms.
1) μ_i'	:	.07534	.06468
		) 3.1055	
2) $\bar{\mu}$	:	.024259	.0367
		) 1.122	
3) Stein	:	.021611	.0346
		3.486	

Get $\frac{1}{k}\Sigma(\mu_i' - \bar{\mu})^2$

$\frac{1}{k}\Sigma \frac{\mu_i'(1-\mu_i')}{N}$

I don't at default.

Do: Give details on just how well "Baseball" analysis did:

ms err for

1) $(\overline{U_i'} - U_i)^2$

2) $(\overline{U_\phi} - U_i')^2$

3) $(\bar{U} - U_{\text{stein}})^2$

Give P.O. w. source eqs. & Refs.

Define

μ_i, $\bar{\mu}$, τ^2

μ_i', σ_i^2, Σ^2.

See ST 109

using $\Sigma \mu_i'$ & $\Sigma \mu_i'^2$

we may have

for Σ^2 & $\Sigma \frac{\mu_i'(1-\mu_i')}{N}$

$$\frac{1}{k}\underset{\alpha}{\Sigma}\mu_i'^2 - \underset{\beta^2}{\left(\frac{\Sigma\mu_i}{k}\right)^2}$$

$= \Sigma^2$

$$\frac{1}{k}\Sigma\sigma_i^2 = \frac{1}{k}\Sigma\frac{\mu_i'(1-\mu_i')}{N}$$

$$= \frac{1}{N}\frac{\Sigma\mu_i}{k} - \frac{1}{N}\frac{\Sigma\mu_i'^2}{k}$$

$$= \frac{1}{N}\beta - \frac{1}{N}\alpha$$

v.s. $\alpha - \beta^2$

7/17/99 STEIN 146½

42.01! ST42

Some time ago, I ~~compar~~ did a simplified ALP model w. only 2 dim! (a) using $\bar{\mu}$ as estimate of μ_i (b) using μ_i^c as estimate of μ_i. From the relative pc's, I got relative wts. The value of R ~~or~~ ½ was plotted as a funct of N the no. of times "At Bat". The wts changed from 0 to 1 over a narrow range of N values, so I felt it was not such a good idea; Now, hvr. It may be that the RMS Difference betw it & conventional Stein are quite small. Look at this in more detail!

4305-.30 The transition betw. the 2 methods was very sharp in N

for N = 211.4 wts were about = but changed by factor of ($e^{1.68}$ = ×5.4 or $\frac{1}{5.4}$) for N = 211.4±10.

See if regular Stein does this also. — ~~Not easy to do, since "regular Stein" only has N = 45 only~~

ST103:

N	w
150	.2373
210	.167
2.70	.1318
100	.356
45	.791
20	1.779 !
30	1.18
35	1.017

strangely weird! as N goes thru 35 W changes sign, but general character of the prediction is about the same. at N = 35, W ≅ 1 & all predns of μ_i are $\approx \bar{\mu}$ (as expected)

Hvr, that W = 0 for N as large as 35 Seems inappropriate.

$(1-w)$ is <0!

Using "k−1" instead of "k−3" in the calcns.

K−3 → k−1

N	w
35	1.15
→ 40	1.008
45	.85

So the W≅0 pt. goes from N = 35 to N = 40!

So ~~the~~ k−1 is perhaps worse: we would like W → 1 as N → 0.

see 7/15/99 notes for talk at OxB..

So: use ST42, & ST103 to get the 2 estimates & get their rms difference! compare w. rms errors of the pgms — say simple rms of predictg "next 4 yr" estimates)

Easiest to do: Get rms. error of Both methods as funct of N.

I am a bit (surprised) disturbed ~~that~~ at ST42: That the transition was so sharp! —

Suggesting error in pgm.

7799 GXB

1) Stern Effect: Relative to Evaluating "Bunch of" strats.

2) try: 1984 → 1994 to get less curve fitted.

3) Try other cmb (time of day, cross commodities, other Mkts)

Protein from "Self" is heavily Mined.

Bugs in "Trading Pgm": $5 – 5%/yr. 20 trades/day #100/d → 25%/yr.

Avg 15/trade is due to big losses. (~~may be~~ small size statistically untractable!) say $\sigma^2 = 80$!

1) porting present strats to new Mkts, Mkts

~~2)~~ Tweaking current Strats w. originally desired (More complexity)

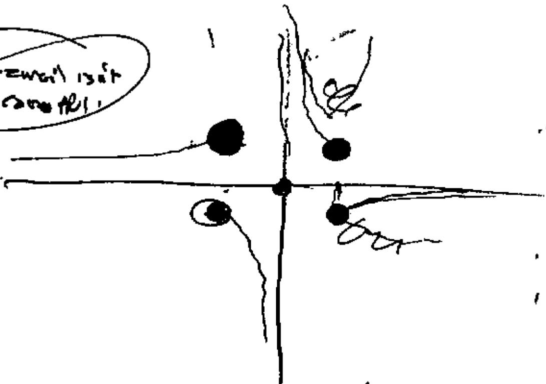
3) Really new strats:

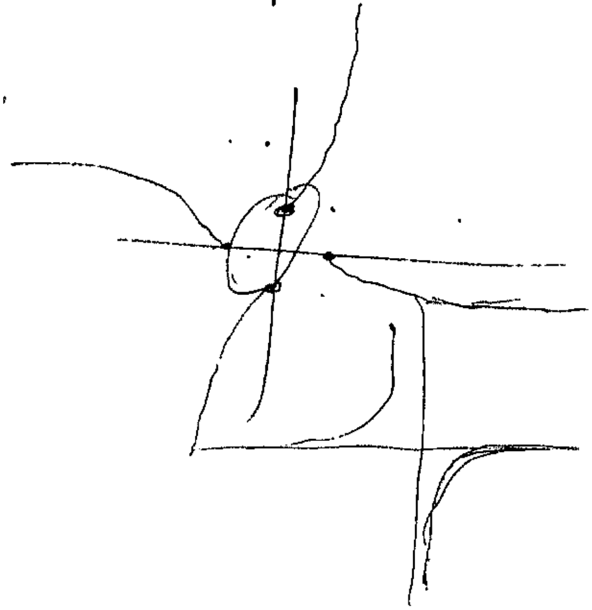
Kurzweil says 5% of Mkt is Genetic Algos. ~~Kurzweil isn't very smart!~~

④ Margin put up can be inept.

$f(f(f(x)))f(-x) =$
$= f(x)$

$f(-x) = f(x)$
$f(x) = -f(-x)$
$f(f(f(x))) = f(-x)$
$f(x) = f^{-1}(-x)$
$f(-x) = f^{2}(x)$





7/15/99 STEIN Paradox:

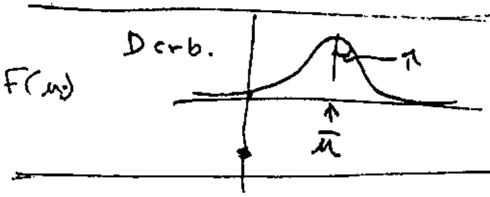
Descrb. Baseball situation : k players = 18; n times at bat = 45.

"Hit or not hit".

Describe N = 500 cases

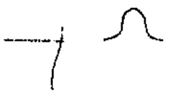
n = 1 case.

Descrb.

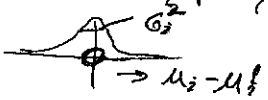


Start with all $\sigma_i^2 = 1$ later approx how to deal w. $\sigma_i^2 \neq 1$.

Explain about final D.f. $f_i(\mu_f)$



$f_i(\mu_i) \cdot F(\mu_i)$. Prob'ty of observation = prob'ty of μ_i being chosen times prob'ty of observed μ_i^f given μ_i which is given by



(BAYES Thm.

Using different wts. than $e^{-\frac{t}{\pi}}$ for smoothing

Say Σ of several exponential(s) may be useful.

Mixing of .. Baseball, basketball.

Gaussian:

Basket Ball: Mix r.z:

Bimodal d.f.?

look at previous yr(s) for effectiveness. — e.g. 2 yrs means closer?

7/15/99

STEIN PARADOX

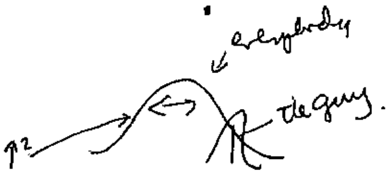
μ_i – batting average – true mean for i^{th} player

μ_i' apparent mean for i^{th} player based batting average. $M_i' - M_i$ deviation from observed batting average

σ_i^2 Variance of data for i^{th} player.

$\bar{\mu}$ average of all μ_i' : $\bar{\mu} = \frac{1}{k} \sum \mu_i'$

τ^2 variance of distribution generating $[\mu_i]$



Σ^2 Variance of observed distribution of $[\mu_i']$:

$$\Sigma^2 = \frac{1}{k-1} \sum (\mu_i' - \bar{\mu})^2$$?

$\overline{\sigma^2}$ mean of σ_i^2 : $\overline{\sigma^2} = \frac{1}{k} \sum \sigma_i^2$

k number of players : ($k=18$ in this Example)

n number of times at bat (same for each player : $n=45$ in this example)

$$\sigma_i^2 = \frac{\mu_i'(1-\mu_i')}{n}$$

$$\tau^2 \approx \Sigma^2 - \overline{\sigma^2}$$

Stein's [redacted] estimate of μ_i : $\left(\frac{\bar{\mu}}{\tau^2} + \frac{\mu_i'}{\sigma_i^2}\right) // \left(\frac{1}{\tau^2} + \frac{1}{\sigma_i^2}\right)$

Empirical rms errors : (error with respect to data on about 400 more "at bat's" for rest of year (1970))

μ_i' as estimate of μ_i : .06468

$\bar{\mu}$ " " " " : .0367

Stein " " " " : .0346

STEIN's Paradox : Efron, Morris, Sci Amer, May 1977 p 119-127
More refs in ≈ last page of issue

Data Analysis Using Stein's Estimator and Its Generalizations : Efron, Morris
Journ. Am. Stat. Assoc Jun 1995 p311

Empirical Bayes Methods Applied to Estimating Fire Alarm Probabilities
Carter, Rolph ; Journ Amer. Stat. Assoc, Dec 1974

7/15/99

STEIN. PARADOX

original.

μ_i true mean for i^{th} player

μ_i' apparent mean for i^{th} player

σ_i^2 Variance of data for i^{th} player.

$\bar{\mu}$ average of all μ_i' : $\bar{\mu} = \frac{1}{k} \sum \mu_i'$

τ^2 variance of distribution generating $[\mu_i]$

Σ^2 Variance of observed distribution of $[\mu_i']$:

$$\Sigma^2 = \frac{1}{k-1} \sum (\mu_i' - \bar{\mu})^2$$

$\overline{\sigma_i^2}$ mean of σ_i^2 : $\overline{\sigma_i^2} = \frac{1}{k} \sum \sigma_i^2$

k number of players : ($k = 18$ in this Example)

n number of times at bat (same for each player : $n = 45$ in this example)

$$\sigma_i^2 = \frac{\mu_i'(1-\mu_i')}{n}$$

$$\tau^2 \approx \Sigma^2 - \overline{\sigma_i^2}$$

Stein's [redacted] estimate of μ_i : $\left(\frac{\bar{\mu}}{\tau^2} + \frac{\mu_i'}{\sigma_i^2}\right) \Big/ \left(\frac{1}{\tau^2} + \frac{1}{\sigma_i^2}\right)$

Empirical rms errors: (error with respect to data on about 400 more "at bat's" for rest of year (1970))

μ_i' as estimate of μ_i : .06468

$\bar{\mu}$ " " " " : .0367

Stein " " " " : .0346

STEIN's Paradox : Efron, Morris, Sci Amer, May 1977 p 119-127
More refs on ≈ last page of issue

Data Analysis Using Stein's Estimator and Its Generalizations: Efron, Morris
Journ. Am. Stat. Assoc Jun 1995 p311

Empirical Bayes Methods Applied to Estimating Fire Alarm Probabilities
Carte, Rolph; Journ Amer. Stat. Assoc, Dec 1974

This is fine, except for the fact that the observed variance τ' is subject to random variations and could potentially be much smaller than it is supposed to be. In that case, $\tau'^2 - 1$ could potentially be negative. Even if it is positive, if it were tiny, because of random variations, the estimate of τ would be very much in error.

A more careful way of estimating τ from τ' would be an improvement.

We could use approx d.f. for $\hat{\tau}^2_w$. cutoff for $\hat{\tau}^2 < 0$.

1.2 Fixing up to remove the unit variances

1.3 Bayes formulation

1.4 Tree version of the problem

This is a variation on the original setting which is relevant to clustering.

We have a large set of random variables, some of which are very similar to each other, and others are more different. We have observations of each of them, and we want to estimate the mean of each random variable. The problem is that we only have a few observations of each RV, not enough to do a good job of estimating its mean. So we would like to also use observations from other RV. The observations from very similar RVs would have a large weight, and observations from less similar RVs would have a smaller weight. We want to use some Stein-like formulation to decide how the different observations should be weighted.

We will assume that the similarity of the RVs has been already determined somehow, and expressed in a tree. Two random variables that are very similar will have a near common ancestor, perhaps even the same parent or grandparent. Two RVs that are less similar will have a more distant common ancestor.

To go along with the tree of random variables (and their observations) we will imagine another parallel tree of random variables, related to the Stein estimation setting. At the root of the tree will be a random variable U_0, with mean u_0 and variance τ_0. It will have several children U_{0i} and their means u_{0i} will be samples taken from their parent. I don't know where their variance will come from. They will in turn have children U_{0ij}, whose means u_{0ij} and (variances τ_{0ij}?) will be samples from their parents U_{0i}. This will be repeated all the way down to the leaves, which will correspond to the actual random variables whose means we want to estimate.

Since the means (and variances?) of each node in the tree are themselves random variables, the RVs can be thought of as several different random variables, each will different means and variances, depending on now many times we compose it with its ancestors.

For example, a RV at the bottom of the tree can be thought of as having mean $u_{0ij\ldots yz}$ and variance $\tau^2_{0ij\ldots yz}$ or mean $u_{0ij\ldots y}$ and variance $\tau^2_{0ij\ldots yz} + \tau^2_{0ij\ldots y}$ or $\ldots$ mean u_0 and variance $\tau^2_0 + \tau^2_{0i} + \ldots + \tau^2_{0ij\ldots yz}$.

2 Conclusions

References

[1] A.U.Thor, Ph.D., "Big Hairy Equations", Faceless Corporate Press, 1988

7/14/99 Alex: Voice Recogn.

They use "Triphones" (a phoneme in "before" & "after" context).

every $\boxed{\text{10 ms}}$ they get a kepstrum: which is spectrum of log of spectrum? Perhaps they should take spectrum w. log in ω direction — take spectrum of that — anyway, the kepstrum is supposed to give a picture of the ω response of the filter that the "buzz" is subject to ... leaving out all Buzz information. Each kepstrum is represented by 45 numbers. So they get 1, 45 vector every 10 ms.

They have a HMM w. 5 states that models a triphone. The 5 states are constrained, so they have to go L → R, but may "skip" 1 or more states.

The p.d.'s in 45 space are modeled as "Gaussian Mixtures". The Gaussians are MV d.f.'s w. only diag. terms. They sum up to 256 of these Gaussians to get the "mixture" d.f.

Guess that the Triphones are each a sequence of 5 (or less) states & each state is a p.d. on this Mix of 256 Gaussians (≡ Multivariate d.f.'s).

Also they have a "language model" using "unigrams" (word freqs) "digrams" & "trigrams". SSZ ↓ rapidly from uni – di – tri: So how can we pool data & what sort of Stein Model can we use? One way to pool words is Parts of Speech. "Parts of speech" can be "defined" by: ① what comes after "the"

They have a corpus of ~150 hrs of speech (over telephones): at ≈ 1 sec/word this is ~ .5 M words — so SSZ for digrams & trigrams would be rather small.