Uses, say, 3 coiffs (quadratic regression, if a function of 1 variable), is a special way of inserting apri. info.

It seems ~~very~~ probable that if one stated exactly, what all of one's important apri assumptions were, that one could get optimum regression curves and their variances.

Stated exactly, it means that one has an apri density in function space, and a "probability of error" function [This means that the probability that if th. "true" function were $y=f(x)$, that th. pair x_0, y_0 could occure.], then one could easily find th. wite regression curve, using Bayes inverse probty.

.16 From (15).07 I guess: that expected fractional error in variance = expected fractional ··· — No, its a rather complex problem, but th. answ. can be found in any statistics book.

→ From Cramer: p 349 eqn 27.4.5:

Expected (observed value of variance (for normal distrib) of true variance σ^2, is $\frac{n-1}{n}\sigma^2$

Expected second moment of this variance $= \frac{2(n-1)}{n^2}\sigma^4$ $\left(\approx \frac{2}{n}\sigma^4\right)$

Th. s.d. of this variance $\approx \frac{\sqrt{2}\,\sigma^2}{\sqrt{n}}$ for large n.

On P. 382 (Cramer) $\frac{n s^2}{\sigma^2}$ has a χ^2 distrib of $n-1$ °'s of freedom. I guess s^2 is th. observed variance.

probably

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2$$

We can, then, look upon th. variance as a quantity of mean σ^2 and s.d. $\frac{\sqrt{2}\,\sigma^2}{\sqrt{n}}$.

~~I think it is th. variance of a quantity that determines~~

I think th. statistical wt. of an observation is directly prop. to ~~its~~ th. reciprocal of its variance.

→ It may be possl. to arrange th. meanings of one's coiffs, → as one's sample size ↑ one tends to use more coiffs — in a certain order.

(Notes)

Th. prob. of page 13) is this: From a "certain source" one gets U_{aprip} and $\frac{1}{\sigma^2}$ ($\equiv$ wt. to be assigned to U_{aprip}). One observes some statistical data on U, which we will call S — this data is to be gu. wt. n, since there were n observations. Th. best estimate for U is then

$$\frac{\frac{U_{aprip}}{\sigma^2} + nS}{\frac{1}{\sigma^2} + n}.$$

We want to know how $U_{ultimate}$ differs from U_{aprip}, on th. average for many estimates of U_{aprip} that come from this "certain source". When we find this $\overline{(U_{ult} - U_{aprip})^2}$ ← taken over th. output of th. "certain source", we will set this = to σ^2 for th. entire source.

Let U_{aprip_i}, n_i, S_i $(i = 1 / r)$ be th. U_{aprip}, n and S, assoc. with th. i^{th} output of th. "certain source".

then $$\sigma^2 \approx \frac{1}{r}\sum_{i=1}^{r}\left(\frac{\frac{U_{aprip_i}}{\sigma^2} + n_i S_i}{\frac{1}{\sigma^2} + n_i} - U_{aprip_i}\right)^2$$

($\approx$ because ~~σ^2~~ there may be some $n-1$'s or $\sqrt{n-1}$'s in this eqn.).

$$\sigma^2 \approx \frac{1}{r}\sum_{i=1}^{r}\frac{(S_i - U_{aprip_i})^2}{\left(\frac{1}{\sigma^2 n_i} + 1\right)^2}$$

Th. formula isn't exactly rite. There is, I think, some kind of proportionality constant betw. th. σ^2 of $\frac{1}{\sigma^2}$ weight, and th. σ^2 due to deviations betw. U and U_{aprip}.

Let us get more specific: Say $U \equiv$ th. mean no. of "cases" per unit time (actually, this may not be such a good def. of what I want, but let it go for awhile).

(Notes)

Say we have ~~had~~ had T q. elements up to time t.

If we have had c cases for a certain pngm$_i$ then

$s_i = \frac{c}{T}$ and the stand. dev. is $\sqrt{c}\,\frac{c}{T} = \cdot\frac{\sqrt{c}}{T}$

Th. variance is $\frac{c}{T^2}$

~~The weight gn. to this~~

Now if $\sigma^2 = \overline{\left(U_{aprip} - U_{actual}\right)^2}$

Then the estimate must be $\left[\dfrac{\frac{U_{aprip}}{\sigma^2} + \frac{T^2}{c}\cdot\frac{c}{T}}{\frac{1}{\sigma^2}+\frac{T^2}{c}}\right]$

$$= \frac{\frac{U_{aprip}}{\sigma^2}+T}{\frac{1}{\sigma^2}+\frac{T^2}{c}} = \boxed{\frac{U_{aprip}\frac{c}{T\sigma^2}+c}{\frac{c}{T\sigma^2}+T}}$$

Say $U_{aprip} \approx \frac{c}{T}$ th. no. of cases wt. given to U_{aprip} is $\approx \frac{U_{aprip}}{\sigma^2}$

Then $\left(\dfrac{\frac{U_{apri}}{\sigma^2}+T}{\frac{1}{\sigma^2}+\frac{T^2}{c}} - U_{apri}\right)^2 = \left(\dfrac{\frac{U_{aprip}}{\sigma^2}+T-\frac{U_{aprip}}{\sigma^2}-U_{aprip}\frac{T^2}{c}}{\frac{1}{\sigma^2}+\frac{T^2}{c}}\right)^2$

This is probably th. best estimate of U_{actual}.

$$= \left(\frac{\frac{T^2}{c}\left(\frac{c}{T}-U_{aprip}\right)}{\frac{1}{\sigma^2}+\frac{T^2}{c}}\right)^2$$

So

.30 $$\sigma^2 \approx \frac{1}{r}\sum_{i=1}^{r}\frac{\left(\frac{c_i}{T_i}-U_{aprip_i}\right)^2}{\left(1+\frac{c_i}{T_i^2\sigma^2}\right)^2}$$

This equ. is of th. form

$$x = \sum_{i=1}^{r}\frac{a_i}{\left(1+\frac{b_i}{x}\right)^2}$$

x, a_i, b_i all >0.

.38 Actually, th. equ. should have more in it. In determining σ^2, each pngm. cannot be gn. = wt. — because first, th. approxs. of $U_{ultimate_i}$ are not equally good, and secondly, all pngms are not of = importance.

(Notes)

This ~~seems~~ kind of wrong — i.e. I feel that σ^2 should be "open loop" and not "~~depend upon~~ be a function of itself. In a sense, it isn't much of a function of itself — assuming ~~xx~~ 18).30 is correct (~~xx~~ ignoring the comments of 18).30).

σ^2 is almost an average of $\left(\frac{c_i}{\tau_i} - U_{aprip}\right)^2$

when ~~xxx~~ $\frac{\tau_i^2}{c_i} >> \frac{1}{\sigma^2}$ [i.e. $\frac{c_i}{\tau_i^2\sigma^2} << 1$]

say $\frac{\tau_i}{c_i} \approx \frac{1}{f_i}$ [$f_i \to U_{ultimate_i}$ at $\tau_i \to \infty$]

then we want $\frac{\tau_i}{f} >> \frac{1}{\sigma^2}$ for σ^2 to be not hyly dependent on σ^2, and ∴ for equ 18).30 to be easilly solvable.

More exactly, we want ~~xx~~ $\frac{\tau_i}{f_i} >> \frac{1}{\sigma^2}$ for most pugms

---

Woops! $\tau_i = \tau$ is the same for all pugms.

---

or ~~xxx~~ $\tau\sigma^2 >> f_i$ for most pugms.

note $\tau\sigma^2$ is indip. of i.

Of course if $\tau\sigma^2$ is $>> f_i$ for most i, then

$$\sigma^2 \approx \frac{1}{r}\sum_{i=1}^{r}\left(f_i - U_{aprip_i}\right)^2$$ is good enuf!!!

What we really want, is to be able to get a reasonably good σ^2 as soon as possl. — i.e. with as small τ as possl. — which is the pt. of all this mathical refinement. The q. is — is formula ~~(18).38)~~ 18).30 ever any better than — ?

{ Well, our main goal ~~xxx~~ here, is to ~~xxx~~ define and use σ^2 in such a way as to minimize prediction errors.

This seems like a fairly straight forward problem in statistics — I suspect that 18).30 with, perhaps, corrections in line with (8).38, and also corrections to give various pngms different weights since they ~~occur~~ occur with different frequencies and are of different "importance"

Statistical problem: There exists a platonic set of numbers, f_i. $(i=1|r)$.

We have one source that give us U_{apri_i} which form a normal distrib. about f_i's — all with th. same unknown σ^2.

We have ~~one~~ another source that gives us S_i's that are also normal about f_i's but each has a known $\sigma^2 = \frac{1}{n_i}$

Given all the U_{apri_i}'s, $\frac{1}{n_i}$'s and S_i's: to find σ^2.

This seems O.K., except for the reservations and probably 18).30 is a reasonable soln., within th. reservations.

Thinkiup about it some more: what 18).30 does is assume $f_i = a\,U_{apri_i} + b\,S_i$ (a and b are proper wts). and then get $\overline{(U_{apri_i} - f_i)^2}$. This is very probably

.30 not correct in general. Th. correct way, is to try all f_i's and σ^2's and find th. $\sigma^2 \ni$ th. observed phenomena are most probable (integrating over all possl. f_i's with, perhaps, suitable wts. for various f_i regions). There is a more extensive discn. of this on 1α122 and α121

This may not be so hard: fix σ^2, then integrate over all $\vec{f}$ space to find out how probable th. observed info is. This integral will be a funct. of σ^2. Find $\sigma^2 \ni$ th. integral is max. The integration isn't dift. — Use r dim. vectors, and multidim. gaussian distribs. — However, each dim. appears to be indip. of all th. rest.

Notes

I don't think I should bother with Th. ~~McCarthy~~ integrations and extremizations (they mite be done simultaneously) except, perhaps, after lunch. They are of the type $\int_{-\infty}^{+\infty}\left(\frac{1}{\sigma}e^{-\frac{x^2}{\sigma^2}}\cdot\frac{1}{\sigma_1}e^{-\frac{(x-a)^2}{\sigma_1^2}}\right)dx$
It mite be more imp. to look into 18).38. Th. first q. of 18).38 is probably included in Th. soln. of Th. problem as stated in 20).30. Th. second q. in 18).38 isn't so clear. I must state exactly what I want σ to be.

But I think that I have gone far enuf into this problem for Th. present. Th. imp. pt., is that it is a definite math. problem, ~~which~~ and no additional assumptions need be made.

At Th. present time, Th. most imp. single problem seems.

What is U? Some discn. on pps 6), 7), 8), 9)

Also, try to define U for non-Math-T.M. — i.e. probablistic prediction. Try to fix it so it is easy to isolate out Th. U. of a partic pngm. — ~~um~~ since this ~~~~ solves problem ~~~~ of ~~whether~~ how to retain a word that is not Th. best, but still o.k.. ~~Make~~ Also B.G. becomes easy: i.e. by making each prediction depend on a single pngm.

A way: In Th. gen. T.M. (not Math-T.M.) case: Each pngm. has a prediction vector assoc. with it. To find Th. U of ~~~~ a ~~~~ pngm. ~~~~ take a McCarthy-like eval. over all cases in which it applied. This gives a certain amt. of money ~~~~ for Th. value of that pngm. Then chose a set of pngms. ∋ their total worth is maximized. In cases where ~~they~~ 2 pngms that were chosen, conflict, then take ~~av~~ wtd. average of their prediction vectors. Wts. to be decided on Th. basis of some sort of U criterion — tho one should try to chose a set of pngms so that they do not often cover Th. same case. If 2 pngms have Th. same prediction vector, mite as well $\beta\leq$ them.

How does one get prediction vectors? — simply Th. ~~~~ distribution over Th. history of that pngm.

So Th. big problem, then, is to decide which ~~~~ set of pngms to use, and how to weight them when 2 apply to Th. same ~~~~ ϕ-element.

(Notes)

This isn't far from th. ~~U~~ U problem in Math T.M. as discussed on pages 8) and 9)

What one will get, is several sets of pugms. Each set can be used fairly well for prediction. What happends when new pugms are ~~de~~ devised, however? — Don't they have a very poor chance of being useful, since most possibilities are already covered by th. old pugms — ~~tho~~ even tho some are not covered very well.

Will each set ~~pt~~ of pugms have its own set of U's for all of th. ugms, strs and ~~w~~ tps?

But th. imp. q. is : Just how are new pugms introduced?

.12 from .15 $\boxed{SN}$: In $U_{ij} \approx U_{si} + U_{nj}$ $(i,j = 1|n)$
(2).40

we wrote $U_{si} \approx \frac{1}{n}\sum_j U_{ij} - \frac{1}{2}\sum\sum U_{ij}$

$U_{nj} \approx \frac{1}{n}\sum_i U_{ij} - \frac{1}{2}\sum\sum U_{ij}$

It must be noted that th. criterion for fit is not for th. goodness of th. U_{si}'s and U_{nj}'s as individuals, but for how good $U_{si} + U_{nj}$ is — which is easier. It may be possl. to show that somthink like , perhaps itself, is, indeed optimum.

Also, ~~✗~~ How to form th. sets in a convenient way.

There is th. q. of whether to take all ~~data~~ pugms and select out th. "good" ones and use them only, or use all of th. pugms, but use wts.

In this new method, there is th. q. of how to group together pugms that don't have th. same prediction vector?

Th. q's are 1)

2) It is felt better, epistemologically, to use as few words as possl., to cover as wide a range of phenomena as possl.

3) This conflicts with th. idea that one can usually get very good prediction by th. use of very small sets of phenomena, so that th. members of each set are very much ~~a~~ alike, and so that

Mon July 23, 1956

one can get very fine differences ~~b~~ between cases by using many very small sets.

Discu: It may be that what/ one wants to be few in number, are th. objects from which one forms th. ~~s~~ sets. ("sets" ≡ pngms) ("objects" ≡ ngms, strs ntps). Examples: Red bike, old orange.

.07 (from Methodology note: ~~eat~~ work untill 11 P.M.: eat 11 P.M. - Mid.
sleep Mid - 8 AM: work 8 AM - Noon; Eat Noon - 1 PM:
~~mm~~ rec. 1 PM. - 5 PM: Work 4 P.M. - 11 P.M.

4) On th. other hand, if many interesting statements ~~statements, there~~ have ~~swamped~~ few words in them, then ~~th words are~~ there must be many words and ∴ small sets corresponding to them (?).

I think this / is forgoing all being very el.. Th. final def. of U, is how useful th. object is in prediction. I am afraid, however, that this ends us up in th. "swamping out" problem. Th. way to avoid this problem, may be to put one's words into gps., so that each gp. has a U. - But then there th. prob. of how to introduce new words into th. gps.

A reasonably good method of prediction for th. **Non-Math T.M.** case that suffers from "swamp out".

This method turns out to look pretty bad. Most objections seem overcomable.

Each pngm is gn. an apri prediction vector, and an apri. U. (U is a funct. of prediction vector, and freq of occurance of th. pngm.). For a gn. prediction, th. "best" pngm. that applies, is used.

To get away from "Swamp Out": every pngm (good or bad) will have assoc. with it, a param. that tells how useful it would have been, if it were used every time it was applicable (whether it was "best" pngm or not). This param. ~~may be ...~~ will be used to determine th. aprip. that this pngm will be useful in

~~Also~~ forming new pngms — via, ptps or ngms.

~~Also perhaps~~ This U' will be

$f \sum p_i \ln p_i$, where $\vec{p}$ is th. prediction vector

and f is th. freq. of ~~occurrence of~~ applicability of this pngm.

Also, perhaps, when all pngms are formed, their apriv U, will be th. apriv of this U'. In this way, one will, as before, try to get pngms with as hy U''s as possl.

But In each case of choice as to which pngm to use, one will chose th. one with greatest $\sum p_i \ln p_i$, not th. one with greatest $f \sum p_i \ln p_i$.

There is a serious Q. here: Do we want to get pngms with hy $\sum p_i \ln p_i$ or with hy $f \sum p_i \ln p_i$?

If $\sum p_i \ln p_i$ is too low, th. pngm will not be used, so that its ΔU will $= 0$. Also if f is too low, even if $\sum p_i \ln p_i$ is hy, th. pngm will not be often used, so $\Delta U \approx 0$.

Both. f and $\sum p_i \ln p_i$ are imp., but it is not clear that their product is th. best representative of probable ΔU.

It may be possl. to show that it should be $\propto f$.

$\therefore$ th. U' would $= f \cdot g(\sum p_i \ln p_i)$, where g is some

.30 function to be determined — it would depend on th. distribution function of $\sum p_i \ln p_i$ for all pngms.

Another ~~Obj~~ Trouble with above: One cannot group together pngms that donot have th. same prediction vector — certainly this is bad. An ad-hock way to avoid this that doesn't look too bad! Each pngm prediction vector has only 2 components — th. component of interest (which we want close to 1), and 1 minus that component — this prediction vector, has, then, only 1 essential component. We can then group any pngms together that we like — say $1\boxed{0}$ and $\overset{\alpha}{\boxed{1}}$. In th. first case, th.

meaning of th. prediction vector is one thing, in th. second case, another. The $g(\sum p_i(\ln p_i))$ ~~[illegible]~~ of 24).30 becomes simply ~~[illegible]~~ $g(p)$.

This all seems to apply very easy to Math T.M., as a special case. If we use $f\,g(p)$ as a measure of utility, then in math T.M., $g(p)$ = constant, so utility is $\propto f$, which is th. criterion we were thinking of using before. Since all ~~[illegible]~~ "consistant" pgms have th. same $g(p)$, then ~~both~~ all pgms of $g(p) <$ max., are discarded. Th. only choice ~~[illegible]~~ of are = value for making a decision. lies in th. probability that they are, indeed, both of $\neq g(p)$ — i.e. one or both may not be "consistant" — This is a q. of sample size — ~~[illegible] but, as was discussed bef~~ or, more exactly, th. "expected error" in $g(p)$ — which would be gn. by th. expected mortality ~~of~~ a pgm for a gn. sample size.

As was discussed before, however, th. likelyhood ~~of~~ such a conflict is small, and ∴ unimp. in contributing to th. mean U of T.M.

However, th. q's ~~[illegible]~~ raised here in T.M. on th. variance of th. $g(p)$'s of th. 2 pgms. may be ~~[illegible]~~ important when applied to Non-Math T.M.

Def. of Non-Math T.M.: In Math T.M. there is always ~~[illegible]~~ a rite answ. for each q. In Non-Math T.M., there may not be a single rite answer — ~~[illegible]~~ i.e. even with th. very best pgms, for a gn. □ there is not necessarily 100% prediction accuracy.

Question: Still isn't clear as to whether to use pgm count for anything — i.e. in addition to no. of "cases". Try to see if Non-Math T.M. sheds any lite on this problem. Review previous writing in these notes, on this problem.

It seems clear that U should be ~~oc~~ "cases", but th. question of variance of ~~[illegible]~~ p is th. ~~[illegible]~~ problem here. How does th. "count" size / help determine ~~consistancy~~ consistancy proby ?

Th. suggestion of 7).22 : to plot mortality of consistancy as a funct of "count" and "cases", after some theoretical work to ~~[illegible]~~ get gen. functional form.

There is no "hurry" with th. above plot, since it ~~[illegible]~~ is n't ~~[illegible]~~ too imp. any way. ~~[illegible]~~ ~~[illegible]~~ Algms of hy count may be imp., ~~[illegible]~~, and aprip U in forming new pngms may be a funct of ~~[illegible]~~ pngms count as well as no. of cases.

Discuss Attempts to predict / both U and U'.
Real U, due to its being "th. chosen one"
U due to no. of cases
(that U' is imp. is suggested in th. foll. discussion, up to 27).27

In Math T.M., there is also th. same problem of over lap of pngms in assigning U. ~~[illegible]~~ If pngm A completely overlaps B, then ~~[illegible]~~ th. existance of B is ~~[illegible]~~ (wasteful, and certainly pointless) ? except that A may "die" and leave B surviving. For this reason, it appears O.K. to ~~[illegible]~~ score both

~~Th. ultimate goal of Math T.M. is to have only 1 consistant pngm that covers all possl. cases.~~

A and B whenever ~~[illegible]~~ a case of B occures. Another possy is to score th. one with largest U. At first this would give th. one with largest Uaprip a large lead — which it would always retain. If this was B, then A would occasionally score, when a case occured in A but not in B — but only then.

Probably this : that actually one should score both A and B whenever B is scored — regardless of th. aprip. U's of A and B. ~~However from a p~~ Th. fact that either A or B mite have slitely more U than th. other, ~~probably isn't~~ ~~[illegible]~~ and ~~[illegible]~~ was ~~[illegible]~~ ∴ the one that was chosen, probably isn't too relevant. A slite difference in aprip U batw. A and B mite give A practically no "cases" — yet the U of A and B mite really be about =. Also, they should have about = U in creating new pagms.

Now what is done in R.W. language is that often A or B would be chosen and th. other "swamped out." The reasons may be 1) that A and B are used in a lang. for communication — not only for prediction 2) That "swamping out" one, makes more economical use of th. available memory. It does this without much loss in efficiency (usually), since if A and B overlap and occure usually simultaneously, then usually either one can substitute for th. other in creating about the same new pagms.

"Swamping out" may not usually be so bad, but I seem to have encountered one seriously bad case in th. prevalence of ptetragrams over ptrigms. in th. prediction of ν. Also I think it tends to make getting stuck in blind alleys more likely.

Probably th. best thing to do: keep counts on all pagms in both Math and Non-math T.M.s. In non-math T.M, th. predictions will be made on th. basis of th. "best pagm", but th. fact that a gn. pagm. was "best" very often, will reflect in no way in its prediction U or

.27 in its U for creation of new pagms.

Later, the discarding of pagms that are not "best" often, or are "best" only by a small margin if ever, will be done as a memory economy measure.

This decision may cause T.M. to diverge from intuitive methods of problem solving, so we must remember this decision, if any divergencies or difficulties appear.

About th. biggest Q. at th. present time: What to do about "cases" and "counts" in evaluating probability of death? (U evaluation was decided to be on th. basis of "cases"). What I have done so far is to keep both "count" and "cases", but discard (≡ becomes inconsistent) pagms on th. basis of of "count". ~~T~~his isn't a bad idea — since T.M. need never get answers to its □'s. These direct answs would be nec, if we are to be able to discard pugms of "inconsistancy" on th. basis of "cases". This arguement is analogously true in non-Math T.M. But it suggest a soln. to th. other part of th. Math T.M. problem!

In NMTM (≡ Non Math T.M.) [~~M~~ MTM ≡ Math TM], the prediction vector is determined on th. basis of "count", as th. U simp., since hy U tend to bias th. approp probty. distrib. of th. probty. vector

.10 is th. expected accuracy of that vector. Analogously, th. expected "accuracy" of a prediction vector in MTM should be on th. basis of "count"! This is kind of reasonable in after that, any way. If one discards on th. basis of "count", and A has lasted n_A counts, B, n_B counts, ($n_A \gg n_B$), then if A is to die next ~~...~~ "count" ~~...~~ than it is much less probable than that B should die on th. next "count". Remmember, we have only ~~...~~ "count" data, as far as discarding criteria is concerned. It seems fairly certain that th. probability of discard on th. next count would be a function of total count. There may however, be some correlation betw. ~~...~~ "case no." and probty of discard on th. next count. (there is, because of →)

Useful Defs: 1) An occurance of a count for pugm P ≡ α
2) " " " case. ≡ β

1) ~~≡~~ then P counts ← verb
2) then P cases ← verb (can only occure at time of a q. element.)

count no. of P ≡ no. of times P has counted
case no. "P ≡ " " " " cased.

.38 Great!! So If we have 2 conflicting pugms for a □ in MTM, then th. pugm with th. largest count no. wins. Th. probty ratio will be something like th. count ratio. ~~if th. sq. root of th. count ratio~~.

Notes

Well, this looks like th. last unsolved problem for a while!!

Trouble is, it seems to solve a B. G. problem, which I Know to have no such simple soln! — i.e. th. Mexican Gold coin problem.

Anyway, if an event can happend in one of 2 ways, and its proby. can be any where betw. 0 and 1, and it happends in ~~[illegible]~~ th. first way n times in succession, what is th. probty. it will happend th. same way on th. n+1 th time?

distrib. of U not Poission

Actually, there is more info. If th. apri U is hy, then it is ~~[illegible]~~ more likely that th. pngm will not have a counter case at n+1

It must be noted that ~~[illegible]~~ in Math T.M. th. ~~distribution~~ probty. distrib. of U_{aprip} is not uniform or smooth, since it has a big bump at zero. There is some q. about whether one shouldn't give pngms $U>0$ if they have "lasted" for a long time — i.e. remained "consistant" for a long time.

If one uses uniform apripd betw. 0 and 1, then using "Laplaces rule", th. "Expected value" of th. probty. is $\frac{1}{2+n}$. To use th. U_{aprip}, one mite want to use $\frac{\alpha}{\beta+n}$, ~~where~~ where α and β depend upon U_{aprip}.

To get α and β: Find out th. expected life of any pngm as a function of ~~[illegible]~~ its U_{aprip}.

Another poss. that seems non-ad hoch. Take th. pngmst that is th. "meet" of A and B. If A and B have much U, then their meet will have much ~~[illegible]~~ U_{aprip}.

It may be that this looks ad-hoch, ~~[illegible]~~ — particularly if one didn't try to form th. meet of A and B until an unruly member of this class appeared. Actually, this is O.K. One could assume that all "such" "meets" are made, whether cases of them occuro or not, but that this "making th. class after th. case occures", is done as an economy

Wed July 25, 1956

(Notes)

measure — economy of both time and memory

General Principal: Any apparently ad-hock method may be used, providing it is shown to be a time or memory economy measure, rather than purely a specific problem solving device.

That the above problem of "counter overlap" between **A** and **B** that has been discussed above, is unimportant: In General, there will be many cases in which T.M. will make very poor predictions — and ~~the~~ it will be apparent to T.M. apriori, that they are expected to be poor. Usually this poorness will be because of small sample. In the present case, this is approximately true. — Any of the prediction refinements I have devised do not ↑ prediction accuracy much in this case. The only value of them is that they may be useful in designing NMTM.

$\boxed{SN}$ Case no. may be > Count in some tug sequs. If this is so, then the tug seq. is probably poor.

Whether $U_{empirical} = \frac{\text{no. of cases for that pugm}}{\Sigma \text{ no. of cases}}$ or $\frac{\text{no cases pugm}}{\Sigma \text{ no. of q. elements}}$ or etc., is irrelevant, since we only want these U empiricals's to compare pugms; and the denominators are the same for all pugms. However, we would like to chose a denominator ∋ this ratio tends to remain constant for a gn. pugm. Probably Σ no. of cases for all pugms would be best. — i.e. Σ no. of □'s.

(Notes)

1. = lg 10
1.097 = lg 12.5
1.176 = lg 15
1.243 = lg 17.5
1.301 = lg 20

Organization of Report:

1) Contents: With adequate descriptions of each section.

2) Introduction: a) Explaination that this is a prelim. report, that several terms used will probably be soon ~~e~~ eliminated or revised. b) That / This T.M. is mean to correspond to the operation of methods by which humans solve probs. That correspondance will be drawn in a later section (perhaps this should come here)

3) General description of operation. Examine this carefully, so that no wrong ideas are gotten by reader. Note to reader that section on intuitive corresp. may now be read - or preferably after definitions.

4) Definitions

5) Operation: a) How new pugms, ngms, strs, ntps are formed from old. How U's are determined. How U of pugms are computed from Uaprip and empirical data b) How U aprip for pugms is computed. How U's of ngms, strs, ntps are computed.

6)

Examples of operation: How ~~...~~ T.M. deals with Th. tng. sequence =, ~, ⊕, ⊗, +, −. complementation (?)

7) How one ~~...~~ uses ordinary language in induction. How pugms, ngms, ntps and strs. correspond to objects in intuitive induction.

8) Criterion of effectiveness of present program

9) What next on program: + without carry, ~~...~~ .... Language: simple Mult, divide, linear notation, algebra. Q's and A's about Math, then perhaps about other Things. — That there is some reason to believe T.M. will not be able to ans. q's about Th. real world, unless it has a very different kind of input (i.e. continuous)

10) Use of same tng. sequence on children.