Wed July 18, 1956   T.M.

means. However, it is clear from th. prob. of 82.15, that occasionally one will want to ~~average~~ categorize th. □ sq. with <u>all</u> sqs. (because one <u>has</u> to). At other times, if <u>very</u> many □ cases are available, it is clearly best to cat. it with □ sqs. only, ~~and~~ or mostly/. since they are clearly more "similar" to th. prob. of interest. These are 2 extremes. It would seem that for intermediate cases, we would want a mixture of th. 2 methods, or that we mite compute a ~~transi~~ transition point giving th. rel. goodness of th. 2 methods, and telling us when to use one <u>rather than</u> the other.

This ~~new~~ q. is very much like ~~my~~ old concept of what th. B.G. problem was.

I am fairly sure that T.M. should keep' abss. devised by <u>both</u> methods. Even if th. □ method should have a large enuf sample size, th. ngms devised by averaging over all sqs. would be useful in providing ngms to be made into ntpsts. — i.e. these would be "words of th. 2nd kind". (see 80.09 for def. of).

This is a case where we have 2 methods of prediction. We <u>use</u> th. "best" of th. 2 for prediction. We use th. ngms. of <u>both</u> to form new trial ngms via ntpsts. and strs.

What T.M. will have to do, is take a long time history of predictions via □ and via all sqs., and find their relative relyability. Q. of "relyability vs. sample size". "Sample" size operates as foll: One has an ngmst. One is gn. r examples (no 2 th. same), and all of them "fit" i.e. no exceptions. Say these are ngms with □ in them. Say there are M members in this ngmst. One has picked r of these M at random, and found them O.K.; What is th. proby that ~~only~~ at least s of the M are "good"? ($\equiv f(s) : f(r) = 1$) — This is th. problem that I solved α 84 array they are. For this reason, if they are in th. "wrong" part of th. array (distant from where □'s mite be) they will not "fit" or "apply to" th. □'s that are asked.

.05 (from α81.11) Th. method of 69.10 does apply to all

.06 cases and is automatically not nec. automatically used in th. method of α81.16.

According to 69.10 we collect data on all points and make up ngmsts. We then keep freq. counts on only those ngmsts which have at least one ngm in them that has been applicable to an example with a □ in it.

* I'm still not entirely satisfied that α81.16 is not ad-hock, or that it is in any sense optimum.

.15 Consider th. foll. prob: One is gn. a very big set of examples with □'s in them. Then one is gn. a Q with a □ in it — what is th. optimum prediction? It would appear that 81.16 is inapplicable — Th. q. is — is this a "legit" problem to ask of T.M.?

Perhaps: It is a q. of whether we want to categorize th. □ sq. with only other □ sqs. or with all sqs.. In this case, we have no choice; we must cat. it will all sqs., since □ occures only once. In general, it is best to cat. it with □, if th. sample size is large enuf. If not, we must cat. it with all sqs.. We may look upon ~~each~~ each of th. 2 types of cat'ization as having a certain ~~predictability~~ mean relyability (for ∞ sample size) along with a lack of relyability due to finite sample size. A good way to use both kinds of data is to use a weighted mean of th. 2 kinds of data. Th. rel. wts. will depend on both sample size and expected relyability. I think th. □ type should be expected to give zero mean error, whatever that

T.M.

.01 Before we start on this review, let us clear up this
□ problem. Th. discussion on α73.13 and α73.31 is
not correct. Th. method of α69.10 is not as good as was
that on α69.10 — on th. other hand, it may, in th. long run,
be best. Th. reason why th. discn. of α73.31 is wrong!
In lite of th. new way ~~[illegible]~~ of defining ngmsts:

$\frac{1}{S}$ is of hy freq. ∴ th. digm $\frac{1}{\boxed{1}}$ will not be very
useful in prediction.

.11 On th. other hand use of th. method of α69.10,
will cause T.M. to use trigrams to solve "=" problems. It
isn't clear that th. method of α69.10 is meaningful, if
applied to ngmsts, other than ngms. (e.g. digms, trigms, tetragms, etc)

see α82.05

.16 A possible way: Use ngmst. freqs. averaged over all
possl. points in th. array, but use th. U's obtained with only
tentative finality. For th. final judgement of th. U of an
~~[illegible]~~ ngmst, use its average over only interrogation positions
i.e. its "score" in answering questions actually asked.
This sounds not too bad. Th. "all-over averaging" gives a kind of
apriori U estimate.

It must be noted that th. questions ~~[illegible]~~ raised here,
are essential Q's. This is because ultimately, T.M.
will be shown many examples with no □'s in them,
and then, occasionally ~~a~~ a □ will be given.
I think that if T.M. uses sufficiently good ~~[illegible]~~ ngmsts,
of suff. large n, it will be possl. for T.M. to
get good predictions by averaging all data over all
squares in th. array. I think, however, that th. ultimate U
assignments must give much weight to th. efficacy of
th. ngms in ~~[illegible]~~ answering actual questions.

Th. reason why I think "good" and "large n"
ngms. ~~[illegible]~~ can be used in this way: "Good" ngms,
will have in them, info about just where in th. α82

Also, this enables one to evaluate strs. and ntpsts, to some extent, indipendantly of th. presence of ~~other~~ good ngms that they did not have a hand in th. creation of.

If this solution to th. problem proves unsatisfactory, try to find th. equivalent problem in real Language and science, and solve it. — Then translate back to Math T.M.

We still have 2 kinds of words: from α77.28

.09 first kind: Ngms that ~~...~~ consist of ngms that are divided into 2 parts. (See α79.25 for description)

second kind: Ngms that may be combined to form Ngms of th. above type. — e.g. one may concatenate 2 of them, and have th. first predict th. second.

Words of th. first kind (after th. division is taken out) may be used as trial words of th. second kind. — And, of course words of th. sec. kind are normally combined by str. xfmns. to give words of th. first kind.

Well, it looks like I am now ready to do a detailed description of Math T.M. up to + with carry line). When going past this point, remember, if there is a certain abs. you want T.M. to make, then if it is a really good abs., there will be many problems that are either unsolvable without it, or solvable with much difficulty without it. If these problems are gn. to T.M., and there exists a sequence of abstractions that will lead T.M. to it, then T.M. will eventually get there. — So take heart, oh intrepid voyager into lands unknown! — All you have to do is really understand th. probs. you want T.M. to solve — know the abss. you want T.M. to use as you would know thyself!

T. M.

relyable this implication is.

This seems to solve th. problem. There is some Q. about book keeping — and whether one should count certain cases twice or not. ∝ 80.09

Actually, this implication "both ways" is unnec. — we do, however, have th. epist. rule that if $a \to b$, then it is not unlikely that $b \to a$. In terms of ngms: If $\begin{matrix} a & b \\ c & \boxed{d} \end{matrix}$ is a hy U ngm. i.e. if $\begin{matrix} a & b \\ c & \end{matrix}$ usually implies $\begin{matrix} a & b \\ c & d \end{matrix}$, then $\begin{matrix} \boxed{a} & b \\ c & d \end{matrix}$, $\begin{matrix} a & \boxed{b} \\ c & d \end{matrix}$ and $\begin{matrix} a & b \\ \boxed{c} & d \end{matrix}$ are good trials for useful ngms.

SN Incidently, this ~~view~~ way of assigning U's of an ngm on th. basis of th. ~~a~~ predicting power of that ngm alone, seems to be in th. spirit of th. B. G. soln. that simply choses th. "Best" ~~a~~ ngmst.

.25 An ngmst, then, is a set of objects, ~~that~~ ~~a~~ Each object (≡ ngm) may be divided into 2 parts. Th. ~~predic~~ 1) predictor, and 2) th. predicted. In each case, th. presence of 1) implies th. presence of 2). Th. ~~strength of~~ relyability of this prediction determines th. "goodness" of an ngmst.

.34 ~~Then~~ from ∝ 73.12

Also, this tends to solve th. "swamping out" problem: i.e. how can one recognize ~~the presence of~~ a good ngm. in th. presence of another ngm. that is as good or better? Th. answer is, that often and perhaps always, th. ~~goodness~~ U of an ngm. will be indip. of what other ngms one uses along with it. ∝ 80 of them, is that one of these words may be related, in an implicative way, to another. ~~Th~~ Several of these words may form a space-time config., so that if one observes one [or more] of them, then th. rest are likely to follow.

It may, indeed, be, that [a set of] ~~th~~ words, of th. second type always form a config. and that this config. of words is a word of th. first type. Also these words of th. second type may be ~~also~~ also of th. first type.

This suggests that a, say digm [say 10] is "good" if

a) 1 usually implies 10 and b) if 0 usually implies 10

It would seem that there would be "one way" implications for some types of ngms — e.g. if a) were true, but not b).

What mite happend: That one finds some ngms that are reasonable "good" — that th. implications work both ways or one way for a natural division of that ngm into parts. That ~~—~~ eventually, if one is to make much use of this ngm, one will make a statistical study of th. implication relations betw. its parts.

A "good" ngm, (a type 2 word), then, is one (with clearly defined parts?) ~~—~~ in which th. implication relations betw. its parts, (as found by a statistical study) prove to be useful in prediction.

What this implies for Math T.M.:

One has all of these trial digms. One tries them for implication both ways. Th. one's that are good at least one way, are kept, and indexed as to th. way in which th. implication ~~—~~ works, and how ∝79

T.M.

e.g. say 3 completions are possl. :

1 $\boxed{\alpha}$, 1 $\boxed{\beta}$, 1 $\boxed{\gamma}$. If 1 $\boxed{\alpha}$ <u>and</u> 1 $\boxed{\beta}$ have zero freq., then th. freq. of 1 $\boxed{\gamma}$ isn't very critical.

A possy: Just retain ngms. of hyest freq. — discard those that are of low U.

Another possy: <u>Use</u> th. zero freq. ngms and give them hy U. This will be opposite to language, but may turn out o.k. in th. end.

An ngmst. can be characterized in th. foll. way:

say A $\boxed{\alpha}$, B $\boxed{\beta}$ and G $\boxed{\gamma}$ are members of th. same ngmst. Then this means that if A, then A $\boxed{\alpha}$ is probable, that if B then B $\boxed{\beta}$ is probable, and if G, then G $\boxed{\gamma}$ is probable. This conforms, I think, to th. idea of a word corresponding to an ngmst.

.28 This discn. seems to clear up a lot.

Perhaps I am confusing 2 or more kinds of "words". One e.g. is th. use of th. term car (≡ automobile). One sees part of th. car, and one can extrapolate to th. rest.

Anther kind of "word" is "cloudy day". A similar word is "rain". We have learned to correlated "cloudy day" and "rain".

Th. word "car" is a space-time config. of elements ⇒ if most of them are present, th. rest will probably be present.

Th. "cloudy day", "rain" type words, are also space time configs. — they <u>may</u> be words of this type (and usually are), but th. imp. characteristic α78

T. M.

~~Th. forgoing is ... for a 1 dimens. stoch. process.~~

~~For a 2 dim. case~~:

I is partitioned into r sub sets,

s_i ($i = 1/n$) Each subset s_i is divided into s_i^1 and s_i^2. If one starts th. system at any digit in s_i ($s_i \equiv s_i^1 \overset{\text{Boolean}}{\oplus} s_i^2$) one will, at $t=\infty$, end up in s_i^1.

There is a good discn. of Markov chains in Feller: p 307 ~~ff.~~ on p 318 starts discn. of accessability of states to each other.

Th. forgoing is for 1 dim. case! For a 2 dim. stoch. process: Tho this is an interesting problem, I don't think it is relevant to what I need just now. What I want is, indeed, th. relation of digram freqs. to implication probabilities (correspond to transition probabilities), but this can be done without any of th. Theorems on Markov chains!

Suppose I want to predict $\sim \begin{matrix} 0 & 1 & 1 \\ 1 & \square & 0 \end{matrix}$.

I have th. freqs. of

11, 10, $\begin{matrix}0\\1\end{matrix}$, $\begin{matrix}0\\0\end{matrix}$, $\begin{matrix}1\\1\end{matrix}$, $\begin{matrix}1\\0\end{matrix}$ etc.

[This problem was treated on α52, 53. — no really good answ. was obtained. — th. difficulty was that th. digms of zero freq. proved to have hyest U. This is not in accord with common language, where word names are given to objects of hy freq. and hy U.]

If there are > 2 possl. completion digits (which is usually th. case) then th. ngm. of zero freq. is U, but th. other ngms are also imp. — Not necissarily:

Stoch. Processes. T. M.

compute transition probabilities.

say D_{ij} is th. rel. freq. of th. i^{th} and j^{th} digit.

say $P_i = \sum_j D_{ij} = \sum_j D_{ji}$ is th. long time rel. freq. of th. i^{th} digit. $\sum P_i = \sum_i \sum_j D_{ij} = 1$.

say T_{ij} is th. proby that if i at time t, then j will follow at $t+1$. $\sum_j T_{ij} = 1$.

Then $P_i T_{ij} = D_{ij}$

and $\sum_k D_{ki} T_{ij} = D_{ij}$

so $T_{ij} = \dfrac{D_{ij}}{\sum_k D_{ki}}$ but to go from T_{ij} to D_{ij} one must solve n^2 equs in n^2 unknowns. (There are n differnt digits)

say $\overset{\Rightarrow}{T}{}^* = (\overset{\Rightarrow}{T})^\infty$. than

$\vec{P} = \overset{\Rightarrow}{T}{}^* \vec{P}$ also $\vec{P} = \overset{\Rightarrow}{T}{}^* \cdot$ (any vector), providing any state can go to any other state

T^* has th. same eivectors as T.

T^* has only eivalue of absolute value 1. They come in conjugate pairs, since T^* and T have only real components. If T has complex eivals. of abs. value 1, then its eivects are complex conjugate and $T^\infty \equiv T^*$ does not converge. It oscillates periodically, or "almost periodically". If T^∞ does converge than T has only eivals of value 1, if their eiv is of abs. value 1.

If T has 2 or more eivals. of value 1, than T is degenerate, and any linear comb. of th. corresp. eivecs. is an eivec. of $\lambda = 1$.

There is some Q. about just how one gets th. digits partitioned into sets that cannot go to one another. What must happend, is that the set of digits d_i $(i = 1/n)$