CLM

1372, 3182,6602,6126

>>>

>>>#773.5 1 2 9 114

>>>Hierarchical models with linear terms forced

Statisticians often fit multilinear models (which are the only things they can fit when each factor in the design has only two levels ). Often it makes sense to consider only hierarchical models, i.e. ones in which if a term say xi\*xj\*xk appears, then all the terms

xi\*xj, xi\*xk, xj\*xk, xi, xj, xk and the constant also appear. If this is not required, then the model depends on the origin of measurement of the factors, which doesn't usually make sense. Also usually we would require there to be a constant term, to make the model independent of the origin of measurement of the response. The question arises, how many different hierarchical models are there, as a function of the number of factors? We can

describe them by listing a minimal set of terms; thus e.g. for n=3 factors we have

Total 9 types, 19 different models in all.

The "types" sequence is 2 4 9 29 (I think) and the "models" sequence is 2 5 19 167 (ditto) Pedak (1 all 1)

Now count only the models that force all the linear terms in. So for n=3

we now drop the 1,2,3,5-th types above. This leaves 5 types, 9 models.

The types sequence is now 1 2 5 20 and the models sequence is 1 2 9 114

Again, I have no refrence and no more terms. (But I'll ask my colleagues). This was all suggested by something I wrote on counting kinds of independence hyrotheses, published in J.Statist. Comput. Simul.(1979). But that is not a good ref. for this (It is a Good ref.)

Later: Daryl tells me Anne Freeny and A.McIntosh (Bellcore) may have done something. I'll ask McIntosh tomorrow. (Anne is on vacation).

Mollows June - Tuly 1991

Cc: schupiemath9.Uni-Bielefeld.DE

Status: R

I would appreciate a comment whether you think that there is but I guess you will see what is going on. cyclotomic lattices. It is handwritten and in German, I'll also enclose a copy of my writeup concerning the I'll put a reprint of my paper in the mail for you. that's going to be published). Yes, the second sequence is on page 87, more detailed tables are in a preprint of Hashimoto (rom 1990 (I don't know where and when Dear Neil,

just Thompson's old stuff written up in more detail. something new (and publishable) in it or whether it is

Yours, Rainer.

F6602

6126 cont (2

Counting models again: One can also count hierarchical models of second or higher degree: (i.e. if a term  $x^a*y^b*z^c$  appears in the model, then all terms that are factors of this must appear. This is to make the form of the model invariant under changes of the origins of the factors).

Hierarchical quadratic models:

2 4 14 94(?)

e.g. the 14 counts the models with leading terms

```
0 c x (x,y) xy x^2 (x^2,y) (x^2,xy) (x^2,y^2) (x^2,xy,y^2)
y (x,y^2) (xy,y^2)
```

From mhuxo!gauss!clm Fri Jun 28 13:43:40 EDT 1991 Status: R

> 1+A+B+C \*\*
> 1+A+B+C+A.B \*\*
> 1+A+B+C+A.C
> 1+A+B+C+B.C
> 1+A+B+C+A.B+A.C \*\*
> 1+A+B+C+A.B+B.C
> 1+A+B+C+A.B+B.C
> 1+A+B+C+A.B+B.C
> 1+A+B+C+A.B+A.C+B.C \*\*
> 1+A+B+C+A.B+A.C+B.C \*\*

> then your numbers for 4 are also correct. I can get the numbers for 5, if you > are interested.

gauss\$

please be consistent about

Asom you

with the side condition  $\sum P_i a_i \leq a_i$  say, where  $P_i$  is the probability of using input quantity. Thus input letter i might have cost a; and we wish to find the capacity capacity subject to the constraint that the expected cost not exceed a certain "cost" associated with the different input letters, and it is desired to find the channel. This duality is enhanced if we consider channels in which there is a between the properties of a source with a distortion measure and those of a Duality of a Source and a Channel There is a curious and provocative duality

channel and the desired cost. problem corresponds, in a sense, to finding a source that is just right for the can be shown readily that this function is concave downward. Solving this solution of this problem leads to a capacity cost function C(a) for the channel. It information under variation of the P, with a linear inequality as constraint. The This problem amounts, mathematically, to maximizing a mutual

not have knowledge of it. have knowledge of the past but cannot control it; we may control the future but between past and future and the notions of control and knowledge. Thus we may distortion level. This duality can be pursued further and is related to a duality corresponds to finding a channel that is just right for the source and allowed leads to a function R(d) which is convex downward. Solving this problem variation of the  $q_i(j)$ , again with a linear inequality as constraint. The solution source amounts, mathematically, to minimizing a mutual information under In a somewhat dual way, evaluating the rate distortion function K(d) for a

which a simple explicit solution can be given. The R(d) curve, in fact, is distortion measure is the error probability (per digit). This falls into the class for binary independent letter source with equiprobable letters and suppose that the results will be given for certain simple channels and sources. Consider, first, the Numerical Results for Some Simple Channels In this section some numerical

solves the minimizing problem. and (1-d), the reason being that these are the probability assignment  $q_i(j)$  which This, of course, is the capacity of a symmetric binary channel with probabilities d

approximated by simple means. One point, d=0, is obtained at rate R=1channel. These will give some idea of how well the lower bound may be corresponding to specific simple codes, with the assumption of a noiseless binary This R(d) curve is shown in Fig. 5. Also plotted are a number of points

& sint ]

3

> This result seems to be part of the statistical folklore (e.g. Terry > Speed knows it) but I have never seen it in print anywhere.

> fit models - loglinear and Gaussian. It got started one day when I was in > Ed Fowlkes's office, and he was trying to enumerate and fit all 167 models by > hand.

> Anne Freeny is the author of record on the TM - what happened was that it
> was nearly finished when I left Bell Labs, and then the approval fell through
> the cracks. Anne had a number of requests for the TM, and finally issued
> it herself.

> I can send you the code to print out all models if you are interested - BTL > owns it too. It wouldn't be hard to turn it into a new S function to > take a set of variable names and produce a list of model formulae.

gauss\$

From mhuxo!gauss!clm Mon Jun 24 09:38:22 EDT 1991 Status: R

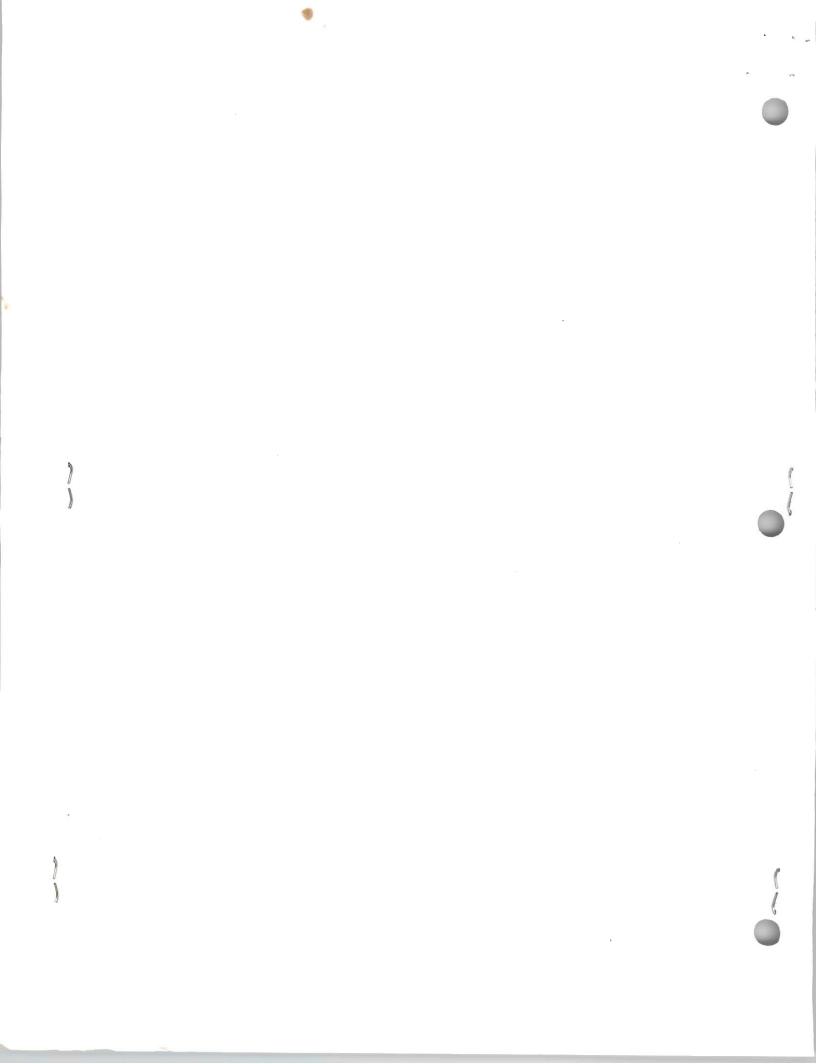
Hi. Here is Allan Mcintosh's reply. You remember I asked him about counting nierarchical models. I should have realised that the unrestricted case is monotone Boolean fns.

> From larch.bellcore.com!mcintosh Thu Jun 20 10:39:00 0400 1991
> Received: by gauss; Thu Jun 20 10:40:02 EDT 1991
> Received: by inet.att.com; Thu Jun 20 10:39 EDT 1991
> Received: by larch.bellcore.com (5.61/1.34)
> id AA11502; Thu, 20 Jun 91 10:39:00 -0400
> Date: Thu, 20 Jun 91 10:39:00 -0400
> From: mcintosh@larch.bellcore.com (Allen Mcintosh)
> Message-Id: <9106201439.AA11502@larch.bellcore.com>
> To: clm@research.att.com
> Subject: hierarchical models
> Cc: mcintosh@larch.bellcore.com

> I can check on this later today - I'll have to grab a program from
> tape, and right now I'm at home. Meanwhile, if you have a copy of
> Neil Sloane's book handy, you can check the number of models yourself.
> The number of hierarchical models on k variables is the same as (you may
> have to subtract 1, depending on your definition)
> - the number of monotone boolean functions on k variables

> - the number of monotone boolean randership of a free lattice with k generators (I may have the terminology wrong here - it's been 9 years).

The paper by Lunon cited in Sloane's book may have the number of orbits under the permutation group - I'm not sure.



From mhuxo!gauss!clm Wed Jul 3 10:01:06 EDT 1991 Status: R

Hi. Message from Allen McIntosh at Bellcore: Column 2 is seq309 monotone Boolean fns.

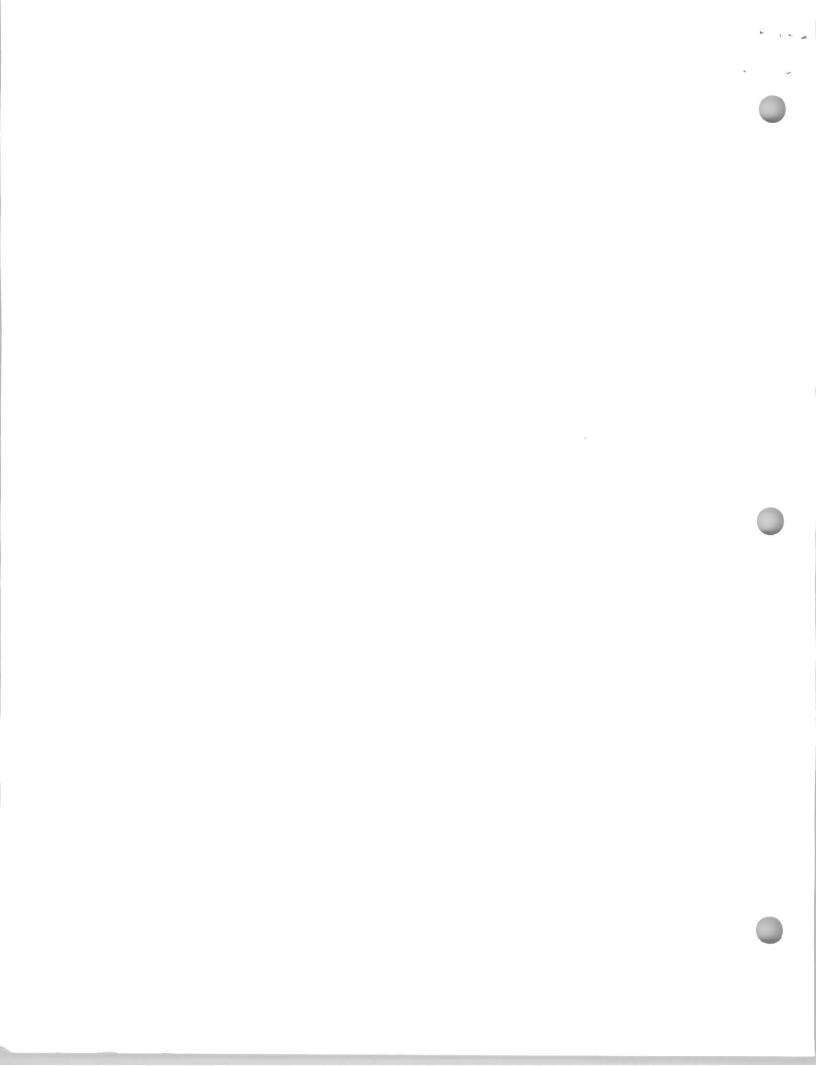
```
> Received: by gauss; Tue Jul 2 18:26:08 EDT 1991
> Received: by inet.att.com; Tue Jul 2 16:25 EDT 1991
> Received: by larch.bellcore.com (5.61/1.34)
> id AA11718; Tue, 2 Jul 91 16:25:00 -0400
> Date: Tue, 2 Jul 91 16:25:00 -0400
> From: mcintosh@larch.bellcore.com (Allen Mcintosh)
> Message-Id: <9107022025.AA11718@larch.bellcore.com>
> To: clm@research.att.com
> Subject: model counts
> Column
           Description
> 1
           number of factors
> 2
           number of hierarchical models
> 3
           number of orbits of same under permutation group
> 4
           number of hierarchical models containing all 1-factor terms
> 5
           number of orbits of same under permutation group
>
                                    1
> 1
                   3
                            1
> 2
           6
                   5
                                    2
> 3
           20
                            9
                                    5
                   10
                                    20
> 4
           168
                   30
                            114
> 5
           7581
                   210
                            6894
                                    180
> 6
           7828354
                                                       660V
           2414682040998
> 7
> Notes:
```

> From larch.bellcore.com!mcintosh Tue Jul 2 16:25:00 0400 1991

> 1) Numbers for 6 and 7 obtained using a modification of the procedure described > by Lunon (I generate all the hierarchical models on n-2 factors). I don't know > of a clever way (short of enumeration) to fill in the rest of the entries.

> 2) I have complete lists for 1 to 5 - if you want them, I can arrange for > you to pick them up from flash by anonmyous ftp.

> 3) 8 (and hence complete table for 6) might be doable with our Maspar. > someday in my spare time...



From mhuxo!gauss!clm Fri Jul 5 16:38:28 EDT 1991 Status: R

Hi. The most explicit reference I can find is

Bishop, Fienberg and Holland, "Discrete Multivariate Analysis", MIT Press 1975, page 34.

They state "The hierarchy principle [...] if any u-term is set equal to zero, all its higher-order relatives must also be set equal to zero. Conversely, if any u-term is not zero, its lower-order relatives must be present in the [...] model". (They defined "relatives" just before this).

This book is a standard reference.

The models that are being discussed relate to the case where the factors are discrete (finitely many levels) and the observations are counts. In this case it doesn't make much sense to use ordinary polynomial response functions, since counts are necessarily non-negative; the trick is to model the log of the mean count as a polynomial function of the factor levels. It is my belief that the "hierarchical" idea is also standard in the kind of polynomial models you have been thinking about, where the response is real-valued, but I can't come up with a good ref. for this.

