Approximate Models and Regularization

Talk 1: Location and the Analysis of Variance

Laurie Davies

University Duisburg-Essen

YES III: Paradigms of Model Choice

Eurandom, Eindhoven, 5th-7th October, 2009

Location

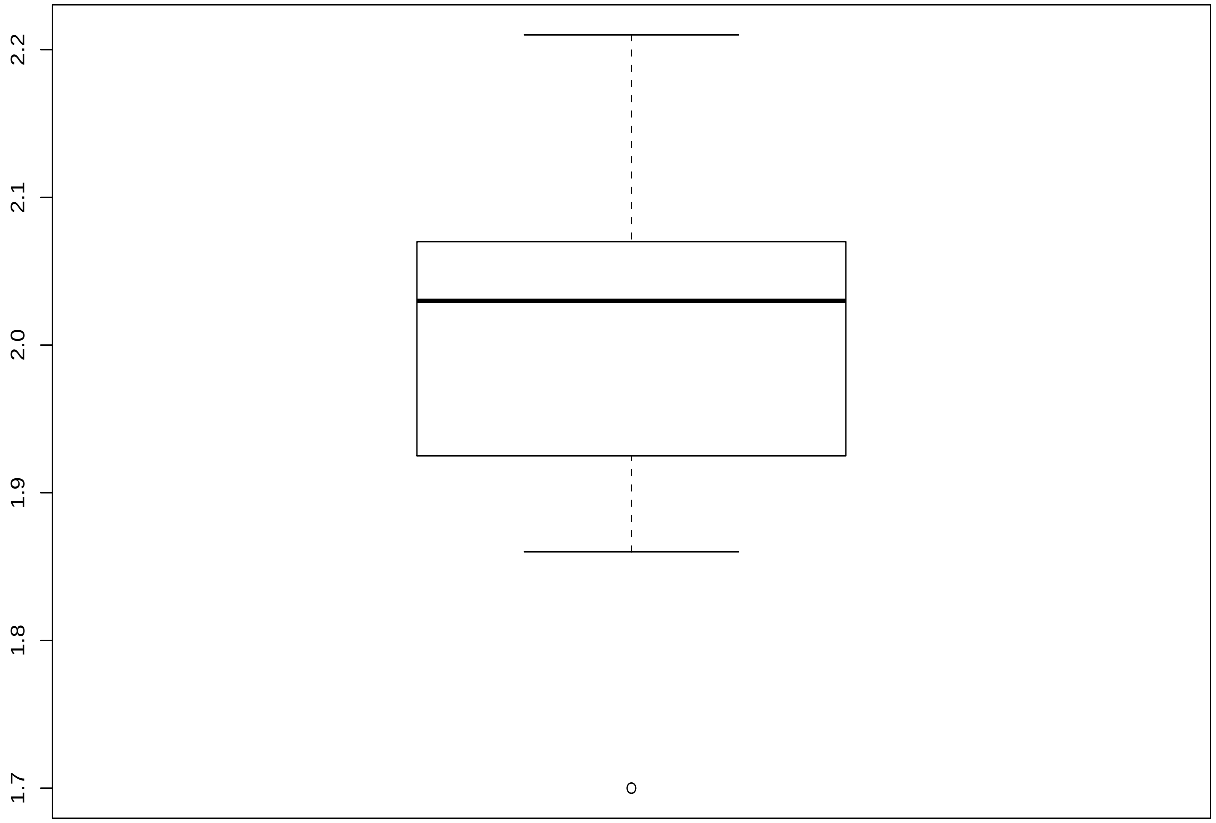
27 measurements of copper (milligrams per litre) in samples of drinking water

2.16	2.21	2.15	2.05	2.06	2.04	1.90	2.03	2.06
2.02	2.06	1.92	2.08	2.05	1.88	1.99	2.01	1.86
1.70	1.88	1.99	1.93	2.2	2.02	1.92	2.13	2.13

We require a point estimate and a range of reasonable values for the concentration of the copper in the water.

Location

Boxplot of the data



Location

How do we proceed? Which model do we choose and why?

Location

Maximum likelihood for Gauss and Laplace models

	Kuiper	log–like.	95%–conf. int.	length
Gauss	0.171	20.31	$[1.970, 2.062]$	0.092
Laplace	0.163	20.09	$[1.989, 2.071]$	0.082

Location

Tukey:

TINSTAAFL

There In No Such Thing As A Free Lunch

Here, of course, 'FREE LUNCH' means 'usefulness of a model that is locally easy to make inferences from'.

What is the (location) model for which it is most difficult to make inferences?

Location

Minimize Fisher information subject to variance 1:

- $N(0, 1)$, a worst case distribution.

Minimize Fisher subject to $d_{ko}(F, N(0, 1)) < \varepsilon$:

- Huber distribution.

Tukey calls such distributions ‘bland’ or ‘hornless’.

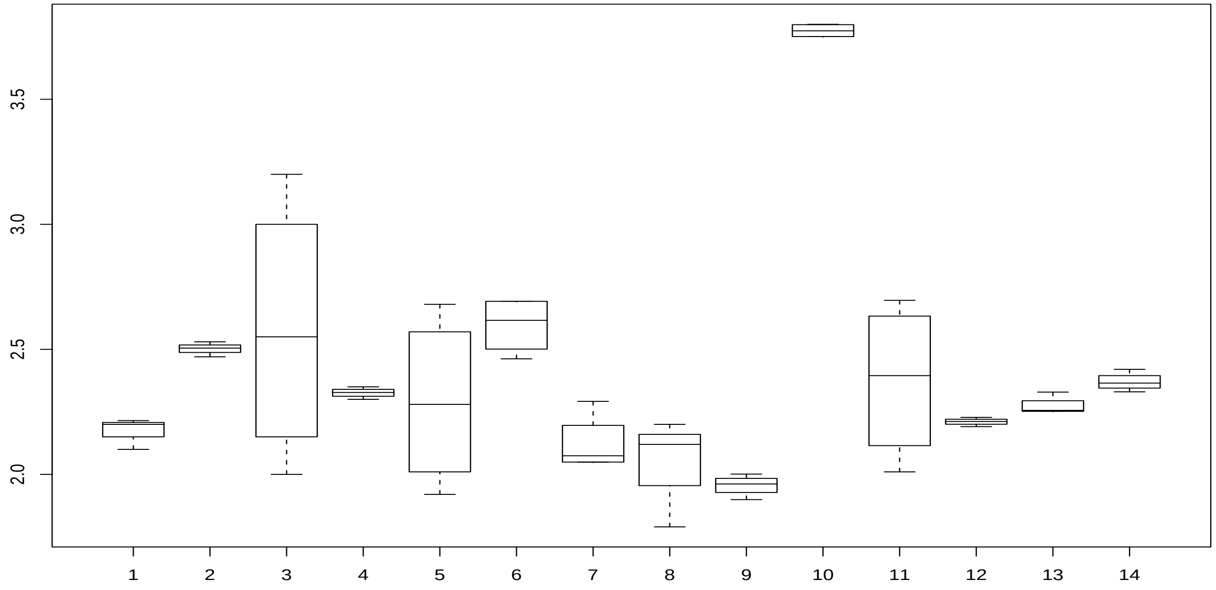
Location

Is the following strategy reasonable?

- Determine the least informative model consistent with the data.
- Calculate the optimal estimator for this model.
- Apply it to the data.

Location

Copper data come from routine control of drinking water. Results of an interlaboratory test, quantity of mercury in drinking water.



Location

Partly for reasons of cost, partly for legal reasons and partly for reasons of comparison such data have to evaluated by a fixed procedure.

As an expert statistician you are given the job of devising a reasonable procedure. How would you approach the problem?

Analysis of Variance

Restrict attention to the two-way table.
Model

$$\begin{aligned} Y_{ijk} &= \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \\ & \quad k = 1, \ldots, K_{ij},\ j = 1, \ldots, J,\ i = 1, \ldots, I \end{aligned}$$

Main effect μ, row effects α_i, column effects β_j and interactions γ_{ij}.
Identifiability restrictions

$$\sum_i \alpha_i = 0,\ \sum_j \beta_j = 0,\ \sum_j \gamma_{ij} = 0\,\forall i,\ \sum_i \gamma_{ij} = 0\,\forall j.$$

Analysis of Variance

Three different medicaments are applied to three different forms of cancer cells. Percentage of cancer cells killed averaged over the three observations per cell.

	1	2	3
1	87.2	31.3	29.6
2	30.7	27.6	28.4
3	28.5	32.3	31.4

Interpretation?

Analysis of Variance

Standard analysis gives $\mu = 36.33$ and

$$\alpha_1 = 13.03, \alpha_2 = -7.43, \alpha_3 = -5.60$$

$$\beta_1 = 12.47, \beta_2 = -5.93, \beta_3 = -6.53$$

and the interactions

$$\gamma = \begin{pmatrix} \text{25.37} & \text{-12.13} & \text{-13.23} \\ \text{-10.67} & \text{4.63} & \text{6.03} \\ \text{-14.70} & \text{7.50} & \text{7.20} \end{pmatrix}$$

Interpretation?

Analysis of Variance

Here the L_1–analysis: $\mu = 31.3$

$$\alpha_1 = 0.0, \alpha_2 = -2.9, \alpha_3 = 0.1$$
$$\beta_1 = 2.3, \beta_2 = 0.0, \beta_3 = 0.0$$

and interactions

$$\gamma = \begin{pmatrix} \text{53.6} & \text{0.0} & \text{-1.7} \\ \text{0.0} & \text{-0.8} & \text{0.0} \\ \text{-5.2} & \text{0.9} & \text{0.0} \end{pmatrix}$$

Interpretation?

Analysis of Variance

The restrictions

$$\sum_i \alpha_i = 0,\ \sum_j \beta_j = 0,\ \sum_j \gamma_{ij} = 0\,\forall i,\ \sum_i \gamma_{ij} = 0\,\forall j.$$

are not data analytically neutral.

They are more than simply identifiability conditions.

What is the minimum number of non-zero interactions they allow?

Analysis of Variance

A sparsity approach: consider exact data x_{ij} without noise. Minimize the number of non-zero interactions

$$\min_{\alpha_i,\beta_j} |\{(i,j) : |x_{ij} - \alpha_i - \beta_j| > 0\}|.$$

We say the solution is unique if the interactions

$$c_{ij} = x_{ij} - \alpha_i - \beta_j$$

are the same for all minimizing values α_i and β_j.

Analysis of Variance

The solution is unique if for example

$$x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{or} \quad x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

but not if

$$x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Analysis of Variance

To see the latter we note that two is the minimum number of non-zero interactions. We have the sequence

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 0 & -1 & -1 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

An interaction pattern Γ^* is a matrix whose elements are either $*$ or 0.

$$\Gamma^* = \begin{pmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Analysis of Variance

An interaction matrix γ is obtainable from an interaction pattern Γ^* by replacing the $*$ by some numerical values.

An interaction pattern Γ^* is called *unconditionally identifiable* if the minimization problem has a unique solution for every interaction matrix γ obtainable from Γ^*.

For an unconditionally identifiable interaction pattern uniqueness depends only on the positions of the interactions and not on their values.

Analysis of Variance

The interaction pattern

$$\Gamma^* = \begin{pmatrix} * & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

is unconditionally identifiable. The interaction pattern

$$\Gamma^* = \begin{pmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

is not unconditionally identifiable.

Analysis of Variance

Given an interaction pattern Γ^* denote by Γ the matrix obtained from Γ^* by replacing all the $*$ by 1.

The following operations are allowed.

To each row or column of Γ we are allowed to add a row of 1s or column of 1s but the addition is $\mod 2$, that is $1 + 1 = 0$, to produce a matrix $\tilde{\Gamma}$.

Theorem

If there is no $\tilde{\Gamma}$ with fewer zeros or with the same number of zeros but in different places, then Γ^* is unconditionally identifiable

Analysis of Variance

The interaction pattern

$$\Gamma^* = \begin{pmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

is not unconditionally identifiable. The interaction pattern

$$\Gamma^* = \begin{pmatrix} * & 0 & 0 & 0 & 0 \\ 0 & * & 0 & 0 & 0 \\ 0 & 0 & * & 0 & 0 \\ 0 & 0 & 0 & * & 0 \\ 0 & 0 & 0 & * & * \end{pmatrix}$$

is unconditionally identifiable.

Analysis of Variance

Theorem

Let γ be an interaction pattern deriving from an unconditionally identifiable interaction patter Γ^* and let x derive from γ by adding arbitrary row and column effects. Then for any solution of

$$\min_{\alpha_i,\beta_j} \sum_{ij} |x_{ij} - \alpha_i - \beta_j|$$

we have

$$x_{ij} - \alpha_i - \beta_j = \gamma_{ij}.$$

In other words, the solution of the L_1-problem solves the sparsity problem.

Analysis of Variance

Moral of the tale:

L_1 is better suited to the analysis of variance than the traditional L_2.

References

Davies, P. L. (1995) Data Features. *Statistica Neerlandica* **49**, 185–245

Davies, P. L. (1998) On locally uniformly linearizable high breakdown location and scale functionals. *Annals of Statistics*, **26**, 1103-1125.

Davies, P. L. (2008) Approximating data (discussion paper: Rudolf Beran, Frank Hampel, Nils Hjort, Benedikt Pötcher), *Journal of the Korean Statistical Society*, **37**, 191–211, 231–240

DIN 38402-45:2003-09 German standard methods for the examination of water, waste water and sludge - General information (group A) - Part 45: Interlaboratory comparisons for proficiency testing of laboratories (A 45). *Deutsches Institut für Normierung*.

Dudley, R. M., Sidenko, S. and Wang, Z. (2009) Differentiability of t-functionals of location and scatter. *Annals of Staistics*, **37**, 939–960.

References

Terbeck, W. and Davies, P. L. (1998) Interactions and Outliers in the Two-way Analysis of Variance. *Annals of Statistics*, **26**, 1279-1305.

Tukey, J. W. (1993) Issues relevant to an honest account of data-based inference, partially in the light of Laurie Davies's paper. *Technical report, Princeton University, Princeton*.