Using chain graph models for structural inference with an application to linguistic data

Craig Alexander[1], Jane Stuart-Smith[2], Tereza Neocleous[1], Ludger Evers [1]

[1] School of Mathematics and Statistics, University of Glasgow, UK
[2] School of Critical Studies, University of Glasgow, UK

E-mail for correspondence: `c.alexander.1@research.gla.ac.uk`

Abstract: Graphical models provide a visualisation of the conditional dependence structure between variables, making them an attractive inference tool. The improved readability makes this an appealing approach to represent complex model output to non-statisticians. In this paper, we introduce a novel approach using graphical models to visualise the output of a mixed effects model with multivariate response with an application to linguistic data.

Keywords: Hierarchical models; Graphical models; Linguistic change.

1 Introduction

In this work, we discuss the use of a chain graph model structure to represent the output from a hierarchical regression model with multivariate response. The chain graph model is inferred in three parts. The dependency structure of the covariates is modelled independently using standard structural inference methods in graphical models. The relationship between the response and explanatory variables and the dependence structure between response variables is jointly inferred using a multivariate Bayesian hierarchical model, where the precision estimates of the residuals and random effects are assumed to conform to an undirected graphical model. We report an application of this model to linguistic data obtained from the Sounds of the City corpus, consisting of Glaswegian speech recordings from the 1970's to the 2000's. From this data, we look to recover the underlying chain graph model detailing which factors affect vowel change.

This paper was published as a part of the proceedings of the 32nd International Workshop on Statistical Modelling (IWSM), Johann Bernoulli Institute, Rijksuniversiteit Groningen, Netherlands, 3–7 July 2017.

2 Chain graph model structure

Implementing a chain graph (Lauritzen & Wermuth, 1989) structure allows the use of directed and undirected edges. Nodes are partitioned into blocks, one for explanatory variables and one for response variables. Edges within blocks are undirected and edges between blocks are directed.
The directed edges between blocks are modelled using a Bayesian hierarchical model. The response is defined as y^l_{ij}, which is the j^{th} measurement from the i^{th} group on the l^{th} response. $\boldsymbol{\beta}^l$ is the corresponding vector of regression coefficients. The random effects coefficients are denoted by $\mathbf{b}^l$.
In vector-matrix notation, the likelihood is defined as:

$$p(\boldsymbol{y} \mid \boldsymbol{\beta}, \mathbf{b}, \boldsymbol{\Omega}_{\boldsymbol{\epsilon}}, \mathbf{X}) = \mathcal{N}\left(\boldsymbol{y} \mid \mathbf{X}\boldsymbol{\beta} + \mathbf{U}\mathbf{b}, \left(\boldsymbol{\Omega}_{\boldsymbol{\epsilon}}^{-1} \otimes \mathbf{I}\right)\right). \tag{1}$$

As conjugate priors have been specified for each parameter where possible, a Gibbs sampler can be used for parameter inference.
The presence of a directed edge in the graphical model corresponds to the value of the respective β^j_i coefficient:

- $\beta^j_i \neq 0 \rightarrow$ an edge is present between variable i and response j.
- $\beta^j_i = 0 \rightarrow$ no edge is present between variable i and response j.

If any interactions are in the model, a factor graph structure is used to represent these, with the interacting explanatory variables connecting to a factor node, which is connected to the relevant response variable.
At each step of the sampler, a new candidate model is proposed by either adding or removing an explanatory variable. By integrating out $\boldsymbol{\beta}$, model evidences are then computed to determine whether we accept or reject the candidate model.
For random effect parameters $\mathbf{b}$ a Gaussian prior is specified of the form:

$$\mathbf{b} \mid \boldsymbol{\Omega}_{\mathbf{b}} \sim \mathcal{N}\left(\mathbf{0}, \left(\boldsymbol{\Omega}_{\mathbf{b}}^{-1} \otimes \mathbf{I}\right)\right). \tag{2}$$

To maintain conjugacy, the random effect and model error precision matrices have G-Wishart (Dobra *et al*, 2011) hyperpriors placed on them:

$$\boldsymbol{\Omega}_{\mathbf{b}}, \boldsymbol{\Omega}_{\boldsymbol{\epsilon}} | \mathrm{G} \sim \mathcal{W}_{\mathrm{G}}(\nu_{\mathbf{b}}, \mathbf{S}_{\mathbf{b}}). \tag{3}$$

where G is a defined graph.
The precision estimates from the hierarchical model are taken as input to a zero mean Gaussian graphical model, defined as:

$$\mathcal{M}_G = \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}^{-1}). \tag{4}$$

where $\omega_{i,j} = 0$ corresponds to a missing edge between response i and j.

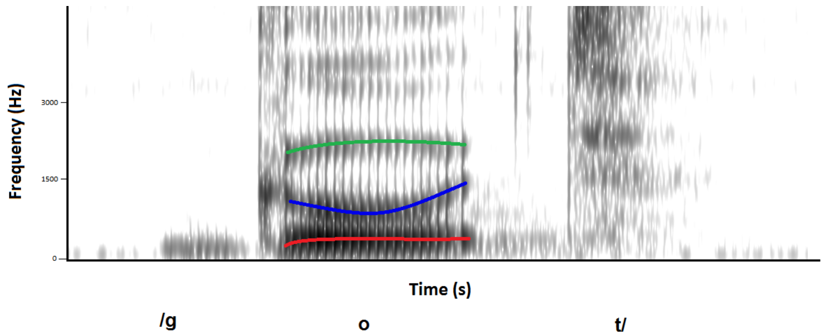

FIGURE 1. Spectrogram of the word *GOAT* spoken by a male Glaswegian speaker. The three coloured lines show the first three formants of the vowel /o/.

The normalising constant, $I_{\mathrm{G}}(\nu, \mathbf{S})$, for chordal graphs can be obtained via a closed form solution by factorising it into a product of density functions:

$$I_{\mathrm{G}}(\nu, \mathbf{S}) = \frac{\prod_{i=1}^{d} I_{T_i}(\nu, S_{T_i,T_i})}{\prod_{j=1}^{d-1} I_{S_i}(\nu, S_{S_i,S_i})}. \quad (5)$$

where T_i are cliques and S_i the separators of G. In the case of a non-chordal graph, model selection can be performed by using trans-dimensional MCMC methods such as those discussed in Mohammadi & Wit (2015).

3 Application: vowel change in Glasgow

We apply the methodology to data from the Sounds of the City (Stuart-Smith, 2017) project, which is a study modelling vowel change in the Glaswegian dialect over the 20th century. Recordings of speakers over various decades are used and vowel measurements are taken along with phonetic quantities of interest and social and biological factors.

Acoustically, a vowel can be characterised by its main resonances, known as formants (Hz). This is illustrated in Figure 1, with the /o/ vowel formants represented by the three coloured lines. Vowel change is studied in terms of how such frequencies alter over time.

Figure 2 shows the best posterior model selected for the vowel /o/, found in words like GOAT, HOPE, etc, selected with posterior probability 0.414.

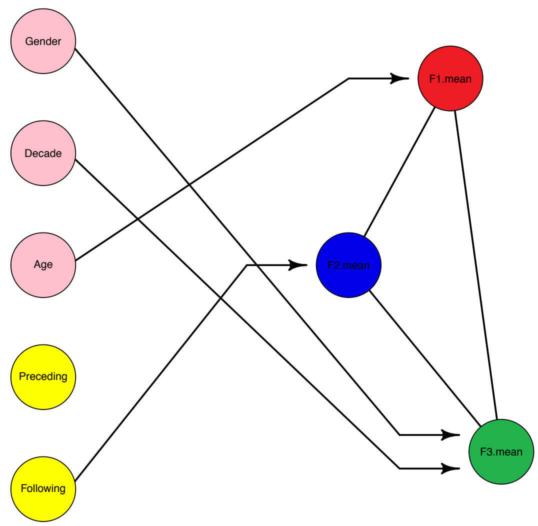

FIGURE 2. Best posterior model selected for the *GOAT* vowel.

From the graphical model, we observe the following significant predictors:

- Age for F1; younger speakers are leading the change in vowel quality. Shifts in F1 relate to raising of the vowel quality, so /o/ becoming more like /u/.
- Shifts in F2 relate to fronting/backing of the vowel quality according to the nature of following consonant.
- Decade and Gender for F3, evidence of vowel change over time and additional differences due to speaker gender.

4 Conclusions & future work

In this work, we have extended beyond previous modelling of sociolinguistic data, by considering multiple formants within the one model, accounting for the high correlation between formants. This approach can also be extended beyond linguistic data, with several academic trials using datasets with similar multivariate and nested features.

In order to promote the usability of our method, we aim to turn it into an online application, allowing users to input their own data.

Acknowledgments: CA is grateful to the Lord Kelvin Adam Smith scholarship for his research studentship.

References

Lauritzen, S.L. and Wermuth, N. (1989). Graphical models for associations between variables, some of which are qualitative and some quantitative. *The Annals of Statistics*, **17**, 31-57.

Dobra, A., *et al* (2011). Bayesian inference for general Gaussian graphical models with application to multivariate lattice data. *JASA* , **106**.

Stuart-Smith, J., *et al* (2017). *Changing sounds in a changing city. Language and a sense of place*, E. Moore & C. Montgomery, CUP, 38-64.

Mohammadi, A. and Wit, E.C. (2015) Bayesian Structure Learning in Sparse Gaussian Graphical Models. *Bayesian Analysis*, **10**, 109-138.