Estimation of Means in Graphical Gaussian Models with Symmetries

Steffen Lauritzen[1]
University of Oxford

Fields Institute

April 16, 2012

[1] Representing joint work with Helene Gehrmann

Consider

$$Y = (Y_\alpha)_{\alpha \in V} \sim \mathcal{N}_{|V|}(0, \Sigma)$$

and let let $K = \Sigma^{-1}$ be the *concentration* matrix.
The *partial correlation* between Y_α and Y_β given all other variables is

$$\rho_{\alpha\beta \mid V \setminus \{\alpha,\beta\}} = -k_{\alpha\beta} / \sqrt{k_{\alpha\alpha} k_{\beta\beta}}. \tag{1}$$

Thus

$$k_{\alpha\beta} = 0 \iff Y_\alpha \perp\!\!\!\perp Y_\beta \mid Y_{V \setminus \{\alpha,\beta\}}.$$

A *graphical Gaussian model* is represented by an undirected graph $\mathcal{G} = (V, E)$ with Y as above and $K \in \mathcal{S}^+(\mathcal{G})$, the set of (symmetric) positive definite matrices with

$$\alpha \not\sim \beta \Rightarrow k_{\alpha\beta} = 0.$$

We shall be interested in also adding means so that $Y \sim \mathcal{N}(\mu, \Sigma)$ with $\mu \in \Omega$, where Ω is a linear subspace of $\mathbb{R}^V$.

Based on observations $Y^1, \dots, Y^n$ the likelihood function is

$$L(\mu, K) \propto \det K^{n/2} \exp^{-\sum_{1 \le i \le n} (y^i - \mu)^T K (y^i - \mu)/2}. \qquad (2)$$

If μ is unrestricted so that $\mu \in \Omega = \mathbb{R}^V$, L is maximised over μ for fixed K by $\hat{\mu} = \mu^* = \bar{y}$ and inference about K can be based on

$$L(\hat{\mu}, K; y) \propto \det K^{n/2} \exp\{-\operatorname{tr}(KW)/2\}, \qquad (3)$$

where $W = \sum_{i=1}^n (y^i - \mu^*)(y^i - \mu^*)^T$ is the matrix of sums of squares and products of the residuals.

In general the situation is more complex. Consider the graph

representing two independent Gaussian variables with unknown variances σ_1^2 and σ_2^2. The Behrens–Fisher problem (Scheffé, 1944) occurs when estimating $\mu = (\mu_1, \mu_2)$ under the restriction $\mu_1 = \mu_2$. The least squares estimator (LSE) $\mu^* = (\bar{y}_1, \bar{y}_2)$ is then not the MLE, the likelihood function (2) under the hypothesis $\mu_1 = \mu_2$ may have multiple modes (Drton, 2008), and there there is no similar test for the hypothesis.

Kruskal (1968) found the following necessary and sufficient condition for the LSE μ^* and MLE $\hat{\mu}$ to agree for a fixed Σ:

Theorem (Kruskal)

Let $Y \sim \mathcal{N}(\mu, \Sigma)$ with unknown mean $\mu \in \Omega$ and known Σ. Then the estimators μ^* and $\hat{\mu}$ coincide if and only if Ω is invariant under $K = \Sigma^{-1}$, i.e. if and only if

$$K\Omega \subseteq \Omega. \tag{4}$$

As $K\Omega \subseteq \Omega$ if and only if $\Sigma\Omega \subseteq \Omega$ this can equivalently be expressed in terms of Σ.

Consequently, if $K \in \Theta$ is unknown and *$K\Omega \subseteq \Omega$ for all $K \in \Theta$* we also have $\mu^* = \hat{\mu}$ and inference on K can be based on the profile likelihood function (3)

$$L(\hat{\mu}, K) \propto \det K^{n/2} \exp\{-\operatorname{tr}(KW)/2\}.$$

The Behrens–Fisher problem is then resolved if we also restrict the variances $\sigma_1^2 = \sigma_2^2 = \sigma^2$ since

$$\begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \begin{pmatrix} \alpha \\ \alpha \end{pmatrix} = \begin{pmatrix} \sigma^2\alpha \\ \sigma^2\alpha \end{pmatrix} = \begin{pmatrix} \beta \\ \beta \end{pmatrix},$$

so the mean space is stable under Σ.

The additional symmetry in the concentration matrix induced by the restriction $\sigma_1^2 = \sigma_2^2$ is represented by a coloured graph

where nodes of same colour have identical elements in their concentration matrix.

Three types of symmetry restrictions:

- RCON restricts concentration matrix;

Three types of symmetry restrictions:

- RCON restricts concentration matrix;
- RCOR restricts partial correlations;

Three types of symmetry restrictions:

- RCON restricts concentration matrix;
- RCOR restricts partial correlations;
- RCOP has restrictions generated by permutation symmetry.

Three types of symmetry restrictions:

- RCON restricts concentration matrix;
- RCOR restricts partial correlations;
- RCOP has restrictions generated by permutation symmetry.

In principle one could/should also study RCOV-models given by symmetry restrictions in the covariance matrix. These are in general different from any of the above but have as far as I know not been investigated at this point in time.

Graph colouring

Undirected graph $\mathcal{G} = (V, E)$.

Colouring vertices of $\mathcal{G}$ with different colours induces partitioning of V into *vertex colour classes.*

Colouring edges E partitions E into disjoint *edge colour classes*

$$V = V_1 \cup \cdots \cup V_p, \quad E = E_1 \cup \cdots \cup E_q.$$

$\mathcal{V} = \{V_1, \ldots, V_p\}$ is a *vertex colouring,*

$\mathcal{E} = \{E_1, \ldots, E_q\}$ is an *edge colouring,*

$\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a *coloured graph.*

- Models with symmetry in covariance are classical and admit unified theory (Wilks, 1946; Votaw, 1948; Olkin and Press, 1969; Andersson, 1975; Andersson et al., 1983);

- Models with symmetry in covariance are classical and admit unified theory (Wilks, 1946; Votaw, 1948; Olkin and Press, 1969; Andersson, 1975; Andersson et al., 1983);
- Stationary autoregressions (circular) (Anderson, 1942; Leipnik, 1947);

- Models with symmetry in covariance are classical and admit unified theory (Wilks, 1946; Votaw, 1948; Olkin and Press, 1969; Andersson, 1975; Andersson et al., 1983);
- Stationary autoregressions (circular) (Anderson, 1942; Leipnik, 1947);
- Spatial Markov models (Whittle, 1954; Besag, 1974; Besag and Moran, 1975);

- Models with symmetry in covariance are classical and admit unified theory (Wilks, 1946; Votaw, 1948; Olkin and Press, 1969; Andersson, 1975; Andersson et al., 1983);
- Stationary autoregressions (circular) (Anderson, 1942; Leipnik, 1947);
- Spatial Markov models (Whittle, 1954; Besag, 1974; Besag and Moran, 1975);
- General combinations with conditional independence are more recent (Hylleberg et al., 1993; Andersson and Madsen, 1998; Madsen, 2000).

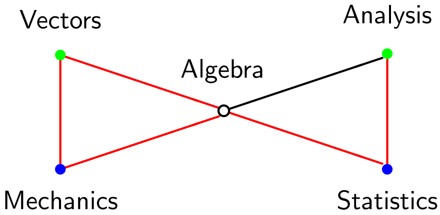
Empirical concentration matrix (inverse covariance) of examination marks of 88 students in 5 mathematical subjects.

	Mechanics	Vectors	Algebra	Analysis	Statistics
Mechanics	5.24	−2.44	−2.74	0.01	−0.14
Vectors	−2.44	10.43	−4.71	−0.79	−0.17
Algebra	−2.74	−4.71	26.95	−7.05	−4.70
Analysis	0.01	−0.79	−7.05	9.88	−2.02
Statistics	−0.14	−0.17	−4.70	−2.02	6.45

Data reported in Mardia et al. (1979)

RCON model

Data support model with symmetry restrictions as in figure:



Elements of concentration matrix corresponding to same colours are identical.

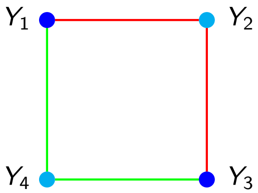
Black or white neutral and corresponding parameters vary freely.

RCON model since restrictions apply to *concentration* matrix

RCON model

1. *Diagonal elements K corresponding to vertices in the same vertex colour class must be identical.*
2. *Off–diagonal entries of K corresponding to edges in the same edge colour class must be identical.*

Diagonal of K thus specified by Tp-dimensional vector η and off-diagonal elements by a q dimensional vector δ so $K = K(\eta, \delta)$. The set of positive definite matrices which satisfy these restrictions is denoted $\mathcal{S}^+(\mathcal{V}, \mathcal{E})$.



Corresponding RCON model will have concentration matrix

$$K = \begin{pmatrix} k_{11} & k_{12} & 0 & k_{14} \\ k_{21} & k_{22} & k_{23} & 0 \\ 0 & k_{32} & k_{33} & k_{34} \\ k_{41} & 0 & k_{43} & k_{44} \end{pmatrix} = \begin{pmatrix} \eta_1 & \delta_1 & 0 & \delta_2 \\ \delta_1 & \eta_2 & \delta_1 & 0 \\ 0 & \delta_1 & \eta_1 & \delta_2 \\ \delta_2 & 0 & \delta_2 & \eta_2 \end{pmatrix}.$$

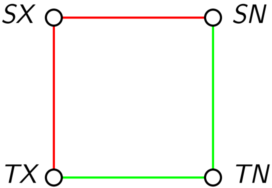
Cox and Wermuth (1993) report data on personality characteristics on 684 students:

Table below shows empirical concentrations ($\times 100$) (on and above diagonal), partial correlations (below diagonal), and standard deviations for personality characteristics of 684 students.

	SX	SN	TX	TN
SX (State anxiety)	0.58	−0.30	−0.23	0.02
SN (State anger)	0.45	0.79	−0.02	−0.15
TX (Trait anxiety)	0.47	0.03	0.41	−0.11
TN (Trait anger)	−0.04	0.33	0.32	0.27
Standard deviations	6.10	6.70	5.68	6.57

RCOR model

Data strongly support conditional independence model displayed below with *partial correlations* strikingly similar in pairs:



Scales for individual variables may not be compatible. *Partial correlations invariant under changes of scale*, and more meaningful. Such symmetry models are denoted *RCOR models.*

RCOR models

1. Diagonal elements of K corresponding to vertices in same vertex colour class must be identical.
2. *partial correlations* along edges in the same edge colour class must be identical.

The set of positive definite matrices which satisfy the restrictions of an $RCOR(\mathcal{V}, \mathcal{E})$ model is denoted $\mathcal{R}^{+}(\mathcal{V}, \mathcal{E})$.

Define A as diagonal matrix with

$$a_\alpha = \sqrt{k_{\alpha\alpha}} = \eta_u, \alpha \in u \in \mathcal{V}$$

We can uniquely represent $K \in \mathcal{R}^+(\mathcal{V}, \mathcal{E})$ as

$$K = ACA = A(\eta)C(\delta)A(\eta),$$

where C has all diagonal entries equal to one and off-diagonal entries are negative partial correlations

$$c_{\alpha\beta} = -\rho_{\alpha\beta \,|\, V\setminus\{\alpha,\beta\}} = k_{\alpha\beta}/\sqrt{k_{\alpha\alpha}k_{\beta\beta}} = k_{\alpha\beta}/(a_\alpha a_\beta).$$

Vertex colour classes restrict A, whereas edge colour classes restrict C.

Although restrictions linear in each of A and C, they are in general not linear in K.

For unrestricted mean, or mean zero *RCOR models are curved exponential families.*

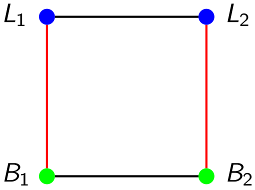
Letting $\lambda_u = \log \eta_u$ the profile likelihood function becomes

$$\log L = \frac{f}{2} \log \det\{C(\delta)\} + f \sum_{u \in \mathcal{V}} \lambda_u \operatorname{tr}(K^u) - \frac{1}{2} \operatorname{tr}\{C(\delta)A(\lambda)WA(\lambda)\}$$

$\log L$ concave in λ for fixed δ and vice versa, but *not in general jointly.*

RCOP model

Data from Frets (1921). Length and breadth of the heads of 25 pairs of first and second sons. Data support the model



Assume distribution unchanged if sons are switched. *RCOP model* as determined by permutation of labels.

Both RCON and RCOR because all aspects of the joint distribution are unaltered when labels are switched.

Let G be permutation matrix for elements of V. If $Y \sim \mathcal{N}_{|V|}(0, \Sigma)$ then $GY \sim \mathcal{N}_{|V|}(0, G\Sigma G^\top)$.

Let $\Gamma \subseteq S(V)$ be a subgroup of such permutations.

Distribution of Y *invariant under the action of* Γ if and only if

$$G\Sigma G^\top = \Sigma \text{ for all } G \in \Gamma. \tag{5}$$

Since G satisfies $G^{-1} = G^\top$, (5) is equivalent to

$$G\Sigma = \Sigma G \text{ for all } G \in \Gamma, \tag{6}$$

i.e. that G commutes with Σ or, equivalently, that G commutes with K:

$$GK = KG \text{ for all } G \in \Gamma.$$

We must insist that zero elements of K are preserved, i.e. G is *automorphism* of the graph, mapping edges to edges:

$$G(\alpha) \sim G(\beta) \iff \alpha \sim \beta \text{ for all } G \in \Gamma,$$

An RCOP model $RCOP(\mathcal{G}, \Gamma)$ generated by $\Gamma \subseteq \mathrm{Aut}(\mathcal{G})$ is given by assuming

$$K \in \mathcal{S}^+(\mathcal{G}, \Gamma) = \mathcal{S}^+(\mathcal{G}) \cap \mathcal{S}^+(\Gamma)$$

where $\mathcal{S}^+(\Gamma)$ is the set of positive definite matrices satisfying

$$GK = KG \text{ for all } G \in \Gamma.$$

An RCOP model can also be represented by a graph colouring:

If $\mathcal{V}$ denotes the *vertex orbits* of Γ, i.e. the equivalence classes of

$$\alpha \equiv_\Gamma \beta \iff \beta = G(\alpha) \text{ for some } G \in \Gamma,$$

and similarly $\mathcal{E}$ the *edge orbits*, i.e. the equivalence classes of

$$\{\alpha, \gamma\} \equiv_\Gamma \{\beta, \delta\} \iff \{\beta, \delta\} = \{G(\alpha), G(\gamma)\} \text{ for some } G \in \Gamma,$$

then we have

$$\mathcal{S}^+(\mathcal{G}, \Gamma) = \mathcal{S}^+(\mathcal{V}, \mathcal{E}) = \mathcal{R}^+(\mathcal{V}, \mathcal{E}).$$

Hence an RCOP model can also be represented as an RCON or an RCOR model with vertex orbits as vertex colour classes and edge orbits as edge colour classes.

For each vertex colour class $v \in \mathcal{V}$ let T^v be the $|V| \times |V|$ diagonal matrix with entries $T^v_{\alpha\alpha} = 1$ if $\alpha \in u$ and 0 otherwise,i.e. T^v is the *indicator* for v.

Similarly, for each edge colour class $e \in \mathcal{E}$ let T^e have entries $T^e_{\alpha\beta} = 1$ if $\{\alpha, \beta\} \in e$ and 0 otherwise, i.e. T^e is the *adjacency* matrix for e.

Now any $K \in \mathcal{S}^+(\mathcal{V}, \mathcal{E})$ can in a unique way be written as

$$K = \sum_{u \in \mathcal{V} \cup \mathcal{E}} \theta_u T^u$$

and any $K \in \mathcal{R}^+(\mathcal{V}, \mathcal{E})$ as ACA with

$$A = \sum_{u \in \mathcal{V}} \eta_u T^u, \quad C = I + \sum_{u \in \mathcal{E}} \delta_u T^u.$$

Gehrmann and Lauritzen (2012) now show that for a given colored graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ we have:

Lemma

The following are equivalent

$$K\Omega \subseteq \Omega \text{ for all } K \in \mathcal{S}^+(\mathcal{V}, \mathcal{E});$$
$$K\Omega \subseteq \Omega \text{ for all } K \in \mathcal{R}^+(\mathcal{V}, \mathcal{E});$$
$$T^u\Omega \subseteq \Omega \text{ for all } u \in \mathcal{V} \cup \mathcal{E}.$$

Thus, by Kruskal's theorem, we can check stability of mean spaces in both RCON and RCON models by checking stability under the action of the model generators $T^u, u \in \mathcal{V} \cup \mathcal{E}$.

We shall be particularly interested in mean spaces generated by a *partition* $\mathcal{M} = \{m\}$ of the vertex set V, so that for

$$\Omega = \Omega(\mathcal{M}) = \{\mu : \mu_\alpha = \mu_\beta \text{ whenever } \alpha, \beta \in m.\}.$$

It is straightforward to show (Gehrmann and Lauritzen, 2012) that

Proposition (Vertex stability)

The space $\Omega(\mathcal{M})$ is stable under $T^v, v \in \mathcal{V}$ if and only if the partition $\mathcal{M}$ is finer than $\mathcal{V}$.

The Behrens–Fisher problem represents a case where this condition is violated unless variances are assumed identical.

To discuss stability under edge colour classes we require the notion of an equitable partition.

A partition $\mathcal{M}$ of V is *equitable* w.r.t. a graph $G = (V, E)$ if for any $\alpha, \beta \in n \in \mathcal{M}$ it holds that

$$|\operatorname{ne}_E(\alpha) \cap m| = |\operatorname{ne}_E(\beta) \cap m| \text{ for all } m \in \mathcal{M}.$$

In words, any two vertices in the same partition set have the same number of neighbours in any other partition set. So in particular, all subgraphs induced by partition sets are regular graphs.

Vertex regular graphs

We say that a coloured graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is *vertex regular* if $\mathcal{V}$ is an equitable partition of the subgraph $G^e = (V, e)$ induced by the edge colour class e for all $e \in \mathcal{E}$.

It now follows easily and also from a result of Chan and Godsil (1997) that

Proposition (Edge stability)

The space $\Omega(\mathcal{M})$ is stable under $T^e, e \in \mathcal{V}$ if and only if the coloured graph $(\mathcal{M}, \mathcal{E})$ is vertex regular.

Combining the propositions with the lemma and Kruskal's theorem we find that for both RCON and RCOP models we have

Theorem

The LSE and MLE for μ under the assumption that $\mu \in \Omega(\mathcal{M})$ are identical if and only if both of the following hold:

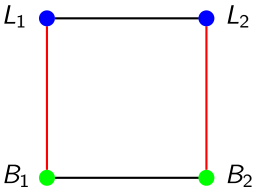
(i) *$\mathcal{M}$ is finer than $\mathcal{V}$;*

(ii) *The coloured graph $(\mathcal{M}, \mathcal{E})$ is vertex regular.*

If both the mean symmetry and concentration symmetry is determined by group invariance, we have $\mathcal{M} = \mathcal{V}$. Gehrmann and Lauritzen (2012) show that *if $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ represents and RCOP model, $\mathcal{G}$ is necessarily vertex regular.*

Hence *for RCOP models with natural mean restrictions we have $\mu^* = \hat{\mu}$.*

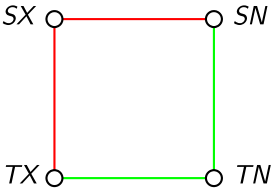
Frets' heads revisited



For the mean partition to be finer than the concentration partition we can either have different mean lengths, or different mean breadths, or both, or none of these.

For the mean partition to be vertex regular we need to have either both means identical or all means different. Thus there are two benign possibilities.

Anxiety and anger revisited



Here there are no benign mean hypotheses as the individual concentrations are all different.

Anderson, R. L. (1942). Distribution of the serial correlation coefficient. *The Annals of Mathematical Statistics*, 13:1–13.

Andersson, S. A. (1975). Invariant normal models. *The Annals of Statistics*, 3:132–154.

Andersson, S. A., Brøns, H., and Jensen, S. T. (1983). Distribution of eigenvalues in multivariate statistical analysis. *The Annals of Statistics*, 11:392–415.

Andersson, S. A. and Madsen, J. (1998). Symmetry and lattice conditional independence in a multivariate normal distribution. *The Annals of Statistics*, 26:525–572.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B*, 36:302–309.

Besag, J. E. and Moran, P. A. P. (1975). On the estimation and testing of spatial interaction in Gaussian lattice processes. *Biometrika*, 62:555–562.

Chan, A. and Godsil, C. (1997). Symmetry and eigenvectors. In Hahn, G. and Sabidussi, G., editors, *Graph Symmetry. Algebraic Methods and Applications*, volume 497 of *NATO ASI Series C: Mathematical and Physical Sciences*, pages 75–106. Kluwer Academic Publishers.

Cox, D. R. and Wermuth, N. (1993). Linear dependencies represented by chain graphs (with discussion). *Statistical Science*, 8:204–218; 247–277.

Drton, M. (2008). Multiple solutions to the likelihood equations in the Behrens–Fisher problem. *Statist. Probab. Lett.*, 78:3288–3293.

Frets, G. P. (1921). Heredity of head form in man. *Genetica*, 3:193–400.

Gehrmann, H. and Lauritzen, S. (2012). Estimation of means in graphical Gaussian models with symmetries. *The Annals of Statistics*, 40:To appear.

Hylleberg, B., Jensen, M., and Ørnbøl, E. (1993). Graphical symmetry models. Master's thesis, Aalborg University, Aalborg, Denmark.

Kruskal, W. (1968). When are Gauss–Markov and least squares estimators identical? A coordinate-free approach. *Ann. Math. Statist.*, 39:70–75.

Leipnik, R. B. (1947). Distribution of the serial correlation coefficent in a circularly correlated universe. *The Annals of Mathematical Statistics*, 18:80–87.

Madsen, J. (2000). Invariant normal models with recursive graphical Markov structure. *The Annals of Statistics*, 28(4):1150–1178.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. Academic Press.

Olkin, I. and Press, S. J. (1969). Testing and estimation for a circular stationary model. *The Annals of Mathematical Statistics*, 40:1358–1373.

Scheffé, H. (1944). A note on the Behrens–Fisher problem. *Ann. Math. Statist.*, 15:430–432.

Votaw, D. F. (1948). Testing compound symmetry in a normal multivariate distribution. *The Annals of Mathematical Statistics*, 19:447–473.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, 41:439–449.

Wilks, S. S. (1946). Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution. *The Annals of Mathematical Statistics*, 17:257–281.