Graphical Models with Symmetry

Steffen Lauritzen, University of Oxford

Wald Lecture, World Meeting on Probability and Statistics

Istanbul 2012

Sparse graphical models with few parameters can describe complex phenomena.

Introduce *symmetry to obtain further parsimony* so models can be well estimated when number of variables $|V|$ higher than number of observed units n, $n << |V|$.

Also, sometimes there are *natural and inherent symmetries* in problems under study, e.g. when these involve twins, measurements on right and left sides, dimensions of a starfish, etc.

Gaussian graphical models with symmetry

Several possible types of restriction:

- RCON restricts concentration matrix;

Gaussian graphical models with symmetry

Several possible types of restriction:

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

Gaussian graphical models with symmetry

Several possible types of restriction:

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

Gaussian graphical models with symmetry

Several possible types of restriction:

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

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

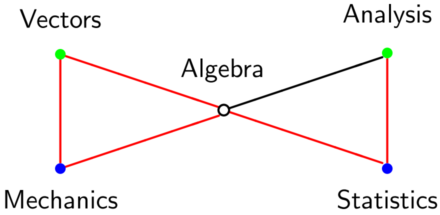
	Mechanics	Vectors	Algebra	Analysis	Statistics
Mechanics	5.24	−2.44	−2.74	0.01	−0.14
Vectors	−2.44	10.43	−4.71	−0.79	−0.17
Algebra	−2.74	−4.71	26.95	−7.05	−4.70
Analysis	0.01	−0.79	−7.05	9.88	−2.02
Statistics	−0.14	−0.17	−4.70	−2.02	6.45

.

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.

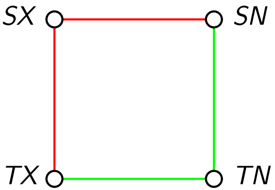
RCON model since restrictions apply to *concentration* matrix

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

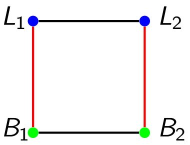
Data strongly support conditional independence model displayed below with *partial correlations* strikingly similar in pairs:



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.*

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, RCOV, and RCOR because all aspects of the joint distribution are unaltered when labels are switched.

- 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);

General combinations with conditional independence are more recent:

(Hylleberg et al., 1993; Andersson and Madsen, 1998; Madsen, 2000; Drton and Richardson, 2008; Højsgaard and Lauritzen, 2008; Gehrmann, 2011b; Gottard et al., 2011; Gehrmann, 2011a; Gehrmann and Lauritzen, 2012).

Although *literarure is steadily growing.*

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 $\mathcal{G} = (V, E)$ with Y as above and $K \in \mathcal{S}^+(\mathcal{G})$, the set of (symmetric) positive definite matrices with

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

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*

$$V = V_1 \cup \cdots \cup V_T, \quad E = E_1 \cup \cdots \cup E_S.$$

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

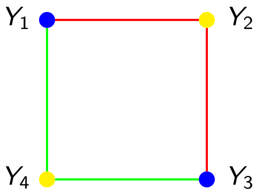
$\mathcal{E} = \{E_1, \ldots, E_S\}$ is an *edge colouring,*

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

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 T dimensional vector η and off-diagonal elements by an S 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

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

Likelihood function

Consider a sample $Y^1 = y^1, \ldots, Y^n = y^n$ of n observations of Y and let W denote the matrix of sums of squares and products

$$W = \sum_{\nu=1}^{n} Y^\nu (Y^\nu)^\top.$$

The log-likelihood function based on the sample is

$$\log L = \frac{n}{2} \log \det(K) - \frac{1}{2} \operatorname{tr}(KW) \tag{2}$$

Note that the restrictions defined are *linear in the concentration matrix K* so *RCON model is linear exponential model.*

Likelihood equations

For each vertex colour class $u \in \mathcal{V}$ let T^u be the $|V| \times |V|$ diagonal matrix with entries $T^u_{\alpha\alpha} = 1$ if $\alpha \in u$ and 0 otherwise.

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

$$T^{blue} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; T^{red} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Likelihood function then becomes

$$\log L(K) \quad = \quad \frac{n}{2}\log(\det K) - \sum_{u\in\mathcal{V}\cup\mathcal{E}} \theta_u \operatorname{tr}\{T^u W\}/2.$$

MLE is obtained by *equating canonical sufficient statistics to their expectation*, i.e.

$$\operatorname{tr}(T^u W) = n \operatorname{tr}(T^u \Sigma), \quad u \in \mathcal{V}\cup\mathcal{E}, \tag{3}$$

provided such a solution exists.

Fitted concentrations ($\times 1000$) for the examination marks in five mathematical subjects assuming the RCON model displayed.

	Mechanics	Vectors	Algebra	Analysis	Statistics
Mechanics	6.30	−3.38	−3.38	0	0
Vectors	−3.38	10.29	−3.38	0	0
Algebra	−3.38	−3.38	24.21	−6.65	−3.38
Analysis	0	0	−6.65	10.29	−3.38
Statistics	0	0	−3.38	−3.38	6.30

The model displayed earlier yields an excellent fit with a likelihood ratio of $-2 \log \mathrm{LR} = 7.2$ on 7 degrees of freedom, when compared to the butterfly model without symmetry restrictions

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

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

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

Likelihood equations

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

RCOR models are curved exponential families.

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

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

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

Differentiation yields the likelihood equations

$$\operatorname{tr}(K^u AWA) = f\operatorname{tr}(K^u C^{-1}), u \in \mathcal{E};\ \operatorname{tr}(K^u ACAW) = f\operatorname{tr}(K^u), u \in \mathcal{V}.$$

MLE is not necessarily unique.

If the MLE is unique, *an alternating algorithm converges to the MLE,* alternating between maximizing in λ for fixed δ and conversely.

Anxiety and anger

Fitted concentrations ($\times 100$) (on and above diagonal) and partial correlations (below diagonal) for RCOR model:

	SX	SN	TX	TN
SX (State anxiety)	0.59	-0.31	-0.22	0
SN (State anger)	0.46	0.78	0	-0.15
TX (Trait anxiety)	0.46	0	0.40	-0.10
TN (Trait anger)	0	0.31	0.31	0.28

Fitting the RCOR model yields likelihood ratio $-2 \log \text{LR} = 0.22$ on 2 d.o.f. comparing with the model without symmetry.

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

$$G\Sigma G^\top = \Sigma \text{ for all } G \in \Gamma. \tag{4}$$

Since G satisfies $G^{-1} = G^\top$, (4) is equivalent to

$$G\Sigma = \Sigma G \text{ for all } G \in \Gamma, \tag{5}$$

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

$$GK = KG.$$

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

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

Identifying the graph colouring

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.

Frets' heads

Observed concentrations ($\times 100$) (on and above diagonal) together with fitted concentrations for RCOP model.

	$L1$	$B1$	$L2$	$B2$
$L1$ (Length of head of first son)	3.21	-1.16	-0.78	-1.11
$B1$ (Breadth of head of first son)	-1.71	2.21	-0.50	0.48
$L2$ (Length of head of second son)	-1.42	0	2.67	-1.89
$B2$ (Breadth of head of second son)	0	-1.83	-1.71	3.37
Fitted concentrations	2.89	2.44	2.89	2.44

The likelihood ratio comparing to model without symmetries is equal to $-2 \log \mathrm{LR} = 5.18$ on 5 degrees of freedom.

Any RCOP model is automatically also RCON and RCOR whereas the converse is false.

If $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a coloured graph we say that $(\mathcal{V}, \mathcal{E})$ is *edge regular* if *any pair of edges in the same colour class in $\mathcal{E}$ connects the same vertex colour classes.*

It then holds (Højsgaard and Lauritzen, 2008) that

The RCON and RCOR models determined by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ yield identical restrictions, i.e.

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

if and only if $(\mathcal{V}, \mathcal{E})$ is edge regular.

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.

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 $e \in \mathcal{E}$.

A graph colouring that is both vertex regular and edge regular is *regular* (Siemons, 1983).

The graph of an RCOP model is always regular

Not all regular colourings correspond to permutation symmetry.
Gehrmann (2011a) uses `nauty` to calculate automorphism group of coloured graph and check that it acts transively on colour classes.

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 $\mathcal{R}^V$.

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

$$L(\mu, K) \propto \det K^{n/2} \exp^{-\sum_{1 \le i \le n}(y^i - \mu)^T K (y^i - \mu)/2}. \tag{6}$$

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

$$L(\hat{\mu}, K; y) \propto \det K^{n/2} \exp\{-\operatorname{tr}(KW)/2\}, \tag{7}$$

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

$$Y_1 \qquad Y_2$$

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 (6) 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

$$K\Omega \subseteq \Omega. \tag{8}$$

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 (7)

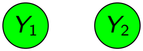
$$L(\hat{\mu}, K) \propto \det K^{n/2} \exp\{-\operatorname{tr}(KW)/2\}.$$

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.

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

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

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

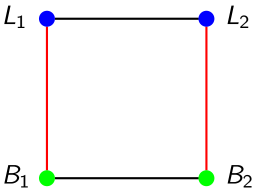
The space $\Omega(\mathcal{M})$ is stable under $T^v, v \in \mathcal{V}$ if and only if the partition $\mathcal{M}$ is finer than $\mathcal{V}$.

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

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.

Note in particular that for an RCOP model generated by permutation symmetry, we have $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is regular and hence a mean partition $\mathcal{M}$ is stable if and only if it is finer than the vertex orbit partition $\mathcal{V}$.

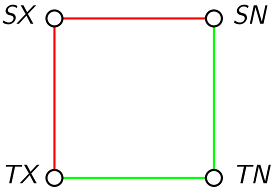
Frets' heads revisited



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.

- Basic theory described in Højsgaard and Lauritzen (2008) and Gehrmann (2011b);
- Clarify conditions for existence of MLE
- Extension to discrete symmetry models should be developed based on graphical log-linear models, extending classic models of symmetry, marginal homogeneity and quasi-symmetry. Some significant progress made by Gottard et al. (2011);
- Application to gene expression data? Efficient and principled methods for model selection and structure estimation needed.

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.

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.

Drton, M. and Richardson, T. S. (2008). Binary models for marginal independence. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 41(2):287–309.

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

Gehrmann, H. (2011a). *Graphical Gaussian Models with*

Symmetries. PhD thesis, Department of Statistics, University of Oxford.

Gehrmann, H. (2011b). Lattices of graphical Gaussian models with symmetries. *Symmetry*, 3(3):653–679.

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

Gottard, A., Marchetti, G. M., and Agresti, A. (2011). Quasi-symmetric graphical log-linear models. *Scandinavian Journal of Statistics*, 38(3):447–465.

Højsgaard, S. and Lauritzen, S. L. (2007). Inference in graphical Gaussian models with edge and vertex symmetries with the gRc package for R. *Journal of Statistical Software*, 23.

Højsgaard, S. and Lauritzen, S. L. (2008). Graphical Gaussian models with edge and vertex symmetries. *Journal of the Royal Statistical Society, Series B*, 68:in press.

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, London.

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.

Siemons, J. (1983). Automorphism groups of graphs. *Arch. Mathematics*, 41:379–384.

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.