Tree-structured covariance matrices

Peter McCullagh

Department of Statistics
University of Chicago

UIC, September 2006

Outline

Structured covariance matrices

q response variables $Y^1, \ldots, Y^q$ observed on n units
$Y_i = (Y_i^1, \ldots, Y_i^q)$ in $\mathcal{R}^q$ observed for unit i

$\Sigma = \mathrm{cov}(Y_i)$ a matrix of order q

Model: $\Sigma \in \Theta_q \subset \mathcal{PD}_q$
SCM: $\Theta = \{\Theta_q \subset \mathcal{PD}_q\}$ for $q = 0, 1, 2, \ldots$

(i) closed under permutation
(ii) closed under deletion $\Theta_{q+1} \to \Theta_q$

Examples:
(i) $\Theta_q = \{\sigma^2 I_q : \sigma > 0\}$
(ii) $\Theta_q = \{\sigma_1^2 I_q + \sigma_2^2 J_q : \sigma_1, \sigma_2 > 0\}$
(iii) $\Theta_q = \mathcal{PD}_q$

Ehrenberg's viewer preference data

Table 1a. Viewer preference correlations for 10 programmes

PrB	1.0000	0.1064	0.0653	0.5054	0.4741	0.0915	0.4732	0.1681	0.3091	0.1242
ThW	0.1064	1.0000	0.2701	0.1424	0.1321	0.1885	0.0815	0.3520	0.0637	0.3946
Tod	0.0653	0.2701	1.0000	0.0926	0.0704	0.1546	0.0392	0.2004	0.0512	0.2437
WoS	0.5054	0.1474	0.0926	1.0000	0.6217	0.0785	0.5806	0.1867	0.2963	0.1403
GrS	0.4741	0.1321	0.0704	0.6217	1.0000	0.0849	0.5932	0.1813	0.3412	0.1420
LnU	0.0915	0.1885	0.1546	0.0785	0.0849	1.0000	0.0487	0.1973	0.0969	0.2661
MoD	0.4732	0.0815	0.0392	0.5806	0.5932	0.0487	1.0000	0.1314	0.3267	0.1221
Pan	0.1681	0.3520	0.2004	0.1867	0.1813	0.1973	0.1314	1.0000	0.1469	0.5237
RgS	0.3091	0.0637	0.0512	0.2963	0.3412	0.0969	0.3261	0.1469	1.0000	0.1212
24H	0.1242	0.3946	0.2432	0.1403	0.1420	0.2661	0.1211	0.5237	0.1212	1.0000

Table 1b. Viewer preference correlations reordered

WoS	1.000	0.581	0.622	0.505	0.296	0.140	0.187	0.145	0.093	0.078
MoD	0.581	1.000	0.593	0.473	0.326	0.121	0.131	0.082	0.039	0.049
GrS	0.622	0.593	1.000	0.474	0.341	0.142	0.181	0.132	0.070	0.085
PrB	0.505	0.473	0.474	1.000	0.309	0.124	0.168	0.106	0.065	0.092
RgS	0.296	0.327	0.341	0.309	1.000	0.121	0.147	0.064	0.051	0.097
24H	0.140	0.122	0.142	0.124	0.121	1.000	0.524	0.395	0.243	0.266
Pan	0.187	0.131	0.181	0.168	0.147	0.524	1.000	0.352	0.200	0.197
ThW	0.145	0.082	0.132	0.106	0.064	0.395	0.352	1.000	0.270	0.188
ToD	0.093	0.039	0.070	0.065	0.051	0.243	0.200	0.270	1.000	0.155
LnU	0.078	0.049	0.085	0.092	0.097	0.266	0.197	0.188	0.155	1.000

Structure: Markovian

Graphical representation: $Y^1 \to Y^2 \to Y^3$ (or reverse arrows)
Conditional independence: $Y^1 \perp\!\!\!\perp Y^3 \mid Y^2$

$$\Sigma = \begin{pmatrix} \Sigma^{11} & \Sigma^{12} & \Sigma^{13} \\ \Sigma^{21} & \Sigma^{22} & \Sigma^{23} \\ \Sigma^{31} & \Sigma^{32} & \Sigma^{33} \end{pmatrix}$$

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{pmatrix}$$

C.I. implies $\Sigma_{13} = 0$ in inverse covariance matrix.

BIG literature on graphical models.....

Structure: additive common factors

Suppose $q = 3$

$$\begin{pmatrix} Y^1 \\ Y^2 \\ Y^3 \end{pmatrix} = \begin{pmatrix} X^0 \\ X^0 \\ X^0 \end{pmatrix} + \begin{pmatrix} X^1 \\ X^1 \\ X^2 \end{pmatrix} + \begin{pmatrix} \epsilon^1 \\ \epsilon^2 \\ \epsilon^3 \end{pmatrix}$$

associated with decreasing partition seq 123, 12|3, 1|2|3
X^0 common to all components
X^1 additional component common to 1&2
All components independent.

$$\begin{pmatrix} Y^1 \\ Y^2 \\ Y^3 \\ Y^4 \end{pmatrix} = \begin{pmatrix} X^0 \\ X^0 \\ X^0 \\ X^0 \end{pmatrix} + \begin{pmatrix} X^1 \\ X^2 \\ X^1 \\ X^1 \end{pmatrix} + \begin{pmatrix} X^3 \\ X^4 \\ X^5 \\ X^3 \end{pmatrix} + \begin{pmatrix} \epsilon^1 \\ \epsilon^2 \\ \epsilon^3 \\ \epsilon^4 \end{pmatrix}$$

partition seq 1234, 134|2, 13|4|2, 1|2|3|4

Consequence of additive common factors

$$\begin{pmatrix} Y^1 \\ Y^2 \\ Y^3 \end{pmatrix} = \begin{pmatrix} X^0 \\ X^0 \\ X^0 \end{pmatrix} + \begin{pmatrix} X^1 \\ X^1 \\ X^2 \end{pmatrix} + \begin{pmatrix} \epsilon^1 \\ \epsilon^2 \\ \epsilon^3 \end{pmatrix}$$

$$\begin{aligned}
\Sigma^{11} &= \mathrm{var}(X^0) + \mathrm{var}(X^1) + \mathrm{var}(\epsilon^1) \\
\Sigma^{22} &= \mathrm{var}(X^0) + \mathrm{var}(X^1) + \mathrm{var}(\epsilon^2) \\
\Sigma^{11} &= \mathrm{var}(X^0) + \mathrm{var}(X^2) + \mathrm{var}(\epsilon^3) \\
\Sigma^{12} &= \mathrm{var}(X^0) + \mathrm{var}(X^1) \\
\Sigma^{13} &= \mathrm{var}(X^0) \\
\Sigma^{23} &= \mathrm{var}(X^0)
\end{aligned}$$

Consequences:

(i) $\Sigma^{ij} \geq 0$

(ii) $\Sigma^{ij} \geq \min(\Sigma^{ik}, \Sigma^{jk})$

$\{\Sigma^{ij}, \Sigma^{ik}, \Sigma^{jk}\}$ contains a duplicate minimum

Similarity matrices

Defn: S is a similarity matrix if

(i) S is symmetric and non-negative

(ii) $S_{ij} \geq \min\{S_{ik}, S_{jk}\}$ for all i, j, k

Consequences:

(a) $S_{ii} \geq S_{ij}$: max occurs on diagonal

(b) S is positive semi-definite

(c) closed under permutation $S \mapsto \sigma^{-1} S \sigma$

(d) closed under deletion $\Theta_{q+1} \to \Theta_q$

(e) closed under monotone transformation $S_{ij} \mapsto g(S_{ij})$

(f) commutative semi-group under $\wedge$: $\Theta_q \times \Theta_q \to \Theta_q$

(g) $\wedge$ commutes with deletion

Similarity matrices (contd)

Semi-group property:
$S, S' \in \Theta_q$ and $T_{ij} = \min(S_{ij}, S'_{ij})$
$T = S \wedge S'$ is in Θ_q

$\wedge$ commutes with deletion (obvious)

Topological structure of $\Theta_q \in \mathcal{PD}_q$:
(a) not a manifold
(b) is a finite union of manifolds (ribs) of dim $2q - 1$
(c) ribs intersect on the spine of dimension $q + 1$

Non-standard sort of parameter space
No universal coordinate system

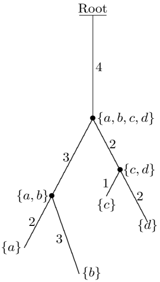
Varieties of mathematical trees

Distinctions: Rooted versus unrooted
Measured edges versus unmeasured (Boolean) edges
Leaf-labelled versus node-labelled versus unlabelled
Coalescent trees versus fragmentation trees
Binary versus non-binary (splits or coalescences)

Examples of trees (diagrams):

Matrix representation of a leaf-labelled tree

$T =$



$$T = \begin{pmatrix} 9 & 7 & 4 & 4 \\ 7 & 10 & 4 & 4 \\ 4 & 4 & 7 & 6 \\ 4 & 4 & 6 & 8 \end{pmatrix}$$

T_{ij} is the first splitting time;

—the distance from the root to first node at which i, j occur on separate branches.

Consequently $T_{ij} \geq \min\{T_{ik}, T_{jk}\}$.

Unrooted version: $D_{ij} = T_{ii} + T_{jj} - 2T_{ij}$

$$D = \begin{pmatrix} 0 & 5 & 8 & 9 \\ 5 & 0 & 9 & 10 \\ 8 & 9 & 0 & 3 \\ 9 & 10 & 3 & 0 \end{pmatrix}$$

Unrooted trees: $D_{ij} > 0$, $D_{ii} = 0$ and

Matrix representation of coalescent trees

Contrast rooted with coalescent trees

$T_{ij} =$ time to the most recent common ancestor
time measured backwards from present

Implies $T_{ii} = 0$ and $T_{ij} \le \max\{T_{ik}, T_{jk}\}$

T is a distance function and
$\{T_{12}, T_{13}, T_{23}\}$ contains a duplicate max pair

Each triangle is isoceles and acute

Rooted trees and similarity matrices

Observe

(i) set of rooted trees = set of similarity matrices
(ii) similarity matrix has graphical representation
(iii) leaf permutation = conjugation
$T \mapsto \sigma^{-1} T \sigma$
(iv) leaf deletion = column deletion
$\mathcal{RT}_{n+1} \to \mathcal{RT}_n$

Minimax projection

A is a symmetric matrix
$A^+ = \max(A, 0)$ is a non-negative symmetric matrix

Minimax projection P

$$(PA)_{ij} = \min_{I,J} \max_{r\in I, s\in J} A^+_{rs}$$

such that $i \in I$ and $j \in J$.

Furthermore

$$P^2A = PA \quad \text{and} \qquad A = PA \text{ implies } A \in \mathcal{RT}$$

PA is not a good approx to A but ...

Maximum likelihood for Ehrenberg's data

R command: `round(TreeFit(corr)[perm, perm], 2)`
Or: `round(TreeFit(corr, sph=1)[perm, perm], 2)`

Table 1c. Fitted viewer preference correlations (reordered)

WoS	0.99	0.59	0.61	0.48	0.32	0.10	0.10	0.10	0.10	0.1
MoD	0.59	1.01	0.59	0.48	0.32	0.10	0.10	0.10	0.10	0.1
GrS	0.61	0.59	0.99	0.48	0.32	0.10	0.10	0.10	0.10	0.1
PrB	0.48	0.48	0.48	1.00	0.32	0.10	0.10	0.10	0.10	0.1
RgS	0.32	0.32	0.32	0.32	1.00	0.10	0.10	0.10	0.10	0.1
24H	0.10	0.10	0.10	0.10	0.10	0.96	0.51	0.36	0.25	0.2
Pan	0.10	0.10	0.10	0.10	0.10	0.51	1.01	0.36	0.25	0.2
ThW	0.10	0.10	0.10	0.10	0.10	0.36	0.36	0.99	0.25	0.2
ToD	0.10	0.10	0.10	0.10	0.10	0.25	0.25	0.25	1.03	0.2
LnU	0.10	0.10	0.10	0.10	0.10	0.20	0.20	0.20	0.20	1.0

Graphical representation:

Statistical issues

(1) Computation of mle
– finding the right foliation of $\mathcal{RT}_q$
– not a manifold so no universal coordinate system

(2) Distribution theory
– for estimates
– for likelihood ratio statistic
– Is the fitted tree adequate?

(3) Graphical representation of $\hat{\Sigma}$
– finding a good permutation of variables

(4) Bayesian methods and prior distributions

Exchangeable random trees

Need, for each integer $n \geq 1$ a prob distn P_n on $\mathcal{RT}_n$
Such that

(i) P_n invariant under permutation
(ii) P_n is marginal distribution of P_{n+1}

Kolmogorov extension implies an infinite exch tree

Do they exist?
How to construct?
How to specify distributions?
References: Kingman, Aldous, Bertoin, Pitman

Example of inf exch frag tree

Described by its finite restrictions to $[n] = \{1, \ldots, n\}$.

(1) Begin at $t = 0$ with intact set $[n]$.
Survival time to fragmentation $T_n \sim exp(\varphi_n)$
$\varphi_n = sum_{j=1}^{n-1} 1/j$

(2) Fragmentation distribution: binary, sizes $r, n - r$

$$p_n(r, n - r) = \frac{1}{2\varphi_n r(n - r)}$$

choose SRS of size r and complementary SRS

(3) Repeat recursively on each branch Homogeneous

(self-similar) Markov fragmentation tree
Can compute height distribution of leaves,....
Height $\sim \log(n)$.

Infinite divisibility

With respect to semi-group operation $\wedge$
Q is prob distn on $\mathcal{RT}_n$
$T_1, \ldots, T_k$ are iid with distn Q
$T_1 \wedge T_2 \wedge \cdots \wedge T_k$ has distn $Q^{\star k}$

Definition: P is inf div if for each $k \geq 1$
there exists Q such that $P = Q^{\star k}$.

Similar definition for random partitions

Observation:
Ewens distribution is not i.d.
But there exist i.d. random partitions
and i.d. random trees.

Lévy fragmentations

Construction is same as for addition of real r.v.s
Replace addition by $\wedge$

Process is non-increasing, and decreases to 0.