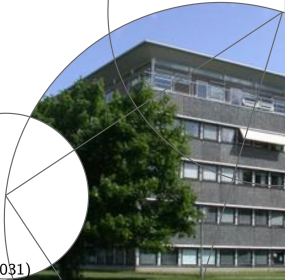
Maximum likelihood estimation of totally positive Gaussian distributions.

Steffen Lauritzen[1]
Department of Mathematical Sciences

Mathematical Methods of Modern Statistics, Luminy 2017



[1] based on joint work with Caroline Uhler and Piotr Zwiernik (arXiv:1702.04031)

Positive dependence and Simpson's paradox

Two real-valued random variables X and Y are *positively associated* if if $\text{Cov}\{f(X), g(Y)\} \geq 0$ for all f, g which are non-decreasing.

Positive dependence and Simpson's paradox

Two real-valued random variables X and Y are *positively associated* if if $\mathrm{Cov}\{f(X), g(Y)\} \geq 0$ for all f, g which are non-decreasing.

The Yule-Simpson paradox says that we may have X and Y positively associated but X and Y negatively associated conditionally on a third variable Z.

Positive dependence and Simpson's paradox

Two real-valued random variables X and Y are *positively associated* if if $\mathrm{Cov}\{f(X), g(Y)\} \geq 0$ for all f, g which are non-decreasing.

The Yule-Simpson paradox says that we may have X and Y positively associated but X and Y negatively associated conditionally on a third variable Z.

Multivariate total positivity (MTP_2) ensures this not to happen: *associations can never change sign due to changes of context.*

Multivariate total positivity for functions

Let f be a function $f : \mathcal{X} = \times_{v \in V} \mathcal{X}_v \to \mathbb{R}$ where $\mathcal{X}_v$ are either discrete or open subsets of $\mathbb{R}$.

Multivariate total positivity for functions

Let f be a function $f : \mathcal{X} = \times_{v \in V} \mathcal{X}_v \rightarrow \mathbb{R}$ where $\mathcal{X}_v$ are either discrete or open subsets of $\mathbb{R}$.

Definition

f is *multivariate totally positive of order two* (MTP_2) if

$$f(x)f(y) \leq f(x \wedge y)f(x \vee y) \text{ for all } x, y \in \mathcal{X}.$$

Here $\wedge$ and $\vee$ should be applied coordinatewise.

Multivariate total positivity for functions

Let f be a function $f : \mathcal{X} = \times_{v \in V} \mathcal{X}_v \to \mathbb{R}$ where $\mathcal{X}_v$ are either discrete or open subsets of $\mathbb{R}$.

Definition

f is *multivariate totally positive of order two* (MTP_2) if

$$f(x)f(y) \leq f(x \wedge y)f(x \vee y) \text{ for all } x, y \in \mathcal{X}.$$

Here $\wedge$ and $\vee$ should be applied coordinatewise.

In the bivariate case, this property is known simply as *total positivity* or TP_2 (Karlin and Rinott, 1980).

Example

For $d = 2$, $x_1 \leq x_2, y_1 \leq y_2$ the condition for MTP_2 simply becomes

$$f(x_1, y_2)f(x_2, y_1) \leq f(x_1, y_1)f(x_2, y_2),$$

Example

For $d = 2$, $x_1 \leq x_2, y_1 \leq y_2$ the condition for MTP_2 simply becomes

$$f(x_1, y_2)f(x_2, y_1) \leq f(x_1, y_1)f(x_2, y_2),$$

or, alternatively

$$\det \left\{ \begin{array}{cc} f(x_1, y_1) & f(x_1, y_2) \\ f(x_2, y_1) & f(x_2, y_2) \end{array} \right\} \geq 0.$$

Multivariate total positivity for distributions

For $\mathcal{X} = \times_{v \in V} \mathcal{X}_v$ as before we adopt a *standard base measure* $\mu = \otimes_{v \in V} \mu_v$ where μ_v is counting measure if $\mathcal{X}_v$ is discrete and Lebesgue measure if $\mathcal{X}_v$ is an open subset of $\mathbb{R}$.

Multivariate total positivity for distributions

For $\mathcal{X} = \times_{v \in V} \mathcal{X}_v$ as before we adopt a *standard base measure* $\mu = \otimes_{v \in V} \mu_v$ where μ_v is counting measure if $\mathcal{X}_v$ is discrete and Lebesgue measure if $\mathcal{X}_v$ is an open subset of $\mathbb{R}$.

We then define

Definition

A distribution P is said to be *multivariate totally positive of order two* (MTP_2) if its density w.r.t. the standard base measure μ is MTP_2.

Multivariate total positivity for distributions

For $\mathcal{X} = \times_{v\in V}\mathcal{X}_v$ as before we adopt a *standard base measure* $\mu = \otimes_{v\in V}\mu_v$ where μ_v is counting measure if $\mathcal{X}_v$ is discrete and Lebesgue measure if $\mathcal{X}_v$ is an open subset of $\mathbb{R}$.

We then define

Definition

A distribution P is said to be *multivariate totally positive of order two* (MTP_2) if its density w.r.t. the standard base measure μ is MTP_2.

Introduced and studied by Karlin and Rinott (1980) using results (FKG inequality) from fundamental paper by Fortuin et al. (1971).

Examples of MTP_2 distributions

Mostly from Karlin and Rinott (1980):

Examples of MTP_2 distributions

Mostly from Karlin and Rinott (1980):

- Characteristic roots of a Wishart matrix W, or of $W_1 W_2^{-1}$, or $W_1(W_1 + W_2)^{-1}$, where $W_1 \perp\!\!\!\perp W_2$ (Dykstra and Hewett, 1978);

Examples of MTP_2 distributions

Mostly from Karlin and Rinott (1980):

- Characteristic roots of a Wishart matrix W, or of $W_1 W_2^{-1}$, or $W_1(W_1 + W_2)^{-1}$, where $W_1 \perp\!\!\!\perp W_2$ (Dykstra and Hewett, 1978);
- Ferromagnetic (attractive) Ising models (Lebowitz, 1972);

Examples of MTP_2 distributions

Mostly from Karlin and Rinott (1980):

- Characteristic roots of a Wishart matrix W, or of $W_1 W_2^{-1}$, or $W_1(W_1 + W_2)^{-1}$, where $W_1 \perp\!\!\!\perp W_2$ (Dykstra and Hewett, 1978);
- Ferromagnetic (attractive) Ising models (Lebowitz, 1972);
- Multivariate logistic density (Gumbel, 1961);

Examples of MTP_2 distributions

Mostly from Karlin and Rinott (1980):

- Characteristic roots of a Wishart matrix W, or of $W_1 W_2^{-1}$, or $W_1(W_1 + W_2)^{-1}$, where $W_1 \perp\!\!\!\perp W_2$ (Dykstra and Hewett, 1978);
- Ferromagnetic (attractive) Ising models (Lebowitz, 1972);
- Multivariate logistic density (Gumbel, 1961);
- Gaussian free fields (random height landscapes) (Dynkin, 1980);

Examples of MTP_2 distributions

Mostly from Karlin and Rinott (1980):

- Characteristic roots of a Wishart matrix W, or of $W_1 W_2^{-1}$, or $W_1(W_1 + W_2)^{-1}$, where $W_1 \perp\!\!\!\perp W_2$ (Dykstra and Hewett, 1978);
- Ferromagnetic (attractive) Ising models (Lebowitz, 1972);
- Multivariate logistic density (Gumbel, 1961);
- Gaussian free fields (random height landscapes) (Dynkin, 1980);
- Markov chains with TP_2 transition densities;

Examples of MTP_2 distributions

Mostly from Karlin and Rinott (1980):

- Characteristic roots of a Wishart matrix W, or of $W_1 W_2^{-1}$, or $W_1(W_1 + W_2)^{-1}$, where $W_1 \perp\!\!\!\perp W_2$ (Dykstra and Hewett, 1978);
- Ferromagnetic (attractive) Ising models (Lebowitz, 1972);
- Multivariate logistic density (Gumbel, 1961);
- Gaussian free fields (random height landscapes) (Dynkin, 1980);
- Markov chains with TP_2 transition densities;
- Order statistics $(X_{(1)}, \dots, X_{(n)})$ if $X_1 \dots, X_n$ are i.i.d. with density f;

Examples of MTP_2 distributions

Mostly from Karlin and Rinott (1980):

- Characteristic roots of a Wishart matrix W, or of $W_1 W_2^{-1}$, or $W_1(W_1+W_2)^{-1}$, where $W_1 \perp\!\!\!\perp W_2$ (Dykstra and Hewett, 1978);
- Ferromagnetic (attractive) Ising models (Lebowitz, 1972);
- Multivariate logistic density (Gumbel, 1961);
- Gaussian free fields (random height landscapes) (Dynkin, 1980);
- Markov chains with TP_2 transition densities;
- Order statistics $(X_{(1)}, \ldots, X_{(n)})$ if $X_1 \ldots, X_n$ are i.i.d. with density f;
- Gaussian latent tree models as in phylogenetics (Zwiernik, 2015);

Examples of MTP_2 distributions

Mostly from Karlin and Rinott (1980):

- Characteristic roots of a Wishart matrix W, or of $W_1 W_2^{-1}$, or $W_1(W_1 + W_2)^{-1}$, where $W_1 \perp\!\!\!\perp W_2$ (Dykstra and Hewett, 1978);
- Ferromagnetic (attractive) Ising models (Lebowitz, 1972);
- Multivariate logistic density (Gumbel, 1961);
- Gaussian free fields (random height landscapes) (Dynkin, 1980);
- Markov chains with TP_2 transition densities;
- Order statistics $(X_{(1)}, \ldots, X_{(n)})$ if $X_1 \ldots, X_n$ are i.i.d. with density f;
- Gaussian latent tree models as in phylogenetics (Zwiernik, 2015);
- Many other examples...

Fundamental properties

A wealth of probability inequalities are satisfied for MTP_2 distributions (Karlin and Rinott, 1980).

Fundamental properties

A wealth of probability inequalities are satisfied for MTP_2 distributions (Karlin and Rinott, 1980). Also

Proposition

Assume X is MTP_2. Then

Fundamental properties

A wealth of probability inequalities are satisfied for MTP_2 distributions (Karlin and Rinott, 1980). Also

Proposition

Assume X is MTP_2. Then

- *If $A \subseteq V$, then the marginal $X_A = (X_v)_{v \in A}$ is MTP_2;*

Fundamental properties

A wealth of probability inequalities are satisfied for MTP_2 distributions (Karlin and Rinott, 1980). Also

Proposition

Assume X is MTP_2. Then

- *If $A \subseteq V$, then the marginal $X_A = (X_v)_{v \in A}$ is MTP_2;*
- *If $C \subseteq V$ then the conditional distribution $\mathcal{L}(X_{V \setminus C} \mid X_C = x^*_C)$ is MTP_2 for almost all $x^*_C \in \mathcal{X}_C$;*

Fundamental properties

A wealth of probability inequalities are satisfied for MTP_2 distributions (Karlin and Rinott, 1980). Also

Proposition

Assume X is MTP_2. Then

- *If $A \subseteq V$, then the marginal $X_A = (X_v)_{v \in A}$ is MTP_2;*
- *If $C \subseteq V$ then the conditional distribution $\mathcal{L}(X_{V \setminus C} \mid X_C = x_C^*)$ is MTP_2 for almost all $x_C^* \in \mathcal{X}_C$;*
- *If $\phi = (\phi_v)_{v \in V}$ are non-decreasing, then $Y = \phi(X)$ is MTP_2.*

Positive association and MTP_2

Proposition

If X is MTP_2 and f and g are non-decreasing in each of its arguments, then X is positively associated

$$\mathrm{Cov}\{f(X), g(X)\} \geq 0.$$

Positive association and MTP_2

Proposition

If X is MTP_2 and f and g are non-decreasing in each of its arguments, then X is positively associated

$$\mathrm{Cov}\{f(X), g(X)\} \geq 0.$$

Proof.

Discrete case by Fortuin et al. (1971). General case by Sarkar (1969). □

Covariance and independence

Proposition

If X is positively associated and $A, B \subseteq V$ are disjoint, then

$$X_A \perp\!\!\!\perp X_B \iff \operatorname{Cov}(X_u, X_v) = 0 \text{ for all } u \in A, v \in B.$$

Covariance and independence

Proposition

If X is positively associated and $A, B \subseteq V$ are disjoint, then

$$X_A \perp\!\!\!\perp X_B \iff \mathrm{Cov}(X_u, X_v) = 0 \text{ for all } u \in A, v \in B.$$

Proof.

Shown in Lebowitz (1972). □

Covariance and independence

Proposition

If X is positively associated and $A, B \subseteq V$ are disjoint, then

$$X_A \perp\!\!\!\perp X_B \iff \mathrm{Cov}(X_u, X_v) = 0 \text{ for all } u \in A, v \in B.$$

Proof.

Shown in Lebowitz (1972). □

Such a result is usually special for the Gaussian distribution.

Covariance and independence

Proposition

If X is positively associated and $A, B \subseteq V$ are disjoint, then

$$X_A \perp\!\!\!\perp X_B \iff \mathrm{Cov}(X_u, X_v) = 0 \text{ for all } u \in A, v \in B.$$

Proof.

Shown in Lebowitz (1972). □

Such a result is usually special for the Gaussian distribution.

So learning MTP_2 structure may be based on correlation analysis.

Markov properties

Let P be a probability distribution on $\mathcal{X} = \times_{v \in V} \mathcal{X}_v$. The *pairwise independence graph* $\mathcal{G}(P) = (V, E)$ is defined through the relation

$$uv \notin E \iff u \perp\!\!\!\perp_P v \mid V \setminus \{u, v\}.$$

Markov properties

Let P be a probability distribution on $\mathcal{X} = \times_{v \in V} \mathcal{X}_v$. The *pairwise independence graph* $\mathcal{G}(P) = (V, E)$ is defined through the relation

$$uv \notin E \iff u \perp\!\!\!\perp_P v \mid V \setminus \{u, v\}.$$

In other words, $\mathcal{G}(P)$ is the smallest graph $\mathcal{G}$ such that P is pairwise Markov w.r.t. $\mathcal{G}$.

Markov properties

Let P be a probability distribution on $\mathcal{X} = \times_{v \in V} \mathcal{X}_v$. The *pairwise independence graph* $\mathcal{G}(P) = (V, E)$ is defined through the relation

$$uv \notin E \iff u \perp\!\!\!\perp_P v \mid V \setminus \{u, v\}.$$

In other words, $\mathcal{G}(P)$ is the smallest graph $\mathcal{G}$ such that P is pairwise Markov w.r.t. $\mathcal{G}$.

We say that P is *globally Markov* w.r.t. a graph $\mathcal{G}$ if

$$A \perp_{\mathcal{G}} B \mid S \implies A \perp\!\!\!\perp_P B \mid S$$

where $\perp_{\mathcal{G}}$ is separation in the graph $\mathcal{G}$.

Markov properties

Let P be a probability distribution on $\mathcal{X} = \times_{v \in V} \mathcal{X}_v$. The *pairwise independence graph* $\mathcal{G}(P) = (V, E)$ is defined through the relation

$$uv \notin E \iff u \perp\!\!\!\perp_P v \mid V \setminus \{u, v\}.$$

In other words, $\mathcal{G}(P)$ is the smallest graph $\mathcal{G}$ such that P is pairwise Markov w.r.t. $\mathcal{G}$.

We say that P is *globally Markov* w.r.t. a graph $\mathcal{G}$ if

$$A \perp_{\mathcal{G}} B \mid S \implies A \perp\!\!\!\perp_P B \mid S$$

where $\perp_{\mathcal{G}}$ is separation in the graph $\mathcal{G}$.

Further, we say that P is *faithful* to $\mathcal{G}$ if

$$A \perp_{\mathcal{G}} B \mid S \iff A \perp\!\!\!\perp_P B \mid S$$

i.e. if the independence models $\perp\!\!\!\perp_P$ and $\perp_{\mathcal{G}}$ are identical.

A main result

Theorem (Fallat et al. (2017))

Assume the distribution P of X is MTP_2 with strictly positive density $f > 0$. Then P is faithful to $\mathcal{G}(P)$.

A main result

Theorem (Fallat et al. (2017))

Assume the distribution P of X is MTP_2 with strictly positive density $f > 0$. Then P is faithful to $\mathcal{G}(P)$.

In other words, *for MTP_2 distributions, the pairwise independence graph yields a complete 'picture'* of the independence relations in P.

Multivariate Gaussian MTP_2 distributions

Proposition

Let $X \sim \mathcal{N}_V(0, \Sigma)$. Then X is MTP_2 if and only if $K = \Sigma^{-1}$ is a positive definite M-matrix i.e. iff

$$k_{uv} \leq 0 \text{ for } u \neq v \text{ and } u, v \in V.$$

Multivariate Gaussian MTP_2 distributions

Proposition

Let $X \sim \mathcal{N}_V(0, \Sigma)$. Then X is MTP_2 if and only if $K = \Sigma^{-1}$ is a positive definite M-matrix i.e. iff

$$k_{uv} \leq 0 \text{ for } u \neq v \text{ and } u, v \in V.$$

Proof.

See Bølviken (1982) and Karlin and Rinott (1983). □

Multivariate Gaussian MTP_2 distributions

Proposition

Let $X \sim \mathcal{N}_V(0, \Sigma)$. Then X is MTP_2 if and only if $K = \Sigma^{-1}$ is a positive definite M-matrix i.e. iff

$$k_{uv} \leq 0 \text{ for } u \neq v \text{ and } u, v \in V.$$

Proof.

See Bølviken (1982) and Karlin and Rinott (1983). □

Since k_{uv} is proportional to the negative partial correlation between X_u and X_v, *X is MTP_2 if and only if all partial correlations are non-negative*.

Multivariate Gaussian MTP_2 distributions

Proposition

Let $X \sim \mathcal{N}_V(0, \Sigma)$. Then X is MTP_2 if and only if $K = \Sigma^{-1}$ is a positive definite M-matrix i.e. iff

$$k_{uv} \leq 0 \text{ for } u \neq v \text{ and } u, v \in V.$$

Proof.

See Bølviken (1982) and Karlin and Rinott (1983). □

Since k_{uv} is proportional to the negative partial correlation between X_u and X_v, *X is MTP_2 if and only if all partial correlations are non-negative*.

Note also that this is a convex restriction in K.

Carcass data

Measurements of the thickness of meat and fat layers at different locations on the back of a slaughter pig on each of 344 carcasses.

Carcass data

Measurements of the thickness of meat and fat layers at different locations on the back of a slaughter pig on each of 344 carcasses.

The sample correlation matrix for this data set is given by

$$S = \begin{array}{c} \\ \end{array} \begin{matrix} \text{Fat11} & \text{Meat11} & \text{Fat12} & \text{Meat12} & \text{Fat13} & \text{Meat13} \end{matrix}$$

$$S = \begin{pmatrix} 1.00 & 0.04 & 0.84 & 0.08 & 0.82 & -0.03 \\ 0.04 & 1.00 & 0.04 & 0.87 & 0.13 & 0.86 \\ 0.84 & 0.04 & 1.00 & 0.01 & 0.83 & -0.03 \\ 0.08 & 0.87 & 0.01 & 1.00 & 0.11 & 0.90 \\ 0.82 & 0.13 & 0.83 & 0.11 & 1.00 & 0.02 \\ -0.03 & 0.86 & -0.03 & 0.90 & 0.02 & 1.00 \end{pmatrix} \begin{matrix} \text{Fat11} \\ \text{Meat11} \\ \text{Fat12} \\ \text{Meat12} \\ \text{Fat13} \\ \text{Meat13} \end{matrix}$$

Empirical concentration for carcass data

and its inverse is

$$S^{-1} = \begin{pmatrix} 4.53 & 0.78 & -2.34 & -1.74 & -1.72 & 1.01 \\ 0.78 & 5.10 & -0.25 & -2.51 & -0.76 & -2.12 \\ -2.34 & -0.25 & 4.52 & 1.40 & -1.93 & -0.94 \\ -1.74 & -2.51 & 1.40 & 7.00 & -0.09 & -4.11 \\ -1.72 & -0.76 & -1.93 & -0.09 & 4.11 & 0.53 \\ 1.01 & -2.12 & -0.94 & -4.11 & 0.53 & 6.51 \end{pmatrix}$$

	Fat11	Meat11	Fat12	Meat12	Fat13	Meat13
Fat11	4.53	0.78	−2.34	−1.74	−1.72	1.01
Meat11	0.78	5.10	−0.25	−2.51	−0.76	−2.12
Fat12	−2.34	−0.25	4.52	1.40	−1.93	−0.94
Meat12	−1.74	−2.51	1.40	7.00	−0.09	−4.11
Fat13	−1.72	−0.76	−1.93	−0.09	4.11	0.53
Meat13	1.01	−2.12	−0.94	−4.11	0.53	6.51

which is clearly not quite an M-matrix.

MLestimates under MTP_2

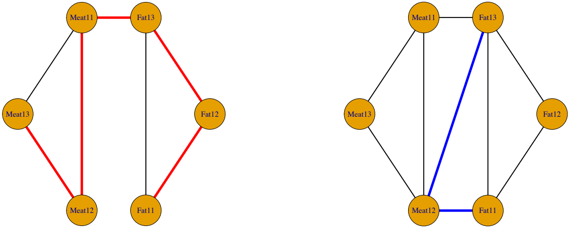
The MLE of the correlation matrix, rounded to 2 decimals, is

$$\hat{\Sigma} = \begin{pmatrix} 1.00 & 0.10 & 0.84 & 0.09 & 0.82 & 0.09 \\ 0.10 & 1.00 & 0.11 & 0.87 & 0.13 & 0.86 \\ 0.84 & 0.11 & 1.00 & 0.09 & 0.83 & 0.09 \\ 0.09 & 0.87 & 0.09 & 1.00 & 0.11 & 0.90 \\ 0.82 & 0.13 & 0.83 & 0.11 & 1.00 & 0.11 \\ 0.09 & 0.86 & 0.09 & 0.90 & 0.11 & 1.00 \end{pmatrix}$$

	Fat11	Meat11	Fat12	Meat12	Fat13	Meat13
Fat11	1.00	0.10	0.84	0.09	0.82	0.09
Meat11	0.10	1.00	0.11	0.87	0.13	0.86
Fat12	0.84	0.11	1.00	0.09	0.83	0.09
Meat12	0.09	0.87	0.09	1.00	0.11	0.90
Fat13	0.82	0.13	0.83	0.11	1.00	0.11
Meat13	0.09	0.86	0.09	0.90	0.11	1.00

Here blue entries are those that have changed in the estimation process.

Independence graph of estimates



(a) MTP_2 constraint (b) Graphical lasso

Figure: Undirected Gaussian graphical models for `carcass` data. The thick red edges in (a) correspond to the MWSF of the correlation matrix and the blue edges in (b) to edges in the model chosen by the graphical lasso which are not in the ML graph of the MTP_2 solution. *Note that the MTP_2 constraint automatically induces sparsity.*

Basic optimization problem

Let $\mathcal{M}$ denote the convex cone of M-matrices. Then the MLE is the solution to the following optimization problem:

$$\begin{aligned} \underset{K}{\text{maximize}} \quad & \log\det(K) - \operatorname{trace}(KS) \\ \text{subject to} \quad & K \in \mathcal{M} \end{aligned} \tag{1}$$

This is a convex optimization problem, since the objective function is concave on $\mathbb{S}^p_{\succeq 0}$.

Convex duality

The dual cone to $\mathcal{M}$ is the cone $\mathcal{N}$ given as

$$\mathcal{N} = \{X \in \mathbb{S}^p \mid \exists Y \in \mathbb{S}^p_{\succ 0} \text{ with } X \leq Y \text{ and } \mathrm{diag}(X) = \mathrm{diag}(Y)\}.$$

Convex duality

The dual cone to $\mathcal{M}$ is the cone $\mathcal{N}$ given as

$$\mathcal{N} = \{X \in \mathbb{S}^p \mid \exists Y \in \mathbb{S}^p_{\succ 0} \text{ with } X \leq Y \text{ and } \mathrm{diag}(X) = \mathrm{diag}(Y)\}.$$

Here $A \leq B$ means $a_{ij} \leq b_{ij}$ for all entries in A and B.

Convex duality

The dual cone to $\mathcal{M}$ is the cone $\mathcal{N}$ given as

$$\mathcal{N} = \{X \in \mathbb{S}^p \mid \exists Y \in \mathbb{S}^p_{\succ 0} \text{ with } X \leq Y \text{ and } \operatorname{diag}(X) = \operatorname{diag}(Y)\}.$$

Here $A \leq B$ means $a_{ij} \leq b_{ij}$ for all entries in A and B.

In other words, it holds that

$$\overline{\mathcal{N}} = \left\{ S \in \mathbb{S} \mid \langle S, K \rangle \geq 0 \text{ for all } K \in \mathcal{M} \right\} \qquad (2)$$

where $\langle A, B \rangle = \operatorname{tr}(AB)$ is the trace inner product.

Basic estimation result

Theorem (Lauritzen et al. (2017))

Consider a Gaussian MTP_2 model. Then the MLE $\hat{\Sigma}$ and $\hat{K}$ exists for a given sample covariance matrix S on V if and only if $S \in \mathcal{N}$. It is then equal to the unique element $\hat{\Sigma} \succ 0$ that satisfies the following system of equations and inequalities

$$
\begin{aligned}
\hat{k}_{uv} &\leq 0 \quad \textit{for all } u \neq v, & (3)\\
\hat{\sigma}_{vv} - s_{vv} &= 0 \quad \textit{for all } v \in V, & (4)\\
(\hat{\sigma}_{uv} - s_{uv}) &\geq 0 \quad \textit{for all } u \neq v, & (5)\\
(\hat{\sigma}_{uv} - s_{uv})\hat{k}_{uv} &= 0 \quad \textit{for all } u \neq v, & (6)
\end{aligned}
$$

Sparsity of the solution

The *maximum likelihood graph* (ML graph) $\hat{G}$ is the graph given by non-zero entries of $\hat{K}$.

Sparsity of the solution

The *maximum likelihood graph* (ML graph) $\hat{G}$ is the graph given by non-zero entries of $\hat{K}$.

We then have the following

Corollary (Lauritzen et al. (2017))

Consider the Gaussian graphical model determined by $k_{uv} = 0$ for $uv \notin E(\hat{G})$, where $\hat{G}$ is the ML graph under MTP_2. Let $\bar{\Sigma}$ be the MLE of Σ under that Gaussian graphical model (without the MTP_2 constraint). Then $\hat{\Sigma} = \bar{\Sigma}$.

Proof of corollary

Proof.

The MLE of Σ under the Gaussian graphical model with graph $\hat{G}$ is the unique element $\bar{\Sigma} \succ 0$ that satisfies the following system of equations:

$$\begin{aligned} \bar{\sigma}_{vv} - s_{vv} &= 0 \text{ for all } v \in V, \\ \bar{\sigma}_{uv} - s_{uv} &= 0 \text{ for all } uv \in E(\hat{G}), \\ k_{uv} &= 0 \text{ for all } uv \notin E(\hat{G}). \end{aligned}$$

The estimation theorem says that also $\hat{\Sigma}$ satisfies these equations and hence we must have $\bar{\Sigma} = \hat{\Sigma}$. □

Ultrametric matrices and inverse M-matrices

A non-negative symmetric matrix U is *ultrametric* if

$$u_{ij} \geq \min(u_{ik}, u_{jk}) \text{ for all } i, j, k.$$

Ultrametric matrices and inverse M-matrices

A non-negative symmetric matrix U is *ultrametric* if

$$u_{ij} \geq \min(u_{ik}, u_{jk}) \text{ for all } i, j, k.$$

We then have the following known result.

Theorem (Dellacherie et al. (2014))

Let U be an ultrametric matrix with strictly positive entries on the diagonal. Then U is non-singular if and only if no two rows are equal. Moreover, if U is non-singular, then U^{-1} is an M-matrix.

Single linkage matrix

By invariance, we may without loss of generality assume that instead of S we consider R given as $r_{uv} = s_{uv}/\sqrt{s_{uu}s_{vv}}$ so in particular $r_{uu} = 1$ for all u.

Single linkage matrix

By invariance, we may without loss of generality assume that instead of S we consider R given as $r_{uv} = s_{uv}/\sqrt{s_{uu}s_{vv}}$ so in particular $r_{uu} = 1$ for all u.

If R is positive semidefinite with $r_{ii} = 1$, we define G^+ as the graph determined by positive entries of R. Define further a matrix Z by

$$z_{ij} \quad := \quad \max_P \min_{uv \in P} r_{uv}, \tag{7}$$

where maximum is taken over paths in G^+ between i and j.

Single linkage matrix

By invariance, we may without loss of generality assume that instead of S we consider R given as $r_{uv} = s_{uv}/\sqrt{s_{uu}s_{vv}}$ so in particular $r_{uu} = 1$ for all u.

If R is positive semidefinite with $r_{ii} = 1$, we define G^+ as the graph determined by positive entries of R. Define further a matrix Z by

$$z_{ij} := \max_{P} \min_{uv \in P} r_{uv}, \tag{7}$$

where maximum is taken over paths in G^+ between i and j.

Z is the *single-linkage* matrix based on R and is known from cluster analysis.

Example

Suppose that

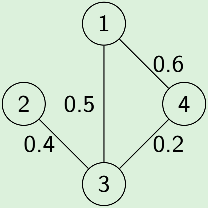
$$R = \begin{bmatrix} 1 & -0.5 & 0.5 & 0.6 \\ -0.5 & 1 & 0.4 & -0.1 \\ 0.5 & 0.4 & 1 & 0.2 \\ 0.6 & -0.1 & 0.2 & 1 \end{bmatrix}$$

Example

Suppose that

$$R = \begin{bmatrix} 1 & -0.5 & 0.5 & 0.6 \\ -0.5 & 1 & 0.4 & -0.1 \\ 0.5 & 0.4 & 1 & 0.2 \\ 0.6 & -0.1 & 0.2 & 1 \end{bmatrix}$$

Then G^+ and Z are given by



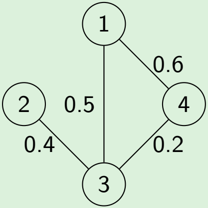
$$Z = \begin{bmatrix} 1 & 0.4 & 0.5 & 0.6 \\ 0.4 & 1 & 0.4 & 0.4 \\ 0.5 & 0.4 & 1 & 0.5 \\ 0.6 & 0.4 & 0.5 & 1 \end{bmatrix}.$$

Example

Suppose that

$$R = \begin{bmatrix} 1 & -0.5 & 0.5 & 0.6 \\ -0.5 & 1 & 0.4 & -0.1 \\ 0.5 & 0.4 & 1 & 0.2 \\ 0.6 & -0.1 & 0.2 & 1 \end{bmatrix}$$

Then G^+ and Z are given by



$$Z = \begin{bmatrix} 1 & 0.4 & 0.5 & 0.6 \\ 0.4 & 1 & 0.4 & 0.4 \\ 0.5 & 0.4 & 1 & 0.5 \\ 0.6 & 0.4 & 0.5 & 1 \end{bmatrix}.$$

Note that $Z \geq R$, Z is invertible, and Z^{-1} is an M-matrix.

Single linkage and ultrametric

Proposition (Lauritzen et al. (2017))

Let R be a symmetric $p \times p$ positive semidefinite matrix satisfying $r_{ii} = 1$ for all $i = 1, \ldots, p$. Then the single-linkage matrix Z based on R is an ultrametric matrix with $z_{ij} \geq r_{ij}$ for all $i \neq j$. If, in addition, $r_{ij} < 1$ for all $i \neq j$, then Z is nonsingular and therefore an inverse M-matrix.

Existence of the MLE

Theorem (Slawski and Hein (2015))

Consider a Gaussian MTP_2 model and let R be the sample correlation matrix. If $r_{ij} < 1$ for all $i \neq j$ then the MLE $\hat{\Sigma}$ (and $\hat{K}$) exists and it is unique. In particular, if the number n of observations satisfies $n \geq 2$, then the MLE exists with probability 1.

Existence of the MLE

Theorem (Slawski and Hein (2015))

Consider a Gaussian MTP_2 model and let R be the sample correlation matrix. If $r_{ij} < 1$ for all $i \neq j$ then the MLE $\hat{\Sigma}$ (and $\hat{K}$) exists and it is unique. In particular, if the number n of observations satisfies $n \geq 2$, then the MLE exists with probability 1.

Proof.

The single linkage matrix Z based on R is feasible both for the primal and dual problem. Hence the optimization problem has a (unique) solution. □

Properties of the maximum likelihood graph

Theorem (Lauritzen et al. (2017))

Let $MWSF(R)$ be the maximum weight spanning forest of R. Then with probability one, it holds that $MWSF(R) \subseteq \hat{G}$, where $\hat{G} = G(\hat{K})$ is the ML graph.

Properties of the maximum likelihood graph

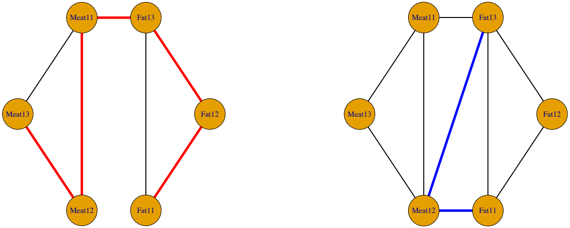
Theorem (Lauritzen et al. (2017))

Let $MWSF(R)$ be the maximum weight spanning forest of R. Then with probability one, it holds that $MWSF(R) \subseteq \hat{G}$, where $\hat{G} = G(\hat{K})$ is the ML graph.

Proof.

(sketch): Add MWSF(R) to $\hat{G}$ and show that $\hat{\Sigma}$ is also the MLE for the graphical model with graph MWSF(R) $\cup\, \hat{G}$. Now this happens with probability zero unless MWSF(R) $\subseteq \hat{G}$. □

Carcass data revisited



(a) MTP_2 constraint (b) Graphical lasso

Figure: Undirected Gaussian graphical models for `carcass` data. The thick red edges in (a) correspond to the MWSF of the correlation matrix and the blue edges in (b) to edges in the model chosen by the graphical lasso which are not in the ML graph of the MTP_2 solution.

Signed MTP_2 distributions

Definition

A distribution of X is *signed MTP_2* if there is a diagonal matrix D with $d_{ii} \in \{-1, 1\}$ so that $Y = DX$ is MTP_2.

Signed MTP_2 distributions

Definition

A distribution of X is *signed MTP_2* if there is a diagonal matrix D with $d_{ii} \in \{-1, 1\}$ so that $Y = DX$ is MTP_2.

Proposition

X is signed MTP_2 if and only if $|X|$ is MTP_2.

Signed MTP_2 distributions

Definition

A distribution of X is *signed MTP_2* if there is a diagonal matrix D with $d_{ii} \in \{-1, 1\}$ so that $Y = DX$ is MTP_2.

Proposition

X is signed MTP_2 if and only if $|X|$ is MTP_2.

Proposition

Any Gaussian latent tree model is MTP_2.

An obvious idea to estimate under signed MTP_2 is first to calculate the MWST with absolute correlations as weights.

Signed MTP_2 distributions

Definition

A distribution of X is *signed MTP_2* if there is a diagonal matrix D with $d_{ii} \in \{-1, 1\}$ so that $Y = DX$ is MTP_2.

Proposition

X is signed MTP_2 if and only if $|X|$ is MTP_2.

Proposition

Any Gaussian latent tree model is MTP_2.

An obvious idea to estimate under signed MTP_2 is first to calculate the MWST with absolute correlations as weights.

Then change signs along the tree to obtain $\tilde{X}$ with all pairwise correlations along edges positive.

Signed MTP_2 distributions

Definition

A distribution of X is *signed MTP_2* if there is a diagonal matrix D with $d_{ii} \in \{-1, 1\}$ so that $Y = DX$ is MTP_2.

Proposition

X is signed MTP_2 if and only if $|X|$ is MTP_2.

Proposition

Any Gaussian latent tree model is MTP_2.

An obvious idea to estimate under signed MTP_2 is first to calculate the MWST with absolute correlations as weights.

Then change signs along the tree to obtain $\tilde{X}$ with all pairwise correlations along edges positive.

Finally estimate under MTP_2 based on $\tilde{X}$.

Personality traits

240 individuals were asked to rate themselves on the scale 1-9 with respect to 32 different personality traits.

Personality traits

240 individuals were asked to rate themselves on the scale 1-9 with respect to 32 different personality traits.

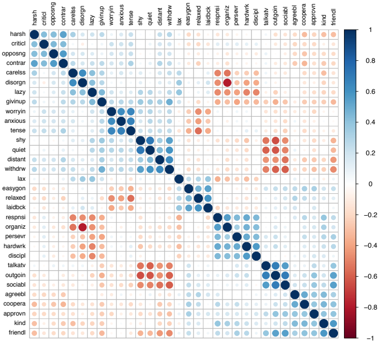


Figure: Correlation matrix of personality traits from the data set described in Malle and Horowitz (1995).

Personality traits

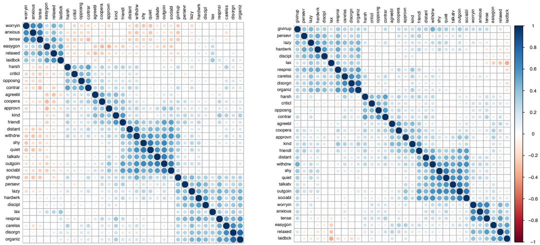


Figure: The correlation matrix of the data set on personality traits after performing the sign switches using the Chow-Liu tree is shown on the left. The correlation matrix resulting from switching the signs of the 16 (negative) traits that constitute the first block of variables in the previous figure is shown on the right.

Personality traits

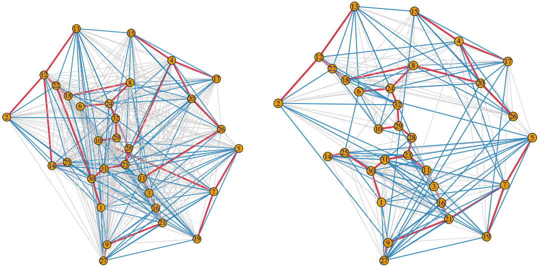


Figure: The graphical models resulting from estimation under MTP_2 under the two correlation matrices. The thin gray edges correspond to the edges of the EC graph that are not part of the ML graph. The blue edges represent edges of the ML graph that are not part of the minimum weight spanning tree. The latter is represented by thick red edges.

Bølviken, E. (1982). Probability inequalities for the multivariate normal with non-negative partial correlations. *Scand. J. Statist.*, 9:49–58.

Dellacherie, C., Martinez, S., and San Martin, J. (2014). *Inverse M-matrices and Ultrametric Matrices*, volume 2118. Springer.

Dykstra, R. L. and Hewett, J. E. (1978). Positive dependence of the roots of a Wishart matrix. *The Annals of Statistics*, 6(1):235–238.

Dynkin, E. (1980). Markov processes and random fields. *Bulletin of the American Mathematical Society*, 3(3):975–999.

Fallat, S., Lauritzen, S., Sadeghi, K., Uhler, C., Wermuth, N., and Zwiernik, P. (2017). Total positivity in Markov structures. *Ann. Statist.*, 45(3):1152–1184.

Fortuin, C. M., Kasteleyn, P. W., and Ginibre, J. (1971).

Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.*, 22(2):89–103.

Gumbel, E. J. (1961). Bivariate logistic distributions. *Journal of the American Statistical Association*, 56(294):335–349.

Karlin, S. and Rinott, Y. (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *J. Multiv. Anal.*, 10(4):467–498.

Karlin, S. and Rinott, Y. (1983). M-matrices as covariance matrices of multinormal distributions. *Linear Algebra Appl.*, 52:419 – 438.

Lauritzen, S., Uhler, C., and Zwiernik, P. (2017). Maximum likelihood estimation in Gaussian models under total positivity. arXiv:1702.04031.

Lebowitz, J. L. (1972). Bounds on the correlations and analyticity properties of ferromagnetic Ising spin systems. *Comm. Math. Phys.*, 28(4):313–321.

Malle, B. F. and Horowitz, L. M. (1995). The puzzle of negative self-views: An exploration using the schema concept. *Journal of Personality and Social Psychology*, 68(3):470.

Sarkar, T. K. (1969). Some lower bounds of reliability. Tech. Report, No. 124, Department of Operations Research and Department of Statistics, Stanford University.

Slawski, M. and Hein, M. (2015). Estimation of positive definite M-matrices and structure learning for attractive Gaussian Markov random field. *Linear Algebra Appl.*, 473:145–179.

Zwiernik, P. (2015). *Semialgebraic Statistics and Latent Tree Models*. Number 146 in Monographs on Statistics and Applied Probability. Chapman & Hall.