Total Positivity in Graphical Models

Master's Thesis

Tryggvi S. Kamban

Department of Mathematical Sciences
University of Copenhagen

Thursday 21 April 2022

Advisor: Steffen L. Lauritzen

Motivation/Objective

The task was to tie together three articles on the subject of *multivariate total positivity of order 2 (MTP$_2$)*.

- Shaun Fallat et al. (2017). "Total positivity in Markov structures"
- Steffen Lauritzen, Caroline Uhler, and Piotr Zwiernik (2017). "Maximum likelihood estimation in Gaussian models under total positivity"
- Steffen Lauritzen, Caroline Uhler, and Piotr Zwiernik (2019). "Total positivity in structured binary distributions"

The articles assume that the reader has spent years on research in theoretical statistics, hence there are many unfamiliar concepts that are not introduced.

Motivation/Objective

We want to write the thesis with master's students as the target audience!

Setup

Chapter 1: Convex Analysis and Optimization

- Intro to Convex Analysis
- Positive Definite Matrices and M-Matrices
- Convex Optimization

Chapter 2: Total Positivity

- Definitions and Basic Results
- Conditional Independence Models
- Markov Properties and Faithfulness

Setup

Chapter 1: Convex Analysis and Optimization

- Intro to Convex Analysis
- Positive Definite Matrices and M-Matrices
- Convex Optimization

Chapter 2: Total Positivity

- Definitions and Basic Results
- Conditional Independence Models
- Markov Properties and Faithfulness

Setup

Chapter 3: Binary Distributions

- Exponential Families
- Binary Distributions
- Binary Models over Graphs

Chapter 4: Quadratic Exponential Families

- Gaussian Graphical Models
- Gaussian Likelihood and Convex Optimization
- Totally Positive Gaussian Graphical Models

Chapter 5: Ising Models and Conditional Gaussian Distributions

- Ising Models
- Conditional Graphical Distributions

Setup

Chapter 3: Binary Distributions

- Exponential Families
- Binary Distributions
- Binary Models over Graphs

Chapter 4: Quadratic Exponential Families

- Gaussian Graphical Models
- Gaussian Likelihood and Convex Optimization
- Totally Positive Gaussian Graphical Models

Chapter 5: Ising Models and Conditional Gaussian Distributions

- Ising Models
- Conditional Graphical Distributions

Setup

Chapter 3: Binary Distributions

- Exponential Families
- Binary Distributions
- Binary Models over Graphs

Chapter 4: Quadratic Exponential Families

- Gaussian Graphical Models
- Gaussian Likelihood and Convex Optimization
- Totally Positive Gaussian Graphical Models

Chapter 5: Ising Models and Conditional Gaussian Distributions

- Ising Models
- Conditional Graphical Distributions

Definition of framework

$V = \{1, ..., d\}$ finite set

$X = (X_v, v \in V)$ random vector with labels in V

$\mathcal{X} = \prod_{v \in V} \mathcal{X}_v$ product space, where $\mathcal{X}_v \subseteq \mathbb{R}$ state space of X_v

$\mathcal{X}_v$ are either discrete finite sets, or open intervals on $\mathbb{R}$.

Definition of framework

$V = \{1, ..., d\}$ finite set
$X = (X_v, v \in V)$ random vector with labels in V
$\mathcal{X} = \prod_{v \in V} \mathcal{X}_v$ product space, where $\mathcal{X}_v \subseteq \mathbb{R}$ state space of X_v
$\mathcal{X}_v$ are either discrete finite sets, or open intervals on $\mathbb{R}$.

Definition of framework

$V = \{1, ..., d\}$ finite set
$X = (X_v, v \in V)$ random vector with labels in V
$\mathcal{X} = \prod_{v \in V} \mathcal{X}_v$ product space, where $\mathcal{X}_v \subseteq \mathbb{R}$ state space of X_v
$\mathcal{X}_v$ are either discrete finite sets, or open intervals on $\mathbb{R}$.

Definition of framework

$V = \{1, ..., d\}$ finite set
$X = (X_v, v \in V)$ random vector with labels in V
$\mathcal{X} = \prod_{v \in V} \mathcal{X}_v$ product space, where $\mathcal{X}_v \subseteq \mathbb{R}$ state space of X_v
$\mathcal{X}_v$ are either discrete finite sets, or open intervals on $\mathbb{R}$.

MTP_2 Property

A function f on $\mathcal{X}$ is *multivariate totally positive of order 2* (MTP_2) if

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

where

$$x \wedge y = (\min(x_v, y_v), v \in V) \text{ and } x \vee y = (\max(x_v, y_v), v \in V)$$

We say that a distribution P is MTP_2 if its density function is MTP_2

MTP$_2$ Property

A function f on $\mathcal{X}$ is *multivariate totally positive of order 2* (MTP$_2$) if

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

where

$$x \wedge y = (\min(x_v, y_v), v \in V) \text{ and } x \vee y = (\max(x_v, y_v), v \in V)$$

We say that a distribution P is MTP$_2$ if its density function is MTP$_2$

Example

Let

$$x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \ y = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}. \text{ Then } x \wedge y = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \ x \vee y = \begin{pmatrix} 3 \\ 2 \\ 3 \end{pmatrix}$$

The function $f : \mathbb{R}^3 \to \mathbb{R}$ defined by

$$f(x_1, x_2, x_3) = x_1 x_2 x_3$$

is MTP_2. Indeed,

$$f(x)f(y) \leq f(x \wedge y)f(x \vee y)$$

$$\Leftrightarrow 6 \times 6 \leq 2 \times 18$$

Cool Properties of MTP_2 Distributions

Proposition 2.2.6

The MTP_2 property is preserved under

- taking products $(f, g \text{ MTP}_2 \Rightarrow fg \text{ MTP}_2)$
- conditioning
- marginalization

Proposition 2.2.7 (Modified)

If f is MTP_2 in any pair of arguments when the remaining arguments are held constant, and if f has full support, then f is MTP_2

A consequence of Proposition 2.2.7 is that in order to check whether a density function f is MTP_2 it is sufficient to consider only 2 variables at a time instead of all variables simultaneously.

Cool Properties of MTP_2 Distributions

Proposition 2.2.6

The MTP_2 property is preserved under

- taking products $\quad (f, g \ \mathrm{MTP}_2 \Rightarrow fg \ \mathrm{MTP}_2)$
- conditioning
- marginalization

Proposition 2.2.7 (Modified)

If f is MTP_2 in any pair of arguments when the remaining arguments are held constant, and if f has full support, then f is MTP_2

A consequence of Proposition 2.2.7 is that in order to check whether a density function f is MTP_2 it is sufficient to consider only 2 variables at a time instead of all variables simultaneously.

Positive Definite Matrices and M-Matrices

Positive Definite Matrices

A symmetric $p \times p$ matrix A is *positive (semi-)definite* ($A \succeq 0, A \succ 0$) if for any $x \in \mathbb{R}^p \backslash \{\mathbf{0}\}$,

$$(x^T Ax) \geq 0, \qquad x^T Ax > 0$$

Example: The $p \times p$ identity matrix I is positive definite.

Lemma 1.2.3

Let A be a symmetric real-valued $p \times p$ matrix with eigenvalues $\lambda_i, i \in \{1, ..., p\}$. Then $A \succeq 0 \iff \lambda_i \geq 0$ for all $i \in \{1, ..., p\}$.

Positive Definite Matrices

A symmetric $p \times p$ matrix A is *positive (semi-)definite* ($A \succeq 0, A \succ 0$) if for any $x \in \mathbb{R}^p \backslash \{\mathbf{0}\}$,

$$(x^T Ax) \geq 0, \qquad x^T Ax > 0$$

Example: The $p \times p$ identity matrix I is positive definite.

Lemma 1.2.3

Let A be a symmetric real-valued $p \times p$ matrix with eigenvalues $\lambda_i, i \in \{1, ..., p\}$. Then $A \succeq 0 \iff \lambda_i \geq 0$ for all $i \in \{1, ..., p\}$.

Positive Definite Matrices

A symmetric $p \times p$ matrix A is *positive (semi-)definite* ($A \succeq 0, A \succ 0$) if for any $x \in \mathbb{R}^p \backslash \{\mathbf{0}\}$,

$$(x^T Ax) \geq 0, \qquad x^T Ax > 0$$

Example: The $p \times p$ identity matrix I is positive definite.

Lemma 1.2.3

Let A be a symmetric real-valued $p \times p$ matrix with eigenvalues $\lambda_i, i \in \{1, ..., p\}$. Then $A \succeq 0 \iff \lambda_i \geq 0$ for all $i \in \{1, ..., p\}$.

M-matrices

Let $M \in \mathbb{R}^{p \times p}$ be a symmetric matrix. We say that M is an *M-matrix* if $M_{ij} \leq 0$ for all $i \neq j$, and all eigenvalues of M are positive.

A matrix A where A^{-1} is an M-matrix is called an *inverse M-matrix*.

Any M-matrix is inverse-positive.

Example:

$$M = \begin{bmatrix} 1 & -0.2 & 0 \\ -0.2 & 1 & -0.4 \\ 0 & -0.4 & 1 \end{bmatrix}, \quad M^{-1} = \begin{bmatrix} 1.05 & 0.25 & 0.1 \\ 0.25 & 1.25 & 0.5 \\ 0.1 & 0.5 & 1.2 \end{bmatrix}$$

Eigenvalues: $\lambda_1 = 1, \lambda_2 = 1/\sqrt{5}, \lambda_3 = 1 - 1/\sqrt{5}$

M-matrices

Let $M \in \mathbb{R}^{p \times p}$ be a symmetric matrix. We say that M is an *M-matrix* if $M_{ij} \leq 0$ for all $i \neq j$, and all eigenvalues of M are positive.
A matrix A where A^{-1} is an M-matrix is called an *inverse M-matrix*.
Any M-matrix is inverse-positive.
Example:

$$M = \begin{bmatrix} 1 & -0.2 & 0 \\ -0.2 & 1 & -0.4 \\ 0 & -0.4 & 1 \end{bmatrix}, \quad M^{-1} = \begin{bmatrix} 1.05 & 0.25 & 0.1 \\ 0.25 & 1.25 & 0.5 \\ 0.1 & 0.5 & 1.2 \end{bmatrix}$$

Eigenvalues: $\lambda_1 = 1, \lambda_2 = 1/\sqrt{5}, \lambda_3 = 1 - 1/\sqrt{5}$

M-matrices

Let $M \in \mathbb{R}^{p \times p}$ be a symmetric matrix. We say that M is an *M-matrix* if $M_{ij} \leq 0$ for all $i \neq j$, and all eigenvalues of M are positive.

A matrix A where A^{-1} is an M-matrix is called an *inverse M-matrix*.

Any M-matrix is inverse-positive.

Example:

$$M = \begin{bmatrix} 1 & -0.2 & 0 \\ -0.2 & 1 & -0.4 \\ 0 & -0.4 & 1 \end{bmatrix}, \quad M^{-1} = \begin{bmatrix} 1.05 & 0.25 & 0.1 \\ 0.25 & 1.25 & 0.5 \\ 0.1 & 0.5 & 1.2 \end{bmatrix}$$

Eigenvalues: $\lambda_1 = 1, \lambda_2 = 1/\sqrt{5}, \lambda_3 = 1 - 1/\sqrt{5}$

M-matrices

Let $M \in \mathbb{R}^{p \times p}$ be a symmetric matrix. We say that M is an *M-matrix* if $M_{ij} \leq 0$ for all $i \neq j$, and all eigenvalues of M are positive.

A matrix A where A^{-1} is an M-matrix is called an *inverse M-matrix*.

Any M-matrix is inverse-positive.

Example:

$$M = \begin{bmatrix} 1 & -0.2 & 0 \\ -0.2 & 1 & -0.4 \\ 0 & -0.4 & 1 \end{bmatrix}, \quad M^{-1} = \begin{bmatrix} 1.05 & 0.25 & 0.1 \\ 0.25 & 1.25 & 0.5 \\ 0.1 & 0.5 & 1.2 \end{bmatrix}$$

Eigenvalues: $\lambda_1 = 1, \lambda_2 = 1/\sqrt{5}, \lambda_3 = 1 - 1/\sqrt{5}$

Gaussian Graphical Models

Totally Positive Gaussian Graphical Models

A multivariate Gaussian graphical model is MTP_2 if and only if the concentration matrix $K = \Sigma^{-1}$ is a symmetric M-matrix. Hence, all partial correlations will be non-negative

$$\rho_{X_i X_j | V \setminus \{i,j\}} = -\frac{K_{ij}}{\sqrt{K_{ii} K_{jj}}} \geq 0 \quad \text{for all} \quad i, j \in \{1, ..., p\}$$

Because of this, Simpson's paradox is impossible in MTP_2 Gaussian graphical models!

Totally Positive Gaussian Graphical Models

A multivariate Gaussian graphical model is MTP_2 if and only if the concentration matrix $K = \Sigma^{-1}$ is a symmetric M-matrix. Hence, all partial correlations will be non-negative

$$\rho_{X_i X_j | V \setminus \{i,j\}} = -\frac{K_{ij}}{\sqrt{K_{ii} K_{jj}}} \geq 0 \quad \text{for all} \quad i, j \in \{1, ..., p\}$$

Because of this, Simpson's paradox is impossible in MTP_2 Gaussian graphical models!

The wildest result (in my opinion)

We are used to the condition for existence of an MLE in Gaussian Models to be that the number of observations n must be at least as large as the number of parameters p.

Theorem 4.5.6 (Slawski and Hein)

Consider a Gaussian MTP_2 model and let S be the sample covariance matrix. If $S_{ij} < \sqrt{S_{ii}S_{jj}}$ 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

The wildest result (in my opinion)

We are used to the condition for existence of an MLE in Gaussian Models to be that the number of observations n must be at least as large as the number of parameters p.

Theorem 4.5.6 (Slawski and Hein)

Consider a Gaussian MTP_2 model and let S be the sample covariance matrix. If $S_{ij} < \sqrt{S_{ii}S_{jj}}$ 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

Thank you for listening, and stay Totally Positive!