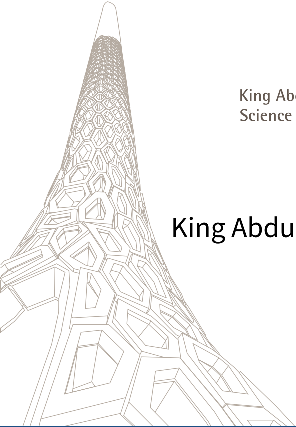
The Role of GMRFs in Statistical Modeling and Scalable Bayesian Inference (part I)



جامعة الملك عبدالله
للعلوم والتقنية

King Abdullah University of
Science and Technology

Håvard Rue
King Abdullah University of Science and Technology
Saudi Arabia

May 2026

I had something else in mind for this talk

- Then I started
- Then I gave up
- I had to email Miguel, saying
- there will be one looong talk instead,
- part I and part II.
- I'll aim for high-level insight, and
- discuss some connections hard to find in writing.

Disclaimer: Neither talks contains anything new...

I had something else in mind for this talk

- Then I started
- Then I gave up
- I had to email Miguel, saying
- there will be one looong talk instead,
- part I and part II.
- I'll aim for high-level insight, and
- discuss some connections hard to find in writing.

Disclaimer: Neither talks contains anything new...

I had something else in mind for this talk

- Then I started
- Then I gave up
- I had to email Miguel, saying
- there will be one looong talk instead,
- part I and part II.
- I'll aim for high-level insight, and
- discuss some connections hard to find in writing.

Disclaimer: Neither talks contains anything new...

I had something else in mind for this talk

- Then I started
- Then I gave up
- I had to email Miguel, saying
- there will be one looong talk instead,
- part I and part II.
- I'll aim for high-level insight, and
- discuss some connections hard to find in writing.

Disclaimer: Neither talks contains anything new...

I had something else in mind for this talk

- Then I started
- Then I gave up
- I had to email Miguel, saying
- there will be one looong talk instead,
- part I and part II.
- I'll aim for high-level insight, and
- discuss some connections hard to find in writing.

Disclaimer: Neither talks contains anything new...

I had something else in mind for this talk

- Then I started
- Then I gave up
- I had to email Miguel, saying
- there will be one looong talk instead,
- part I and part II.
- I'll aim for high-level insight, and
- discuss some connections hard to find in writing.

Disclaimer: Neither talks contains anything new...

I had something else in mind for this talk

- Then I started
- Then I gave up
- I had to email Miguel, saying
- there will be one looong talk instead,
- part I and part II.
- I'll aim for high-level insight, and
- discuss some connections hard to find in writing.

Disclaimer: Neither talks contains anything new...

I had something else in mind for this talk

- Then I started
- Then I gave up
- I had to email Miguel, saying
- there will be one looong talk instead,
- part I and part II.
- I'll aim for high-level insight, and
- discuss some connections hard to find in writing.

Disclaimer: Neither talks contains anything new...

I had something else in mind for this talk

- Then I started
- Then I gave up
- I had to email Miguel, saying
- there will be one looong talk instead,
- part I and part II.
- I'll aim for high-level insight, and
- discuss some connections hard to find in writing.

Disclaimer: Neither talks contains anything new...

Plan for the talk (part I)

- Correlation
- GMRFs
- Why GMRFs?
- Why 'Markov'?
- When is "GMRF + GMRF = GMRF"?
- What about 'non-Markov'?

Russian mathematician Andrey Markov

Plan for the talk (part I)

- Correlation
- GMRFs
- Why GMRFs?
- Why 'Markov'?
- When is "GMRF + GMRF = GMRF"?
- What about 'non-Markov'?

Russian mathematician Andrey Markov

Plan for the talk (part I)

- Correlation
- GMRFs
- Why GMRFs?
- Why 'Markov'?
- When is "GMRF + GMRF = GMRF"?
- What about 'non-Markov'?

Russian mathematician Andrey Markov

Plan for the talk (part I)

- Correlation
- GMRFs
- Why GMRFs?
- Why ‘Markov’?
- When is “GMRF + GMRF = GMRF”?
- What about ‘non-Markov’?

Russian mathematician Andrey Markov

Plan for the talk (part I)

- Correlation
- GMRFs
- Why GMRFs?
- Why 'Markov'?
- When is "GMRF + GMRF = GMRF"?
- What about 'non-Markov'?

Russian mathematician Andrey Markov

Plan for the talk (part I)

- Correlation
- GMRFs
- Why GMRFs?
- Why 'Markov'?
- When is "GMRF + GMRF = GMRF"?
- What about 'non-Markov'?

Russian mathematician Andrey Markov

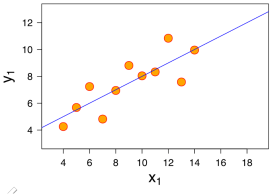
Correlation: definition

Tightly connected to the multivariate Normal distribution

- Need the concept of *correlation*
- Two random variables U and V, with zero mean and unit variance:

$$\mathrm{Corr}(U, V) \stackrel{\text{def}}{=} \mathrm{E}(UV)$$

- Was "invented" in regression late 1800
- Often motivated using regression
- Describes "linear dependence" only

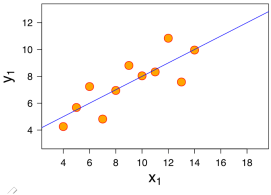


Correlation: definition

Tightly connected to the multivariate Normal distribution

- Need the concept of *correlation*
- Two random variables U and V, with zero mean and unit variance:

$$\mathrm{Corr}(U, V) \overset{\text{def}}{=} \mathrm{E}(UV)$$



- Was "invented" in regression late 1800
- Often motivated using regression
- Describes "linear dependence" only

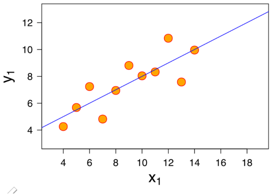
Correlation: definition

Tightly connected to the multivariate Normal distribution

- Need the concept of *correlation*
- Two random variables U and V, with zero mean and unit variance:

$$\text{Corr}(U, V) \stackrel{\text{def}}{=} \text{E}(UV)$$

- Was "invented" in regression late 1800
- Often motivated using regression
- Describes "linear dependence" only



Correlation: definition

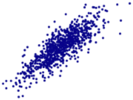
Tightly connected to the multivariate Normal distribution

- Need the concept of *correlation*
- Two random variables U and V, with zero mean and unit variance:

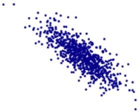
$$\mathrm{Corr}(U, V) \stackrel{\mathrm{def}}{=} \mathrm{E}(UV)$$

- Was “invented” in regression late 1800
- Often motivated using regression
- Describes “linear dependence” only

0.8



-0.8



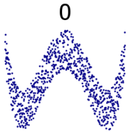
Correlation: definition

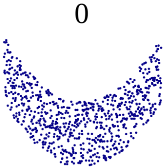
Tightly connected to the multivariate Normal distribution

- Need the concept of *correlation*
- Two random variables U and V, with zero mean and unit variance:

$$\text{Corr}(U, V) \stackrel{\text{def}}{=} \text{E}(UV)$$

- Was "invented" in regression late 1800
- Often motivated using regression
- Describes "linear dependence" only



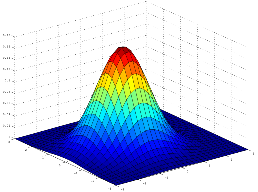


Multivariate Normal distribution

- Random vector $(x_1, \ldots, x_n)$
- (Scaled and centred) density

$$\pi(\boldsymbol{x}) \propto |\boldsymbol{C}|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{C}^{-1}\boldsymbol{x}\right)$$

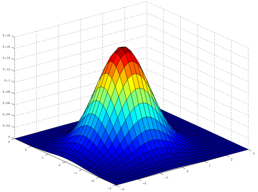
- $\boldsymbol{C}$ is the (SPD) correlation matrix
- $\text{diag}(\boldsymbol{C}) = \boldsymbol{I}$
- $C_{ij} = \text{E}(x_i x_j)$
- $C_{ij} = 0 \Leftrightarrow x_i \perp x_j \ (i \neq j)$



Multivariate Normal distribution

- Random vector $(x_1, \ldots, x_n)$
- (Scaled and centred) density

$$\pi(\boldsymbol{x}) \propto |\boldsymbol{C}|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{C}^{-1}\boldsymbol{x}\right)$$



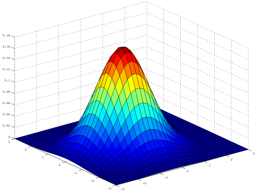
- $\boldsymbol{C}$ is the (SPD) correlation matrix
- $\text{diag}(\boldsymbol{C}) = \boldsymbol{I}$
- $C_{ij} = \text{E}(x_i x_j)$
- $C_{ij} = 0 \Leftrightarrow x_i \perp x_j \ (i \neq j)$

Multivariate Normal distribution

- Random vector $(x_1, \ldots, x_n)$
- (Scaled and centred) density

$$\pi(\boldsymbol{x}) \propto |\boldsymbol{C}|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{x}^T \boldsymbol{C}^{-1}\boldsymbol{x}\right)$$

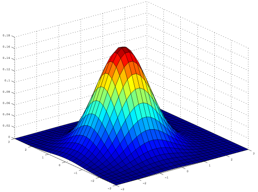
- $\boldsymbol{C}$ is the (SPD) correlation matrix
- $\text{diag}(\boldsymbol{C}) = \boldsymbol{I}$
- $C_{ij} = \text{E}(x_i x_j)$
- $C_{ij} = 0 \Leftrightarrow x_i \perp x_j \; (i \neq j)$



Multivariate Normal distribution

- Random vector $(x_1, \ldots, x_n)$
- (Scaled and centred) density

$$\pi(\boldsymbol{x}) \propto |\boldsymbol{C}|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{C}^{-1}\boldsymbol{x}\right)$$

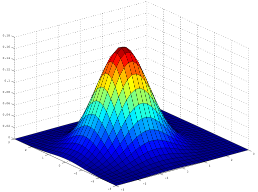


- $\boldsymbol{C}$ is the (SPD) correlation matrix
- $\text{diag}(\boldsymbol{C}) = \boldsymbol{I}$
- $C_{ij} = \text{E}(x_i x_j)$
- $C_{ij} = 0 \Leftrightarrow x_i \perp x_j \; (i \neq j)$

Multivariate Normal distribution

- Random vector $(x_1, \ldots, x_n)$
- (Scaled and centred) density

$$\pi(\boldsymbol{x}) \propto |\boldsymbol{C}|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{C}^{-1}\boldsymbol{x}\right)$$



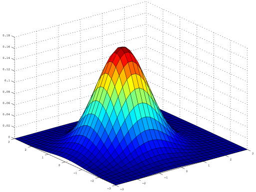
- $\boldsymbol{C}$ is the (SPD) correlation matrix
- $\text{diag}(\boldsymbol{C}) = \boldsymbol{I}$
- $C_{ij} = \text{E}(x_i x_j)$
- $C_{ij} = 0 \Leftrightarrow x_i \perp x_j \ (i \neq j)$

Multivariate Normal distribution

- Random vector $(x_1, \ldots, x_n)$
- (Scaled and centred) density

$$\pi(\boldsymbol{x}) \propto |\boldsymbol{C}|^{-1/2} \exp\left(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{C}^{-1}\boldsymbol{x}\right)$$

- $\boldsymbol{C}$ is the (SPD) correlation matrix
- $\text{diag}(\boldsymbol{C}) = \boldsymbol{I}$
- $C_{ij} = \text{E}(x_i x_j)$
- $C_{ij} = 0 \Leftrightarrow x_i \perp x_j \ (i \neq j)$



How to think about correlation?

- dim$=2$ with correlation ρ, then

$$E(x_1|x_2) = \rho x_2 \qquad \mathrm{Var}(x_1|x_2) = 1-\rho^2$$

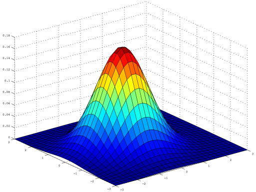
- What value of ρ is “half-correlation”?
- The natural “distance” is $\sqrt{\mathrm{Var}(x_1|x_2)}$, hence

$$\sqrt{1-\rho^2} = \frac{1}{2}$$

gives $\rho = \sqrt{3}/2 = 0.866\ldots$

- A “better” parameterisation:

$$\widetilde{\rho} = 1 - \mathrm{sign}(\rho)\sqrt{1-\rho^2}$$



How to think about correlation?

- dim$=2$ with correlation ρ, then

$$E(x_1|x_2) = \rho x_2 \qquad \text{Var}(x_1|x_2) = 1 - \rho^2$$

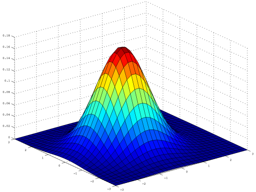
- What value of ρ is "half-correlation"?
- The natural "distance" is $\sqrt{\text{Var}(x_1|x_2)}$, hence

$$\sqrt{1-\rho^2} = \frac{1}{2}$$

gives $\rho = \sqrt{3}/2 = 0.866\ldots$

- A "better" parameterisation:

$$\widetilde{\rho} = 1 - \text{sign}(\rho)\sqrt{1-\rho^2}$$



How to think about correlation?

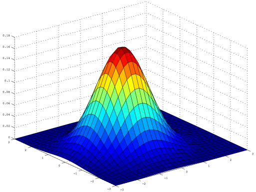
- dim$=2$ with correlation ρ, then

$$\mathrm{E}(x_1|x_2) = \rho x_2 \qquad \mathrm{Var}(x_1|x_2) = 1-\rho^2$$

- What value of ρ is "half-correlation"?
- The natural "distance" is $\sqrt{\mathrm{Var}(x_1|x_2)}$, hence

$$\sqrt{1-\rho^2} = \frac{1}{2}$$

gives $\rho = \sqrt{3}/2 = 0.866\ldots$



- A "better" parameterisation:

$$\widetilde{\rho} = 1 - \mathrm{sign}(\rho)\sqrt{1-\rho^2}$$

How to think about correlation?

- dim$=2$ with correlation ρ, then

$$E(x_1|x_2) = \rho x_2 \qquad \text{Var}(x_1|x_2) = 1 - \rho^2$$

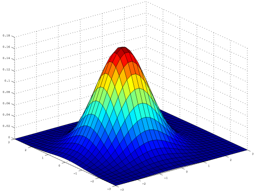
- What value of ρ is "half-correlation"?
- The natural "distance" is $\sqrt{\text{Var}(x_1|x_2)}$, hence

$$\sqrt{1-\rho^2} = \frac{1}{2}$$

gives $\rho = \sqrt{3}/2 = 0.866\ldots$

- A "better" parameterisation:

$$\widetilde{\rho} = 1 - \text{sign}(\rho)\sqrt{1-\rho^2}$$



Conditional independence

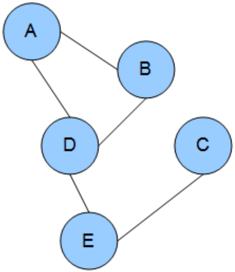
- Conditional independence of x_i and x_j, equals independence in

$$\pi(x_i, x_j \mid \boldsymbol{x}_{-ij})$$

- The result is, with $\boldsymbol{Q} = \boldsymbol{C}^{-1}$, that

$$Q_{ij} = 0 \Leftrightarrow x_i \perp x_j \mid \boldsymbol{x}_{-ij}$$

- This is potentially very very useful, as
  - If $\boldsymbol{Q}$ is fully connected then $\boldsymbol{C}$ is dense (Cayley–Hamilton thm)
  - Describe dependence by a sparse $\boldsymbol{Q}$
  - Sparse matrices: faster computations



Conditional independence

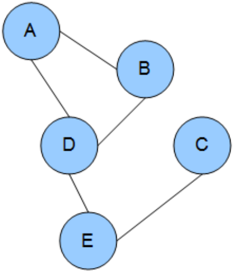
- Conditional independence of x_i and x_j, equals independence in

$$\pi(x_i, x_j \mid \boldsymbol{x}_{-ij})$$

- The result is, with $\boldsymbol{Q} = \boldsymbol{C}^{-1}$, that

$$Q_{ij} = 0 \Leftrightarrow x_i \perp x_j \mid \boldsymbol{x}_{-ij}$$

- This is potentially very very useful, as
  - If $\boldsymbol{Q}$ is fully connected then $\boldsymbol{C}$ is dense (Cayley–Hamilton thm)
  - Describe dependence by a sparse $\boldsymbol{Q}$
  - Sparse matrices: faster computations



Conditional independence

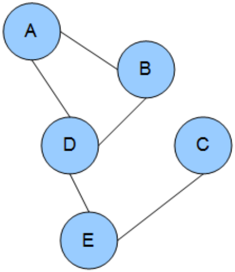
- Conditional independence of x_i and x_j, equals independence in

$$\pi(x_i, x_j \mid \boldsymbol{x}_{-ij})$$

- The result is, with $\boldsymbol{Q} = \boldsymbol{C}^{-1}$, that

$$Q_{ij} = 0 \Leftrightarrow x_i \perp x_j \mid \boldsymbol{x}_{-ij}$$

- This is potentially very very useful, as
  - If $\boldsymbol{Q}$ is fully connected then $\boldsymbol{C}$ is dense (Cayley–Hamilton thm)
  - Describe dependence by a sparse $\boldsymbol{Q}$
  - Sparse matrices: faster computations



Conditional independence

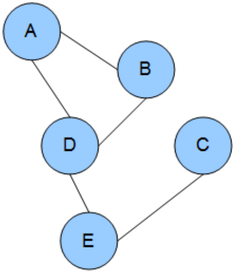
- Conditional independence of x_i and x_j, equals independence in

$$\pi(x_i, x_j \mid \boldsymbol{x}_{-ij})$$

- The result is, with $\boldsymbol{Q} = \boldsymbol{C}^{-1}$, that

$$Q_{ij} = 0 \Leftrightarrow x_i \perp x_j \mid \boldsymbol{x}_{-ij}$$

- This is potentially very very useful, as
  - If $\boldsymbol{Q}$ is fully connected then $\boldsymbol{C}$ is dense (Cayley–Hamilton thm)
  - Describe dependence by a sparse $\boldsymbol{Q}$
  - Sparse matrices: faster computations



Conditional independence

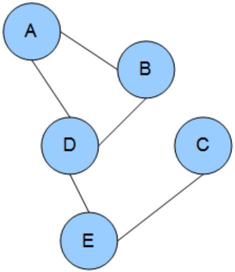
- Conditional independence of x_i and x_j, equals independence in

$$\pi(x_i, x_j \mid \boldsymbol{x}_{-ij})$$

- The result is, with $\boldsymbol{Q} = \boldsymbol{C}^{-1}$, that

$$Q_{ij} = 0 \Leftrightarrow x_i \perp x_j \mid \boldsymbol{x}_{-ij}$$

- This is potentially very very useful, as
  - If $\boldsymbol{Q}$ is fully connected then $\boldsymbol{C}$ is dense (Cayley–Hamilton thm)
  - Describe dependence by a sparse $\boldsymbol{Q}$
  - Sparse matrices: faster computations



Conditional independence

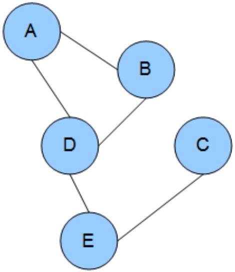
- Conditional independence of x_i and x_j, equals independence in

$$\pi(x_i, x_j \mid \boldsymbol{x}_{-ij})$$

- The result is, with $\boldsymbol{Q} = \boldsymbol{C}^{-1}$, that

$$Q_{ij} = 0 \Leftrightarrow x_i \perp x_j \mid \boldsymbol{x}_{-ij}$$

- This is potentially very very useful, as
    - If $\boldsymbol{Q}$ is fully connected then $\boldsymbol{C}$ is dense (Cayley–Hamilton thm)
    - Describe dependence by a sparse $\boldsymbol{Q}$
    - Sparse matrices: faster computations



Example: Markov processes in time

- Auto-regressive process of order 1

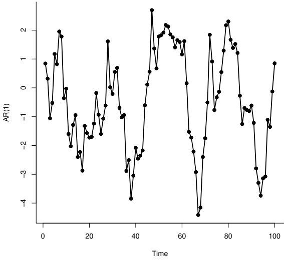
$$x_t = \phi x_{t-1} + \epsilon_t, \quad 0 \leq \phi < 1, \quad t = 1, \ldots, n$$

  with Gaussian noise ϵ_t

- Linear graph
- $\mathbf{Q}$ is tridiagonal (factorisation cost $\mathcal{O}(n)$)
- $\mathbf{C}$ is $\phi^{-|i-j|}$ (exp-correlation function)
- Cont.time: Ornstein-Uhlenbeck process

$$dx_t = -\theta\, x_t\, dt + dW_t,$$

  Wiener process W_t



Example: Markov processes in time

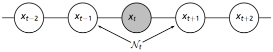
- Auto-regressive process of order 1

$$x_t = \phi x_{t-1} + \epsilon_t, \quad 0 \leq \phi < 1, \quad t = 1, \ldots, n$$

  with Gaussian noise ϵ_t
- Linear graph
- $\mathbf{Q}$ is tridiagonal (factorisation cost $\mathcal{O}(n)$)
- $\mathbf{C}$ is $\phi^{-|i-j|}$ (exp-correlation function)
- Cont.time: Ornstein-Uhlenbeck process

$$dx_t = -\theta\, x_t\, dt + dW_t,$$

  Wiener process W_t



Example: Markov processes in time

- Auto-regressive process of order 1

$$x_t = \phi x_{t-1} + \epsilon_t, \quad 0 \le \phi < 1, \quad t = 1, \ldots, n$$

with Gaussian noise ϵ_t

- Linear graph
- $\boldsymbol{Q}$ is tridiagonal (factorisation cost $\mathcal{O}(n)$)
- **C** is $\phi^{-|i-j|}$ (exp-correlation function)
- Cont.time: Ornstein-Uhlenbeck process

$$dx_t = -\theta\, x_t\, dt + dW_t,$$

Wiener process W_t

$$\begin{pmatrix} 1 & -\phi & & & & \\ -\phi & 1+\phi^2 & -\phi & & & \\ & & \ddots & \ddots & \ddots & \\ & & & -\phi & 1+\phi^2 & -\phi \\ & & & & -\phi & 1 \end{pmatrix}$$

Example: Markov processes in time

- Auto-regressive process of order 1

$$x_t = \phi x_{t-1} + \epsilon_t, \quad 0 \le \phi < 1, \quad t = 1, \ldots, n$$

with Gaussian noise ϵ_t

- Linear graph
- $\boldsymbol{Q}$ is tridiagonal (factorisation cost $\mathcal{O}(n)$)
- $\boldsymbol{C}$ is $\phi^{-|i-j|}$ (exp-correlation function)
- Cont.time: Ornstein-Uhlenbeck process

$$dx_t = -\theta \, x_t \, dt + dW_t,$$

Wiener process W_t

$$\begin{pmatrix} 1 & \phi & \phi^2 & \phi^3 & \phi^4 & \phi^5 & \phi^6 \\ \phi & 1 & \phi & \phi^2 & \phi^3 & \phi^4 & \phi^5 \\ \phi^2 & \phi & 1 & \phi & \phi^2 & \phi^3 & \phi^4 \\ \phi^3 & \phi^2 & \phi & 1 & \phi & \phi^2 & \phi^3 \\ \phi^4 & \phi^3 & \phi^2 & \phi & 1 & \phi & \phi^2 \\ \phi^5 & \phi^4 & \phi^3 & \phi^2 & \phi & 1 & \phi \\ \phi^6 & \phi^5 & \phi^4 & \phi^3 & \phi^2 & \phi & 1 \end{pmatrix}$$

Example: Markov processes in time

- Auto-regressive process of order 1

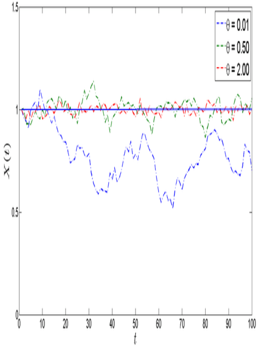
$$x_t = \phi x_{t-1} + \epsilon_t, \quad 0 \leq \phi < 1, \quad t = 1, \ldots, n$$

with Gaussian noise ϵ_t

- Linear graph
- $\boldsymbol{Q}$ is tridiagonal (factorisation cost $\mathcal{O}(n)$)
- $\boldsymbol{C}$ is $\phi^{-|i-j|}$ (exp-correlation function)
- Cont.time: Ornstein-Uhlenbeck process

$$dx_t = -\theta\, x_t\, dt + dW_t,$$

Wiener process W_t



How does correlation work?

- Let

$$x_1 = z_1 + \mu$$

$$x_2 = z_2 + \mu$$

- z_1, z_2, μ are independent zero mean Gaussian's
- $\text{Var}(\mu) = \sigma^2$, $\text{Var}(z_i) = 1$
- Then

$$\text{Corr}(x_1, x_2) = \frac{\sigma^2}{\sigma^2 + 1}$$

- This is an constructive/intuitive way to introduce correlation (which is a 'consequence' of the construction)

How does correlation work?

- Let

$$x_1 = z_1 + \mu$$

$$x_2 = z_2 + \mu$$

- z_1, z_2, μ are independent zero mean Gaussian's
- $\text{Var}(\mu) = \sigma^2$, $\text{Var}(z_i) = 1$
- Then

$$\text{Corr}(x_1, x_2) = \frac{\sigma^2}{\sigma^2 + 1}$$

- This is an constructive/intuitive way to introduce correlation (which is a 'consequence' of the construction)

How does correlation work?

- Let

$$x_1 = z_1 + \mu$$

$$x_2 = z_2 + \mu$$

- z_1, z_2, μ are independent zero mean Gaussian's
- $\text{Var}(\mu) = \sigma^2$, $\text{Var}(z_i) = 1$
- Then

$$\text{Corr}(x_1, x_2) = \frac{\sigma^2}{\sigma^2 + 1}$$

- This is an constructive/intuitive way to introduce correlation (which is a 'consequence' of the construction)

How does correlation work?

- Let

$$x_1 = z_1 + \mu$$

$$x_2 = z_2 + \mu$$

- z_1, z_2, μ are independent zero mean Gaussian's
- $\text{Var}(\mu) = \sigma^2$, $\text{Var}(z_i) = 1$
- Then

$$\text{Corr}(x_1, x_2) = \frac{\sigma^2}{\sigma^2 + 1}$$

- This is an constructive/intuitive way to introduce correlation (which is a 'consequence' of the construction)

How does correlation work?

- Let

$$x_1 = z_1 + \mu$$

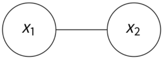
$$x_2 = z_2 + \mu$$

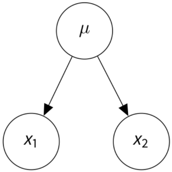
- z_1, z_2, μ are independent zero mean Gaussian's
- $\text{Var}(\mu) = \sigma^2$, $\text{Var}(z_i) = 1$
- Then

$$\text{Corr}(x_1, x_2) = \frac{\sigma^2}{\sigma^2 + 1}$$

- This is an constructive/intuitive way to introduce correlation (which is a 'consequence' of the construction)

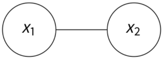
How this influence sparsity?

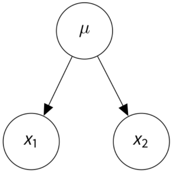




$$\mathbf{Q}_{x_1,x_2} = \begin{pmatrix} \times & \times \\ \times & \times \end{pmatrix} \qquad \mathbf{Q}_{x_1,x_2,\mu} = \begin{pmatrix} \times & & \times \\ & \times & \times \\ \times & \times & \times \end{pmatrix}$$

How this influence sparsity?

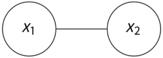


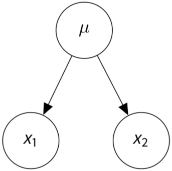


$$\boldsymbol{Q}_{x_1,x_2} = \begin{pmatrix} \times & \times \\ \times & \times \end{pmatrix}$$

$$\boldsymbol{Q}_{x_1,x_2,\mu} = \begin{pmatrix} \times & & \times \\ & \times & \times \\ \times & \times & \times \end{pmatrix}$$

How this influence sparsity?

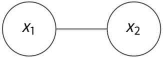


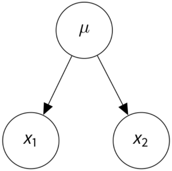


$$\boldsymbol{Q}_{x_1,x_2} = \begin{pmatrix} \times & \times \\ \times & \times \end{pmatrix}$$

$$\boldsymbol{Q}_{x_1,x_2,\mu} = \begin{pmatrix} \times & & \times \\ & \times & \times \\ \times & \times & \times \end{pmatrix}$$

How this influence sparsity?





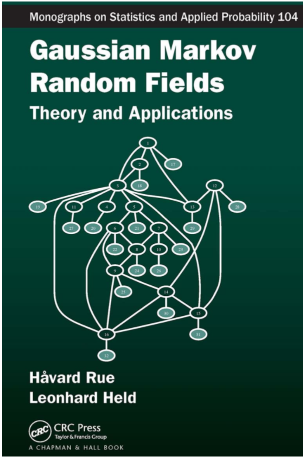
$$\boldsymbol{Q}_{x_1,x_2} = \begin{pmatrix} \times & \times \\ \times & \times \end{pmatrix} \qquad \boldsymbol{Q}_{x_1,x_2,\mu} = \begin{pmatrix} \times & & \times \\ & \times & \times \\ \times & \times & \times \end{pmatrix}$$

What is a GMRF?

- A Gaussian Markov Random Field is a multivariate Gaussian

$$\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}^{-1})$$

  where $\mathbf{Q}$ is the precision matrix ($= \mathbf{\Sigma}^{-1}$)
- $x_i \perp x_j | \mathbf{x}_{-ij} \iff Q_{ij} = 0$
- Form a graph $\mathcal{G}$ with an edge between i and j if $Q_{ij} \neq 0$.
- Then $\mathbf{x}$ is a GMRF wrt $\mathcal{G}$



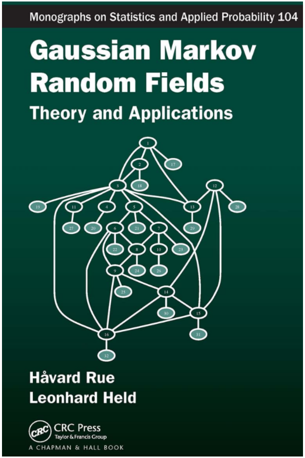
What is a GMRF?

- A Gaussian Markov Random Field is a multivariate Gaussian

$$\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q}^{-1})$$

where $\boldsymbol{Q}$ is the precision matrix $(= \boldsymbol{\Sigma}^{-1})$

- $x_i \perp x_j | \boldsymbol{x}_{-ij} \iff Q_{ij} = 0$
- Form a graph $\mathcal{G}$ with an edge between i and j if $Q_{ij} \neq 0$.
- Then $\boldsymbol{x}$ is a GMRF wrt $\mathcal{G}$

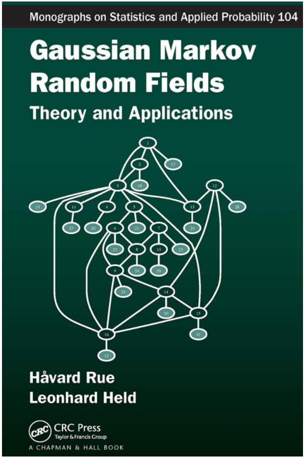


What is a GMRF?

- A Gaussian Markov Random Field is a multivariate Gaussian

$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}^{-1})$$

  where $\boldsymbol{Q}$ is the precision matrix $(= \boldsymbol{\Sigma}^{-1})$
- $x_i \perp x_i | \boldsymbol{x}_{-ij} \iff Q_{ij} = 0$
- Form a graph $\mathcal{G}$ with an edge between i and j if $Q_{ij} \neq 0$.
- Then $\boldsymbol{x}$ is a GMRF wrt $\mathcal{G}$

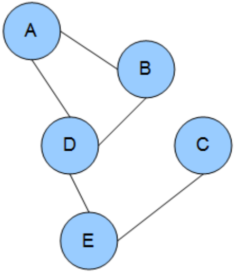


What is a GMRF?

- A Gaussian Markov Random Field is a multivariate Gaussian

$$\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q}^{-1})$$

  where $\boldsymbol{Q}$ is the precision matrix ($= \boldsymbol{\Sigma}^{-1}$)
- $x_i \perp x_j | \boldsymbol{x}_{-ij} \iff Q_{ij} = 0$
- Form a graph $\mathcal{G}$ with an edge between i and j if $Q_{ij} \neq 0$.
- Then $\boldsymbol{x}$ is a GMRF wrt $\mathcal{G}$



What are good "operations" for GMRFs?

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^T & \boldsymbol{\Sigma}_{22} \end{pmatrix} \qquad \boldsymbol{Q} = \begin{pmatrix} \boldsymbol{Q}_{11} & \boldsymbol{Q}_{12} \\ \boldsymbol{Q}_{12}^T & \boldsymbol{Q}_{22} \end{pmatrix}$$

Conditioning $\boldsymbol{x}_1 | \boldsymbol{x}_2$

$$\text{Var}(\boldsymbol{x}_1 | \boldsymbol{x}_2) = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}^T \qquad \text{Prec}(\boldsymbol{x}_1 | \boldsymbol{x}_2) = \boldsymbol{Q}_{11}$$

What are good “operations” for GMRFs?

Updating/Learning:

$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_x^{-1})$$

then we observe

$$\boldsymbol{y}|\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{x}, \boldsymbol{Q}_y^{-1})$$

$$\text{Var}(\boldsymbol{x}|\boldsymbol{y}) = \left(\boldsymbol{\Sigma}_x^{-1} + \boldsymbol{\Sigma}_y^{-1}\right)^{-1} \qquad \text{Prec}(\boldsymbol{x}|\boldsymbol{y}) = \boldsymbol{Q}_x + \boldsymbol{Q}_y$$

What are good "operations" for GMRFs?

Updating/Learning:

$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_x^{-1})$$

then we observe

$$\boldsymbol{y}|\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{x}, \boldsymbol{Q}_y^{-1})$$

$$\text{Var}(\boldsymbol{x}|\boldsymbol{y}) = \left(\boldsymbol{\Sigma}_x^{-1} + \boldsymbol{\Sigma}_y^{-1}\right)^{-1} \qquad \text{Prec}(\boldsymbol{x}|\boldsymbol{y}) = \boldsymbol{Q}_x + \boldsymbol{Q}_y$$

What are good "operations" for GMRFs?

Updating/Learning:

$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_x^{-1})$$

then we observe

$$\boldsymbol{y}|\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{x}, \boldsymbol{Q}_y^{-1})$$

$$\text{Var}(\boldsymbol{x}|\boldsymbol{y}) = \left(\boldsymbol{\Sigma}_x^{-1} + \boldsymbol{\Sigma}_y^{-1}\right)^{-1} \qquad \text{Prec}(\boldsymbol{x}|\boldsymbol{y}) = \boldsymbol{Q}_x + \boldsymbol{Q}_y$$

What are good "operations" for GMRFs?

Distribution of sums:

$$\boldsymbol{\eta} = \boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z}$$

where $\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q}_x^{-1})$, $\boldsymbol{y} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q}_y^{-1})$ and $\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{Q}_z^{-1})$.

$$\boldsymbol{Q}_\eta = \left(\boldsymbol{Q}_x^{-1} + \boldsymbol{Q}_y^{-1} + \boldsymbol{Q}_z^{-1}\right)^{-1}$$

$\boldsymbol{Q}_\eta$ is (in general) dense and it is costly to update $\boldsymbol{Q}_\eta$

The covariance formulation is easy

$$\boldsymbol{\Sigma}_\eta = \boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_y + \boldsymbol{\Sigma}_z$$

Marginalisation destroy sparsity...

What are good "operations" for GMRFs?

Distribution of sums:

$$\eta = \boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z}$$

where $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_x^{-1})$, $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_y^{-1})$ and $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_z^{-1})$.

$$\boldsymbol{Q}_\eta = \left(\boldsymbol{Q}_x^{-1} + \boldsymbol{Q}_y^{-1} + \boldsymbol{Q}_z^{-1}\right)^{-1}$$

$\boldsymbol{Q}_\eta$ is (in general) dense and it is costly to update $\boldsymbol{Q}_\eta$

The covariance formulation is easy

$$\Sigma_\eta = \Sigma_x + \Sigma_y + \Sigma_z$$

Marginalisation destroy sparsity...

What are good "operations" for GMRFs?

Distribution of sums:

$$\eta = \boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z}$$

where $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_x^{-1})$, $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_y^{-1})$ and $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_z^{-1})$.

$$\boldsymbol{Q}_\eta = \left(\boldsymbol{Q}_x^{-1} + \boldsymbol{Q}_y^{-1} + \boldsymbol{Q}_z^{-1}\right)^{-1}$$

$\boldsymbol{Q}_\eta$ is (in general) dense and it is costly to update $\boldsymbol{Q}_\eta$

The covariance formulation is easy

$$\Sigma_\eta = \Sigma_x + \Sigma_y + \Sigma_z$$

Marginalisation destroy sparsity...

What are good “operations” for GMRFs?

Distribution of sums:

$$\eta = \boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z}$$

where $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_x^{-1})$, $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_y^{-1})$ and $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_z^{-1})$.

$$\boldsymbol{Q}_\eta = \left(\boldsymbol{Q}_x^{-1} + \boldsymbol{Q}_y^{-1} + \boldsymbol{Q}_z^{-1}\right)^{-1}$$

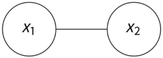
$\boldsymbol{Q}_\eta$ is (in general) dense and it is costly to update $\boldsymbol{Q}_\eta$

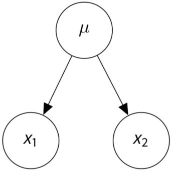
The covariance formulation is easy

$$\boldsymbol{\Sigma}_\eta = \boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_y + \boldsymbol{\Sigma}_z$$

Marginalisation destroy sparsity...

How this influence sparsity?





$$\boldsymbol{Q}_{x_1,x_2} = \begin{pmatrix} \times & \times \\ \times & \times \end{pmatrix} \qquad \boldsymbol{Q}_{x_1,x_2,\mu} = \begin{pmatrix} \times & & \times \\ & \times & \times \\ \times & \times & \times \end{pmatrix}$$

What are good "operations" for GMRFs?

Distribution of sums:

$$\boldsymbol{\eta} = \boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z}$$

where $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_x^{-1})$, $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_y^{-1})$ and $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_z^{-1})$.

- Using "ϵ" with high precision τ

$$\boldsymbol{\eta} = \boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z} + \boldsymbol{\epsilon}$$

- Then

$$\text{Var}\begin{pmatrix}\boldsymbol{\eta}\\ \boldsymbol{x}\\ \boldsymbol{y}\\ \boldsymbol{z}\end{pmatrix} = \begin{pmatrix}\tau\boldsymbol{I} & -\tau\boldsymbol{I} & -\tau\boldsymbol{I} & -\tau\boldsymbol{I}\\ & \boldsymbol{Q}_x + \tau\boldsymbol{I} & \tau\boldsymbol{I} & \tau\boldsymbol{I}\\ & & \boldsymbol{Q}_y + \tau\boldsymbol{I} & \tau\boldsymbol{I}\\ & & & \boldsymbol{Q}_z + \tau\boldsymbol{I}\end{pmatrix}$$

- Good practical solution unless choice of τ becomes an issue

What are good “operations” for GMRFs?

Distribution of sums:

$$\eta = \boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z}$$

where $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_x^{-1})$, $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_y^{-1})$ and $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_z^{-1})$.

- Using “ϵ” with high precision τ

$$\eta = \boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z} + \epsilon$$

- Then

$$\text{Var}\begin{pmatrix}\eta \\ \boldsymbol{x} \\ \boldsymbol{y} \\ \boldsymbol{z}\end{pmatrix} = \begin{pmatrix}\tau \boldsymbol{I} & -\tau \boldsymbol{I} & -\tau \boldsymbol{I} & -\tau \boldsymbol{I} \\ & \boldsymbol{Q}_x + \tau \boldsymbol{I} & \tau \boldsymbol{I} & \tau \boldsymbol{I} \\ & & \boldsymbol{Q}_y + \tau \boldsymbol{I} & \tau \boldsymbol{I} \\ & & & \boldsymbol{Q}_z + \tau \boldsymbol{I}\end{pmatrix}$$

- Good practical solution unless choice of τ becomes an issue

What are good "operations" for GMRFs?

Distribution of sums:

$$\boldsymbol{\eta} = \boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z}$$

where $\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_x^{-1})$, $\boldsymbol{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_y^{-1})$ and $\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_z^{-1})$.

- Using "ϵ" with high precision τ

$$\boldsymbol{\eta} = \boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z} + \epsilon$$

- Then

$$\text{Var}\begin{pmatrix}\boldsymbol{\eta}\\ \boldsymbol{x}\\ \boldsymbol{y}\\ \boldsymbol{z}\end{pmatrix} = \begin{pmatrix}\tau \mathbf{I} & -\tau \mathbf{I} & -\tau \mathbf{I} & -\tau \mathbf{I}\\ & \mathbf{Q}_x + \tau \mathbf{I} & \tau \mathbf{I} & \tau \mathbf{I}\\ & & \mathbf{Q}_y + \tau \mathbf{I} & \tau \mathbf{I}\\ & & & \mathbf{Q}_z + \tau \mathbf{I}\end{pmatrix}$$

- Good practical solution unless choice of τ becomes an issue

What are good "operations" for GMRFs?

Distribution of sums:

$$\eta = \boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z}$$

where $\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_x^{-1})$, $\boldsymbol{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_y^{-1})$ and $\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_z^{-1})$.

- Using cummulative sums

$$\begin{pmatrix} \widetilde{\boldsymbol{x}} \\ \widetilde{\boldsymbol{y}} \\ \widetilde{\boldsymbol{z}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z} \\ \boldsymbol{y} + \boldsymbol{z} \\ \boldsymbol{z} \end{pmatrix}$$

- Then

$$\text{Var}\begin{pmatrix} \widetilde{\boldsymbol{x}} \\ \widetilde{\boldsymbol{y}} \\ \widetilde{\boldsymbol{z}} \end{pmatrix} = \begin{pmatrix} \mathbf{Q}_x & -\mathbf{Q}_x & \\ & \mathbf{Q}_y + \mathbf{Q}_x & -\mathbf{Q}_y \\ & & \mathbf{Q}_z + \mathbf{Q}_y \end{pmatrix}$$

- More efficient, but no access to $\boldsymbol{x}$ and $\boldsymbol{y}$ directly.

What are good "operations" for GMRFs?

Distribution of sums:

$$\eta = \boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z}$$

where $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_x^{-1})$, $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_y^{-1})$ and $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_z^{-1})$.

- Using cummulative sums

$$\begin{pmatrix} \widetilde{\boldsymbol{x}} \\ \widetilde{\boldsymbol{y}} \\ \widetilde{\boldsymbol{z}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z} \\ \boldsymbol{y} + \boldsymbol{z} \\ \boldsymbol{z} \end{pmatrix}$$

- Then

$$\text{Var}\begin{pmatrix} \widetilde{\boldsymbol{x}} \\ \widetilde{\boldsymbol{y}} \\ \widetilde{\boldsymbol{z}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{Q}_x & -\boldsymbol{Q}_x & \\ & \boldsymbol{Q}_y + \boldsymbol{Q}_x & -\boldsymbol{Q}_y \\ & & \boldsymbol{Q}_z + \boldsymbol{Q}_y \end{pmatrix}$$

- More efficient, but no access to $\boldsymbol{x}$ and $\boldsymbol{y}$ directly.

What are good “operations” for GMRFs?

Distribution of sums:

$$\eta = \boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z}$$

where $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_x^{-1})$, $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_y^{-1})$ and $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_z^{-1})$.

- Using cummulative sums

$$\begin{pmatrix} \widetilde{\boldsymbol{x}} \\ \widetilde{\boldsymbol{y}} \\ \widetilde{\boldsymbol{z}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z} \\ \boldsymbol{y} + \boldsymbol{z} \\ \boldsymbol{z} \end{pmatrix}$$

- Then

$$\mathrm{Var}\begin{pmatrix} \widetilde{\boldsymbol{x}} \\ \widetilde{\boldsymbol{y}} \\ \widetilde{\boldsymbol{z}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{Q}_x & -\boldsymbol{Q}_x & \\ & \boldsymbol{Q}_y + \boldsymbol{Q}_x & -\boldsymbol{Q}_y \\ & & \boldsymbol{Q}_z + \boldsymbol{Q}_y \end{pmatrix}$$

- More efficient, but no access to $\boldsymbol{x}$ and $\boldsymbol{y}$ directly.

What are the advantages with GMRFs?

- More focus on a “generative” model, where
- “generative” is interpreted as conditioning
- Better interpretation via conditioning and/or sharing
- The big implication, is that correlation is a **property** of the distribution,
- not how to the distribution is/should-be **defined**
- Superior computational properties
- and others...

What are the advantages with GMRFs?

- More focus on a “generative” model, where
- “generative” is interpreted as conditioning
- Better interpretation via conditioning and/or sharing
- The big implication, is that correlation is a **property** of the distribution,
- not how to the distribution is/should-be **defined**
- Superior computational properties
- and others...

What are the advantages with GMRFs?

- More focus on a “generative” model, where
- “generative” is interpreted as conditioning
- Better interpretation via conditioning and/or sharing
- The big implication, is that correlation is a **property** of the distribution,
- not how to the distribution is/should-be **defined**
- Superior computational properties
- and others...

What are the advantages with GMRFs?

- More focus on a “generative” model, where
- “generative” is interpreted as conditioning
- Better interpretation via conditioning and/or sharing
- The big implication, is that correlation is a **property** of the distribution,
- not how to the distribution is/should-be **defined**
- Superior computational properties
- and others...

What are the advantages with GMRFs?

- More focus on a "generative" model, where
- "generative" is interpreted as conditioning
- Better interpretation via conditioning and/or sharing
- The big implication, is that correlation is a **property** of the distribution,
- not how to the distribution is/should-be **defined**
- Superior computational properties
- and others...

What are the advantages with GMRFs?

- More focus on a “generative” model, where
- “generative” is interpreted as conditioning
- Better interpretation via conditioning and/or sharing
- The big implication, is that correlation is a **property** of the distribution,
- not how to the distribution is/should-be **defined**
- Superior computational properties
- and others...

What are the advantages with GMRFs?

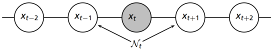
- More focus on a “generative” model, where
- “generative” is interpreted as conditioning
- Better interpretation via conditioning and/or sharing
- The big implication, is that correlation is a **property** of the distribution,
- not how to the distribution is/should-be **defined**
- Superior computational properties
- and others...

Example: dimension $= 1$

Time-series literature is, at best, notoriously confusing

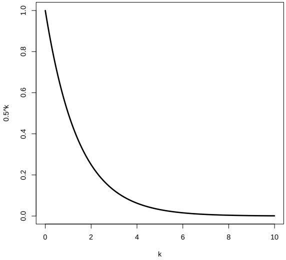
- Auto-regressive processes

$$x_t \mid x_1, \ldots, x_{t-1} \sim \mathcal{N}\Big(\sum_{s=1}^{p} \phi_s x_{t-s}, \sigma^2\Big)$$



- The correlation function is for $p = 1$

$$\text{Corr}(x_t, x_{t+s}) = \phi^{|s|}$$



- Another viewpoint

$$\text{E}(x_t | \boldsymbol{x}_{-t}) = \frac{\phi}{1+\phi^2}(x_{t-1} + x_{t+1})$$

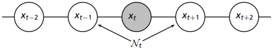
with some conditional variance

Example: dimension$=1$

Time-series literature is, at best, notoriously confusing

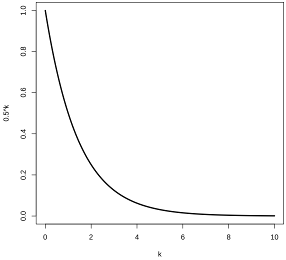
- Auto-regressive processes

$$x_t \mid x_1, \ldots, x_{t-1} \sim \mathcal{N}\left(\sum_{s=1}^{p} \phi_s x_{t-s}, \sigma^2\right)$$



- The correlation function is for $p = 1$

$$\text{Corr}(x_t, x_{t+s}) = \phi^{|s|}$$



- Another viewpoint

$$\text{E}(x_t | \boldsymbol{x}_{-t}) = \frac{\phi}{1+\phi^2}(x_{t-1} + x_{t+1})$$

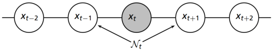
with some conditional variance

Example: dimension$= 1$

Time-series literature is, at best, notoriously confusing

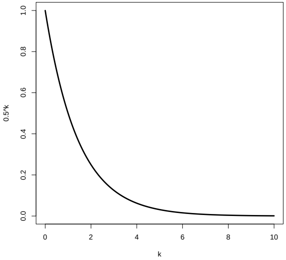
- Auto-regressive processes

$$x_t \mid x_1, \ldots, x_{t-1} \sim \mathcal{N}\Big(\sum_{s=1}^{p} \phi_s x_{t-s}, \sigma^2\Big)$$



- The correlation function is for $p = 1$

$$\mathrm{Corr}(x_t, x_{t+s}) = \phi^{|s|}$$



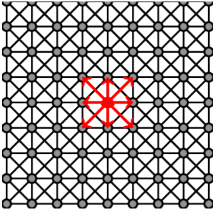
- Another viewpoint

$$\mathrm{E}(x_t|\boldsymbol{x}_{-t}) = \frac{\phi}{1+\phi^2}(x_{t-1} + x_{t+1})$$

with some conditional variance

Example: dimesion$=2$

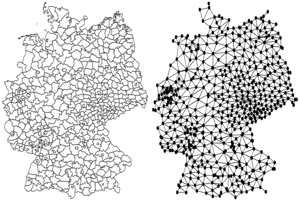
"standard" model in spatial statistics++ with regional data



- Traditionally written as

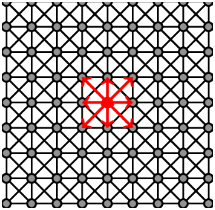
$$\pi(\boldsymbol{x}) \propto \tau^{\frac{n-k}{2}} \exp(-\frac{\tau}{2}\sum_{i\sim j} w_{ij}(x_i - x_j)^2)$$

- (By why should how data are collected determine what model to use?)



Example: dimesion= 2

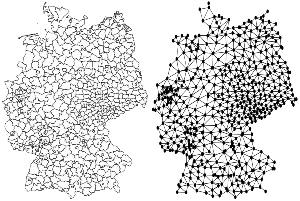
"standard" model in spatial statistics++ with regional data



- Traditionally written as

$$\pi(\boldsymbol{x}) \propto \tau^{\frac{n-k}{2}} \exp(-\frac{\tau}{2} \sum_{i \sim j} w_{ij}(x_i - x_j)^2)$$

- (By why should how data are collected determine what model to use?)



Additive regression models/GLM++

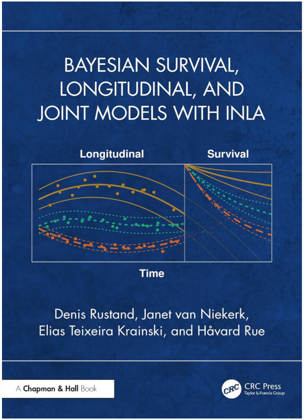
- Linear predictor

$$\eta_i = \mu + \sum_j \beta_j z_{ij} + \boldsymbol{f}_{1i} + \boldsymbol{f}_{2i} + \cdots$$

with observations

$$y_i | \ldots \sim \pi(y_i | \eta_i, \ldots)$$

- $\boldsymbol{f}_k$ is a Gaussian process, think times-series, spatial-model, spline, some kind of dependent 'random effect', measurement error model, etc...



Additive regression models/GLM++

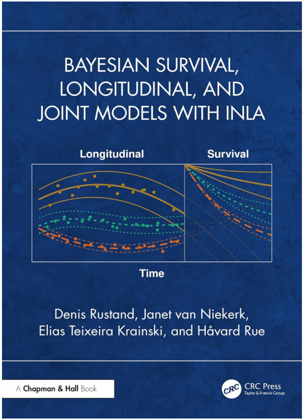
- Linear predictor

$$\eta_i = \mu + \sum_j \beta_j z_{ij} + \boldsymbol{f}_{1i} + \boldsymbol{f}_{2i} + \cdots$$

with observations

$$y_i | \ldots \sim \pi(y_i | \eta_i, \ldots)$$

- $\boldsymbol{f}_k$ is a Gaussian process, think times-series, spatial-model, spline, some kind of dependent 'random effect', measurement error model, etc...



Key observation

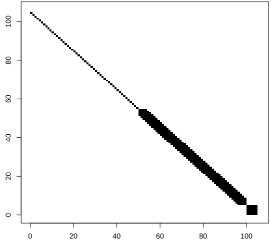
- If each $\boldsymbol{f}_k$ is a GMRF, then

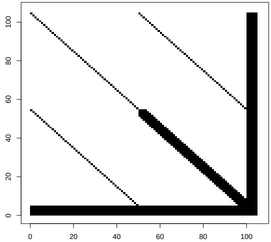
$$\boldsymbol{x} = (\beta_1, \beta_2, \ldots, \boldsymbol{f}_1, \boldsymbol{f}_2, \ldots)$$

is also a GMRF with precision matrix $\boldsymbol{Q}$

- Let $\eta = \boldsymbol{A}\boldsymbol{x}$. Assume Gaussian data with unit variance, then $\boldsymbol{x}|\boldsymbol{y}$ is a GMRF with precision matrix

$$\boldsymbol{Q} + \boldsymbol{A}^T\boldsymbol{A}$$





Key observation

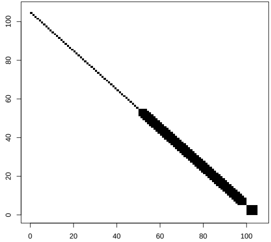
- If each $\boldsymbol{f}_k$ is a GMRF, then

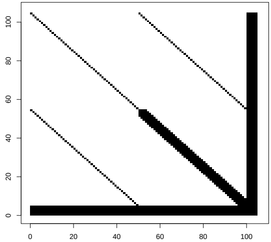
$$\boldsymbol{x} = (\beta_1, \beta_2, \ldots, \boldsymbol{f}_1, \boldsymbol{f}_2, \ldots)$$

is also a GMRF with precision matrix $\boldsymbol{Q}$

- Let $\eta = \boldsymbol{A}\boldsymbol{x}$. Assume Gaussian data with unit variance, then $\boldsymbol{x}|\boldsymbol{y}$ is a GMRF with precision matrix

$$\boldsymbol{Q} + \boldsymbol{A}^T\boldsymbol{A}$$





Keep it in the family!

GMRFs is closed

- sums
- conditioning

iff done right: no marginalisation!

Keep it in the family!

GMRFs is closed

- sums
- conditioning

iff done right: no marginalisation!

Keep it in the family!

GMRFs is closed

- sums
- conditioning

iff done right: no marginalisation!

Keep it in the family!

GMRFs is closed

- sums
- conditioning

iff done right: no marginalisation!

What about non-Markov processes?

- Does the outcome of tomorrow depends on what happed when “Eve ate the apple”?
- Math versus reality
- Non-Markov processes we are interested in, are “almost always”, my claim, just a sum of simpler Markov processes.

What about non-Markov processes?

- Does the outcome of tomorrow depends on what happed when "Eve ate the apple"?
- Math versus reality
- Non-Markov processes we are interested in, are "almost always", my claim, just a sum of simpler Markov processes.

What about non-Markov processes?

- Does the outcome of tomorrow depends on what happed when “Eve ate the apple”?
- Math versus reality
- Non-Markov processes we are interested in, are “almost always”, my claim, just a sum of simpler Markov processes.

What about non-Markov processes?

- AR1 processes: $(1-\phi_1 B)y_t = \epsilon_t$ and $(1-\phi_2 B)z_t = \nu_t$
- Their sum is $x_t = y_t + z_t$:

$$x_t = (1-\phi_1 B)^{-1}\epsilon_t + (1-\phi_2 B)^{-1}\nu_t$$

- This is an ARMA(2,1) process

$$(1-\phi_1 B)(1-\phi_2 B)x_t = (1-\phi_2 B)\epsilon_t + (1-\phi_1 B)\nu_t$$

What about non-Markov processes?

- AR1 processes: $(1-\phi_1 B)y_t = \epsilon_t$ and $(1-\phi_2 B)z_t = \nu_t$
- Their sum is $x_t = y_t + z_t$:

$$x_t = (1-\phi_1 B)^{-1}\epsilon_t + (1-\phi_2 B)^{-1}\nu_t$$

- This is an ARMA(2,1) process

$$(1-\phi_1 B)(1-\phi_2 B)x_t = (1-\phi_2 B)\epsilon_t + (1-\phi_1 B)\nu_t$$

What about non-Markov processes?

- AR1 processes: $(1-\phi_1 B)y_t = \epsilon_t$ and $(1-\phi_2 B)z_t = \nu_t$
- Their sum is $x_t = y_t + z_t$:

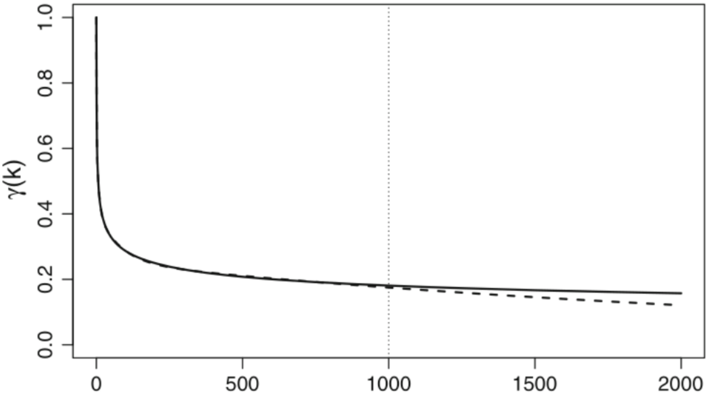
$$x_t = (1-\phi_1 B)^{-1}\epsilon_t + (1-\phi_2 B)^{-1}\nu_t$$

- This is an ARMA(2,1) process

$$(1-\phi_1 B)(1-\phi_2 B)x_t = (1-\phi_2 B)\epsilon_t + (1-\phi_1 B)\nu_t$$

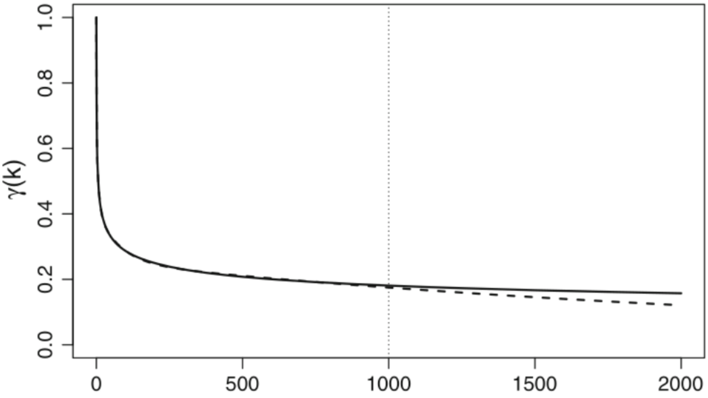
Example: Fractional Gaussian noise

$$\gamma_{\boldsymbol{x}}(k) = \frac{1}{2}\left(|k-1|^{2H} - 2|k|^{2H} + |k+1|^{2H}\right), \quad k = 0, 1, \ldots, n-1.$$



Example: Fractional Gaussian noise

$$\gamma_{\boldsymbol{x}}(k) = \frac{1}{2}\left(|k-1|^{2H} - 2|k|^{2H} + |k+1|^{2H}\right), \quad k = 0, 1, \ldots, n-1.$$



Continue...

$$\tilde{x}_m = \sigma \sum_{j=1}^{m} \sqrt{w_j} z^{(j)},$$

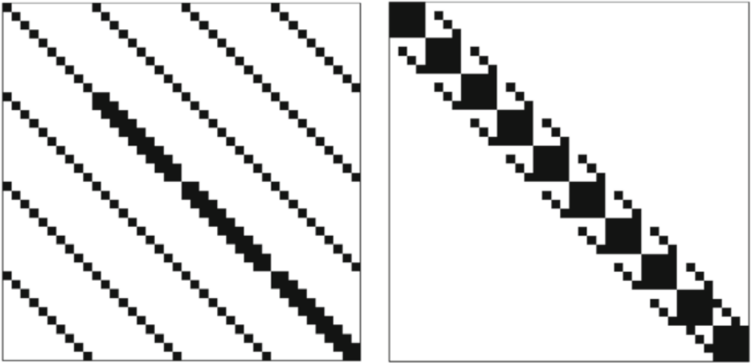
```
> round(dig=4, inla.fgn(0.9)[,-1])
  phi1   phi2   phi3   phi4    w1    w2    w3    w4
0.9996 0.9871 0.8869 0.3766 0.253 0.157 0.217 0.371
```

Continue...

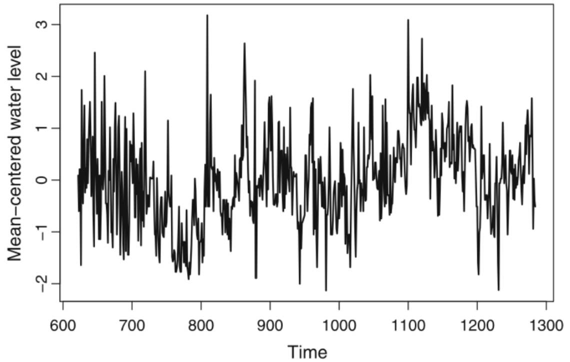
$$\tilde{x}_m = \sigma \sum_{j=1}^{m} \sqrt{w_j} z^{(j)},$$

```
> round(dig=4, inla.fgn(0.9)[,-1])
  phi1   phi2   phi3   phi4     w1    w2    w3    w4
0.9996 0.9871 0.8869 0.3766  0.253 0.157 0.217 0.371
```

Continue...

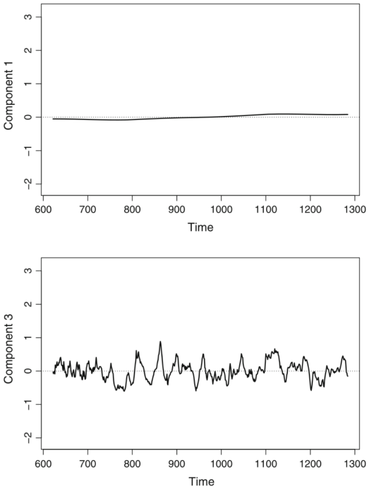


Continue...



The Nile River data, particularly the yearly minima at Roda gauge (622–1284 AD) and annual streamflow (1871–1970), is a classic example of long-memory (or long-range dependence) in hydrology. Harold Edwin Hurst (1950s) identified that Nile fluctuations show high correlation over long periods, with a Hurst exponent $H \approx 0.9$, indicating persistent alternating periods of high and low flow, contrary to random, short-term models. AGU Publications +5

Continue...


Component 1
Component 2
Component 3
Component 4
Time
3
2
1
0
−1
−2
600
700
800
900
1000
1100
1200
1300

FGN

- The FGN is just a mathematical model
- The Hurst parameter H determine the properties of the simpler/Markov components
- The interpretation is through the simpler/Markov components, not that this is “long memory”

FGN

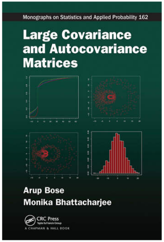
- The FGN is just a mathematical model
- The Hurst parameter H determine the properties of the simpler/Markov components
- The interpretation is through the simpler/Markov components, not that this is “long memory”

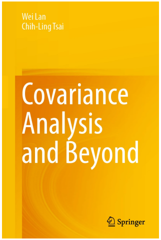
FGN

- The FGN is just a mathematical model
- The Hurst parameter H determine the properties of the simpler/Markov components
- The interpretation is through the simpler/Markov components, not that this is “long memory”

Covariance or Precision, does it matter?

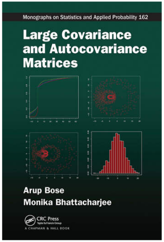
- Yes!
- Covariance is over-sold in the classical literature
- Precision isn't sold (nearly) at all!
- Covariance relates to independence
- Precision relates to conditional independence
- Correlations induced through conditioning, or sharing random variables
- If your model is *not* Markov, then be sure that you are doing it "right"
- The computational cost really depends on the covariance/precision choice

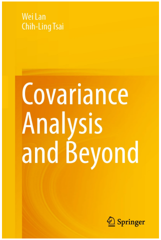




Covariance or Precision, does it matter?

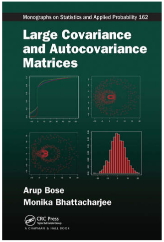
- Yes!
- Covariance is over-sold in the classical literature
- Precision isn't sold (nearly) at all!
- Covariance relates to independence
- Precision relates to conditional independence
- Correlations induced through conditioning, or sharing random variables
- If your model is *not* Markov, then be sure that you are doing it "right"
- The computational cost really depends on the covariance/precision choice

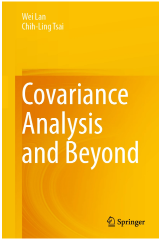




Covariance or Precision, does it matter?

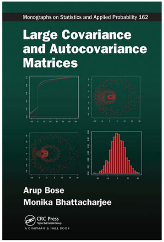
- Yes!
- Covariance is over-sold in the classical literature
- Precision isn't sold (nearly) at all!
- Covariance relates to independence
- Precision relates to conditional independence
- Correlations induced through conditioning, or sharing random variables
- If your model is *not* Markov, then be sure that you are doing it "right"
- The computational cost really depends on the covariance/precision choice

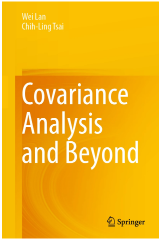




Covariance or Precision, does it matter?

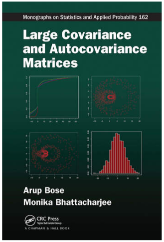
- Yes!
- Covariance is over-sold in the classical literature
- Precision isn't sold (nearly) at all!
- Covariance relates to independence
- Precision relates to conditional independence
- Correlations induced through conditioning, or sharing random variables
- If your model is *not* Markov, then be sure that you are doing it "right"
- The computational cost really depends on the covariance/precision choice

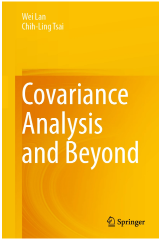




Covariance or Precision, does it matter?

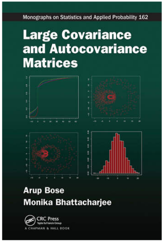
- Yes!
- Covariance is over-sold in the classical literature
- Precision isn't sold (nearly) at all!
- Covariance relates to independence
- Precision relates to conditional independence
- Correlations induced through conditioning, or sharing random variables
- If your model is *not* Markov, then be sure that you are doing it "right"
- The computational cost really depends on the covariance/precision choice

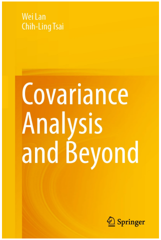




Covariance or Precision, does it matter?

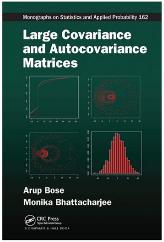
- Yes!
- Covariance is over-sold in the classical literature
- Precision isn't sold (nearly) at all!
- Covariance relates to independence
- Precision relates to conditional independence
- Correlations induced through conditioning, or sharing random variables
- If your model is *not* Markov, then be sure that you are doing it "right"
- The computational cost really depends on the covariance/precision choice

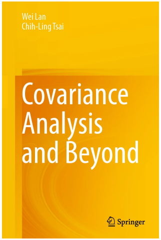




Covariance or Precision, does it matter?

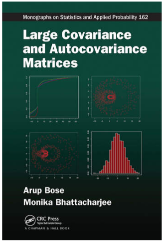
- Yes!
- Covariance is over-sold in the classical literature
- Precision isn't sold (nearly) at all!
- Covariance relates to independence
- Precision relates to conditional independence
- Correlations induced through conditioning, or sharing random variables
- If your model is *not* Markov, then be sure that you are doing it "right"
- The computational cost really depends on the covariance/precision choice

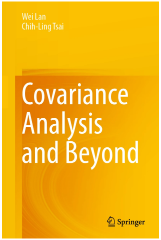




Covariance or Precision, does it matter?

- Yes!
- Covariance is over-sold in the classical literature
- Precision isn't sold (nearly) at all!
- Covariance relates to independence
- Precision relates to conditional independence
- Correlations induced through conditioning, or sharing random variables
- If your model is *not* Markov, then be sure that you are doing it "right"
- The computational cost really depends on the covariance/precision choice





What's tomorrow?

- Continous time
- Continous space
- Where do GMRFs come now?
- Generalisation: how simple ideas can solve hard problems
- ...and the solution is still a GMRF
- (Approximate) Bayesian inference
- Variational Bayes
- Numerics for sparse (precision-)matrices
- New sparse solver: *sTiles*

What's tomorrow?

- Continous time
- Continous space
- Where do GMRFs come now?
- Generalisation: how simple ideas can solve hard problems
- ...and the solution is still a GMRF
- (Approximate) Bayesian inference
- Variational Bayes
- Numerics for sparse (precision-)matrices
- New sparse solver: *sTiles*

What's tomorrow?

- Continous time
- Continous space
- Where do GMRFs come now?
- Generalisation: how simple ideas can solve hard problems
- ...and the solution is still a GMRF
- (Approximate) Bayesian inference
- Variational Bayes
- Numerics for sparse (precision-)matrices
- New sparse solver: *sTiles*

What's tomorrow?

- Continous time
- Continous space
- Where do GMRFs come now?
- Generalisation: how simple ideas can solve hard problems
- ...and the solution is still a GMRF
- (Approximate) Bayesian inference
- Variational Bayes
- Numerics for sparse (precision-)matrices
- New sparse solver: *sTiles*

What's tomorrow?

- Continous time
- Continous space
- Where do GMRFs come now?
- Generalisation: how simple ideas can solve hard problems
- ...and the solution is still a GMRF
- (Approximate) Bayesian inference
- Variational Bayes
- Numerics for sparse (precision-)matrices
- New sparse solver: *sTiles*

What's tomorrow?

- Continous time
- Continous space
- Where do GMRFs come now?
- Generalisation: how simple ideas can solve hard problems
- ...and the solution is still a GMRF
- (Approximate) Bayesian inference
- Variational Bayes
- Numerics for sparse (precision-)matrices
- New sparse solver: *sTiles*

What's tomorrow?

- Continous time
- Continous space
- Where do GMRFs come now?
- Generalisation: how simple ideas can solve hard problems
- ...and the solution is still a GMRF
- (Approximate) Bayesian inference
- Variational Bayes
- Numerics for sparse (precision-)matrices
- New sparse solver: *sTiles*

What's tomorrow?

- Continous time
- Continous space
- Where do GMRFs come now?
- Generalisation: how simple ideas can solve hard problems
- ...and the solution is still a GMRF
- (Approximate) Bayesian inference
- Variational Bayes
- Numerics for sparse (precision-)matrices
- New sparse solver: *sTiles*

What’s tomorrow?

- Continous time
- Continous space
- Where do GMRFs come now?
- Generalisation: how simple ideas can solve hard problems
- ...and the solution is still a GMRF
- (Approximate) Bayesian inference
- Variational Bayes
- Numerics for sparse (precision-)matrices
- New sparse solver: *sTiles*

Thank you • شكرا

جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology