Lecture 2: Definitions and properties

Gaussian Markov random fields

David Bolin
Chalmers University of Technology
January 20, 2015

Some basic notation

If $x = (x_1, \ldots, x_n)$ and $A \subset I = \{1, \ldots, n\}$, we let

$$x_A = \{x_i : i \in A\}$$
$$x_{-A} = \{x_i : i \in I \setminus A\}$$

If x and y are independent, we have

$$\pi(x, y) = \pi(x)\pi(y),$$

and denote this by $x \perp y$.
If x and y are conditionally independent given z, we have

$$\pi(x, y|z) = \pi(x|z)\pi(y|z),$$

and denote this by $x \perp y|z$.
x and y may be marginally dependent while being conditionally independent given x.We have

$$x \perp y|z \Leftrightarrow \pi(x, y|z) = f(x, z)g(y, z)$$

for some functions f and g for all z with $\pi(z) > 0$.

Undirected graphs

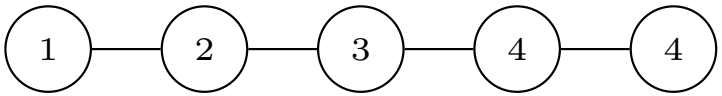
We typically use undirected graphs to represent conditional independence.

An undirected graph $\mathcal{G}$ consists of

$\mathcal{V}$ A set of nodes, $1, \ldots, n$.

$\mathcal{E}$ A set of edges $\{i, j\}$, where $i \neq j \in \mathcal{V}$

Example for AR(1) process



We say that a graph is fully connected if $\{i, j\} \in \mathcal{E} \, \forall \, i \neq j \in \mathcal{V}$.

The neighbors of a node i are all nodes in $\mathcal{G}$ having an edge to i:

$$\text{ne}(i) = \{j \in \mathcal{V} : \{i, j\} \in \mathcal{E}\}.$$

We write $i \sim j$ if i and j are neighbors.

The formal definition of a GMRF

Definition

A random vector $\mathbf{x} = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$ is called a GMRF with respect to a labelled graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ with mean μ and positive definite precision matrix $\mathbf{Q}$ if and only if its density has the form

$$\pi(\mathbf{x}) = (2\pi)^{-n/2}|\mathbf{Q}|^{1/2} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{Q}(\mathbf{x}-\boldsymbol{\mu})\right)$$

and

$$Q_{ij} \neq 0 \Longleftrightarrow \{i,j\} \in \mathcal{E} \text{ for all } i \neq j$$

Note that if $\mathbf{Q}$ is dense, then $\mathcal{G}$ is fully connected. Thus, any normal distribution is a GMRF and vice versa.

The interesting case is when $\mathcal{G}$ is *not* fully connected.

The graph connection

The graph $\mathcal{G}$ is used to represent the conditional independence properties of the GMRF.

We know that the covariance matrix only directly gives us information about *marginal* dependence structure, and not the conditional structure.

It turns out that the precision matrix $\mathbf{Q}$ plays a crucial role for the conditional structure.

Theorem 2.2

Let $x \sim \mathsf{N}(\boldsymbol{\mu}, \mathbf{Q}^{-1})$. Then for $i \neq j$,

$$i \perp j | \mathbf{x}_{-ij} \Longleftrightarrow Q_{i,j} = 0.$$

Thus, the non-zero pattern of $\mathbf{Q}$ both determines $\mathcal{G}$ and the conditional structure: We can simply read of from $\mathbf{Q}$ whether two nodes are conditionally independent.

The precision matrix

The natural parametrisation of a GMRF is in terms of its precision matrix, which has nice conditional interpretations

Theorem 2.3

Let $x \sim \mathsf{N}(\boldsymbol{\mu}, \mathbf{Q}^{-1})$. Then

$$\mathsf{E}(x_i|\mathbf{x}_{-i}) = \mu_i - \frac{1}{Q_{ii}} \sum_{j:j\sim i} Q_{ij}(x_j - \mu_j)$$

$$\mathrm{Prec}(x_i|\mathbf{x}_{-i}) = \mathrm{Var}(x_i|\mathbf{x}_{-i})^{-1} = Q_{ii}$$

$$\mathrm{Corr}(x_i, x_j|\mathbf{x}_{-ij}) = -\frac{Q_{ij}}{\sqrt{Q_{ii}Q_{jj}}}$$

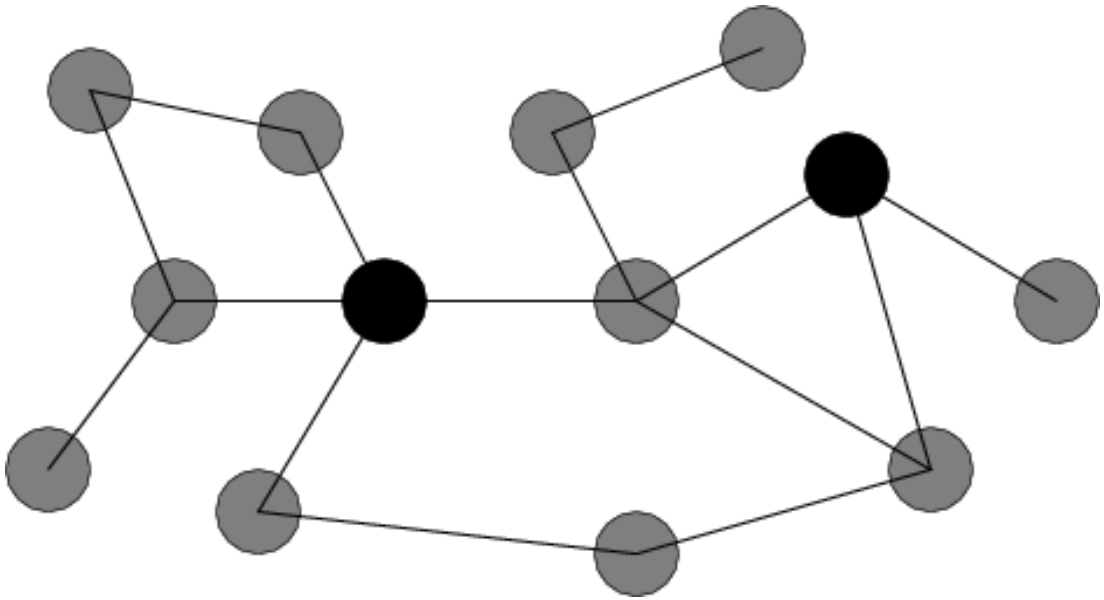
It should be noted that it basically is impossible to interpret the elements in $\mathbf{Q}$ marginally, since $\boldsymbol{\Sigma} = \mathbf{Q}^{-1}$ holds the marginal information.

Markov properties

Let $\mathbf{x}$ be a GMRF wrt $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Then the following are equivalent.

The pairwise Markov property:

$$x_i \perp x_j \mid \mathbf{x}_{-ij} \quad \text{if } \{i, j\} \notin \mathcal{E} \text{ and } i \neq j.$$

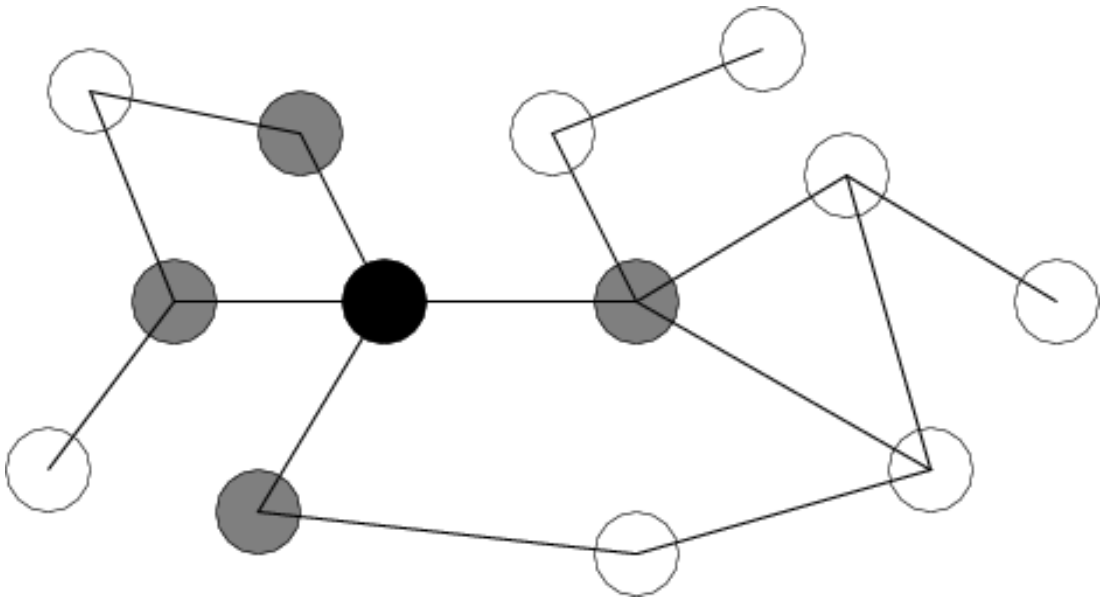


Markov properties

Let $\mathbf{x}$ be a GMRF wrt $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Then the following are equivalent.

The local Markov property:

$$x_i \perp \mathbf{x}_{-\{i,\mathsf{ne}(i)\}} \mid \mathbf{x}_{\mathsf{ne}(i)} \quad \text{for every } i \in \mathcal{V}.$$



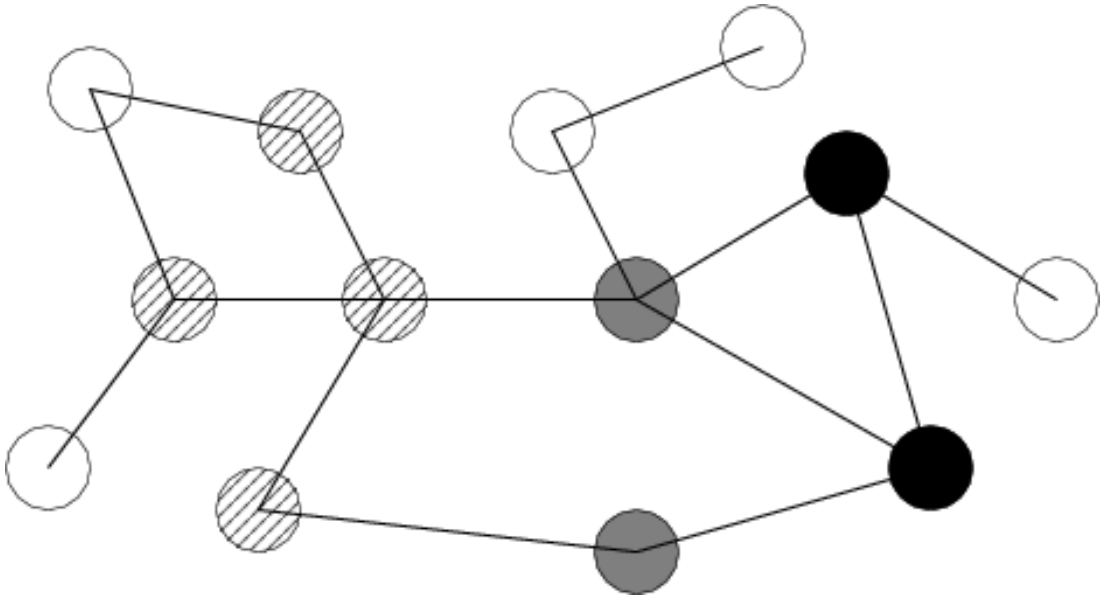
Markov properties

Let $\mathbf{x}$ be a GMRF wrt $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Then the following are equivalent.

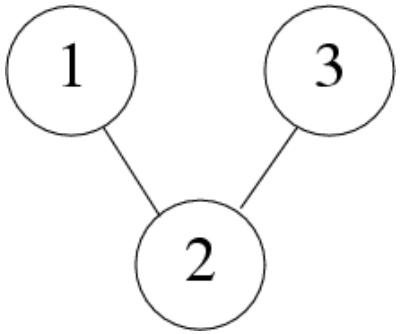
The global Markov property:

$$\mathbf{x}_A \perp \mathbf{x}_B \mid \mathbf{x}_C$$

for all disjoint sets A, B and C where C separates A and B, and A and B are non-empty.



(Induced) subgraph



Let $A \subset \mathcal{V}$

$\mathcal{G}^A$ denote the graph restricted to A.

- remove all nodes not belonging to A, and
- all edges where at least one node does not belong to A

Example

$\mathcal{A} = \{1, 2\}$, then

$$\mathcal{V}^A = \{1, 2\} \quad \text{and} \quad \mathcal{E}^A = \{\{1, 2\}\}$$

Conditional distributions

Theorem 2.5

Let $\mathcal{V} = A \cup B$ where $A \cap B = \emptyset$, and let $\mathbf{x}$ be a GMRF wrt $\mathcal{G}$, with

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_A \\ \mathbf{x}_B \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{pmatrix}, \quad \mathbf{Q} = \begin{pmatrix} \mathbf{Q}_{AA} & \mathbf{Q}_{AB} \\ \mathbf{Q}_{BA} & \mathbf{Q}_{BB} \end{pmatrix}.$$

$\mathbf{x}_A | \mathbf{x}_B$ is then a GMRF wrt the subgraph $\mathcal{G}^A$ with parameters $\boldsymbol{\mu}_{A|B}$ and $\mathbf{Q}_{A|B} > 0$, where

$$\boldsymbol{\mu}_{A|B} = \boldsymbol{\mu}_A - \mathbf{Q}_{AA}^{-1}\mathbf{Q}_{AB}(\mathbf{x}_B - \boldsymbol{\mu}_B) \quad \text{and} \quad \mathbf{Q}_{A|B} = \mathbf{Q}_{AA}.$$

Conditional distributions II

$$\boldsymbol{\mu}_{A|B} = \boldsymbol{\mu}_A - \mathbf{Q}_{AA}^{-1}\mathbf{Q}_{AB}(\mathbf{x}_B - \boldsymbol{\mu}_B) \quad \text{and} \quad \mathbf{Q}_{A|B} = \mathbf{Q}_{AA}.$$

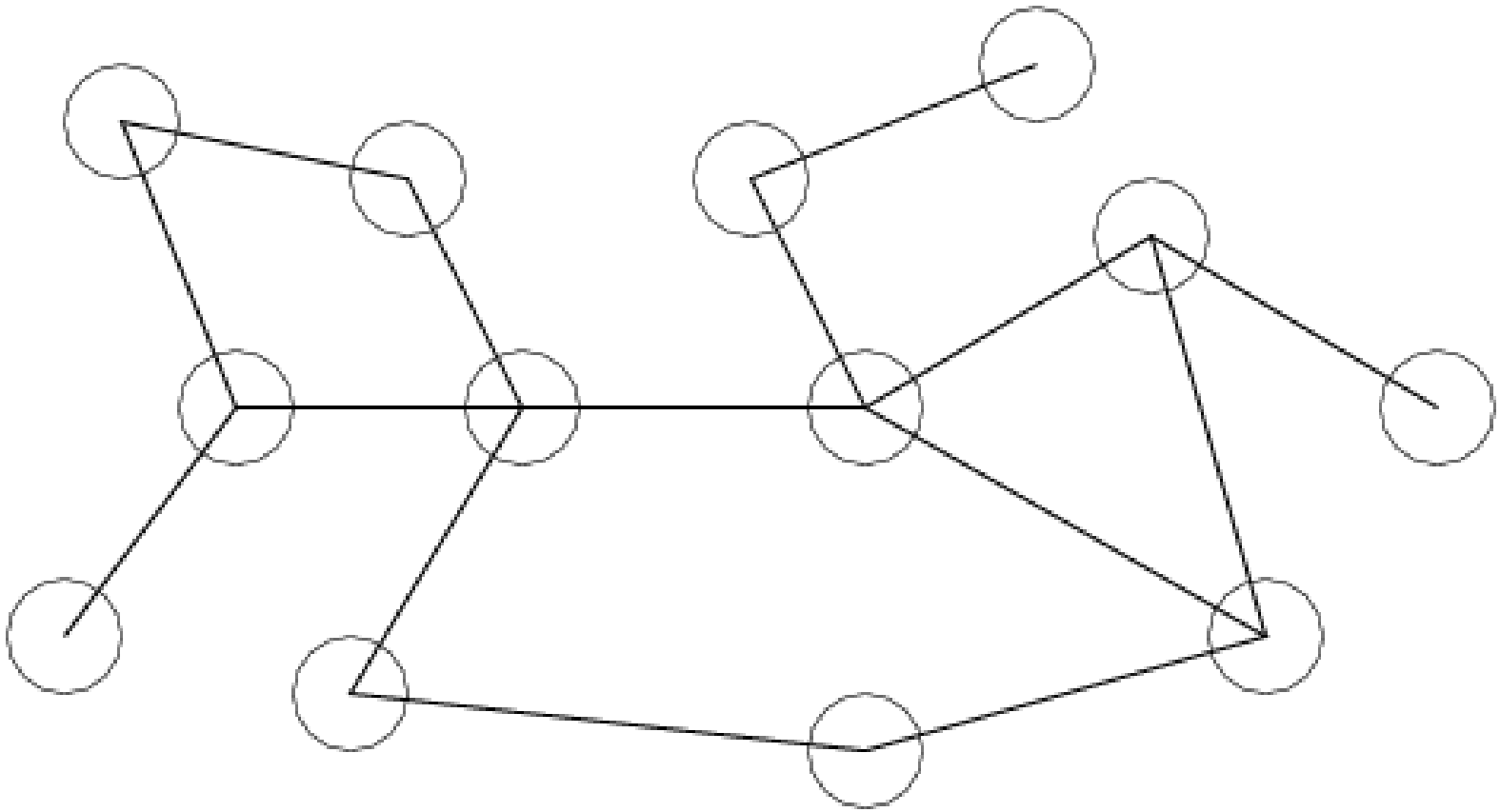
- We have explicit knowledge $\mathbf{Q}_{A|B}$ as the principal matrix $\mathbf{Q}_{AA}$.
- The subgraph $\mathcal{G}^A$ does not change the structure, only removes nodes and edges in A.
- The conditional mean only depends on nodes in $A \cup \mathsf{ne}(A)$.
- If $\mathbf{Q}_{AA}$ is sparse, then $\boldsymbol{\mu}_{A|B}$ is the solution of a sparse linear system

$$\mathbf{Q}_{AA}(\boldsymbol{\mu}_{A|B} - \boldsymbol{\mu}_A) = -\mathbf{Q}_{AB}(\mathbf{x}_B - \boldsymbol{\mu}_B)$$

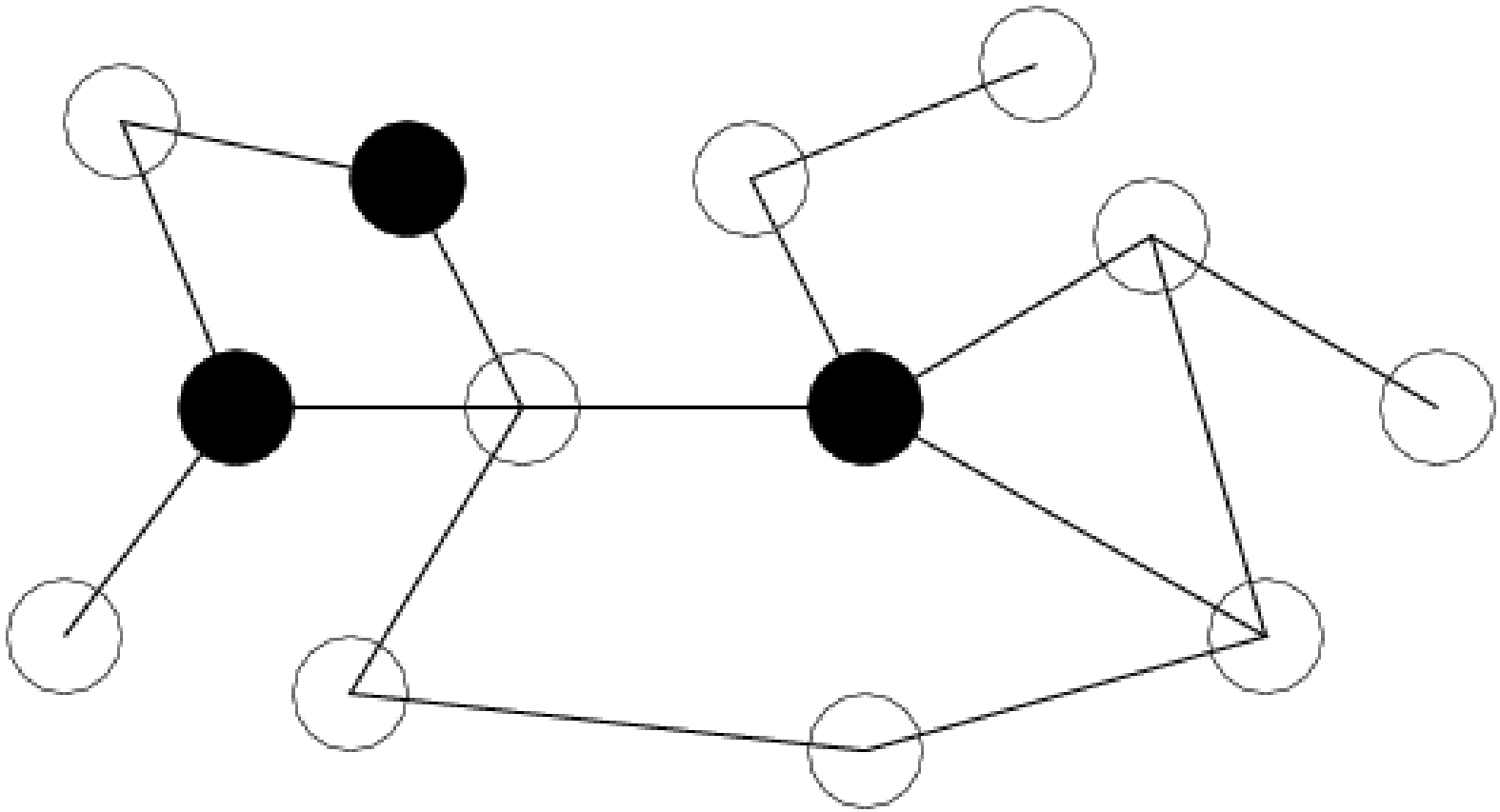
- This result should be compared to the standard covariance-based formula

$$\mathbf{x}_A|\mathbf{x}_B \sim \mathsf{N}(\boldsymbol{\mu}_A + \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}(\mathbf{x}_B - \boldsymbol{\mu}_B), \boldsymbol{\Sigma}_{AA} - \boldsymbol{\Sigma}_{AB}\boldsymbol{\Sigma}_{BB}^{-1}\boldsymbol{\Sigma}_{BA})$$

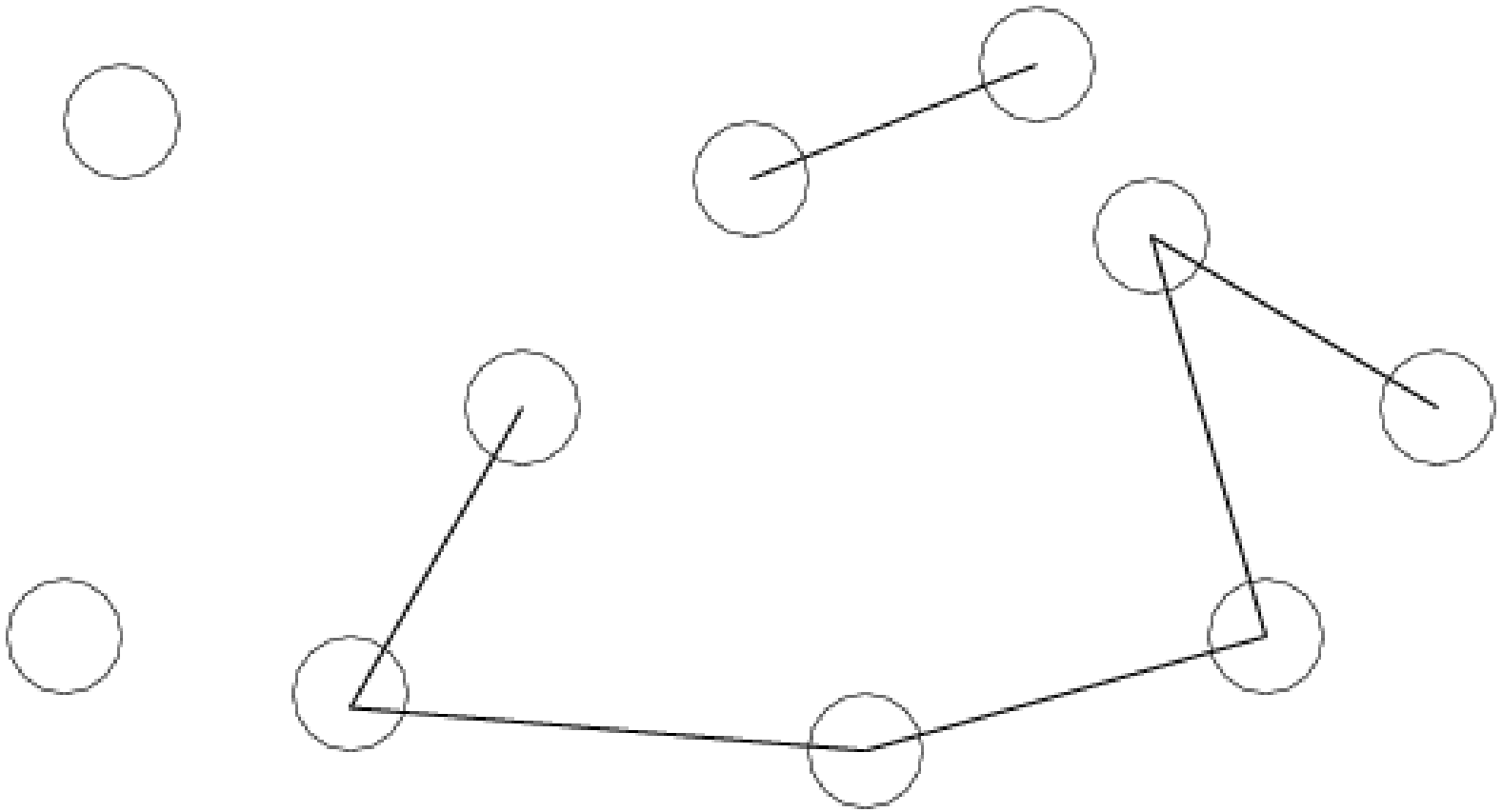
The induced subgraph



The induced subgraph



The induced subgraph



The canonical parameterisation

Definition

A GMRF $\mathbf{x}$ wrt $\mathcal{G}$ having a canonical parameterisation $(\mathbf{b}, \mathbf{Q})$, has density

$$\pi(\mathbf{x}) \propto \exp\Big(-\frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{b}^T\mathbf{x}\Big) \tag{1}$$

ie, precision matrix $\mathbf{Q}$ and mean $\mu = \mathbf{Q}^{-1}\mathbf{b}$. We write this as

$$\mathbf{x} \sim \mathsf{N}_C(\mathbf{b}, \mathbf{Q}). \tag{2}$$

The relation to the Gaussian distribution, is that

$$\mathcal{N}(\boldsymbol{\mu}, \mathbf{Q}^{-1}) = \mathsf{N}_C(\mathbf{Q}\boldsymbol{\mu}, \mathbf{Q})$$

The canonical parameterisation II

The following properties of the canonical parameterisation are highly useful in practise

- If $\mathbf{x} \sim \mathsf{N}_C(\mathbf{b}, \mathbf{Q})$, then

$$\mathbf{x}_A | \mathbf{x}_B \sim \mathsf{N}_C(\mathbf{b}_A - \mathbf{Q}_{AB}\mathbf{x}_B, \mathbf{Q}_{AA})$$

- Let $\mathbf{x} \sim \mathsf{N}_C(\mathbf{b}, \mathbf{Q})$ and $\mathbf{y}|\mathbf{x} \sim \mathsf{N}(\mathbf{x}, \mathbf{P}^{-1})$, then

$$\mathbf{x}|\mathbf{y} \sim \mathsf{N}_C(\mathbf{b} + \mathbf{P}\mathbf{y}, \mathbf{Q} + \mathbf{P})$$

During successive conditioning, we can update the canonical parameterisation without explicitly computing the mean.

Conditional autoregressions (CAR models)

- Recall the conditional specification of the AR(1) process:

$$x_t = \phi x_{t-1} + \epsilon_t, \qquad t = 1, 2, \ldots, \qquad \epsilon_t \sim \mathcal{N}(0, 1)$$

- The joint density for $\mathbf{x}$ could then be calculated using

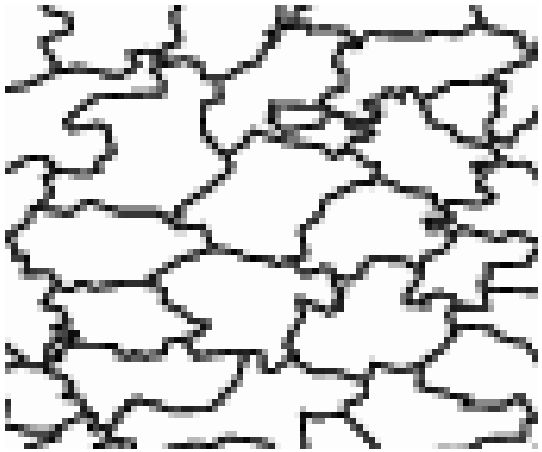
$$\pi(\mathbf{x}) = \pi(x_0)\pi(x_1|x_2) \cdots \pi(x_{n-1}|x_{n-2}) \tag{3}$$

- Often, there is no natural ordering of the nodes in $\mathbf{x}$. For example if x_i is the value of pixel i in an image or the relative risk of a disease in county i.
- It is then no longer useful to use a formulation like (3) to specify the joint distribution.
- A common approach is then to implicitly specify the GMRF through the full conditionals

$$\{\pi(x_i|\mathbf{x}_{-i}),\ i = 1, \ldots, n\}$$

- These models are called conditional autoregressions, or CAR models, and was pioneered by Julian Besag in the mid 70s.

In general



Specify the full conditionals as normals with

$$\mathrm{E}(x_i \mid \mathbf{x}_{-i}) = \mu_i - \sum_{j:j\sim i} \beta_{ij}(x_j - \mu_j) \quad \text{and} \tag{4}$$

$$\mathrm{Prec}(x_i \mid \mathbf{x}_{-i}) = \kappa_i > 0 \tag{5}$$

for $i = 1, \ldots, n$, for some $\{\beta_{ij}, i \neq j\}$, and vectors $\boldsymbol{\mu}$ and κ.
Clearly, $\sim$ is defined implicitly by the nonzero terms of $\{\beta_{ij}\}$.

- Even though the "full conditionals" are well defined, are we sure that the *joint density* exists and is unique?
- What are the compatibility conditions required for a joint GMRF to exists with the prescribed Markov properties?

Theorem 2.6

The normal conditionals with means and precisions

$$\mathrm{E}(x_i \mid \mathbf{x}_{-i}) = \mu_i - \sum_{j:j\sim i} \beta_{ij}(x_j - \mu_j) \quad \text{and} \quad \mathrm{Prec}(x_i \mid \mathbf{x}_{-i}) = \kappa_i > 0$$

specify a GMRF $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and precision $\mathbf{Q}$ with elements $Q_{ii} = \kappa_i$, and $Q_{ij} = \kappa_i\beta_{ij}$ given that $\kappa_i\beta_{ij} = \kappa_j\beta_{ji}$ for all $i \neq j$ and that $\mathbf{Q}$ is positive definite.

A proof based on Brook's lemma

Lemma (Brook's lemma)

Let $\pi(\mathbf{x})$ *be the density for* $\mathbf{x} \in \mathbb{R}^n$ *and define* $\Omega = \{\mathbf{x} \in \mathbb{R}^n : \pi(\mathbf{x}) > 0\}$. *Let* $\mathbf{x}, \mathbf{x}' \in \Omega$, *then*

$$\begin{aligned}\frac{\pi(\mathbf{x})}{\pi(\mathbf{x}')} &= \prod_{i=1}^{n} \frac{\pi(x_i|x_1,\ldots,x_{i-1},x'_{i+1},\ldots,x'_n)}{\pi(x'_i|x_1,\ldots,x_{i-1},x'_{i+1},\ldots,x'_n)} \\ &= \prod_{i=1}^{n} \frac{\pi(x_i|x'_1,\ldots,x'_{i-1},x_{i+1},\ldots,x_n)}{\pi(x'_i|x'_1,\ldots,x'_{i-1},x_{i+1},\ldots,x_n)}.\end{aligned}$$

The valid parameter space

- Positive definiteness requires $\kappa_i > 0$ for all i but also further complicated contraints on β_{ij}.
- In general, if $\mathbf{Q}(\boldsymbol{\theta})$ depends on parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, we need to restrict $\boldsymbol{\theta}$ to the valid parameter space

$$\boldsymbol{\Theta}^+ = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : \mathbf{Q}(\boldsymbol{\theta}) > 0\}$$

- We can always check if $\boldsymbol{\theta} \in \boldsymbol{\Theta}^+$ by direct verification, but this is expensive.
- Finding $\boldsymbol{\Theta}^+$ is in general difficult.
- A common approach is to demand that $\mathbf{Q}$ is diagonal dominant: $Q_{ii} > \sum_{j:j\neq i} |Q_{ij}|$ for each i.
- This sufficient but *not* necessary for positive definiteness, thus it gives a subspace of $\boldsymbol{\Theta}^+$.