Outline

- Template models (6.3-6.5)
- Structural uncertainty (6.6)
- Multivariate Gaussians (7.1)
- Gaussian DAGs (7.2)
- Gaussian MRFs (7.3)
Parameter tying

- A DBN defines a distribution over an unboundedly large number of variables by assuming that they all share the same CPDs.
- This is called parameter tying (weight sharing).
- It is useful even for fixed sized models in order to help learning (pool the sufficient statistics).
- We now discuss notational conventions (“syntactic sugar”) for representing large “unrolled” networks with shared parameters.
Plates

- Plates are useful for specifying simple repetitive patterns, as frequently arise in hierarchical Bayesian models.
Grade(s,c) in \{A,B,C\} is encoded on edges. Cf discrete probabilistic matrix factorization
Limitations of plates

- There are various structures that plates cannot represent
- Eg DBNs
- Eg genotype(x1) depends on genotype(x2), where x2=parent(x1)
- We can write programs to generate graphs of specified structure, but we would like a declarative representation language for such repetitive patterns so that no new code has to be written
Beyond plates

- Probabilistic Relational Models (PRMs) encode large DAG models with tied CPDs
- Relational Markov Networks encode large MRFs with tied factors
- Markov Logic Networks are like RMNs, except the factors are represented in log-linear form, and the features are represented as logical expressions
### Table I. Example of a first-order knowledge base and MLN. Fr() is short for Friends(), Sm() for Smokes(), and Ca() for Cancer().

<table>
<thead>
<tr>
<th>English</th>
<th>First-Order Logic</th>
<th>Clausal Form</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Friends of friends are friends.</td>
<td>$\forall x \forall y \forall z \ Fr(x, y) \land Fr(y, z) \Rightarrow Fr(x, z)$</td>
<td>$\neg Fr(x, y) \lor \neg Fr(y, z) \lor Fr(x, z)$</td>
<td>0.7</td>
</tr>
<tr>
<td>Friendless people smoke.</td>
<td>$\forall x \ (\neg (\exists y Fr(x, y)) \Rightarrow Sm(x))$</td>
<td>$Fr(x, g(x)) \lor Sm(x)$</td>
<td>2.3</td>
</tr>
<tr>
<td>Smoking causes cancer.</td>
<td>$\forall x \ Sm(x) \Rightarrow Ca(x)$</td>
<td>$\neg Sm(x) \lor Ca(x)$</td>
<td>1.5</td>
</tr>
<tr>
<td>If two people are friends, either both smoke or neither does.</td>
<td>$\forall x \forall y Fr(x, y) \Rightarrow (Sm(x) \leftrightarrow Sm(y))$</td>
<td>$\neg Fr(x, y) \lor Sm(x) \lor \neg Sm(y)$, $\neg Fr(x, y) \lor \neg Sm(x) \lor Sm(y)$</td>
<td>1.1</td>
</tr>
</tbody>
</table>

![Markov Logic Network Diagram](image)
Directed vs undirected models

• Undirected models are simpler: no need to worry about cycles, lots of freedom in defining factors
• However, in a UG, the probability of a node depends on the *size* of the graph and/or its connectivity, even if all the other nodes are hidden.
• This may not be desirable.

\[
\begin{align*}
X_1 &\rightarrow X_2 \rightarrow X_3 \\
X_1 &\rightarrow X_2 \rightarrow \ldots \rightarrow X_{10} \\
p(X_2) &\text{ same} \\
X_1 &- X_2 - \ldots - X_{10} \\
p(X_2) &\text{ different}
\end{align*}
\]
Structural uncertainty

- For a fixed domain, if we do not know the graph structure, we may estimate it using model selection.
- But for relational domains, the structure may change depending on the values of the nodes.
- Eg. Genotype(x1) -> genotype(x2) is only active if parent(x1,x2)=true.
- In addition, we may be uncertain about how many objects exist in the world.
- Eg. In tracking, 3 blips on the radar is consistent with {0,1,…, infty} objects in the world!
Data association ambiguity
Citation matching

Are these the same article?
Huge industry concerned with database merging


• Assumes there is an unknown number of authors and papers, which generates the observed set of citation strings.
UG model

• No unknown objects. Just enforce that citations are the same.
• Need 3 way factor to encode transitivity of sameness relation: $S(c_1,c_2)$, and $S(c_2,c_3) \Rightarrow S(c_1,c_3)$
• And if 2 docs are same, text should be similar: $\text{Factor}(s(c_1,c_2), T(c_1), T(c_2))$
MVN: 2 parameterizations

- **Moment form**

\[
\mathcal{N}(x|\mu, \Sigma) \overset{\text{def}}{=} \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right]
\]

- **Information (canonical) form**

\[
\Lambda \overset{\text{def}}{=} \Sigma^{-1} \quad \eta \overset{\text{def}}{=} \Sigma^{-1} \mu
\]

\[
\mathcal{N}(x|\eta, \Lambda) = \frac{|\Lambda|^{1/2}}{(2\pi)^{d/2}} \exp\left[-\frac{1}{2}(x^T \Lambda x + \eta^T \Lambda^{-1} \eta - 2x^T \eta)\right]
\]

\[
= \exp\left[c - \frac{1}{2}x^T \Lambda x + x^T \eta\right]
\]
Moment and canonical form

- Canonical form is denoted
  \[ x \sim \mathcal{N}_C(b, Q) \iff p(x) \propto \exp \left( -\frac{1}{2} x^T Q x + b^T x \right) \]

- Moment form
  \[ \mathcal{N}(\mu, Q^{-1}) = \mathcal{N}_C(Q\mu, Q) \]
Thm 7.1.3. Let $X \sim \text{MVN}$. $X_i \perp X_j$ iff $\Sigma_{i,j}=0$

Thm 7.1.4. Let $X \sim \text{MVN}$ with info matrix $J$. Then $J_{i,j}=0$ iff $X_i \perp X_j | X_{-ij}$

Factorization thm.

$$x \perp y | z \iff p(x, y, z) = f(x, vz)g(y, vz)$$
Indep => uncorrelated

- Ex 7.2.1. For any \( p(X,Y) \), if \( X \perp Y \) then \( \text{Cov}[X,Y]=0 \).

\[
\text{Cov}[x,y] = \int \int p(x,y)(x - \bar{x})(y - \bar{y})dxdy \\
= (\int p(x)(x - \bar{x})dx)(\int p(y)(y - \bar{y})dy) \\
= (\bar{x} - \bar{x})(\bar{y} - \bar{y}) = 0
\]
Uncorrelated & MVN => indep

Ex 7.2.2. If \( p(X,Y) \) is Gaussian, and \( \text{Cov}[X,Y]=0 \), then \( X \perp Y \).

Pf. The bivariate Gaussian can be written as

\[
p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2}
\right.ight.
\]
\[
\left.\left.-2\rho \frac{(x_1 - \mu_1)}{\sigma_1} \frac{(x_2 - \mu_2)}{\sigma_2}\right]\right]
\]

If \( \rho=0 \), then

\[
p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left[-\frac{1}{2}\left(\frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2}\right)\right]
\]
\[
= f(x_1)g(x_2)
\]

Hence by factorization thm, \( x_1 \perp x_2 \).
Uncorrelated not imply independent

• Ex 7.2.3. Find an example where Cov\([X,Y]=0\) yet not \(X \perp Y\).

• Let \(X \sim U(-1,1)\) and \(Y=X^2\). Clearly \(Y\) is dependent on \(X\) yet one can show (exercise) that \(\text{Cov}(X,Y)=0\).

• Let \(X \sim N(0,1)\) and \(Y=W X, p(W=-1)=p(W=1)=0.5\). Clearly \(Y\) is dependent on \(X\), yet one can show (exercise) that \(Y \sim N(0,1)\) and \(\text{Cov}[X,Y]=0\).
Independencies in MVN

- Thm 7.1.3. Let $X \sim \text{MVN}$. $X_i \perp X_j$ iff $\Sigma_{i,j}=0$

- Pf. By ex 7.2.1, we have $\Rightarrow$ direction.
- By ex 7.2.2, we have that $\Rightarrow$ direction.
- By ex 7.2.3, we have that $X \sim \text{MVN}$ is necessary for $\Rightarrow$ direction to work.
Conditional Independencies in MVN

• Thm 7.1.4. let \( X \sim \text{MVN} \) with info matrix \( J \). Then 
  \( J_{i,j}=0 \) iff \( X_i \perp X_j \mid X_{-ij} \)

• Pf. Let \( \mu=0 \).

\[
p(x_i, x_j, x_{-ij}) \propto \exp\left(-\frac{1}{2} \sum_{k,l} x_k Q_{kl} x_l \right)
\]

\[
\propto \exp \left( -\frac{1}{2} x_i x_j (Q_{ij} + Q_{ji}) - \frac{1}{2} \sum_{\{k,l\} \neq \{i,j\}} x_k Q_{kl} x_l \right)
\]

• The second term does not involve \( x_i x_j \), and nor does the first iff \( Q_{ij}=0 \). Hence this factorizes into 
  \( f(x_i, x_{-ij}) \ g(x_j, x_{-ij}) \) iff \( Q_{ij}=0 \). QED.
Structural zeros

Zeros in the precision matrix correspond to missing edges in the UGM

\[
\Sigma = \begin{pmatrix}
4 & 2 & -2 \\
2 & 5 & -5 \\
-2 & -5 & 8
\end{pmatrix}, \quad \Lambda = \Sigma^{-1} = \begin{pmatrix}
0.3125 & -0.125 & 0 \\
-0.125 & 0.5833 & 0.3333 \\
0 & 0.3333 & 0.3333
\end{pmatrix}
\]

\[X_1 - X_2 - X_3\]
# Marginals and conditionals

| Marginal $p(x_2)$ | Conditional $p(x_2|x_1)$ |
|-------------------|--------------------------|
| Marginal $p(x_2)$ | $\mathcal{N}(x_2|\mu_2, \Sigma_2)$ |
| Moment            | $\mathcal{N}(x_2|\eta_2 - \Lambda_{21}\Lambda_{11}^{-1}\eta_1, \Lambda_{22} - \Lambda_{21}\Lambda_{11}^{-1}\Lambda_{12})$ |
| Info              | $\mathcal{N}(x_2|\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$ |
|                   | $\mathcal{N}(x_2|\eta_1 - \Lambda_{12}x_2, \Lambda_{11})$ |

Marginalization easy in moment form.
 Conditioning easy in canonical form.
Conditioning in canonical form

• Thm (Conditioning).

\[ x \sim \mathcal{N}_C(b, Q) \Rightarrow x_A|x_B \sim \mathcal{N}_C(b_A - Q_{AB}x_B, Q_{AA}) \]

• Thm (soft conditioning).

\[ x \sim \mathcal{N}_C(b, Q) \quad \text{and} \quad y|x \sim \mathcal{N}(x, P^{-1}) \]

\[ x|y \sim \mathcal{N}_C(b + Py, Q + P) \quad \text{Precisions add} \]

• We can accumulate evidence by addition of matrix-vector products, and then compute posterior mean at end by solving \( Qb = \mu \).
Partial correlation coefficient

• Let $X \sim \text{Mvn}$ with precision matrix 

$$\Omega = \Sigma^{-1} = \begin{pmatrix} \omega_{11} & \cdots & \omega_{1d} \\ \vdots & \ddots & \vdots \\ \omega_{d1} & \cdots & \omega_{dd} \end{pmatrix}$$

• The conditional distribution $p(x_1, x_2 | x_3, \ldots, x_d)$ is bivariate Gaussian with covariance 

$$\begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix}^{-1} = \frac{1}{\omega_{11}\omega_{22} - (\omega_{12})^2} \begin{pmatrix} \omega_{22} & -\omega_{12} \\ -\omega_{21} & \omega_{11} \end{pmatrix}$$

• The partial correlation coefficient is given by 

$$\rho_{1,2|3,\ldots,d} \overset{\text{def}}{=} \frac{\text{Cov}[X_1, X_2 | X_{3:d}]}{\sqrt{\text{Var}[X_1 | X_{3:d}]\text{Var}[X_2 | X_{3:d}]}} = \frac{-\omega_{21}}{\sqrt{\omega_{11}\omega_{22}}}$$
Conditioning in moment form

• Thm (Rue&Held p26).

\[ \mathbf{x} \sim \mathcal{N}(\mu, \mathbf{Q}^{-1}) \Rightarrow \]
\[ \mathbf{x}_A|\mathbf{x}_B \sim \mathcal{N}(\mu_{A|B}, \mathbf{Q}_{AA}) \]
\[ \mu_{A|B} = \mu_A - \mathbf{Q}_{AA}^{-1} \mathbf{Q}_{AB}(\mathbf{x}_B - \mu_B) \]

• Thus to find the mean we need to solve the linear system

\[ \mathbf{Q}_{AA}\mu_{A|B} = \mathbf{Q}_{AA}\mu_A - \mathbf{Q}_{AB}\mathbf{x}_B + \mathbf{Q}_{AB}\mu_B \]

• Eg if \( A=\{i\} \) we have

\[ E[\mathbf{x}_i|\mathbf{x}_{-i}] = \mu_i - \frac{1}{Q_{ii}} \sum_{j:j\neq i} Q_{ij}(\mathbf{x}_j - \mu_j) \]
\[ \text{prec}(\mathbf{x}_i|\mathbf{x}_{-i}) = Q_{ii} \]
Proof

• Assume $\mu=0$ for simplicity. Then

$$p(x_A|x_B) \propto \exp \left( -\frac{1}{2} (x_A \ x_B) \begin{pmatrix} Q_{AA} & Q_{AB} \\ Q_{BA} & Q_{BB} \end{pmatrix} \begin{pmatrix} x_A \\ x_B \end{pmatrix} \right)$$

$$\propto \exp \left( -\frac{1}{2} x_A^T Q_{AA} x_A - (Q_{AB} x_B)^T x_A \right)$$

• Compare this to a Gaussian with precision $K$ and mean $m$

$$p(z) \propto \exp \left( -\frac{1}{2} z^T K z + (Km)^T z \right)$$

• We see that $Q_{\{AA\}}$ is the conditional precision and the conditional mean is given by

$$Q_{AA} \mu_{A|B} = -Q_{AB} x_B$$

QED
Soft conditioning in moment form

\[
\begin{align*}
  x & \sim \mathcal{N}(\mu, \Sigma) \\
  y \mid x & \sim \mathcal{N}(x, S) \\
  x \mid y & \sim \mathcal{N}(\mu_{x \mid y}, \Sigma_{x \mid y}) \\
  \Sigma_{x \mid y}^{-1} & = \Sigma^{-1} + S^{-1} \\
  \Sigma_{x \mid y}^{-1} \mu_{x \mid y} & = \Sigma^{-1} \mu + S^{-1} y
\end{align*}
\]

Bayes rule for linear Gaussian systems
Linear Gaussian DGMs

• A CPD is linear Gaussian if

\[ p(x_i | x_{\pi_i}) = \mathcal{N}(x_i | \sum_{j \in \pi_i} w_{ij} x_j + b_i, v_i) \]

• A DGM is linear Gaussian if all CPDs are LG.

• Such networks define a joint Gaussian. Each node is given by

\[ x_i = \sum_{j \in \pi_i} w_{ij} x_j + b_i + \sqrt{v_i} \epsilon_i \]

where \( \epsilon_i \sim \mathcal{N}(0,1) \) and \( \mathbb{E}[\epsilon_i \epsilon_j] = I_{i,j} \).

• \( W \) is lower triangular matrix: \( w_{\{i,j\}} = \) weights into \( i \) from \( j \).
LG DGM to MVN

- We can compute the global mean and covariance recursively, in topological order

\[
x_i = \sum_{j \in \pi_i} w_{ij} x_j + b_i + \sqrt{v_i} \epsilon_i
\]

\[
E[x_i] = \sum_{j \in \pi_i} w_{ij} E[x_j] + b_i
\]

\[
\text{Cov}[x_i, x_j] = E[(x_i - E[x_i])(x_j - E[x_j])]
\]

\[
\begin{align*}
&= E \left[ (x_i - E[x_i]) \left\{ \sum_{k \in \pi_j} w_{jk} (x_k - E[x_k]) + \sqrt{v_j} \epsilon_j \right\} \right] \\
&= \sum_{k \in \pi_j} w_{jk} \text{Cov}[x_i, x_k] + I_{i,j} v_j
\end{align*}
\]

Bishop p371
LG DGM to MVN

- Consider a chain $x_1 \rightarrow x_2 \rightarrow x_3$

  \[
  \mu = (b_1, b_2 + w_{21}b_1, b_3 + w_{32}b_2 + w_{32}w_{21}b_1)
  \]

  \[
  \Sigma = \begin{pmatrix}
  v_1 & w_{21}v_1 & w_{32}w_{31}v_1 \\
  w_{21}v_1 & v_2 + w_{21}^2v_1 & w_{32}(v_2 + w_{21}^2v_1) \\
  w_{32}w_{21}v_1 & w_{32}(v_2 + w_{21}^2v_1) & v_3 + w_{32}^2(v_2 + w_{21}^2v_1)
  \end{pmatrix}
  \]

- In general, when adding node $(k+1)$

  \[
  \beta \sim \mathcal{N}(\mathbf{0}, \mathbf{W}(K+1, 1:K)\mathbf{1})
  \]
Alternative parameterization

- The results are much “prettier” if we write

\[ X_j = \mu_j + \sum_{k \in \pi_j} w_{jk}(X_k - \mu_k) + \sqrt{v_j}Z_j \]

where the offset is given by

\[ w_j^{(0)} = \mu_j - \sum_{k \in \pi_j} w_{jk}\mu_k \]

- Then we have

\[
(x - \mu) = W(x - \mu) + S^T z = W(x - \mu) + e
\]

\[
e = S^T z = (I - W)(x - \mu)
\]

\[
\begin{pmatrix}
e_1 \\
e_2 \\
\vdots \\
e_d
\end{pmatrix} = 
\begin{pmatrix}
1 & & & & \\
-w_{21} & 1 & & & \ \\
-w_{32} & -w_{31} & 1 & & \\
& \vdots & & \ddots & \\
-w_{d1} & -w_{d2} & \ldots & -w_{d,d-1} & 1
\end{pmatrix}
\begin{pmatrix}
x_1 - \mu_1 \\
x_2 - \mu_2 \\
\vdots \\
x_d - \mu_d
\end{pmatrix}
\]
DAG weights = Cholesky Decomposition

\[ x - \mu = (I - W)^{-1} e \overset{\text{def}}{=} U e = US^T z \overset{\text{def}}{=} A^T z \]

\[ \Sigma = \text{Var} [x] = \text{Var} [x - \mu] \]
\[ = \text{Var} [A^T z] = A^T \text{Var} [z] A = A^T A \]
\[ = US^T SU^T = UDU^T \]

\[ \Sigma^{-1} = U^{-T} D^{-1} U^{-1} = (I - W)^T D^{-1} (I - W) \overset{\text{def}}{=} T^T D^{-1} T \]

\[
T = \begin{pmatrix}
1 \\
-w_{21} & 1 \\
-w_{32} & -w_{31} & 1 \\
& \ddots & \ddots \\
-w_{d1} & -w_{d2} & \ldots & -w_{d,d-1} & 1
\end{pmatrix}
\]
Chains

• Consider a chain $X_1 \rightarrow X_2 \rightarrow \ldots \rightarrow X_5$

• The DAG and UG are both sparse (same CI)

```matlab
n = 5;
w = randn(n,1);
W = spdiags([w zeros(n,1) zeros(n,1)], -1:1, n, n);
T = eye(n)-W;
D = diag(ones(n,1));
K = T'*D*T;

>> full(W)
ans =
   0         0         0         0    0.0000
   0.0000    1.1909         0         0    0.0000
   0.0000    0.0000    1.1892         0    0.0000
   0.0000    0.0000   -0.0376         0    0.0000
   0.0000    0.0000         0    0.3273         0

>> K
K =
   2.4183   -1.1909         0         0    0.0000
  -1.1909    2.4141   -1.1892         0    0.0000
   0.0000   -1.1892    1.0014   0.0376         0
   0.0000    0.0376    1.1071   -0.3273         0
   0.0000    0.0000   -0.3273    1.0000         0
```
Diamond

- DAG is sparse, Sigma and SigmaInv are dense

\[ W = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0.5488 & 0 & 0 & 0 & 0 \\ 0.7152 & 0 & 0 & 0 & 0 \\ 0 & 0.6028 & 0.5449 & 0 & 0 \end{bmatrix} \]

\[ K = \begin{bmatrix} 1.8127 & -0.5488 & -0.7152 & 0 \\ -0.5488 & 1.3633 & 0.3284 & -0.6028 \\ -0.7152 & 0.3284 & 1.2969 & -0.5449 \\ 0 & -0.6028 & -0.5449 & 1.0000 \end{bmatrix} \]

\[ \text{inv}(K) \]
\[ \text{ans} = \begin{bmatrix} 1.0000 & 0.5488 & 0.7152 & 0.7205 \\ 0.5488 & 1.3012 & 0.3925 & 0.9982 \\ 0.7152 & 0.3925 & 1.5115 & 1.0602 \\ 0.7205 & 0.9982 & 1.0602 & 2.1793 \end{bmatrix} \]
Gaussian MRFs

• Defn. A GMRF is a Gaussian of the form $N(\mu, Q^{-1})$ where $Q_{ij} \neq 0$ iff $G_{ij} \neq 0$ ($Q =$ precision matrix)
• Thm. For a GMRF, the following properties are equivalent.
  • Pairwise Markov: $x_i \perp x_j \mid x_{-ij}$ if $G_{i,j} = 0$ and $i \neq j$
  • Local Markov: $x_i \perp x_{-i,ne(i)} \mid x_{ne(i)}$
  • Global Markov: $x_A \perp x_B \mid x_C$

Rue&Held p25
MVN to Gaussian UGM

- We can convert any MVN into a UGM with pairwise potentials which are quadratics.

\[ J \overset{\text{def}}{=} \Sigma^{-1} \]
\[ h \overset{\text{def}}{=} J \mu \]
\[ \mathcal{N}(x|h, J) = \exp[c - \frac{1}{2} x^T J x + x^T h] \]
\[ \log p(x) = c - \frac{1}{2} \sum_i [J_{i,i} x_i^2 + h_i x_i] - \frac{1}{2} \sum_i \sum_j J_{i,j} x_i x_j \]
\[ = c + \sum_i \phi_i(x_i) + \sum_i \sum_{j>i} \phi_{i,j}(x_i, x_j) \]
\[ \phi_i(x_i) = -\frac{1}{2} [J_{i,i} x_i^2 + h_i x_i] \]
\[ \phi_{i,j}(x_i, x_j) = -J_{i,j} x_i x_j \]
Pairwise UGM to MVN

• Consider a UGM in which the node and edge potentials are quadratics

\[ \epsilon_i(x_i) = d_0^i + d_1^i x_1 + d_2^i x_i^2 \]
\[ \epsilon_{ij}(x_i, x_j) = a_{00}^{i,j} + a_{01}^{i,j} x_i + a_{10}^{i,j} x_j + a_{11}^{i,j} x_i x_j + a_{02}^{i,j} x_i^2 + a_{20}^{i,j} x_j^2 \]

• We can always rewrite the corresponding unnormalized distribution as

\[ p'(x) = \exp[-\frac{1}{2}x^T J x + x^T h] \]

• But the normalization constant Z will only be finite if J is positive definite.
Sufficient conditions on info matrix

• Def 7.3.1. A matrix $J$ is attractive if, for all $i \neq j$, we have that all partial correlations are non-neg

\[- \frac{J_{i,j}}{\sqrt{J_{i,i}J_{j,j}}} \geq 0\]

• Thm 7.3.2. If $J$ is attractive, then $p$ is a valid MVN.

• Def 7.3.1b. A matrix $J$ is diagonally dominant if, for all rows $i$, $J_{ii} > \sum_{j \neq i} |J_{i,j}|$

• Thm 7.3.2b. If $J$ is diagonally dominant, then $p$ is a valid MVN.
• Def 7.3.3. A pairwise MRF with energies of the form

\[
\epsilon_i(x_i) = d_i^0 + d_i^1 x_1 + d_i^2 x_i^2 \\
\epsilon_{ij}(x_i, x_j) = a_{00}^{ij} + a_{01}^{ij} x_i + a_{10}^{ij} x_j + a_{11}^{ij} x_i x_j + a_{02}^{ij} x_i^2 + a_{20}^{ij} x_j^2
\]

is called pairwise normalizable if

\[d_i^2 > 0, \forall i \quad \text{and} \quad \begin{pmatrix} a_{02}^{ij} & a_{11}^{ij}/2 \\ a_{11}^{ij}/2 & a_{20}^{ij} \end{pmatrix} \text{ is PSD for all } i,j\]

• Thm 7.3.4. If the MRF is pairwise normalizable, then it defines a valid Gaussian.
• Sufficient but not necessary eg.

\[
\begin{pmatrix} 1 & 0.6 & 0.6 \\ 0.6 & 1 & 0.6 \\ 0.6 & 0.6 & 1 \end{pmatrix}
\]

May be able to reparameterize the node/edge potentials to ensure pairwise normalized.
Conditional autoregressions (CAR)

• We can parameterize a GMRF in terms of its full conditionals

\[
E[x_i|x_{-i}] = \mu_i - \sum_{j:j \sim i} \beta_{ij} (x_j - \mu_j)
\]

\[
\text{prec}[x_i|x_{-i}] = \kappa_i > 0
\]

• From before, we have

\[
E[x_i|x_{-i}] = \mu_i - \frac{1}{Q_{ii}} \sum_{j:j \neq i} Q_{ij} (x_j - \mu_j)
\]

\[
\text{prec}(x_i|x_{-i}) = Q_{ii}
\]

• To be a valid MVN we must set

\[
\kappa_i = Q_{ii}, \beta_{ij} = \frac{Q_{ij}}{\kappa_i}, \kappa_i \beta_{ij} = \kappa_j \beta_{ji}
\]

\[
Q = \text{diag}(\kappa)(I + \beta)
\]