ECE/CS/ISYE 8803

Exact Inference in Graphical Models

Module 6 (Part B)
Message Passing Algorithm on General Graphs

Faramarz Fekri

Center for Signal and Information Processing

Overview

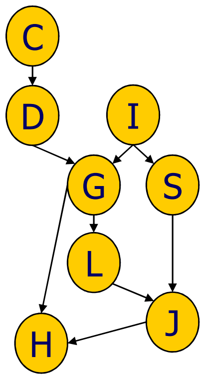
- Exact Inference in general graphs
  - Message passing over junction trees
    - Shafer-Shenoy Algorithm in Junction Trees
      - Example
    - Inference Outside of a Clique in Junction Tree
    - Belief Updating (Hugin Alg.) in Junction tree (Junction tree Calibration)
      - Factor Division
      - Example
- Message Passing for Continuous Ran. Variables

Read Chapter 10 of K&F

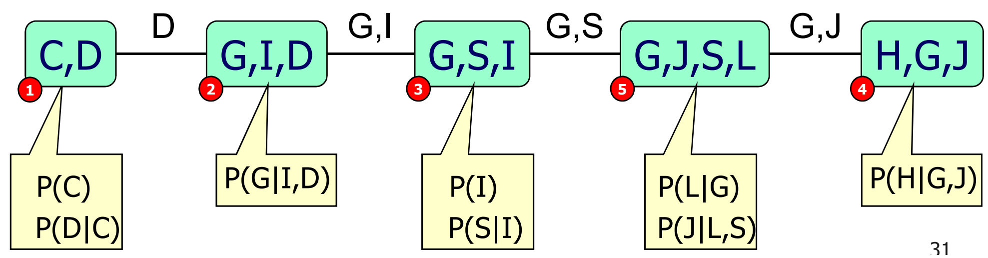
MP Alg. (Shafer-Shenoy) in Junction Trees (I)

Clique initial potentials:

- Multiply factors (CPDs) assigned to each clique, resulting in initial potentials, as each factor is assigned to some clique.



Original graph



Clique initial potential for clique 'GSI':

$$\phi(G,S,I) = P(I)P(S \mid I)$$

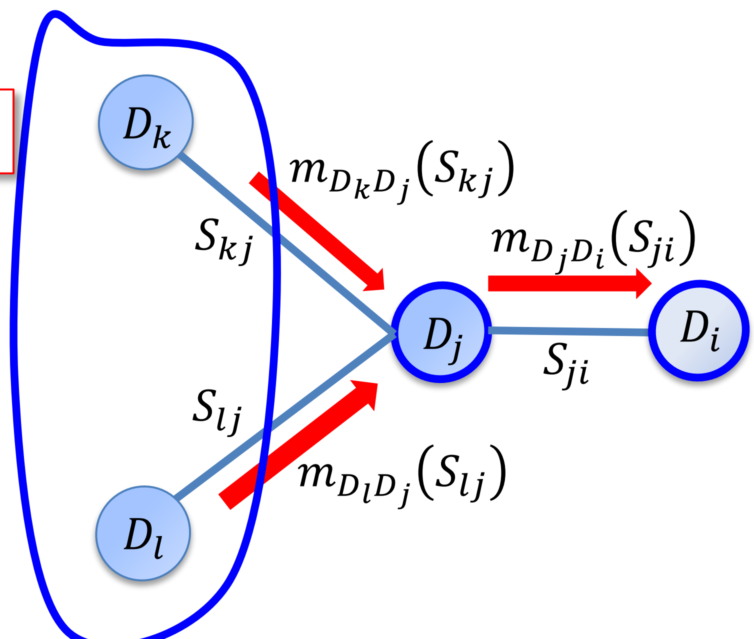
- Any clique containing 'node of interest in query" can be a root. Then start MP from leaves and move upward to the root.

- Message from clique C_i to clique C_j is always the same, regardless of what the query variables are.
  - The message from C_i to C_j is marginalized over all variables in C_i except the variables in sepset S_{ij} of C_i and C_j

Message Passing Alg. In Junction Trees (II)

- $$m_{D_jD_i}(S_{ji}) \propto \underbrace{\sum_{D_j\setminus S_{ji}} \underbrace{\Phi(D_j) \underbrace{\prod_{D_t \in \mathrm{N}(D_j)\setminus D_i} m_{D_tD_j}(S_{tj})}_{\textit{product of incoming messages}}}_{\textit{multiply by local potentials}}}_{\textit{Sum out variables not in separator}}$$

$\mathrm{N}(D_j)\backslash \mathrm{D_i}$



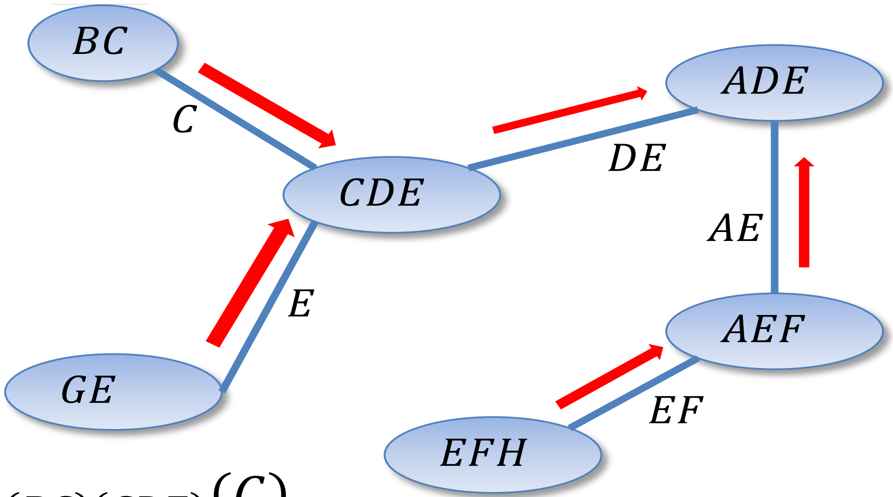
Separator:
$S_{kj} = D_k \cap D_j$

Can also be applied to loopy clique graphs for approximate inference

Just like MP on tress, we can do two-pass MP on clique tree to solve for all variables.

Example: Inference in Junction Tree

- Compute $P(A|h)$
  - $m_{(GE)(CDE)}(E) = \sum_G \Phi(G,E)$
  - $m_{(BC)(CDE)}(C) = \sum_B \Phi(B,C)$
  - $m_{(CDE)(ADE)}(D,E) = \sum_C \Phi(C,D,E) m_{(GE)(CDE)}(E)\, m_{(BC)(CDE)}(C)$
  - $m_{(EFh)(AEF)}(E,F) = \Phi(E,F,h)$
  - $m_{(AEF)(ADE)}(A,E) = \sum_F \Phi(A,E,F)\, m_{(EFh)(AEF)}(E,F)$
  - $P(A,h) \propto \sum_D \sum_E \underbrace{\Phi(A,D,E) m_{(AEF)(ADE)}(A,E) m_{(CDE)(ADE)}(D,E)}$



Any cluster Containing 'A' Can be a root.

renormalize afterwards

- When the root clique C_r has received all messages, it multiplies them with its own initial potential, resulting in a factor called the belief

Inference Outside of a Clique in Junction Tree

Consider a query $P(Y \mid e)$ where the variables in Y are not present together in a single clique (i.e., dispersed in multiple cliques).

- One naive approach is to construct a clique tree where we force one of the cliques to contain all variables in Y.
  - Forces us to tailor our clique tree to different queries, negating many of its advantages.

- A better approach is to perform variable elimination over a calibrated clique tree.

Factor Division

a^1	b^1	0.5
a^1	b^2	0.2
a^2	b^1	0
a^2	b^2	0
a^3	b^1	0.3
a^3	b^2	0.45

a^1	0.8
a^2	0
a^3	0.6

a^1	b^1	0.625
a^1	b^2	0.25
a^2	b^1	0
a^2	b^2	0
a^3	b^1	0.5
a^3	b^2	0.75

Let $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *be disjoint sets of variables, and let* $\phi_1(\boldsymbol{X}, \boldsymbol{Y})$ *and* $\phi_2(\boldsymbol{Y})$ *be two factors. We define the* division $\frac{\phi_1}{\phi_2}$ *to be a factor* ψ *of scope* $\boldsymbol{X}, \boldsymbol{Y}$ *defined as follows:*

$$\psi(\boldsymbol{X}, \boldsymbol{Y}) = \frac{\phi_1(\boldsymbol{X}, \boldsymbol{Y})}{\phi_2(\boldsymbol{Y})},$$

where we define $0/0 = 0$.

Belief Updating (Hugin Alg.) in Junction tree (Junction tree Calibration)

Assume the original model has the following form:

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{c \in \mathcal{C}(G)} \psi_c(\mathbf{x}_c)$$

The clique tree defines a distribution of the following form:

C(T): nodes of the junction tree
S(T): separators of the tree

$$p(\mathbf{x}) = \frac{\prod_{c \in \mathcal{C}(T)} \psi_c(\mathbf{x}_c)}{\prod_{s \in \mathcal{S}(T)} \psi_s(\mathbf{x}_s)}$$

Example:

The graph on the left represents $\phi(\mathcal{X}^1)\phi(\mathcal{X}^2)/\phi(\mathcal{X}^1 \cap \mathcal{X}^2)$.

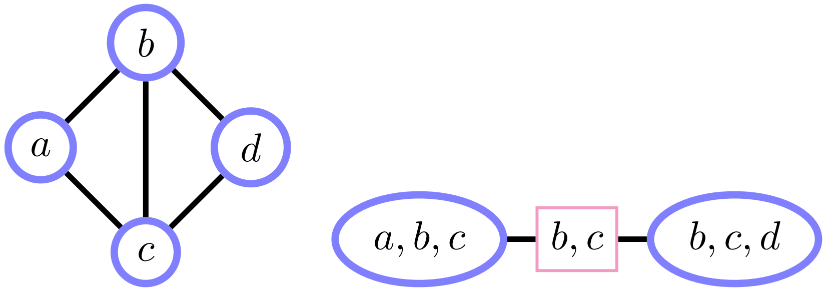
Belief Updating (Hugin Alg.) (II)

Example:

(a) Markov network

$$\phi(a,b,c)\phi(b,c,d)$$

(b) Clique graph



(a) (b)

From Graph (a):

$$p(a,b,c,d) = \frac{\phi(a,b,c)\phi(b,c,d)}{Z} \qquad (1)$$

$$Zp(a,b,c) = \phi(a,b,c)\sum_d \phi(b,c,d), \qquad Zp(b,c,d) = \phi(b,c,d)\sum_a \phi(a,b,c)$$

Multiplying the above two:

Used (1)

$$Z^2 p(a,b,c)p(b,c,d) = \left(\phi(a,b,c)\sum_d \phi(b,c,d)\right)\left(\phi(b,c,d)\sum_a \phi(a,b,c)\right) = Z^2 p(a,b,c,d)\sum_{a,d} p(a,b,c,d)$$

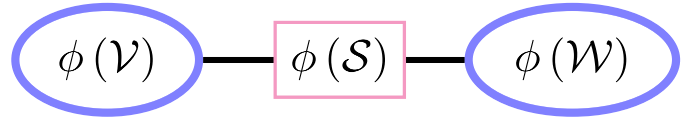
$$\Rightarrow \quad p(a,b,c,d) = \frac{p(a,b,c)p(b,c,d)}{p(c,b)}$$

$$\begin{cases} \phi(a,b,c) \to p(a,b,c) \\ \phi(b,c,d) \to p(b,c,d) \\ Z \to p(c,b) \end{cases}$$

Belief Updating (Hugin Alg.) (III)

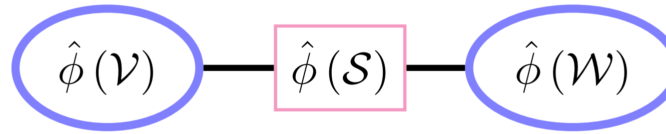
Absorption:

$$p(\mathcal{X}) = \frac{\phi(\mathcal{V})\phi(\mathcal{W})}{\phi(\mathcal{S})}$$



Our aim is to find a new representation:

$$p(\mathcal{X}) = \frac{\hat{\phi}(\mathcal{V})\hat{\phi}(\mathcal{W})}{\hat{\phi}(\mathcal{S})}$$



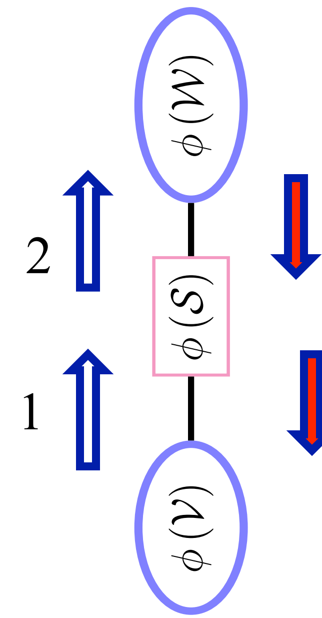
in which the potentials are given by

$$\hat{\phi}(\mathcal{V}) = p(\mathcal{V}), \quad \hat{\phi}(\mathcal{W}) = p(\mathcal{W}), \quad \hat{\phi}(\mathcal{S}) = p(\mathcal{S})$$

$$\phi^*(\mathcal{S}) = \sum_{\mathcal{V}\backslash\mathcal{S}} \phi(\mathcal{V}) \qquad 1$$

$$\phi^*(\mathcal{W}) = \phi(\mathcal{W})\frac{\phi^*(\mathcal{S})}{\phi(\mathcal{S})} \qquad 2$$

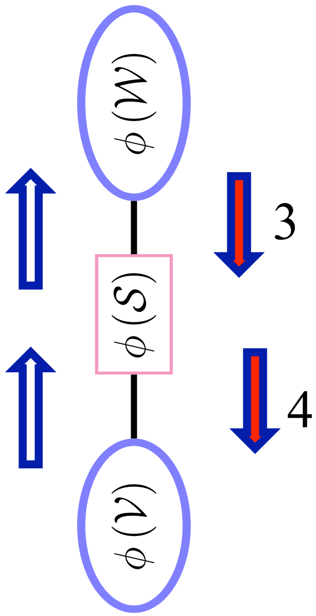
$$\frac{\phi(\mathcal{V})\phi^*(\mathcal{W})}{\phi^*(\mathcal{S})} = \frac{\phi(\mathcal{V})\phi(\mathcal{W})\frac{\phi^*(\mathcal{S})}{\phi(\mathcal{S})}}{\phi^*(\mathcal{S})} = \frac{\phi(\mathcal{V})\phi(\mathcal{W})}{\phi(\mathcal{S})} = p(\mathcal{X})$$



Belief Updating (Hugin Alg.) (IV)

$$\phi^{**}(\mathcal{S}) = \sum_{\mathcal{W}\backslash\mathcal{S}} \phi^{*}(\mathcal{W}) = \sum_{\mathcal{W}\backslash\mathcal{S}} \frac{\phi(\mathcal{W})\phi^{*}(\mathcal{S})}{\phi(\mathcal{S})} \qquad 3$$

$$\phi^{*}(\mathcal{V}) = \frac{\phi(\mathcal{V})\phi^{**}(\mathcal{S})}{\phi^{*}(\mathcal{S})} \qquad 4$$



Now, we see that:

$p(\mathcal{X})$

$$\phi^{**}(\mathcal{S}) = \sum_{\mathcal{W}\backslash\mathcal{S}} \phi^{*}(\mathcal{W}) = \sum_{\mathcal{W}\backslash\mathcal{S}} \frac{\phi(\mathcal{W})\phi^{*}(\mathcal{S})}{\phi(\mathcal{S})} = \sum_{\{\mathcal{W}\cup\mathcal{V}\}\backslash\mathcal{S}} \frac{\phi(\mathcal{W})\phi(\mathcal{V})}{\phi(\mathcal{S})} = p(\mathcal{S})$$

$$\phi^{*}(\mathcal{V}) = \frac{\phi(\mathcal{V})\phi^{**}(\mathcal{S})}{\phi^{*}(\mathcal{S})} = \frac{\phi(\mathcal{V})\sum_{\mathcal{W}\backslash\mathcal{S}}\phi(\mathcal{W})\phi^{*}(\mathcal{S})/\phi(\mathcal{S})}{\phi^{*}(\mathcal{S})} = \frac{\sum_{\mathcal{W}\backslash\mathcal{S}}\phi(\mathcal{V})\phi(\mathcal{W})}{\phi(\mathcal{S})} = p(\mathcal{V})$$

Hence, the new representation is:

$$\hat{\phi}(\mathcal{V}) = \phi^{*}(\mathcal{V}),\ \hat{\phi}(\mathcal{S}) = \phi^{**}(\mathcal{S}),\ \hat{\phi}(\mathcal{W}) = \phi^{*}(\mathcal{W})$$

Belief Updating (Hugin Alg.) (V)

Assume the original model has the following form:

The clique tree defines a distribution of the following form:

$\mathcal{C}(T)$: nodes of the junction tree
$\mathcal{S}(T)$: separators of the tree

$$p(\mathbf{x}) = \frac{\prod_{c \in \mathcal{C}(T)} \psi_c(\mathbf{x}_c)}{\prod_{s \in \mathcal{S}(T)} \psi_s(\mathbf{x}_s)}$$

- Send messages from leaves to the root (upward) and send back downward to leaves (steps of Alg in below)

- Initialize by defining ψs = 1 for all separators, and ψc via potentials in the graphical model, as in Shafer-Shenoy algorithm.

Belief Updating (Hugin Alg.) (VI)

- On the way up, sending from i to j we compute the separator potential:

$$\psi^*_{ij}(S_{ij}) = \sum_{C_i \setminus S_{ij}} \psi_i(C_i)$$

- Then update the recipient potential:

$$\psi^*_j(C_j) \propto \psi_j(C_j) \frac{\psi^*_{ij}(S_{ij})}{\psi_{ij}(S_{ij})}$$

- Recall that we had initialized: $\psi_{ij}(S_{ij}) = 1$

- On the way down to leaves, from i to j, we compute separator potential

$$\psi^{**}_{ij}(S_{ij}) = \sum_{C_i \setminus S_{ij}} \psi^*_i(C_i)$$

- Then update the recipient potential:

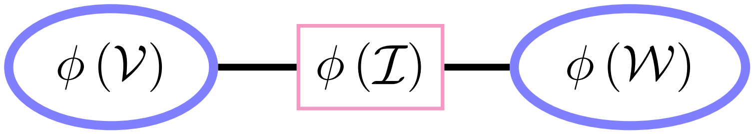
$$\psi^{**}_j(C_j) \propto \psi^*_j(C_j) \frac{\psi^{**}_{ij}(S_{ij})}{\psi^*_{ij}(S_{ij})}$$

- The resulting new potentials are the correct marginal distributions.

Belief Updating (Hugin Alg.) (VII)

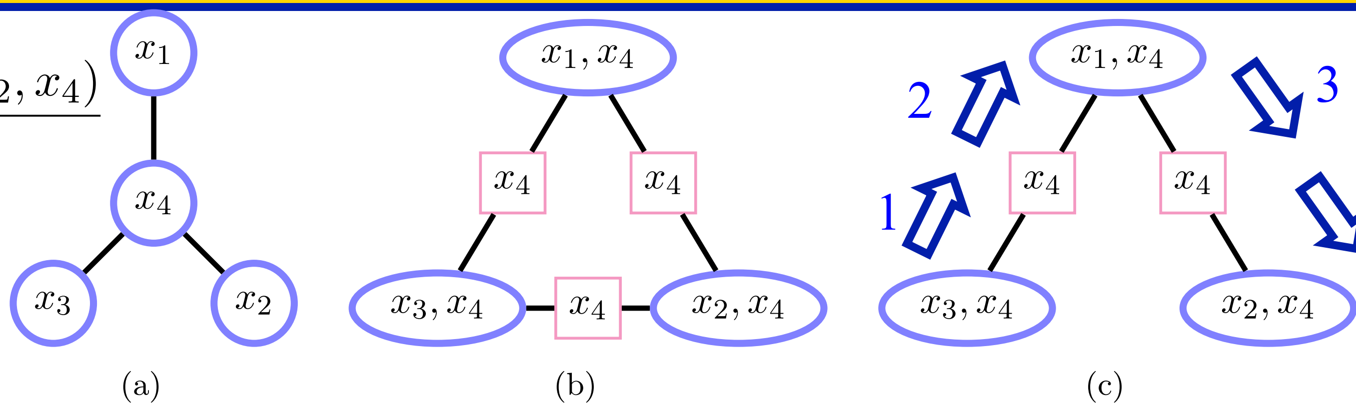
The new representation is *consistent* in the sense that for any (not necessarily neighbouring) cliques $\mathcal{V}$ and $\mathcal{W}$ with intersection $\mathcal{I}$, and corresponding potentials $\phi(\mathcal{V})$ and $\phi(\mathcal{W})$,

$$\sum_{\mathcal{V}\backslash\mathcal{I}} \phi(\mathcal{V}) = \sum_{\mathcal{W}\backslash\mathcal{I}} \phi(\mathcal{W})$$



Example: Belief Updating (Hugin Alg.) (VIII)

$$\frac{\phi(x_3,x_4)\phi(x_1,x_4)\phi(x_2,x_4)}{\phi_1(x_4)\phi_2(x_4)}$$



(a): Singly connected Markov network. **(b)**: Clique graph. **(c)**: Clique tree.

$$\phi_1^*(x_4) = \sum_{x_3} \phi(x_3, x_4) \qquad 1$$

$$\phi^*(x_1,x_4) = \phi(x_1,x_4)\frac{\phi_1^*(x_4)}{\phi_1(x_4)} = \phi(x_1,x_4)\phi_1^*(x_4) \qquad 2$$

$$\phi_2^*(x_4) = \sum_{x_1} \phi^*(x_1, x_4) \qquad 3$$

$$\phi^*(x_2,x_4) = \phi(x_2,x_4)\frac{\phi_2^*(x_4)}{\phi_2(x_4)} = \phi(x_2,x_4)\phi_2^*(x_4) \qquad 4$$

$\phi_1(x_4) = \phi_2(x_4) = 1$

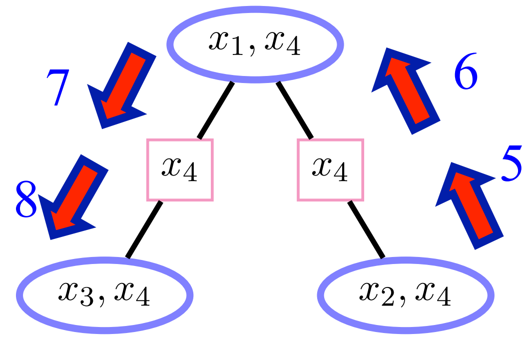
Initialization step

We observe that:

$$\phi^*(x_2,x_4) = \phi(x_2,x_4)\phi_2^*(x_4) = \phi(x_2,x_4)\sum_{x_1}\phi^*(x_1,x_4)$$

$$= \phi(x_2,x_4)\sum_{x_1}\phi(x_1,x_4)\sum_{x_3}\phi(x_3,x_4) = \sum_{x_1,x_3} p(x_1,x_2,x_3,x_4) = p(x_2,x_4)$$

Example: Belief Updating (Hugin Alg.) (IX)



(c)

To complete a full round of message passing we need to have passed messages in a valid schedule along both directions of each separator. To do so, we continue as follows:

We absorb $(x_2, x_4) \rightsquigarrow (x_1, x_4)$. The new separator is

$$\phi_2^{**}(x_4) = \sum_{x_2} \phi^*(x_2, x_4) \qquad 5$$

$$\phi^{**}(x_1, x_4) = \phi^*(x_1, x_4)\frac{\phi_2^{**}(x_4)}{\phi_2^*(x_4)} \qquad 6$$

Note that $\phi_2^{**}(x_4) = \sum_{x_2} \phi^*(x_2, x_4) = \sum_{x_2} p(x_2, x_4) = p(x_4)$

Likewise, one can show that: $\phi^{**}(x_1, x_4) = p(x_1, x_4)$.

$$\phi_1^{**}(x_4) = \sum_{x_1} \phi^{**}(x_1, x_4) = p(x_4) \qquad 7$$

$$\phi^*(x_3, x_4) = \phi(x_3, x_4)\frac{\phi_1^{**}(x_4)}{\phi_1^*(x_4)} = p(x_3, x_4) \qquad 8$$

Message Passing for Continuous Ran. Variables

- For parametric continuous distributions $P(x|\theta_x)$, message passing corresponds to passing parameters θ of the distributions.
- This requires that the operations of multiplication and integration in BP over the variables are closed with respect to the family of distributions.
  - This is the case, for example, for the Gaussian distribution.
    - sum-product algorithm based on passing mean and covariance parameters.
  - For more general "exponential family distributions", message passing is essentially straightforward.
- In cases where the operations of sum and products are not closed within the family, the distributions need to be projected back to the chosen message family. Expectation propagation is relevant in this case.