MPE inference in Conditional Linear Gaussian Networks

Antonio Salmerón[1] Rafael Rumí[1] Helge Langseth[2]
Anders L. Madsen[3,4] Thomas D. Nielsen[4]

[1]Dept. Mathematics, University of Almería, Spain

[2]Dept. Computer and Information Science. Norwegian University of Science and Technology, Trondheim, Norway

[3]Hugin Expert A/S, Aalborg, Denmark

[4]Dept. Computer Science, Aalborg University, Denmark

Introduction

- The AMiDST project: Analysis of MassIve Data STreams
  http://www.amidst.eu

Introduction

- The AMiDST project: Analysis of MassIve Data STreams
  http://www.amidst.eu
- Large number of variables
- Queries to be answered in real time
- Hybrid Bayesian networks (involving discrete and continuous variables)
  - Conditional linear Gaussian networks

Conditional Linear Gaussian networks

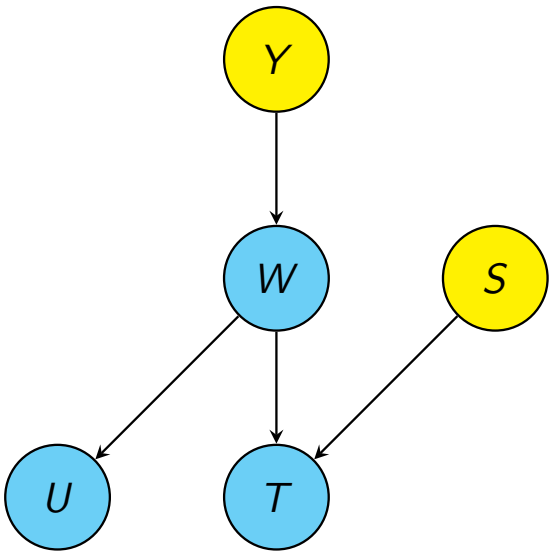
A Conditional Linear Gaussian (CLG) network is a hybrid Bayesian network where

- The conditional distribution of each discrete variable X_D given its parents is a multinomial
- The conditional distribution of each continuous variable Z with discrete parents $\mathbf{X}_D$ and continuous parents $\mathbf{X}_C$, is

$$p(z|\mathbf{X}_D = \mathbf{x}_D, \mathbf{X}_C = \mathbf{x}_C) = \mathcal{N}(z; \alpha(\mathbf{x}_D) + \beta(\mathbf{x}_D)^{\mathsf{T}}\mathbf{x}_C, \sigma(\mathbf{x}_D))$$

for all $\mathbf{x}_D$ and $\mathbf{x}_C$, where α and β are the coefficients of a linear regression model of Z given $\mathbf{X}_C$, potentially different for each configuration of $\mathbf{X}_D$.

Conditional Linear Gaussian networks. Example



$$P(Y) = (0.5, 0.5)$$

$$P(S) = (0.1, 0.9)$$

$$f(w|Y = 0) = \mathcal{N}(w; -1, 1)$$

$$f(w|Y = 1) = \mathcal{N}(w; 2, 1)$$

$$f(t|w, S = 0) = \mathcal{N}(t; -w, 1)$$

$$f(t|w, S = 1) = \mathcal{N}(t; w, 1)$$

$$f(u|w) = \mathcal{N}(u; w, 1)$$

Querying a Bayesian network (I)

- Probabilistic inference: Computing the posterior distribution of a target variable:

$$p(x_i|\mathbf{x}_E) = \frac{\sum_{\mathbf{x}_D} \int_{\mathbf{x}_C} p(\mathbf{x}, \mathbf{x}_E) \mathrm{d}\mathbf{x}_C}{\sum_{\mathbf{x}_{D_i}} \int_{\mathbf{x}_{C_i}} p(\mathbf{x}, \mathbf{x}_E) \mathrm{d}\mathbf{x}_{C_i}}$$

Querying a Bayesian network (II)

- Maximum a posteriori (MAP): For a set of target variables $\mathbf{X}_I$, the goal is to compute

$$\mathbf{x}_I^* = \arg\max_{\mathbf{x}_I} p(\mathbf{x}_I|\mathbf{X}_E = \mathbf{x}_E)$$

where $p(\mathbf{x}_I|\mathbf{X}_E = \mathbf{x}_E)$ is obtained by first marginalizing out from $p(\mathbf{x})$ the variables not in $\mathbf{X}_I$ and not in $\mathbf{X}_E$

Querying a Bayesian network (II)

- Maximum a posteriori (MAP): For a set of target variables $\mathbf{X}_I$, the goal is to compute

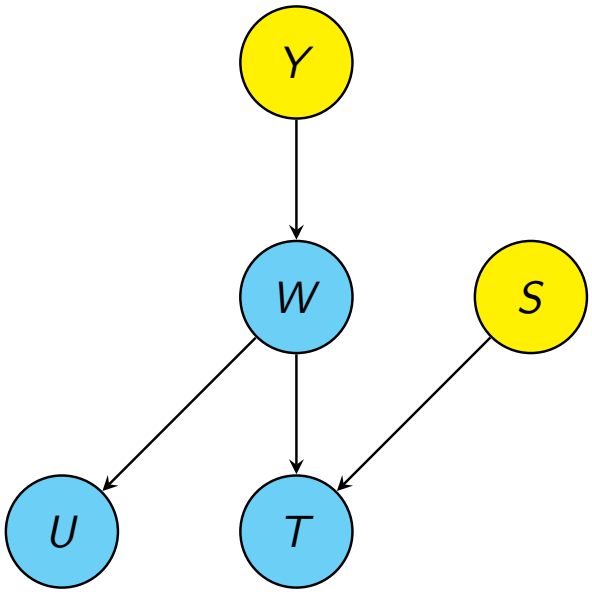
$$\mathbf{x}_I^* = \arg\max_{\mathbf{x}_I} p(\mathbf{x}_I|\mathbf{X}_E = \mathbf{x}_E)$$

  where $p(\mathbf{x}_I|\mathbf{X}_E = \mathbf{x}_E)$ is obtained by first marginalizing out from $p(\mathbf{x})$ the variables not in $\mathbf{X}_I$ and not in $\mathbf{X}_E$
- Most probable explanation (MPE): A particular case of MAP where $\mathbf{X}_I$ includes all the unobserved variables

MPE in CLG networks

- Can be carried out using bucket elimination (Dechter, 1999):
  - A bucket containing probability functions is kept for each variable.
  - Initially, an ordering of the variables in the network is established, and each conditional distribution in the network is assigned to the bucket corresponding to the variable in its domain holding the highest rank.
  - Afterwards, the buckets are processed in a sequence opposite to the initial ordering of the variables.
  - Each bucket is processed by combining all the functions it contains and by marginalizing the main variable in that bucket by maximization.

Example



$$P(Y) = (0.5, 0.5)$$
$$P(S) = (0.1, 0.9)$$
$$f(w|Y=0) = \mathcal{N}(w; -1, 1)$$
$$f(w|Y=1) = \mathcal{N}(w; 2, 1)$$
$$f(t|w, S=0) = \mathcal{N}(t; -w, 1)$$
$$f(t|w, S=1) = \mathcal{N}(t; w, 1)$$
$$f(u|w) = \mathcal{N}(u; w, 1)$$

Elimination order: Y, S, W, T, U

Example

Elimination order: Y, S, W, T, U

$B_Y : P(Y)$

$B_S : P(S)$

$B_W : f(w|Y)$

$B_T\ f(t|w, S)$

$B_U : f(u|w)$

$P(Y) = (0.5, 0.5)$

$P(S) = (0.1, 0.9)$

$f(w|Y = 0) = \mathcal{N}(w; -1, 1)$

$f(w|Y = 1) = \mathcal{N}(w; 2, 1)$

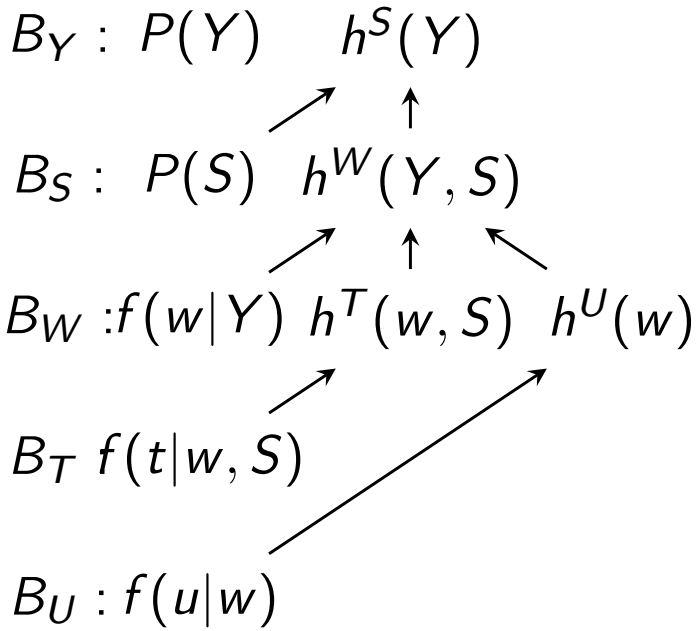
$f(t|w, S = 0) = \mathcal{N}(t; -w, 1)$

$f(t|w, S = 1) = \mathcal{N}(t; w, 1)$

$f(u|w) = \mathcal{N}(u; w, 1)$

Example

Elimination order: Y, S, W, T, U



$P(Y) = (0.5, 0.5)$

$P(S) = (0.1, 0.9)$

$f(w|Y=0) = \mathcal{N}(w; -1, 1)$

$f(w|Y=1) = \mathcal{N}(w; 2, 1)$

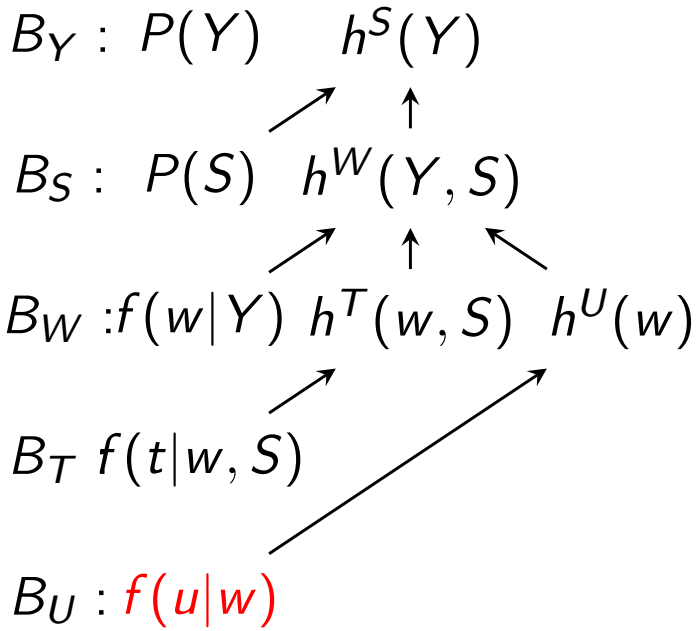
$f(t|w, S=0) = \mathcal{N}(t; -w, 1)$

$f(t|w, S=1) = \mathcal{N}(t; w, 1)$

$f(u|w) = \mathcal{N}(u; w, 1)$

Example

Elimination order: Y, S, W, T, U



$P(Y) = (0.5, 0.5)$

$P(S) = (0.1, 0.9)$

$f(w|Y=0) = \mathcal{N}(w; -1, 1)$

$f(w|Y=1) = \mathcal{N}(w; 2, 1)$

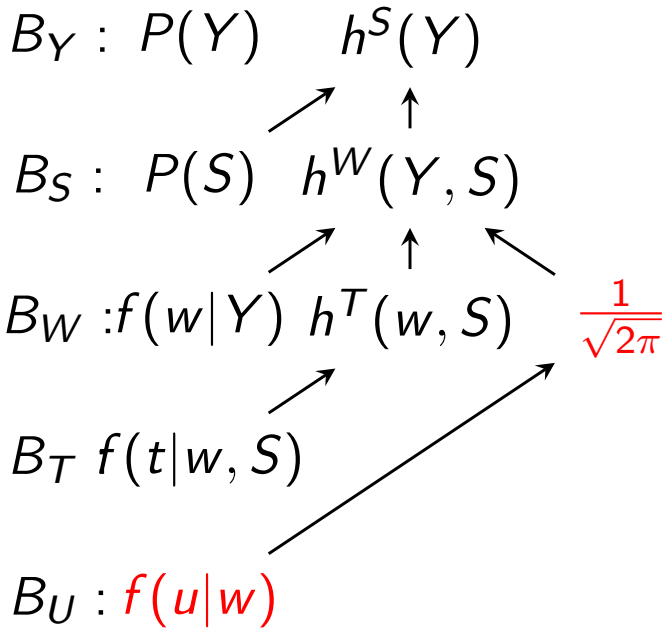
$f(t|w, S=0) = \mathcal{N}(t; -w, 1)$

$f(t|w, S=1) = \mathcal{N}(t; w, 1)$

$f(u|w) = \mathcal{N}(u; w, 1)$

Example

Elimination order: Y, S, W, T, U



$P(Y) = (0.5, 0.5)$

$P(S) = (0.1, 0.9)$

$f(w|Y=0) = \mathcal{N}(w; -1, 1)$

$f(w|Y=1) = \mathcal{N}(w; 2, 1)$

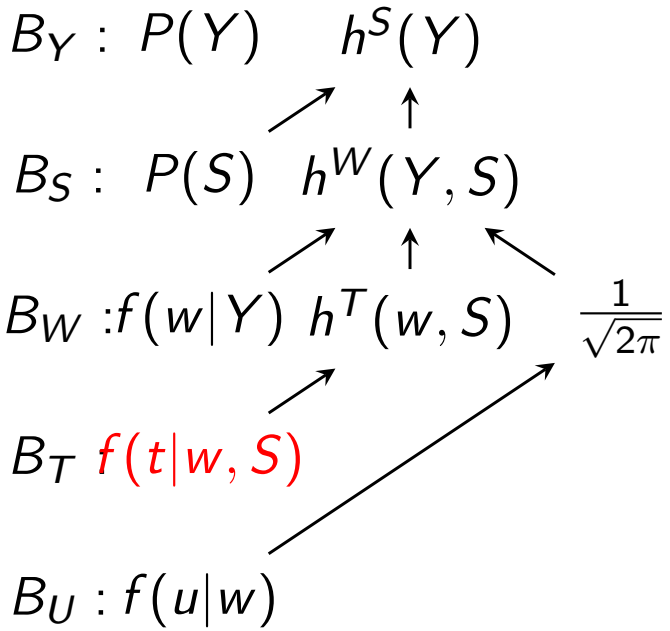
$f(t|w, S=0) = \mathcal{N}(t; -w, 1)$

$f(t|w, S=1) = \mathcal{N}(t; w, 1)$

$f(u|w) = \mathcal{N}(u; w, 1)$

Example

Elimination order: Y, S, W, T, U



$P(Y) = (0.5, 0.5)$

$P(S) = (0.1, 0.9)$

$f(w|Y=0) = \mathcal{N}(w; -1, 1)$

$f(w|Y=1) = \mathcal{N}(w; 2, 1)$

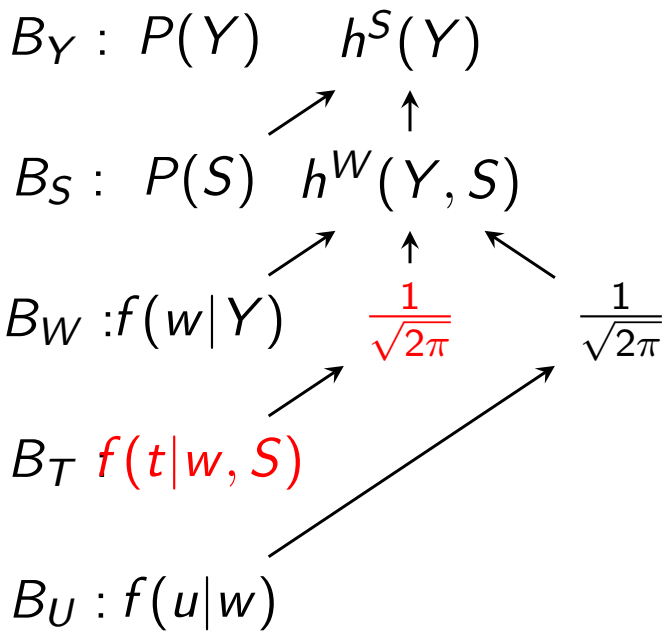
$f(t|w, S=0) = \mathcal{N}(t; -w, 1)$

$f(t|w, S=1) = \mathcal{N}(t; w, 1)$

$f(u|w) = \mathcal{N}(u; w, 1)$

Example

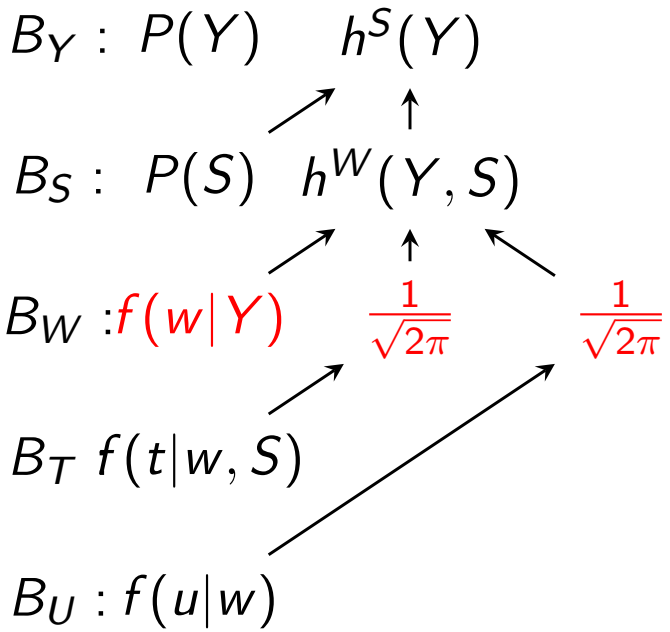
Elimination order: Y, S, W, T, U



$$P(Y) = (0.5, 0.5)$$
$$P(S) = (0.1, 0.9)$$
$$f(w|Y=0) = \mathcal{N}(w; -1, 1)$$
$$f(w|Y=1) = \mathcal{N}(w; 2, 1)$$
$$f(t|w, S=0) = \mathcal{N}(t; -w, 1)$$
$$f(t|w, S=1) = \mathcal{N}(t; w, 1)$$
$$f(u|w) = \mathcal{N}(u; w, 1)$$

Example

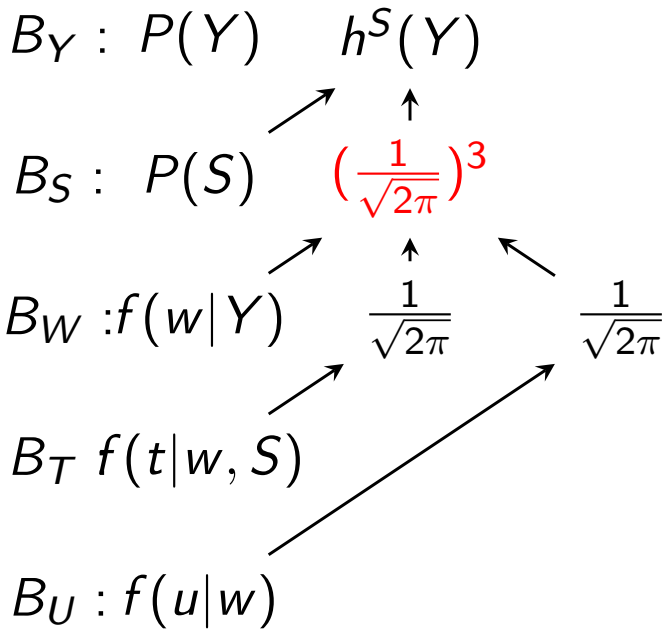
Elimination order: Y, S, W, T, U



$$P(Y) = (0.5, 0.5)$$
$$P(S) = (0.1, 0.9)$$
$$f(w|Y=0) = \mathcal{N}(w; -1, 1)$$
$$f(w|Y=1) = \mathcal{N}(w; 2, 1)$$
$$f(t|w, S=0) = \mathcal{N}(t; -w, 1)$$
$$f(t|w, S=1) = \mathcal{N}(t; w, 1)$$
$$f(u|w) = \mathcal{N}(u; w, 1)$$

Example

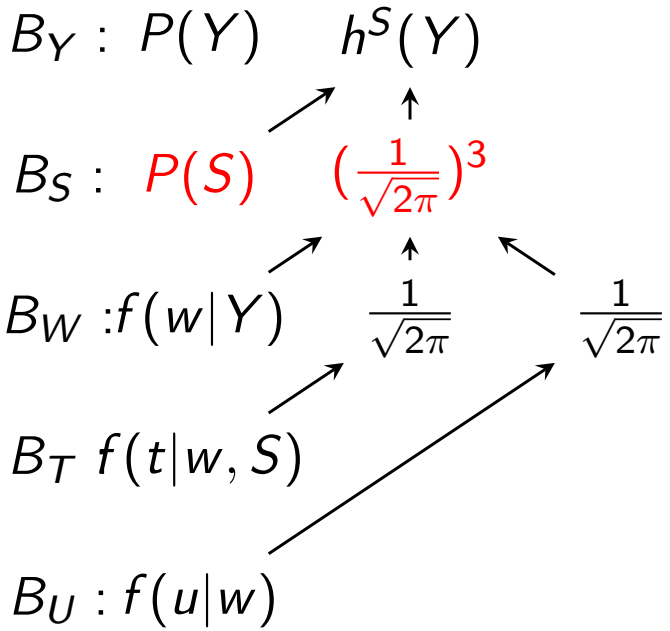
Elimination order: Y, S, W, T, U



$$P(Y) = (0.5, 0.5)$$
$$P(S) = (0.1, 0.9)$$
$$f(w|Y=0) = \mathcal{N}(w; -1, 1)$$
$$f(w|Y=1) = \mathcal{N}(w; 2, 1)$$
$$f(t|w, S=0) = \mathcal{N}(t; -w, 1)$$
$$f(t|w, S=1) = \mathcal{N}(t; w, 1)$$
$$f(u|w) = \mathcal{N}(u; w, 1)$$

Example

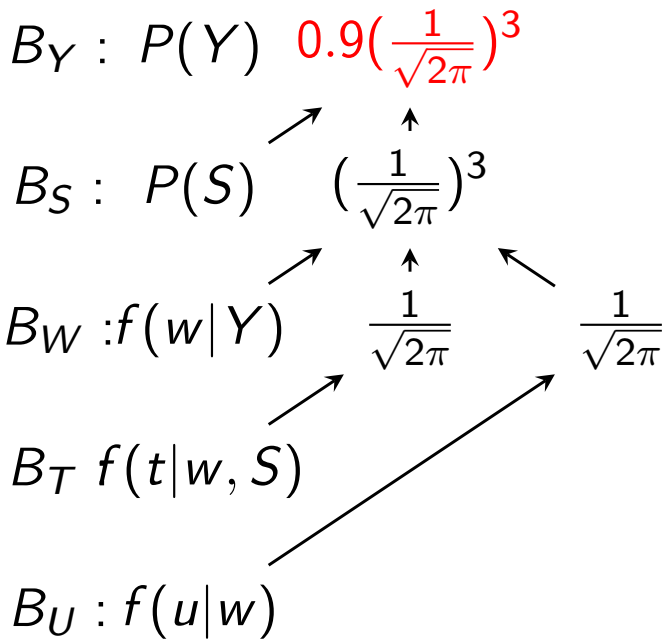
Elimination order: Y, S, W, T, U



$$P(Y) = (0.5, 0.5)$$
$$P(S) = (0.1, 0.9)$$
$$f(w|Y=0) = \mathcal{N}(w; -1, 1)$$
$$f(w|Y=1) = \mathcal{N}(w; 2, 1)$$
$$f(t|w, S=0) = \mathcal{N}(t; -w, 1)$$
$$f(t|w, S=1) = \mathcal{N}(t; w, 1)$$
$$f(u|w) = \mathcal{N}(u; w, 1)$$

Example

Elimination order: Y, S, W, T, U



$P(Y) = (0.5, 0.5)$

$P(S) = (0.1, 0.9)$

$f(w|Y=0) = \mathcal{N}(w; -1, 1)$

$f(w|Y=1) = \mathcal{N}(w; 2, 1)$

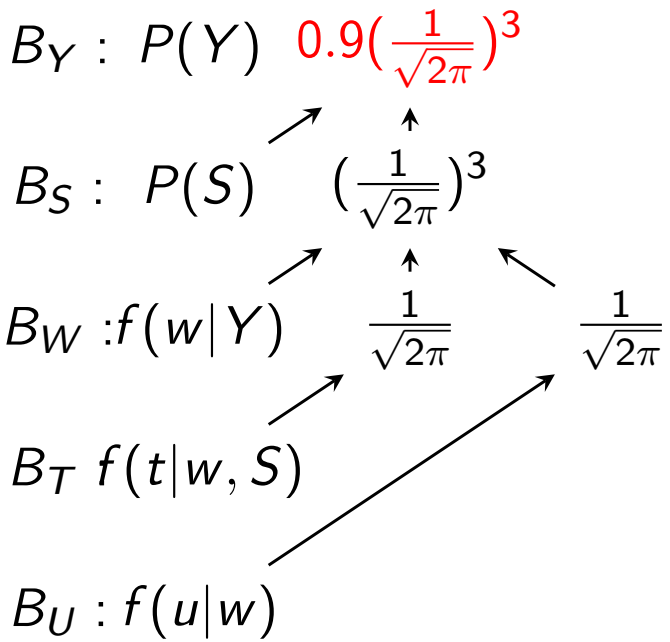
$f(t|w, S=0) = \mathcal{N}(t; -w, 1)$

$f(t|w, S=1) = \mathcal{N}(t; w, 1)$

$f(u|w) = \mathcal{N}(u; w, 1)$

Example

Elimination order: Y, S, W, T, U



$P(Y) = (0.5, 0.5)$

$P(S) = (0.1, 0.9)$

$f(w|Y=0) = \mathcal{N}(w; -1, 1)$

$f(w|Y=1) = \mathcal{N}(w; 2, 1)$

$f(t|w, S=0) = \mathcal{N}(t; -w, 1)$

$f(t|w, S=1) = \mathcal{N}(t; w, 1)$

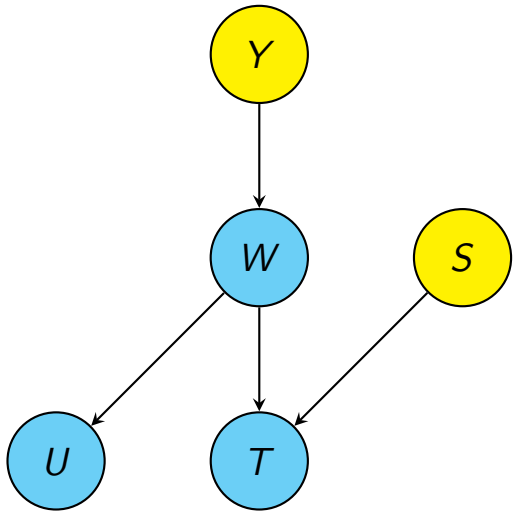
$f(u|w) = \mathcal{N}(u; w, 1)$

The MPE configuration is obtained tracing back the steps

What we have learnt so far

- Marginalizing continuous variables is easy if they are the first to be marginalized out
- The price to pay is that, in the worst case, a function containing all the discrete variables would be created
- This complexity blow-up can be avoided in many cases by allowing orderings for constructing the buckets where discrete and continuous variables can be arranged with no restrictions
- But then a new problem arises, as the maximization operation becomes more complex

Not so smooth max-marginalization



$P(Y) = (0.5, 0.5)$

$P(S) = (0.1, 0.9)$

$f(w|Y = 0) = \mathcal{N}(w; -1, 1)$

$f(w|Y = 1) = \mathcal{N}(w; 2, 1)$

$f(t|w, S = 0) = \mathcal{N}(t; -w, 1)$

$f(t|w, S = 1) = \mathcal{N}(t; w, 1)$

$f(u|w) = \mathcal{N}(u; w, 1)$

- Assume, for instance, that we reach a point where Y is maximized out before W. This amounts to computing

$$h^Y(w) = \max\{0.5\mathcal{N}(w; -1, 1), 0.5\mathcal{N}(w; 2, 1)\}$$

- h^Y is not a function with a single analytical expression, but it is piece-wise defined instead.

Entering evidence

- If a variable is observed, no bucket is created for it and the variable is replaced by its observed value in every function where it appears
- Assume a variable X with parents $Y_1, \ldots, Y_n$ that is observed taking on value $X = x_0$
- Replacing X by value x_0 in its conditional density results in a function

$$\phi(y_1, \ldots, y_n) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp\left\{ -\frac{(x_0 - (\beta_0 + \sum_{i=1}^{n} \beta_i y_i))^2}{2\sigma_x^2} \right\}$$

Entering evidence

- Replacing X by value x_0 in its conditional density results in a function

$$\phi(y_1, \ldots, y_n) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp\left\{ -\frac{(x_0 - (\beta_0 + \sum_{i=1}^{n} \beta_i y_i))^2}{2\sigma_x^2} \right\}$$

- Eventually, function ϕ will be passed to the bucket corresponding to one of its parents, say Y_j, and will be multiplied by the parent's density

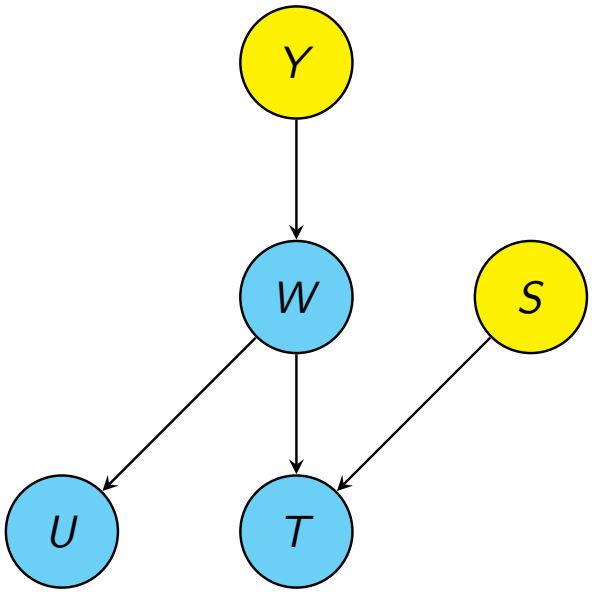
$$f(y_j | Pa(Y_j)) = \frac{1}{\sigma_{y_j} \sqrt{2\pi}} \exp\left\{ -\frac{(y_j - \mu_{y_j|pa(y_j)})^2}{2\sigma_{y_j}^2} \right\}.$$

- Maximizing the product of ϕ and f with respect to y_j is equivalent to maximizing the sum of their respective logarithms. It is obtained by solving

$$\frac{\partial}{\partial y_j} \left(-\frac{(x_0 - (\beta_0 + \sum_{i=1}^{n} \beta_i y_i))^2}{2\sigma_x^2} - \frac{(y_j - \mu_{y_j|pa(y_j)})^2}{2\sigma_{y_j}^2} \right) = 0$$

which simply amounts to maximizing a quadratic function.

MPE with unrestricted elimination order



$$P(Y) = (0.5, 0.5)$$
$$P(S) = (0.1, 0.9)$$
$$f(w|Y=0) = \mathcal{N}(w; -1, 1)$$
$$f(w|Y=1) = \mathcal{N}(w; 2, 1)$$
$$f(t|w, S=0) = \mathcal{N}(t; -w, 1)$$
$$f(t|w, S=1) = \mathcal{N}(t; w, 1)$$
$$f(u|w) = \mathcal{N}(u; w, 1)$$

- Elimination order: W, T, S, Y
- Observation: $U = 1$

MPE with unrestricted elimination order

- Elimination order: W, T, S, Y
- Observation: $U = 1$

$$B_W : \ f(u = 1|w)$$

$$B_T : \ 1$$

$$B_S : \ P(S), f(t|w, S)$$

$$B_Y : \ P(Y), f(w|Y)$$

MPE with unrestricted elimination order

- Elimination order: W, T, S, Y
- Observation: $U = 1$

$$B_W: \quad f(u=1|w) \quad h_1^Y(w)$$

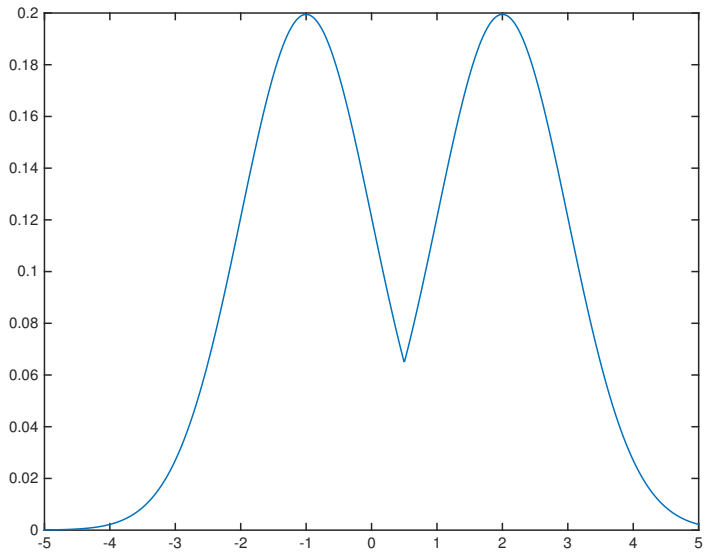
$$B_T: \quad 1$$

$$B_S: \quad P(S), f(t|w,S)$$

$$B_Y: \quad P(Y), f(w|Y)$$

MPE with unrestricted elimination order

$$
\begin{aligned}
h_1^Y(w) &= \max_y P(y) f(w|y) \\
&= \max[P(Y=0) f(w|Y=0), P(Y=1) f(w|Y=1)]
\end{aligned}
$$



We represent it as a list

MPE with unrestricted elimination order

- Elimination order: W, T, S, Y
- Observation: $U = 1$

$$B_W: \quad f(u=1|w) \quad h_1^Y(w)$$

$$B_T: \quad 1 \quad h_2^S(t,w)$$

$$B_S: \quad P(S), f(t|w,S)$$

$$B_Y: \quad P(Y), f(w|Y)$$

$$h_2^S(t,w) = \max[P(S=0)f(t|w,S=0), P(S=1)f(t|w,S=1)]$$

MPE with unrestricted elimination order

- Elimination order: W, T, S, Y
- Observation: $U = 1$

$$B_W: \quad f(u=1|w) \quad h_1^Y(w) \quad h_3(w)$$

$$B_T: \quad 1 \quad h_2^S(t,w)$$

$$B_S: \quad P(S), f(t|w,S)$$

$$B_Y: \quad P(Y), f(w|Y)$$

$$h_3(w) = \max_t \max_s P(s) f(t|w,s) = \max_s P(s) \max_t f(t|w,s)$$

MPE with unrestricted elimination order

- Elimination order: W, T, S, Y
- Observation: $U = 1$

$$B_W : \quad f(u=1|w) \quad h_1^Y(w) \quad h_3(w)$$

$$B_T : \quad 1 \quad h_2^S(t,w)$$

$$B_S : \quad P(S), f(t|w,S)$$

$$B_Y : \quad P(Y), f(w|Y)$$

MPE with unrestricted elimination order

- Elimination order: W, T, S, Y
- Observation: $U = 1$

$$B_W: \quad f(u=1|w) \quad h_1^Y(w)$$

$$B_T: \quad 1 \quad h_2^S(t,w)$$

$$B_S: \quad P(S), f(t|w,S)$$

$$B_Y: \quad P(Y), f(w|Y)$$

MPE with unrestricted elimination order

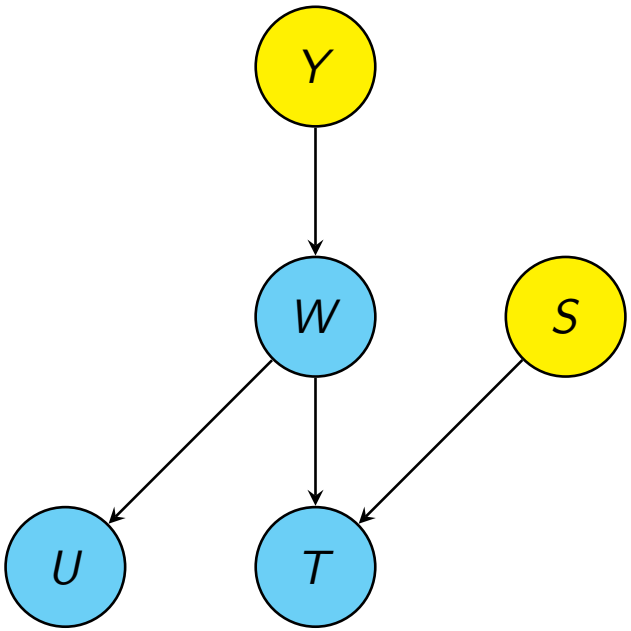
$$
\begin{aligned}
h_4^Y &= \max_w [f(U=1|w)h_1(w)] \\
&= \max_w [f(U=1|w)\max[P(Y=0)f(w|Y=0), P(Y=1)f(w|Y=1)]] \\
&= \max[\max_w f(U=1|w)P(Y=0)f(w|Y=0), \\
&\qquad \max_w f(U=1|w)P(Y=1)f(w|Y=1)]]
\end{aligned}
$$

- The two maximizations over w can easily be solved analytically :

$$\frac{\partial}{\partial w}\left(-\frac{1}{2}(1-\beta_U w)^2 - \frac{1}{2}(w-\mu_{W,Y=i})^2\right) = 0$$

- Again, the MPE configuration is obtained by tracing back the calculations

Extension to MAP inference



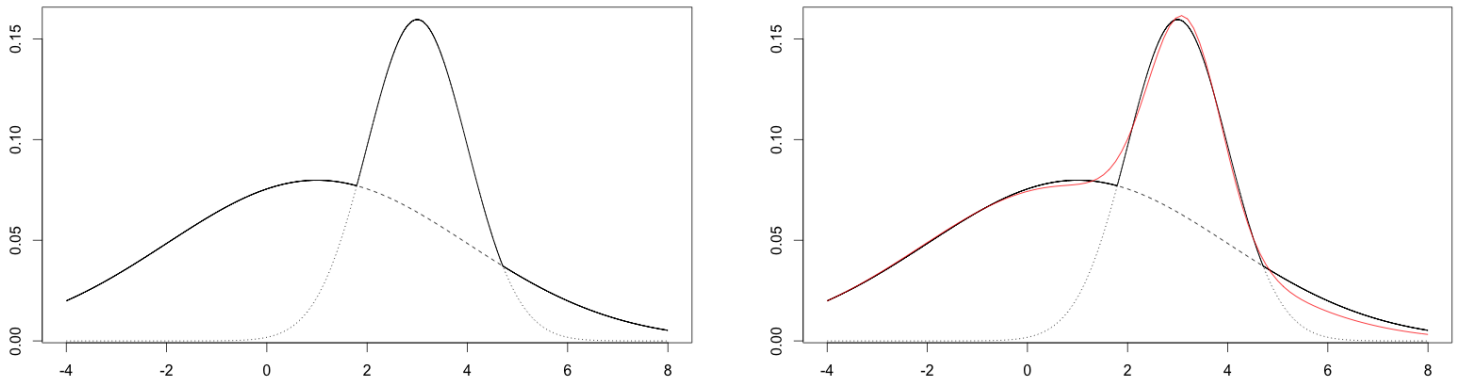
- Assume we are interested in the MAP configuration over Y and T
- Eliminating S (by summation) will result in a mixture of Gaussians potential, while eliminating T (by maximization) results in a maximum of Gaussians potential
- The two potentials should later be combined.
- This is unsatisfactory from a computational point of view.

Conclusions

- Same complexity as (marginal) probabilistic inference
  - The elimination order is able to exploit the conditional independencies in the model structure, and we therefore avoid the computational blow-up of having to consider all combinations of the discrete variables
  - Easy obtention of the MPE configuration of the continuous variables as either corresponding to the conditional means of the densities involved or by maximizing a quadratic function.
- Calculations are exact
- The key contributor to the complexity is maintaining the list of Gaussian components representing the densities of the unobserved continuous variables.

Current work

- A technique to approximate the max-potentials using sum-potentials, which will enable us to do the calculations using a single data structure



- Selecting optimal variable orders for computing the buckets
- Approximation using simulated annealing

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 619209