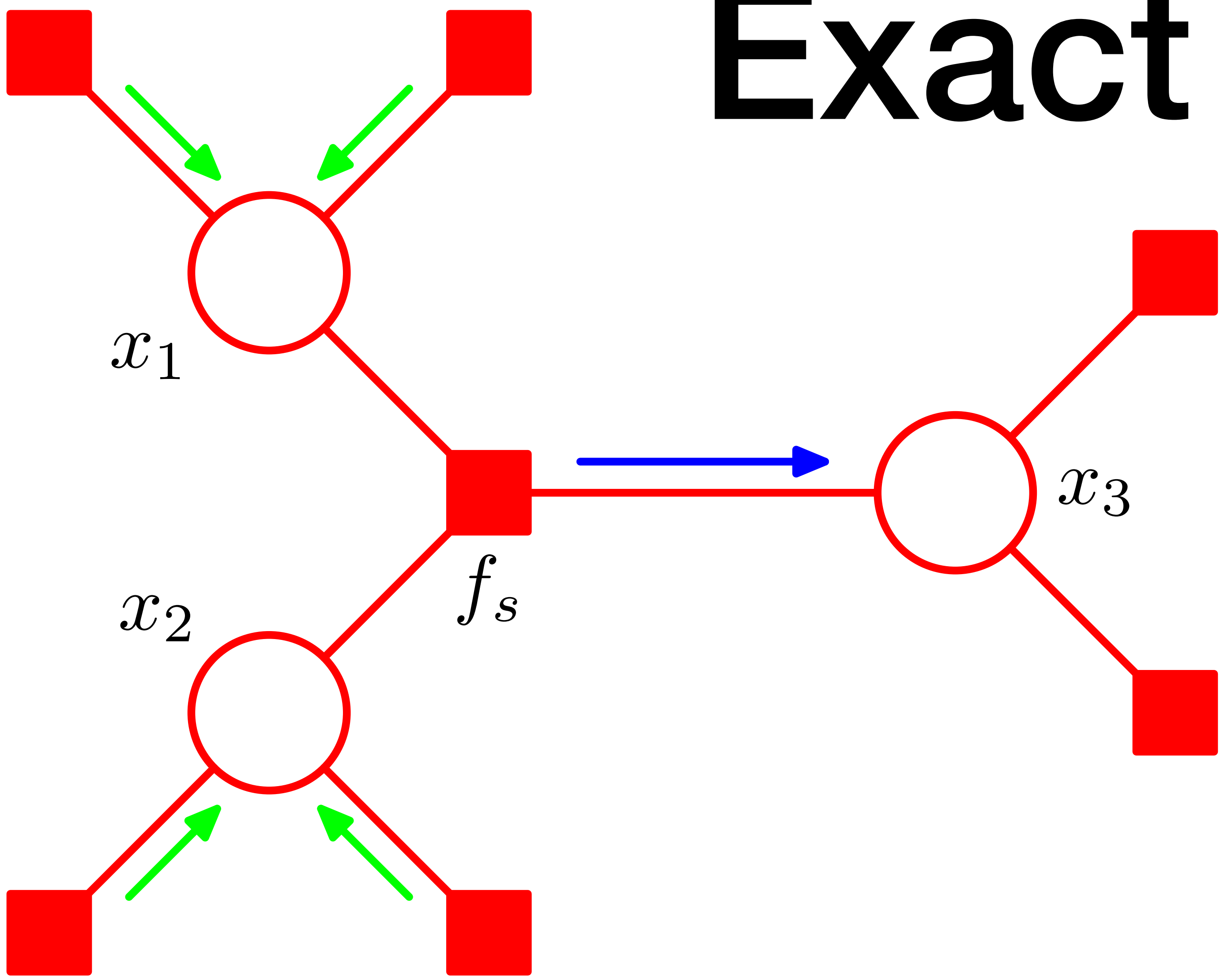
Exact Inference



Factor Graphs through Max-Sum Algorithm

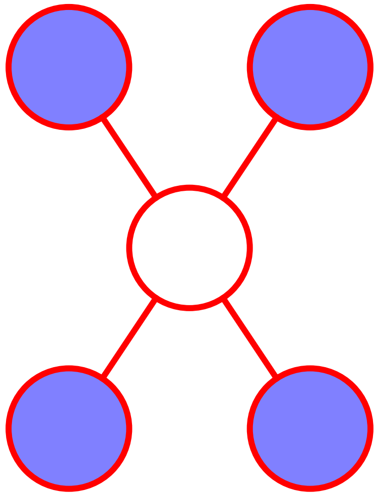
Figures from Bishop PRML Sec. 8.3/8.4

Geoffrey Roeder
roeder@cs.toronto.edu
8 February 2018

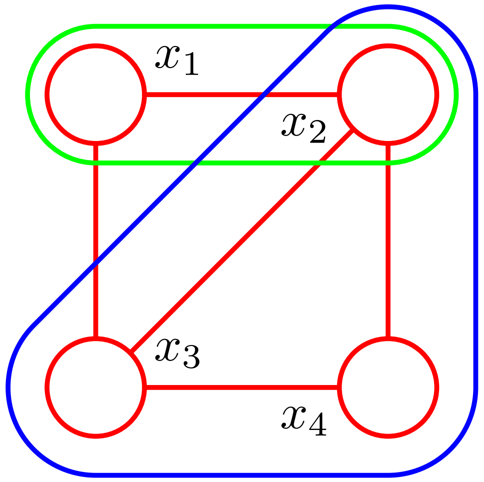
Building Blocks

UGMs, Cliques, Factor Graphs

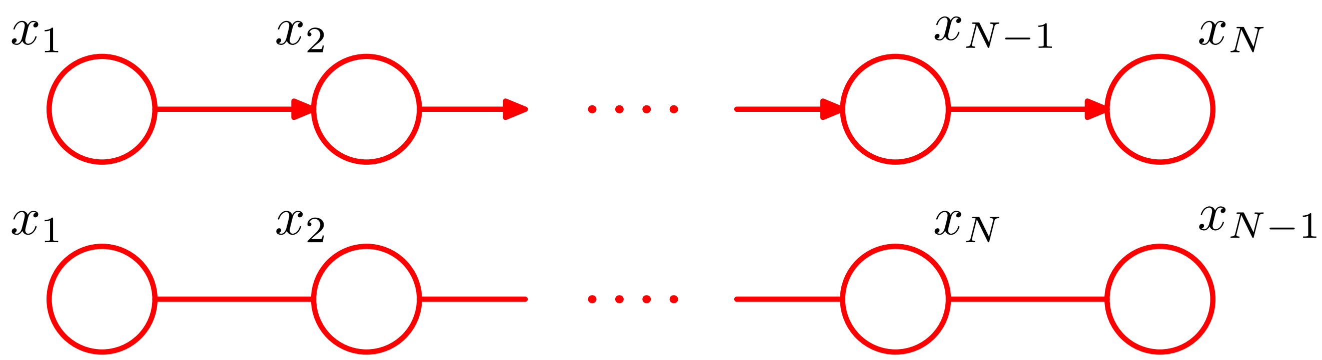
Markov Random Fields / UGMs



Parameterization: maximal cliques



Example: Equivalent DGM and UGM

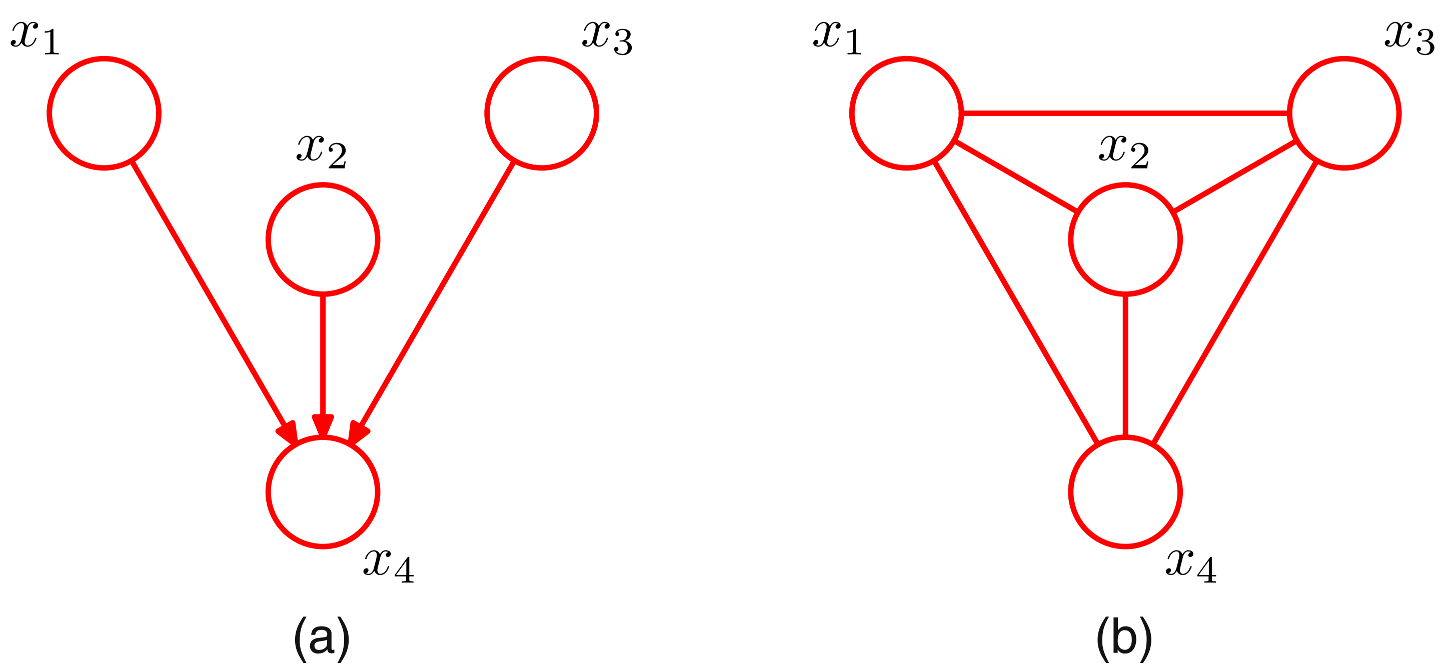


$$p(\mathbf{x}) = p(x_1)p(x_2|x_1)p(x_3|x_2)\cdots p(x_N|x_{N-1})$$

$$p(\mathbf{x}) = \frac{1}{Z}\psi_{1,2}(x_1,x_2)\psi_{2,3}(x_2,x_3)\cdots\psi_{N-1,N}(x_{N-1},x_N)$$

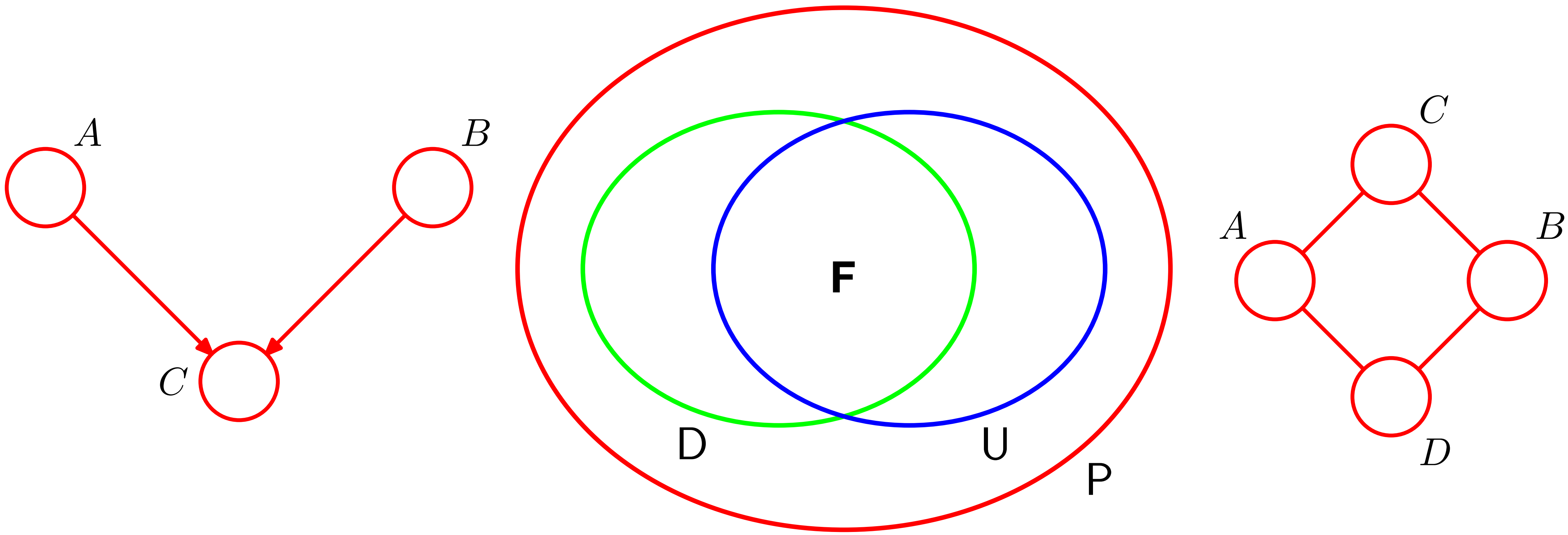
$$\begin{aligned}
\psi_{1,2}(x_1,x_2) &= p(x_1)p(x_2|x_1)\\
\psi_{2,3}(x_2,x_3) &= p(x_3|x_2)\\
&\vdots\\
\psi_{N-1,N}(x_{N-1},x_N) &= p(x_N|x_{N-1})
\end{aligned}$$

Conversion: “Moralization” (Marry the Parents of Every Child)



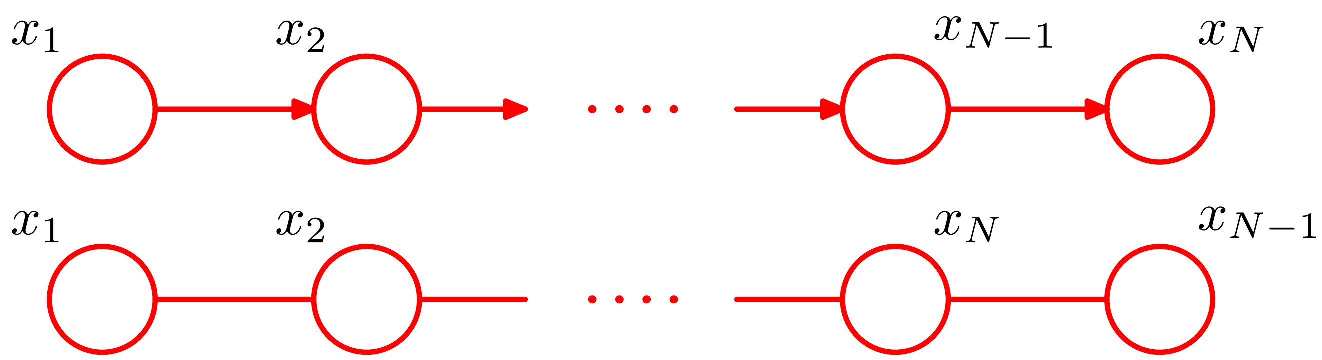
$$p(\mathbf{x}) = p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2, x_3)$$

DGMs and UGMs represent distinct distributions



Motivations:
Exact Inference in a Chain

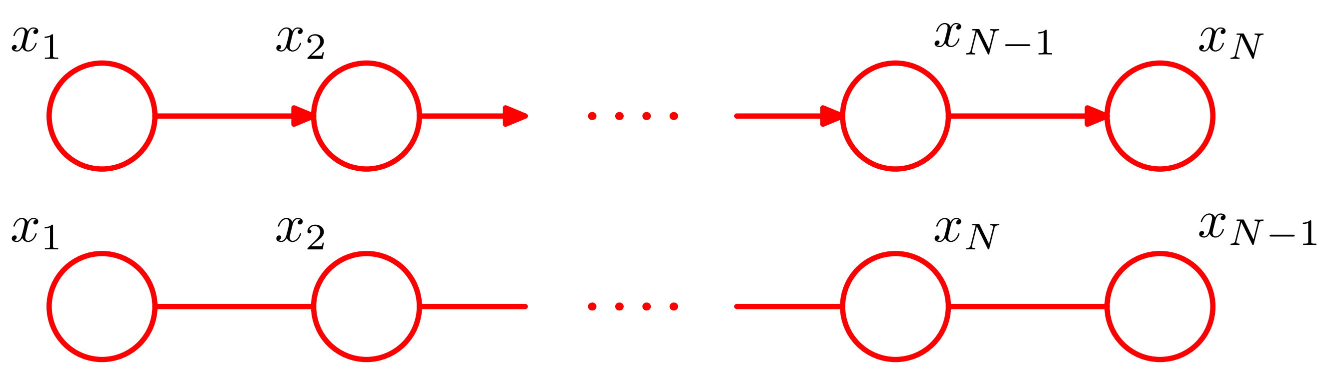
Query probability of a configuration for node X_n: p(x_n)



$$p(\mathbf{x}) = \frac{1}{Z}\psi_{1,2}(x_1, x_2)\psi_{2,3}(x_2, x_3)\cdots\psi_{N-1,N}(x_{N-1}, x_N)$$

$$p(x_n) = \sum_{x_1}\cdots\sum_{x_{n-1}}\sum_{x_{n+1}}\cdots\sum_{x_N} p(\mathbf{x}).$$

Query probability of a configuration for node X_n: p(x_n)



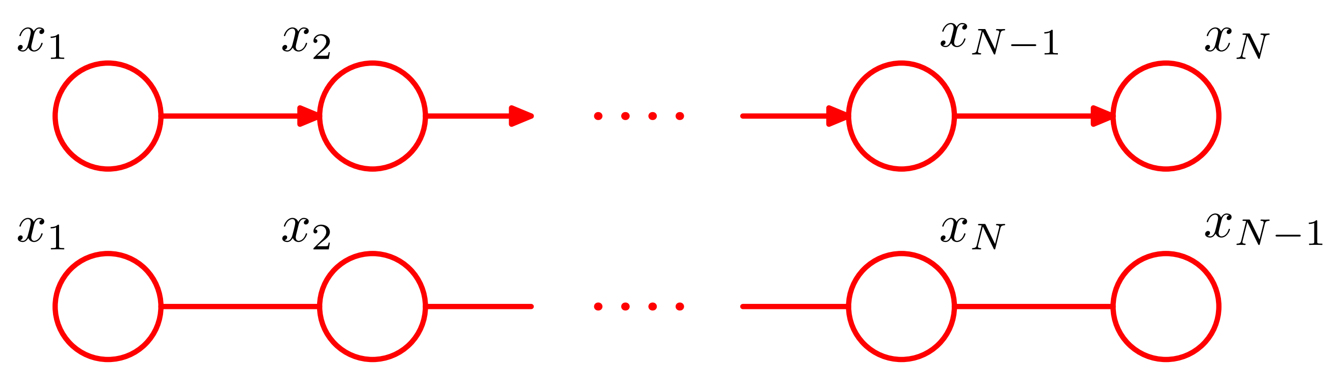
$$p(\mathbf{x}) = \frac{1}{Z}\psi_{1,2}(x_1, x_2)\psi_{2,3}(x_2, x_3)\cdots\psi_{N-1,N}(x_{N-1}, x_N)$$

$$p(x_n) = \sum_{x_1}\cdots\sum_{x_{n-1}}\sum_{x_{n+1}}\cdots\sum_{x_N} p(\mathbf{x}).$$

Naively:
N variables, K states per variable: computation complexity?

$$p(x_n) = \sum_{x_1}\cdots\sum_{x_{n-1}}\sum_{x_{n+1}}\cdots\sum_{x_N} p(\mathbf{x}).$$

Query probability of a configuration for node X_n: p(x_n)



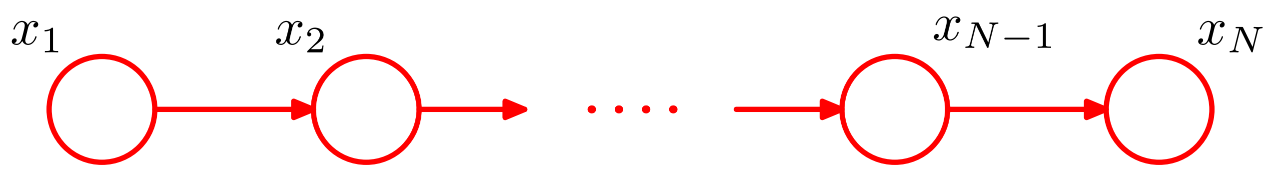
$$p(\mathbf{x}) = \frac{1}{Z}\psi_{1,2}(x_1, x_2)\psi_{2,3}(x_2, x_3)\cdots\psi_{N-1,N}(x_{N-1}, x_N)$$

$$p(x_n) = \sum_{x_1}\cdots\sum_{x_{n-1}}\sum_{x_{n+1}}\cdots\sum_{x_N} p(\mathbf{x}).$$

We ignored the conditional independence! Notice for x_n:

$$\sum_{x_N}\psi_{N-1,N}(x_{N-1}, x_N)$$

Be clever about order of computation:

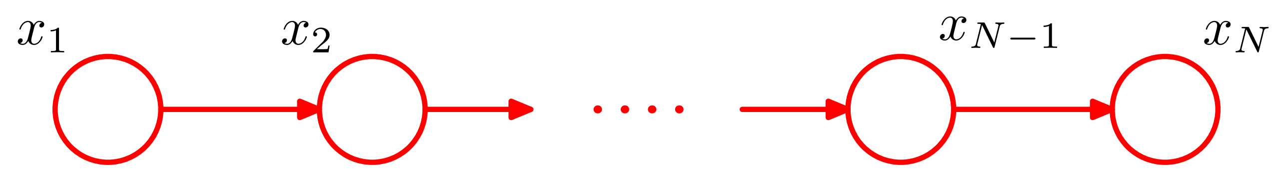


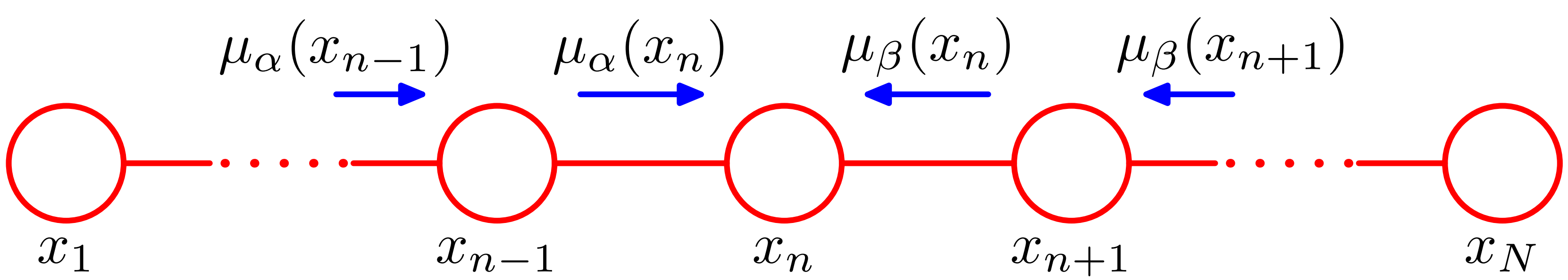
$$p(x_n) = \frac{1}{Z} \underbrace{\left[\sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n) \cdots \left[\sum_{x_2} \psi_{2,3}(x_2, x_3) \left[\sum_{x_1} \psi_{1,2}(x_1, x_2)\right]\right] \cdots \right]}_{\mu_\alpha(x_n)}$$

$$\underbrace{\left[\sum_{x_{n+1}} \psi_{n,n+1}(x_n, x_{n+1}) \cdots \left[\sum_{x_N} \psi_{N-1,N}(x_{N-1}, x_N)\right] \cdots \right]}_{\mu_\beta(x_n)} . \qquad (8.52)$$

$$p(x_n) = \frac{1}{Z} \mu_\alpha(x_n) \mu_\beta(x_n)$$

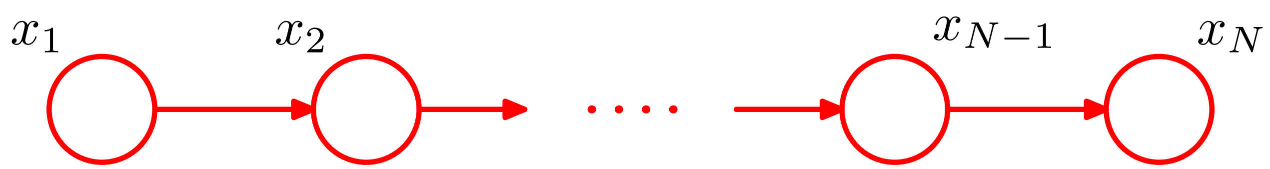
Be clever about order of computation:

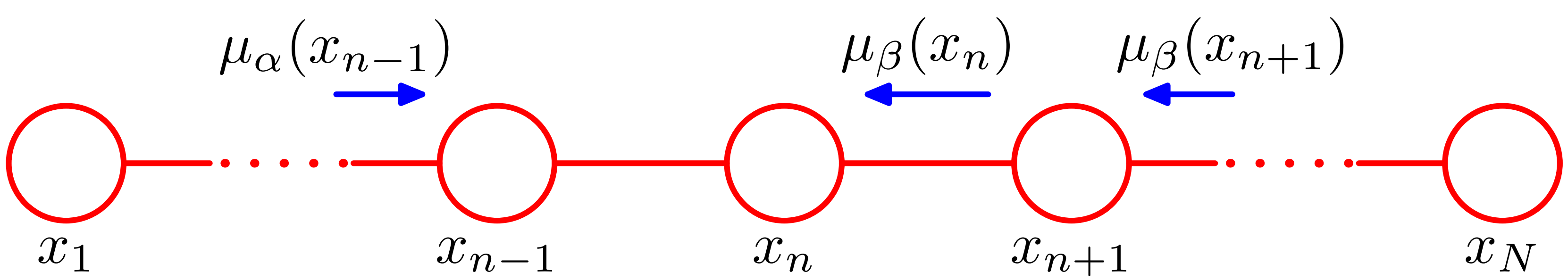




$$p(x_n) = \frac{1}{Z}\mu_\alpha(x_n)\mu_\beta(x_n)$$

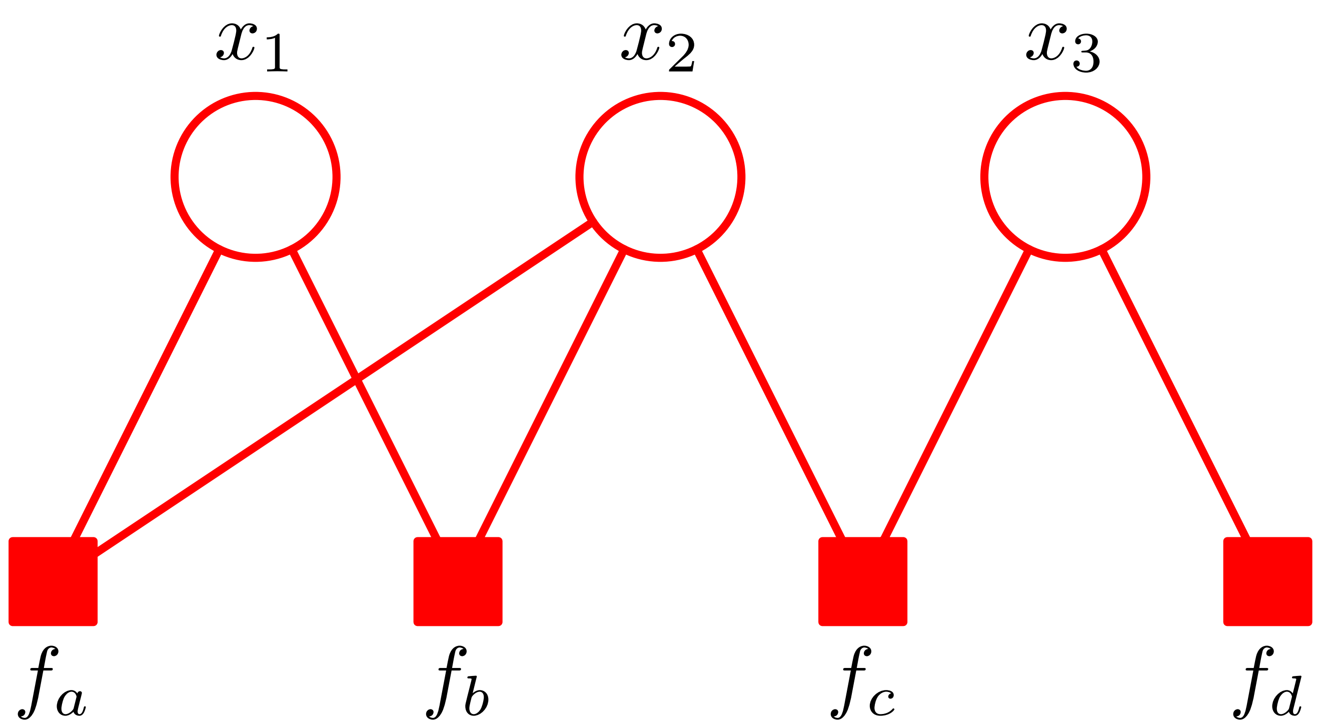
We get joint marginals over variables, too:





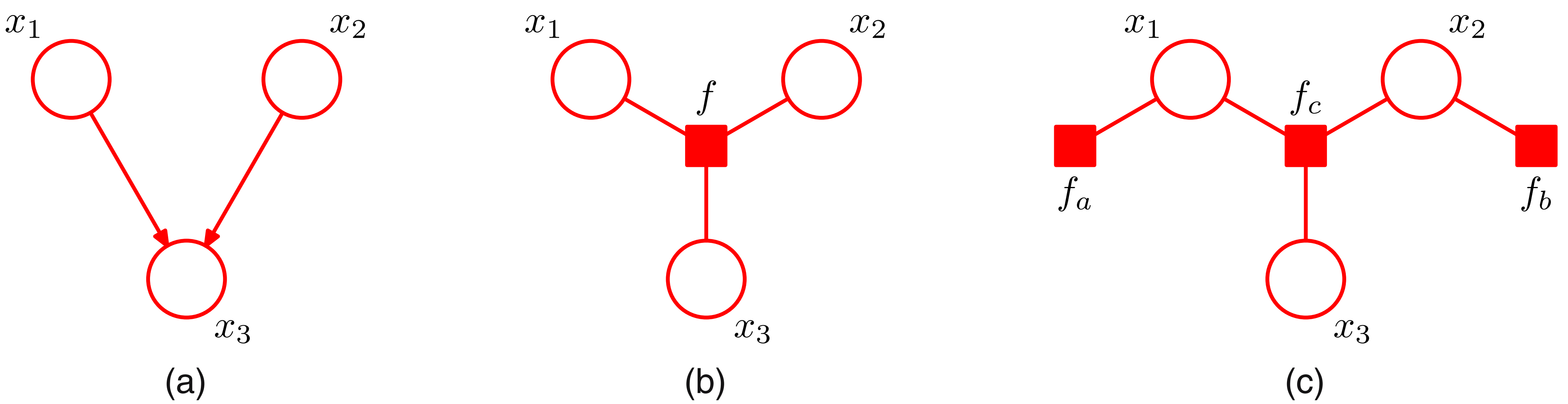
$$p(x_{n-1}, x_n) = \frac{1}{Z}\mu_\alpha(x_{n-1})\psi_{n-1,n}(x_{n-1}, x_n)\mu_\beta(x_n)$$

Factor Graph Review



$$p(\mathbf{x}) = \prod_s f(\mathbf{x}_s)$$

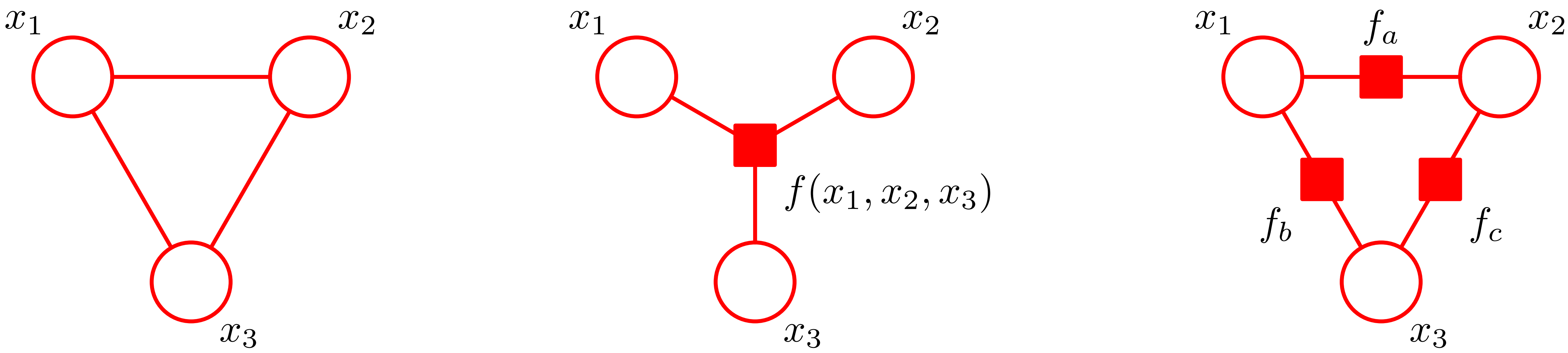
$$p(\mathbf{x}) = f_a(x_1, x_2) f_b(x_1, x_2) f_c(x_2, x_3) f_d(x_3)$$



$$(a) \quad p(x_1)p(x_2)p(x_3|x_1, x_2)$$

$$(b) \quad f(x_1, x_2, x_3) = p(x_1)p(x_2)p(x_3|x_1, x_2)$$

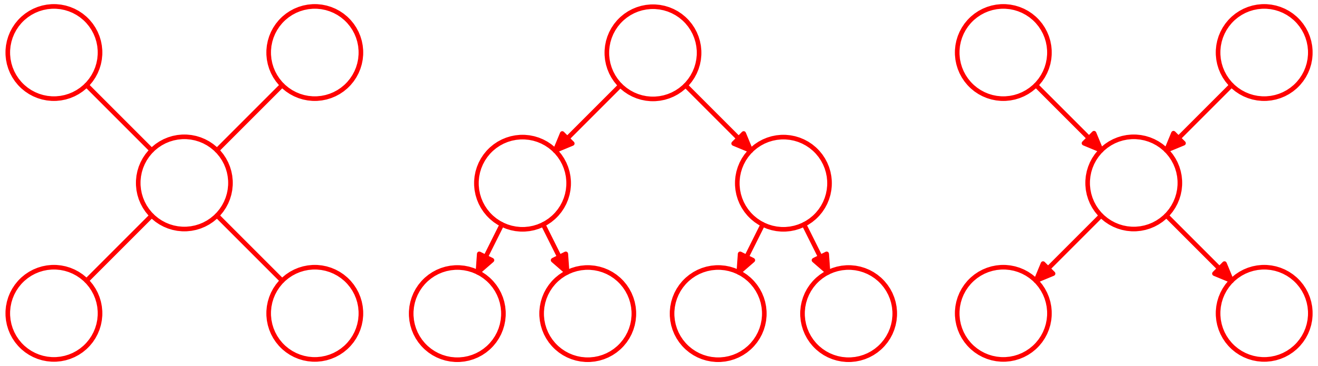
$$(c) \quad f_a(x_1) = p(x_1), f_b(x_2) = p(x_2), f_c(x_1, x_2, x_3) = p(x_3|x_1, x_2)$$



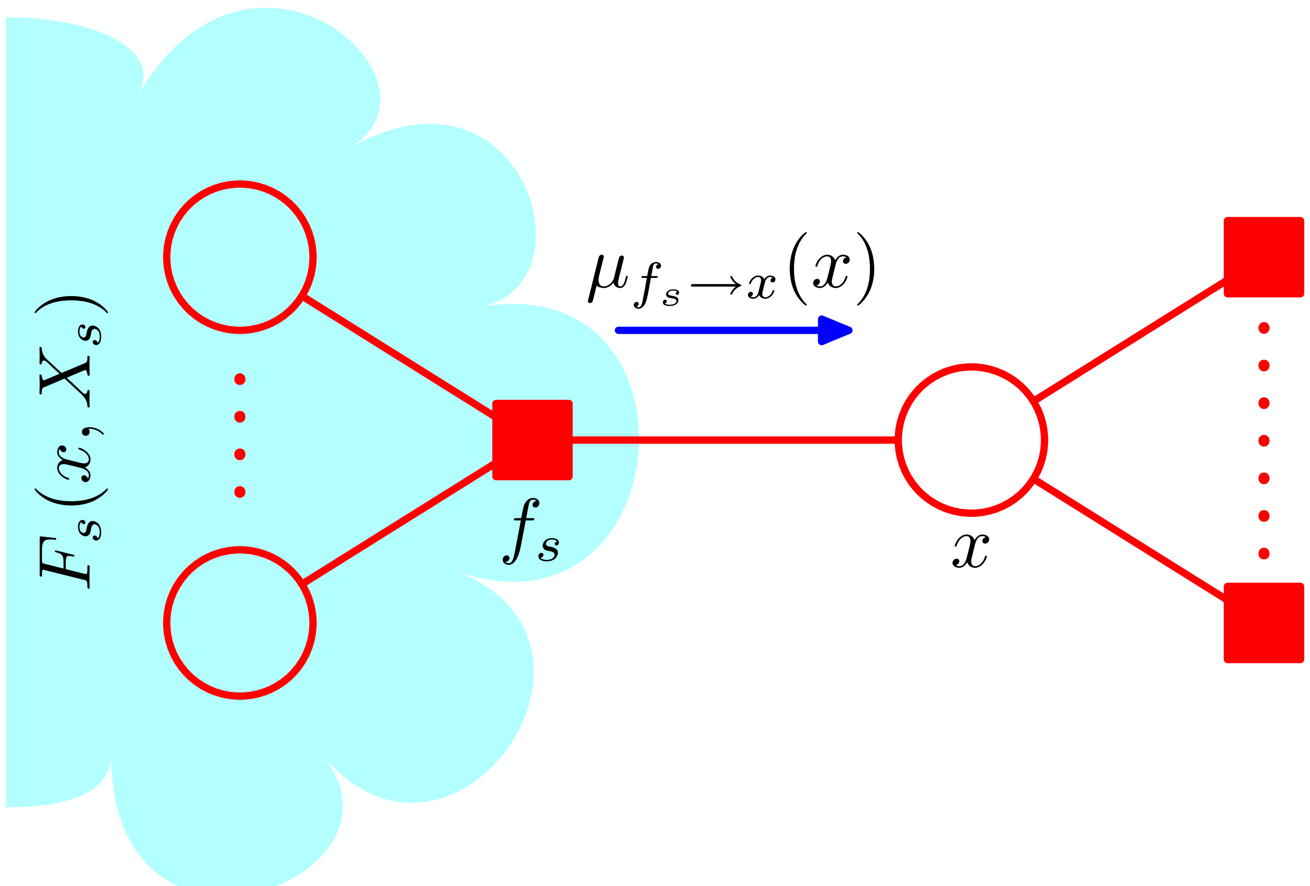
x_1
x_2
x_3
x_1
x_2
$f(x_1, x_2, x_3)$
x_3
x_1
f_a
x_2
f_b
f_c
x_3

Sum-Product Algorithm

Generalize Exact Inference in Chains
to Tree-Structured PGMs



Problem setup: notation



$$p(x) = \sum_{\mathbf{x} \setminus x} p(\mathbf{x})$$

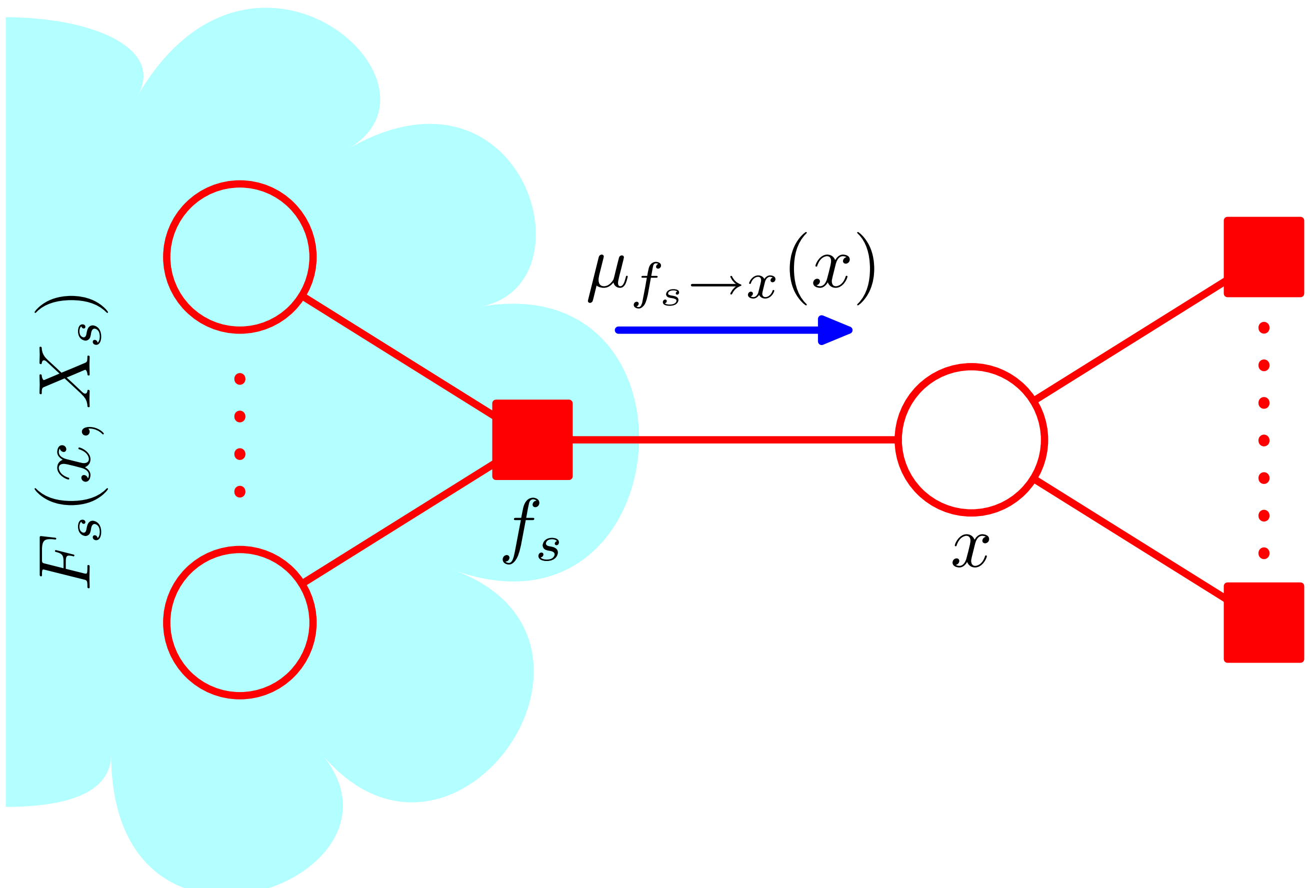
$$p(\mathbf{x}) = \prod_{s \in \mathrm{ne}(x)} F_s(x, X_s)$$

ne(x): set of factor nodes that are neighbours of x

X_s: set of all variables in the subtree connected to the variable node x via factor node f_s

F_s(x, X_s): the product of all the factors in the group associated with factor f_s

Problem setup: notation



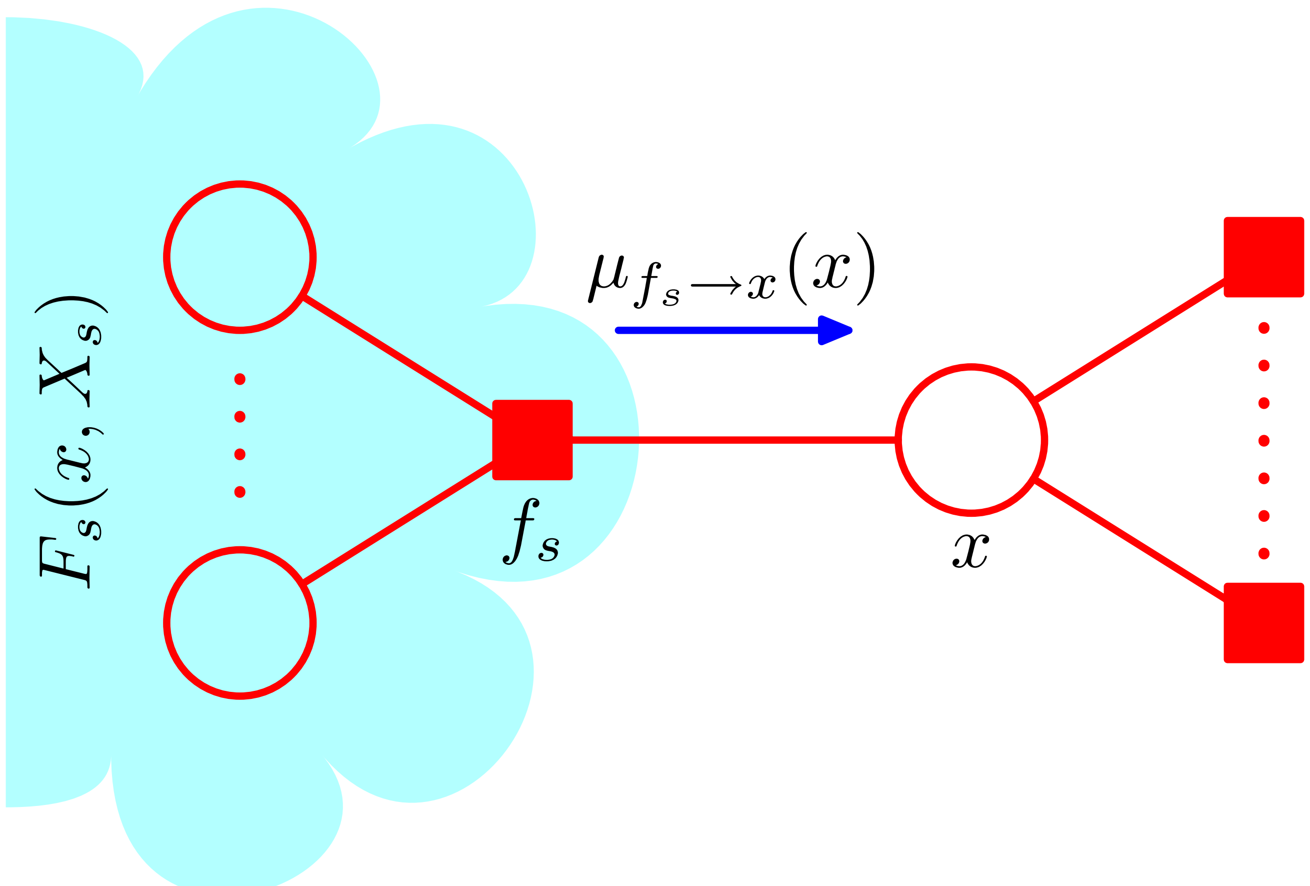
$$
\begin{aligned}
p(x) &= \prod_{s \in \mathrm{ne}(x)} \left[\sum_{X_s} F_s(x, X_s) \right] \\
&= \prod_{s \in \mathrm{ne}(x)} \mu_{f_s \to x}(x).
\end{aligned}
$$

ne(x): set of factor nodes that are neighbours of x

X_s: set of all variables in the subtree connected to the variable node x via factor node f_s

F_s(x, X_s): the product of all the factors in the group associated with factor f_s

Problem setup: notation



$$p(x) = \prod_{s \in \mathrm{ne}(x)} \left[\sum_{X_s} F_s(x, X_s) \right] = \prod_{s \in \mathrm{ne}(x)} \mu_{f_s \to x}(x).$$

We evaluate the marginal p(x) as product of messages from surrounding factors!

Factor messages: decomposition

$$
\begin{aligned}
p(x) &= \prod_{s \in \mathrm{ne}(x)} \left[\sum_{X_s} F_s(x, X_s) \right] \\
&= \prod_{s \in \mathrm{ne}(x)} \boxed{\mu_{f_s \to x}(x).}
\end{aligned}
$$

Each factor is itself described by a factor sub-graph, so we can decompose:

$$
F_s(x, X_s) = f_s(x, x_1, \ldots, x_M) G_1(x_1, X_{s1}) \ldots G_M(x_M, X_{sM})
$$

(Each variable associated with f_x is {x, x_1, ..., x_M})

Rewriting the factor-to-variable message:

$$
\boxed{\mu_{f_s \to x}(x)} = \sum_{x_1} \cdots \sum_{x_M} f_s(x, x_1, \ldots, x_M) \prod_{m \in \mathrm{ne}(f_s) \setminus x} \left[\sum_{X_{xm}} G_m(x_m, X_{sm}) \right]
$$

Factor messages: decomposition

$$p(x) = \prod_{s \in \mathrm{ne}(x)} \left[\sum_{X_s} F_s(x, X_s) \right] = \prod_{s \in \mathrm{ne}(x)} \mu_{f_s \to x}(x).$$

Each factor is itself described by a factor sub-graph, so we can decompose:

$$F_s(x, X_s) = f_s(x, x_1, \ldots, x_M) G_1(x_1, X_{s1}) \ldots G_M(x_M, X_{sM})$$

(Each variable associated with f_x is {x, x_1, …, x_M})

Rewriting the factor-to-variable message:

$$\mu_{f_s \to x}(x) = \sum_{x_1} \ldots \sum_{x_M} f_s(x, x_1, \ldots, x_M) \prod_{m \in \mathrm{ne}(f_s) \setminus x} \left[\sum_{X_{xm}} G_m(x_m, X_{sm}) \right]$$

$$= \sum_{x_1} \ldots \sum_{x_M} f_s(x, x_1, \ldots, x_M) \prod_{m \in \mathrm{ne}(f_s) \setminus x} \boxed{\mu_{x_m \to f_s}(x_m)} \qquad (8.66)$$

Factor-to-variable messages: decomposition

$$p(x) = \prod_{s \in \mathrm{ne}(x)} \left[\sum_{X_s} F_s(x, X_s) \right] = \prod_{s \in \mathrm{ne}(x)} \mu_{f_s \to x}(x).$$

Each factor is itself described by a factor sub-graph, so we can decompose:

$$F_s(x, X_s) = f_s(x, x_1, \ldots, x_M) G_1(x_1, X_{s1}) \ldots G_M(x_M, X_{sM})$$

(Each variable associated with f_x is {x, x_1, …, x_M})

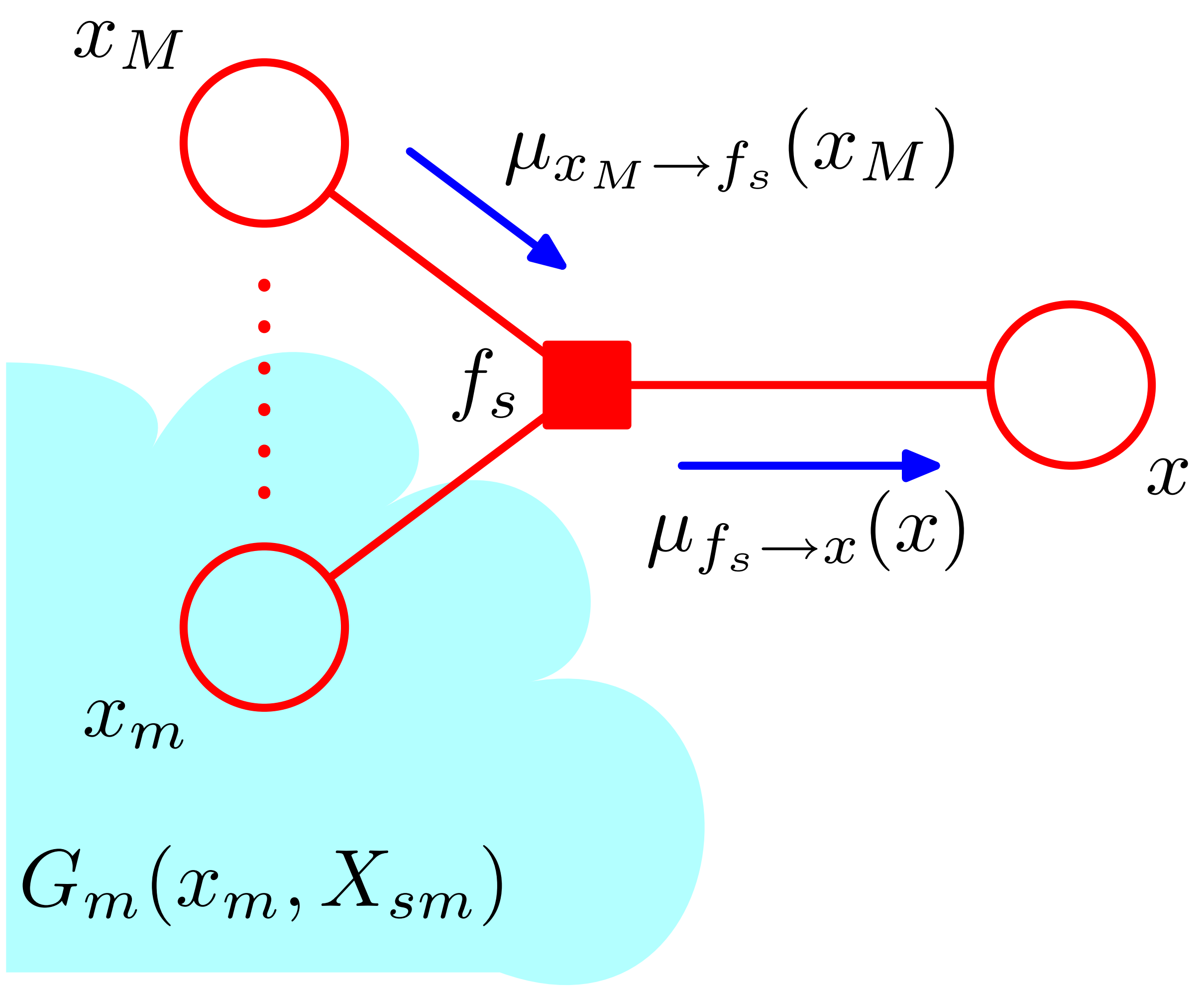
Rewriting the factor-to-variable message:

$$\mu_{f_s \to x}(x) = \sum_{x_1} \ldots \sum_{x_M} f_s(x, x_1, \ldots, x_M) \prod_{m \in \mathrm{ne}(f_s) \setminus x} \left[\sum_{X_{xm}} G_m(x_m, X_{sm}) \right]$$

$$= \sum_{x_1} \ldots \sum_{x_M} f_s(x, x_1, \ldots, x_M) \prod_{m \in \mathrm{ne}(f_s) \setminus x} \mu_{x_m \to f_s}(x_m) \qquad (8.66)$$

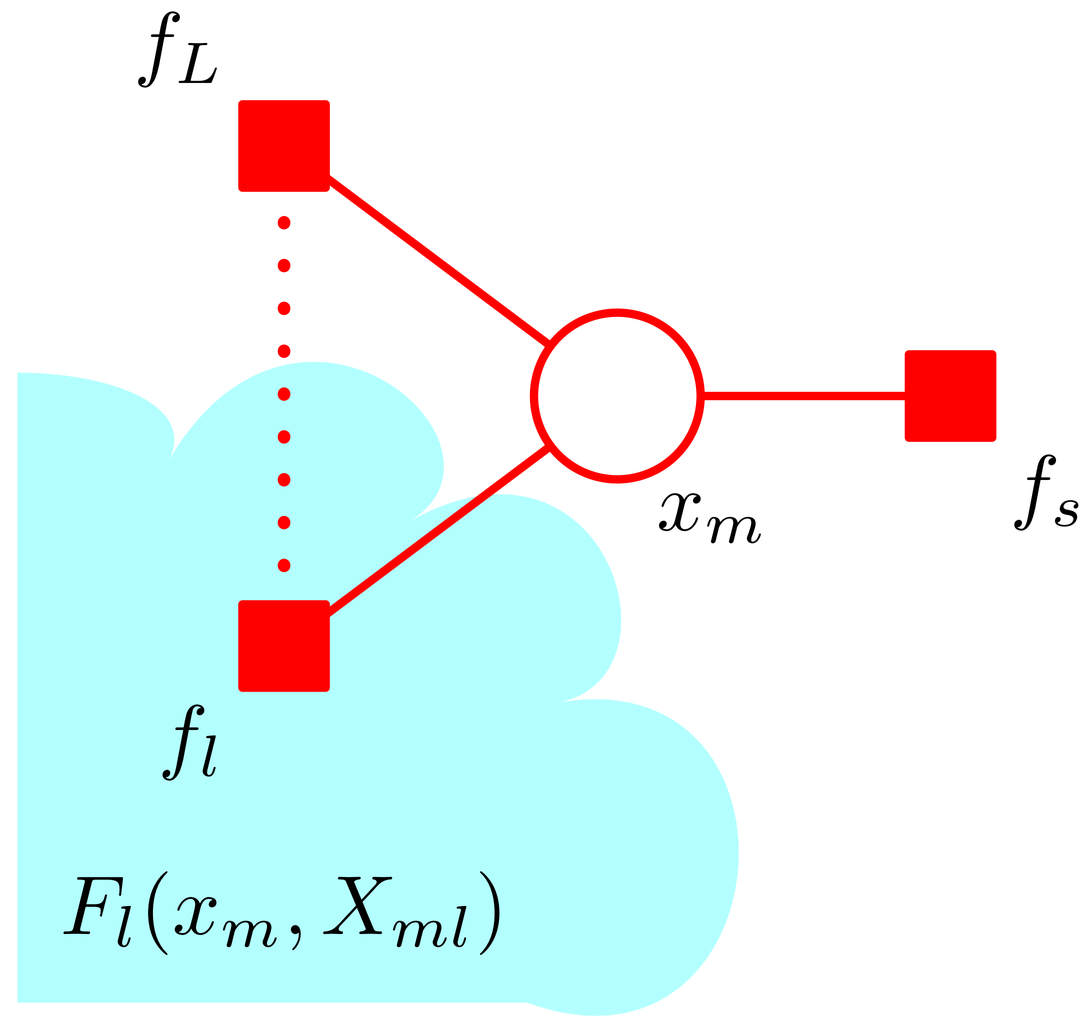
$$\mu_{x_m \to f_s}(x_m) \equiv \sum_{X_{sm}} G_m(x_m, X_{sm})$$

Variable-to-factor messages: decomposition



$$\mu_{x_m \to f_s}(x_m) \equiv \sum_{X_{sm}} G_m(x_m, X_{sm})$$

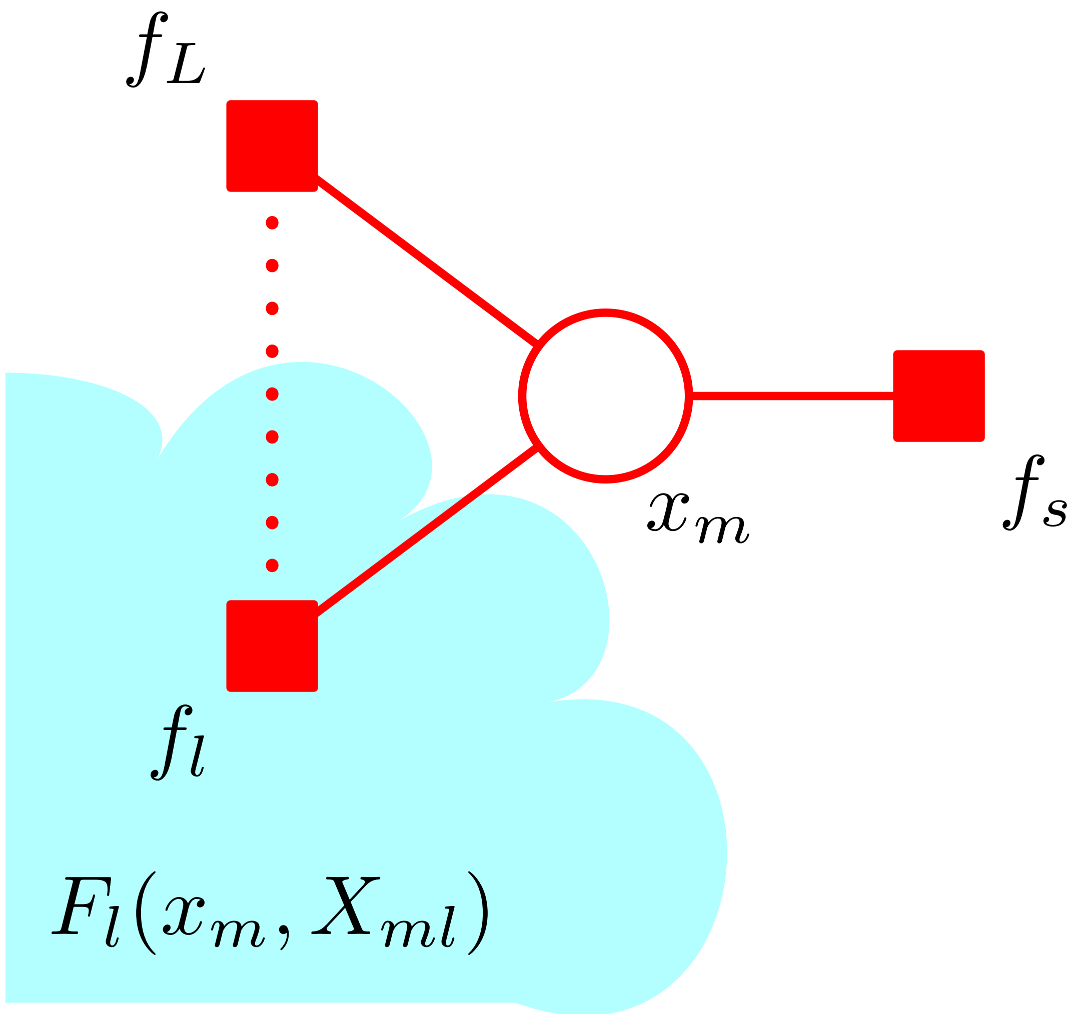
Factor-to-variable messages: one step back towards the leaves



$$F_s(x, X_s) = f_s(x, x_1, \ldots, x_M) G_1(x_1, X_{s1}) \ldots G_M(x_M, X_{sM})$$

$$G_m(x_m, X_{sm}) = \prod_{l \in \mathrm{ne}(x_m) \backslash f_s} F_l(x_m, X_{ml})$$

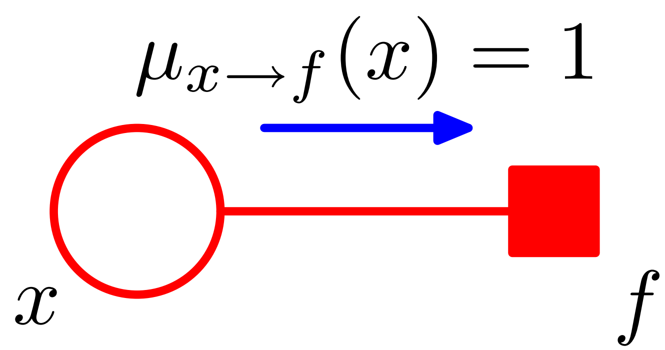
Factor-to-variable messages: one step back towards the leaves

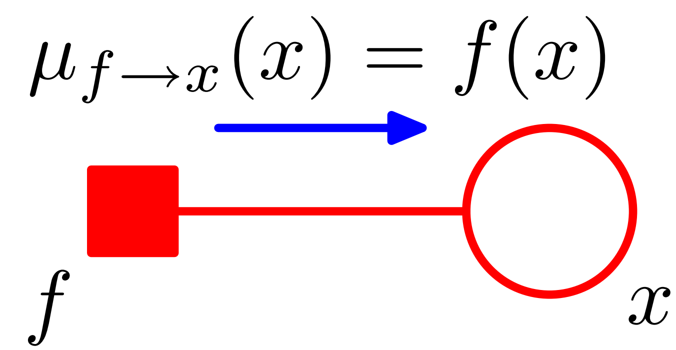


$$\mu_{x_m \to f_s}(x_m) = \prod_{l \in \mathrm{ne}(x_m) \setminus f_s} \left[\sum_{X_{ml}} F_l(x_m, X_{ml}) \right]$$

$$= \prod_{l \in \mathrm{ne}(x_m) \setminus f_s} \mu_{f_l \to x_m}(x_m)$$

Sum-Product Initialization at Leaves





Sum-Product: Marginal distribution over x

$$\begin{aligned} p(x) &= \prod_{s \in \mathrm{ne}(x)} \left[\sum_{X_s} F_s(x, X_s) \right] \\ &= \prod_{s \in \mathrm{ne}(x)} \mu_{f_s \to x}(x). \end{aligned}$$

See Bishop p. 409 for a fully worked, simple example!