Model evidence and gate messages in Infer.NET

Thomas Minka
Microsoft Research Ltd., Cambridge, UK

June 18, 2013

1 Simplifying $\tilde{Z}$

Suppose we have run expectation propagation (EP) to convergence on a factor graph $\prod_a f_a(\mathbf{x})$ and obtained a set of messages $m_{a\to i}(x_i)$ representing a fully factorized approximate distribution $q(\mathbf{x}) = \prod_a \prod_i m_{a\to i}(x_i)$. How can we efficiently compute the EP normalizing constant $\tilde{Z}$? Assuming the messages were properly scaled during message passing, we should have $\tilde{Z} = \int_{\mathbf{x}} q(\mathbf{x})d\mathbf{x}$. But if the messages were normalized or otherwise rescaled during message passing, as is common for numerical stability, we need to do some extra work to reconstruct the optimal scale factors.

Given any set of messages, properly scaled or not, the optimal normalizing constant for EP is (Minka, 2005) ($\alpha = 1$):

$$\tilde{Z} = \left(\int_{\mathbf{x}} q(\mathbf{x})d\mathbf{x}\right) \prod_{a=1}^{A} s_a \tag{1}$$

$$\text{where } s_a = \frac{\displaystyle\int_{\mathbf{x}} \frac{f_a(\mathbf{x})}{\prod_i m_{a\to i}(x_i)} q(\mathbf{x})d\mathbf{x}}{\displaystyle\int_{\mathbf{x}} q(\mathbf{x})d\mathbf{x}} \tag{2}$$

To simplify this formula, let's define the following notation:

$$z_i = \int_{x_i} \prod_a m_{a\to i}(x_i)dx_i \tag{3}$$

$$\bar{q}(x_i) = \frac{1}{z_i} \prod_a m_{a\to i}(x_i) \tag{4}$$

$$z_i^{\setminus a} = \int_{x_i} \prod_{b\neq a} m_{b\to i}(x_i)dx_i \tag{5}$$

$$\bar{q}^{\setminus a}(x_i) = \frac{1}{z_i^{\setminus a}} \prod_{b\neq a} m_{b\to i}(x_i) \tag{6}$$

$$\frac{z_i}{z_i^{\setminus a}} = \int_{x_i} m_{a\to i}(x_i)\bar{q}^{\setminus a}(x_i) \tag{7}$$

$$\frac{z_i^{\setminus a}}{z_i} = \int_{x_i} \frac{\bar{q}(x_i)}{m_{a\to i}(x_i)} dx_i \tag{8}$$

Now we can simplify as follows:

$$\int_{\mathbf{x}} q(\mathbf{x})d\mathbf{x} = \prod_i \int_{x_i} \prod_a m_{a\to i}(x_i)dx_i \tag{9}$$

$$= \prod_i z_i \tag{10}$$

$$s_a = \int_{\mathbf{x}} \frac{f_a(\mathbf{x})}{\prod_i m_{a\to i}(x_i)} \prod_i \bar{q}(x_i) d\mathbf{x} \tag{11}$$

$$= \int_{\mathbf{x}} f_a(\mathbf{x}) \prod_i \frac{\bar{q}(x_i)}{m_{a\to i}(x_i)} d\mathbf{x} \tag{12}$$

$$= \int_{\mathbf{x}} f_a(\mathbf{x}) \prod_i \frac{z_i^{\setminus a}}{z_i} \bar{q}^{\setminus a}(x_i) d\mathbf{x} \tag{13}$$

Let $\mathbf{x}_a$ denote the set of variables used by f_a, and let $i \in a$ be shorthand for $x_i \in \mathbf{x}_a$. Then (13) simplifies to

$$s_a = \left(\prod_{i\in a} \frac{z_i^{\setminus a}}{z_i} \right) \int_{\mathbf{x}_a} f_a(\mathbf{x}_a) \prod_{i\in a} \bar{q}^{\setminus a}(x_i) d\mathbf{x}_a \tag{14}$$

$$= \frac{\int_{\mathbf{x}_a} f_a(\mathbf{x}_a) \prod_{i\in a} \bar{q}^{\setminus a}(x_i) d\mathbf{x}_a}{\prod_{i\in a} \int_{x_i} m_{a\to i}(x_i) \bar{q}^{\setminus a}(x_i) d\mathbf{x}_i} \tag{15}$$

$$= \frac{\int_{\mathbf{x}_a} f_a(\mathbf{x}_a) \prod_{i\in a} \bar{q}^{\setminus a}(x_i) d\mathbf{x}_a}{\int_{\mathbf{x}_a} \tilde{f}_a(\mathbf{x}_a) \prod_{i\in a} \bar{q}^{\setminus a}(x_i) d\mathbf{x}_a} \tag{16}$$

Note that $z_i^{\setminus a}/z_i$ can be computed locally as the normalizer of $\bar{q}(x_i)/m_{a\to i}(x_i)$ (8).

As a computational scheme, each variable collects its messages, computes z_i and $\bar{q}(x_i)$, sending $\bar{q}(x_i)$ to neighboring factors. Each factor then computes s_a from $\bar{q}(x_i)$ and $m_{a\to i}(x_i)$. The final result is $\tilde{Z} = \prod_i z_i \prod_a s_a$.

2 Factor-specific simplications

In the case of a unary factor which is already in the approximating family, we have $f_a(x_i) = m_{a\to i}(x_i)$ so $s_a = 1$. More generally, consider a factor whose message to x_j (for some specific j) satisfies the relation:

$$m_{a\to j}(x_j) = \int_{\mathbf{x}_a \setminus x_j} f_a(\mathbf{x}_a) \prod_{i\in a, i\neq j} (\bar{q}^{\setminus a}(x_i) dx_i) \tag{17}$$

This happens when the marginal for x_j is already in the approximating family. When (17) holds, we say that $m_{a\to j}$ is *exact*. For example, when all variables are Gaussian, the factor $I(x_i = x_j + x_k)$ sends exact messages to all its arguments. Another example is when x_j is discrete so that $m_{a\to j}$ is an exact table.

If $m_{a\to j}$ is exact for some j, then s_a simplifies as follows:

$$s_a = \left(\prod_{i\in a}\frac{z_i^{\setminus a}}{z_i}\right)\int_{\mathbf{x}_a} f_a(\mathbf{x}_a)\prod_{i\in a}\bar{q}^{\setminus a}(x_i)d\mathbf{x}_a \tag{18}$$

$$= \left(\prod_{i\in a}\frac{z_i^{\setminus a}}{z_i}\right)\int_{x_j}\left[\int_{\mathbf{x}_a\setminus x_j} f_a(\mathbf{x}_a)\prod_{i\in a, i\neq j}(\bar{q}^{\setminus a}(x_i)dx_i)\right]\bar{q}^{\setminus a}(x_j)dx_j \tag{19}$$

$$= \left(\prod_{i\in a}\frac{z_i^{\setminus a}}{z_i}\right)\int_{x_j} m_{a\to j}(x_j)\bar{q}^{\setminus a}(x_j)dx_j \tag{20}$$

$$= \left(\prod_{i\in a}\frac{z_i^{\setminus a}}{z_i}\right)\frac{z_j}{z_j^{\setminus a}} \tag{21}$$

$$= \prod_{i\in a, i\neq j}\frac{z_i^{\setminus a}}{z_i} \tag{22}$$

In general, (22) will hold whenever $m_{a\to j}$ has the exact scale factor defined by EP, i.e. $m_{a\to j}$ was not rescaled. But in practice this only happens when $m_{a\to j}$ is exact.

3 Example

Consider the factor graph defined by the following stochastic program:

$$x_1 \sim \mathcal{N}(0,1) \qquad (f_1) \tag{23}$$

$$x_2 \sim \mathcal{N}(0,1) \qquad (f_2) \tag{24}$$

$$x_3 = x_1 - x_2 \qquad (f_3) \tag{25}$$

$$\text{assert}(x_3 > 0) \qquad (f_4) \tag{26}$$

Because f_1 and f_2 are exact unary factors, $s_1 = s_2 = 1$. Because f_3 is exact for x_3, $s_3 = \frac{z_1^{\setminus 3}}{z_1}\frac{z_2^{\setminus 3}}{z_2}$. Because the messages from the unary factors are normalized, $z_1^{\setminus 3} = z_2^{\setminus 3} = 1$. The message from f_3 to x_3 is also normalized so $z_3^{\setminus 4} = 1$. Thus

$$s_4 = \frac{z_3^{\setminus 4}}{z_3}\int_{x_3} f_4(x_3)\bar{q}^{\setminus 4}(x_3)dx_3 \tag{27}$$

$$\tilde{Z} = z_1 z_2 z_3 s_1 s_2 s_3 s_4 \tag{28}$$

$$= z_1 z_2 z_3 (1)(1)\frac{1}{z_1}\frac{1}{z_2}\frac{1}{z_3}\int_{x_3} f_4(x_3)\bar{q}^{\setminus 4}(x_3)dx_3 \tag{29}$$

$$= \int_{x_3} f_4(x_3)\bar{q}^{\setminus 4}(x_3)dx_3 \tag{30}$$

The simplifications achieved in this example suggest another way of computing the evidence, which exploits the directedness of factors. The next section develops this algorithm.

4 Evidence computation on directed graphs

Suppose we have a factor graph with directed edges. The edge directions can be arbitrary, i.e. they do not have to imply any independences. Then for each node we can distinguish its parent nodes from its child nodes. Previously, we computed $\tilde{Z}$ as follows:

$$\tilde{Z} = \prod_i z_i \prod_a s_a \tag{31}$$

$$= \prod_i \left(\int_{x_i} \prod_a m_{a\to i}(x_i) dx_i \right) \prod_a \frac{\int_{\mathbf{x}_a} f_a(\mathbf{x}_a) \prod_{i\in a} \bar{q}^{\setminus a}(x_i) d\mathbf{x}_a}{\prod_{i\in a} \int_{x_i} m_{a\to i}(x_i) \bar{q}^{\setminus a}(x_i) d\mathbf{x}_i} \tag{32}$$

The denominator of this expression has one term for every edge in the factor graph. In the directed case, we will rearrange these denominator terms as follows:

$$z_i' = \frac{\int_{x_i} \prod_a m_{a\to i}(x_i) dx_i}{\prod_{a\in \mathrm{ch}(i)} \int_{x_i} m_{a\to i}(x_i) \bar{q}^{\setminus a}(x_i) d\mathbf{x}_i} \tag{33}$$

$$s_a' = \frac{\int_{\mathbf{x}_a} f_a(\mathbf{x}_a) \prod_{i\in a} \bar{q}^{\setminus a}(x_i) d\mathbf{x}_a}{\prod_{i\in \mathrm{ch}(a)} \int_{x_i} m_{a\to i}(x_i) \bar{q}^{\setminus a}(x_i) d\mathbf{x}_i} \tag{34}$$

$$\tilde{Z} = \prod_i z_i' \prod_a s_a' \tag{35}$$

$$= \prod_i \frac{\int_{x_i} \prod_a m_{a\to i}(x_i) dx_i}{\prod_{a\in \mathrm{ch}(i)} \int_{x_i} m_{a\to i}(x_i) \bar{q}^{\setminus a}(x_i) d\mathbf{x}_i} \prod_a \frac{\int_{\mathbf{x}_a} f_a(\mathbf{x}_a) \prod_{i\in a} \bar{q}^{\setminus a}(x_i) d\mathbf{x}_a}{\prod_{i\in \mathrm{ch}(a)} \int_{x_i} m_{a\to i}(x_i) \bar{q}^{\setminus a}(x_i) d\mathbf{x}_i} \tag{36}$$

Some special cases:

1. If a factor has no children (a leaf factor), then

$$s_a' = \int_{\mathbf{x}_a} f_a(\mathbf{x}_a) \prod_{i\in a} \bar{q}^{\setminus a}(x_i) d\mathbf{x}_a \tag{37}$$

2. If a factor has one child variable x_i, and $m_{a\to i}$ is exact, then $s_a' = 1$.
3. If a variable has one parent and one child factor, then $z_i' = \int_{x_i} m_{\mathrm{par}\to i}(x_i) dx_i$.

Applying these rules to the example in section 3 gives immediately the most compact form of the evidence. This is the approach that Infer.NET uses.

5 Power EP

Given any set of messages, properly scaled or not, the optimal normalizing constant for Power EP is (Minka, 2005):

$$\tilde{Z} = \left(\int_{\mathbf{x}} q(\mathbf{x}) d\mathbf{x} \right) \prod_{a=1}^{A} s_a \tag{38}$$

$$\text{where } s_a = \left(\frac{\displaystyle\int_{\mathbf{x}} \frac{f_a(\mathbf{x})^{\alpha_a}}{\prod_i m_{a\to i}(x_i)^{\alpha_a}} q(\mathbf{x}) d\mathbf{x}}{\displaystyle\int_{\mathbf{x}} q(\mathbf{x}) d\mathbf{x}} \right)^{1/\alpha_a} \tag{39}$$

To simplify this formula, let's define the following notation:

$$z_i = \int_{x_i} \prod_a m_{a\to i}(x_i) dx_i \tag{40}$$

$$\bar{q}(x_i) = \frac{1}{z_i} \prod_a m_{a\to i}(x_i) \tag{41}$$

$$z_i^{\setminus a} = \int_{x_i} m_{a\to i}(x_i)^{1-\alpha_a} \prod_{b\neq a} m_{b\to i}(x_i) dx_i \tag{42}$$

$$\bar{q}^{\setminus a}(x_i) = \frac{1}{z_i^{\setminus a}} m_{a\to i}(x_i)^{1-\alpha_a} \prod_{b\neq a} m_{b\to i}(x_i) \tag{43}$$

$$\frac{z_i}{z_i^{\setminus a}} = \int_{x_i} m_{a\to i}(x_i)^{\alpha_a} \bar{q}^{\setminus a}(x_i) \tag{44}$$

$$\frac{z_i^{\setminus a}}{z_i} = \int_{x_i} \frac{\bar{q}(x_i)}{m_{a\to i}(x_i)^{\alpha_a}} dx_i \tag{45}$$

Now we can simplify as follows:

$$\int_{\mathbf{x}} q(\mathbf{x}) d\mathbf{x} = \prod_i \int_{x_i} \prod_a m_{a\to i}(x_i) dx_i \tag{46}$$

$$= \prod_i z_i \tag{47}$$

$$s_a = \left(\int_{\mathbf{x}} \frac{f_a(\mathbf{x})^{\alpha_a}}{\prod_i m_{a\to i}(x_i)^{\alpha_a}} \prod_i \bar{q}(x_i) d\mathbf{x} \right)^{1/\alpha_a} \tag{48}$$

$$= \left(\int_{\mathbf{x}} f_a(\mathbf{x})^{\alpha_a} \prod_i \frac{\bar{q}(x_i)}{m_{a\to i}(x_i)^{\alpha_a}} d\mathbf{x} \right)^{1/\alpha_a} \tag{49}$$

$$= \left(\int_{\mathbf{x}} f_a(\mathbf{x})^{\alpha_a} \prod_i \frac{z_i^{\setminus a}}{z_i} \bar{q}^{\setminus a}(x_i) d\mathbf{x} \right)^{1/\alpha_a} \tag{50}$$

Let $\mathbf{x}_a$ denote the set of variables used by f_a, and let $i \in a$ be shorthand for $x_i \in \mathbf{x}_a$. Then (50) simplifies to

$$s_a = \left(\prod_{i\in a} \frac{z_i^{\setminus a}}{z_i} \right)^{1/\alpha_a} \left(\int_{\mathbf{x}_a} f_a(\mathbf{x}_a)^{\alpha_a} \prod_{i\in a} \bar{q}^{\setminus a}(x_i) d\mathbf{x}_a \right)^{1/\alpha_a} \tag{51}$$

$$= \left(\frac{\int_{\mathbf{x}_a} f_a(\mathbf{x}_a)^{\alpha_a} \prod_{i\in a} \bar{q}^{\setminus a}(x_i) d\mathbf{x}_a}{\prod_{i\in a} \int_{x_i} m_{a\to i}(x_i)^{\alpha_a} \bar{q}^{\setminus a}(x_i) d\mathbf{x}_i} \right)^{1/\alpha_a} \tag{52}$$

A message in Power EP is exact if

$$m_{a\to j}(x_j)^{\alpha_a} = \int_{\mathbf{x}_a \setminus x_j} f_a(\mathbf{x}_a)^{\alpha_a} \prod_{i\in a, i\neq j} (\bar{q}^{\setminus a}(x_i) dx_i) \tag{53}$$

If $m_{a\to j}$ is exact for some j, then s_a simplifies as follows:

$$s_a = \left(\prod_{i\in a, i\neq j} \frac{z_i^{\setminus a}}{z_i} \right)^{1/\alpha_a} \tag{54}$$

6 Power EP on directed graphs

As with EP, we can redistribute denominator terms according to the edge directions. Previously, we computed $\tilde{Z}$ as follows:

$$\tilde{Z} = \prod_i z_i \prod_a s_a \tag{55}$$

$$= \prod_i \left(\int_{x_i} \prod_a m_{a\to i}(x_i) dx_i \right) \prod_a \left(\frac{\int_{\mathbf{x}_a} f_a(\mathbf{x}_a)^{\alpha_a} \prod_{i\in a} \bar{q}^{\setminus a}(x_i) d\mathbf{x}_a}{\prod_{i\in a} \int_{x_i} m_{a\to i}(x_i)^{\alpha_a} \bar{q}^{\setminus a}(x_i) d\mathbf{x}_i} \right)^{1/\alpha_a} \tag{56}$$

The denominator of this expression has one term for every edge in the factor graph. In the directed case, we will rearrange these denominator terms as follows:

$$z'_i = \frac{\int_{x_i} \prod_a m_{a\to i}(x_i) dx_i}{\prod_{a\in \mathrm{ch}(i)} \left(\int_{x_i} m_{a\to i}(x_i)^{\alpha_a} \bar{q}^{\setminus a}(x_i) d\mathbf{x}_i \right)^{1/\alpha_a}} \tag{57}$$

$$s'_a = \left(\frac{\int_{\mathbf{x}_a} f_a(\mathbf{x}_a)^{\alpha_a} \prod_{i\in a} \bar{q}^{\setminus a}(x_i) d\mathbf{x}_a}{\prod_{i\in \mathrm{ch}(a)} \int_{x_i} m_{a\to i}(x_i)^{\alpha_a} \bar{q}^{\setminus a}(x_i) d\mathbf{x}_i} \right)^{1/\alpha_a} \tag{58}$$

$$\tilde{Z} = \prod_i z'_i \prod_a s'_a \tag{59}$$

$$= \prod_i \frac{\int_{x_i} \prod_a m_{a\to i}(x_i) dx_i}{\prod_{a\in \mathrm{ch}(i)} \left(\int_{x_i} m_{a\to i}(x_i)^{\alpha_a} \bar{q}^{\setminus a}(x_i) d\mathbf{x}_i \right)^{1/\alpha_a}} \prod_a \left(\frac{\int_{\mathbf{x}_a} f_a(\mathbf{x}_a)^{\alpha_a} \prod_{i\in a} \bar{q}^{\setminus a}(x_i) d\mathbf{x}_a}{\prod_{i\in \mathrm{ch}(a)} \int_{x_i} m_{a\to i}(x_i)^{\alpha_a} \bar{q}^{\setminus a}(x_i) d\mathbf{x}_i} \right)^{1/\alpha_a} \tag{60}$$

Some special cases:

1. If a factor has no children (a leaf factor), then

$$s'_a = \left(\int_{\mathbf{x}_a} f_a(\mathbf{x}_a)^{\alpha_a} \prod_{i\in a} \bar{q}^{\setminus a}(x_i) d\mathbf{x}_a \right)^{1/\alpha_a} \tag{61}$$

2. If a factor has one child variable x_i, and $m_{a\to i}$ is exact, then $s'_a = 1$.
3. If a variable has one parent and one child factor, then z'_i does not simplify unless $\alpha_a = 1$.

7 Power plates

A power plate is a factor of the form

$$f_a(\mathbf{x}) = f_c(\mathbf{x})^n \tag{62}$$

By applying power EP with α_a, we reduce the problem to power EP on f_c with $\alpha_c = n\alpha_a$, with the following conversions:

$$\begin{aligned}
m_{a\to i}(x_i) &= m_{c\to i}(x_i)^n s_{ai} && (63)\\
s_{ai} &= \frac{\int_{x_i} m_{a\to i}(x_i) dx_i}{\int_{x_i} m_{c\to i}(x_i)^n dx_i} && (64)\\
\bar{q}^{\setminus c}(x_i) &= \bar{q}^{\setminus a}(x_i) && (65)\\
&\propto m_{a\to i}(x_i)^{1-\alpha_a} \prod_{b\neq a} m_{b\to i}(x_i) = m_{c\to i}(x_i)^{n-\alpha_c} \prod_{b\neq a} m_{b\to i}(x_i) && (66)\\
\frac{z_i}{z_i^{\setminus a}} &= s_{ai}^{\alpha_a} \int_{x_i} m_{c\to i}(x_i)^{\alpha_c} \bar{q}^{\setminus c}(x_i) = s_{ai}^{\alpha_a} \frac{z_i}{z_i^{\setminus c}} && (67)\\
s_a &= \left(\prod_{i\in a} \frac{z_i^{\setminus a}}{z_i} \right)^{1/\alpha_a} \left(\int_{\mathbf{x}_a} f_a(\mathbf{x}_a)^{\alpha_a} \prod_{i\in a} \bar{q}^{\setminus a}(x_i) d\mathbf{x}_a \right)^{1/\alpha_a} && (68)\\
&= \left(\prod_{i\in a} \frac{z_i^{\setminus c}}{z_i} \right)^{n/\alpha_c} \left(\prod_{i\in a} \frac{1}{s_{ai}} \right) \left(\int_{\mathbf{x}_c} f_c(\mathbf{x}_c)^{\alpha_c} \prod_{i\in c} \bar{q}^{\setminus c}(x_i) d\mathbf{x}_c \right)^{n/\alpha_c} && (69)\\
&= s_c^n \prod_{i\in a} \frac{1}{s_{ai}} = s_c^n \prod_{i\in a} \frac{\int_{x_i} m_{c\to i}(x_i)^n dx_i}{\int_{x_i} m_{a\to i}(x_i) dx_i} && (70)
\end{aligned}$$

In the directed case, this becomes

$$\begin{aligned}
s'_a &= \left(\frac{\int_{\mathbf{x}_a} f_a(\mathbf{x}_a)^{\alpha_a} \prod_{i\in a} \bar{q}^{\setminus a}(x_i) d\mathbf{x}_a}{\prod_{i\in \mathrm{ch}(a)} \int_{x_i} m_{a\to i}(x_i)^{\alpha_a} \bar{q}^{\setminus a}(x_i) d\mathbf{x}_i} \right)^{1/\alpha_a} && (71)\\
&= \left(\frac{\int_{\mathbf{x}_c} f_c(\mathbf{x}_a)^{\alpha_c} \prod_{i\in c} \bar{q}^{\setminus c}(x_i) d\mathbf{x}_c}{\prod_{i\in \mathrm{ch}(a)} s_{ai}^{\alpha_a} \int_{x_i} m_{c\to i}(x_i)^{\alpha_c} \bar{q}^{\setminus c}(x_i) d\mathbf{x}_i} \right)^{n/\alpha_c} && (72)\\
&= (s'_c)^n \prod_{i\in \mathrm{ch}(a)} \frac{1}{s_{ai}} = (s'_c)^n \prod_{i\in \mathrm{ch}(a)} \frac{\int_{x_i} m_{c\to i}(x_i)^n dx_i}{\int_{x_i} m_{a\to i}(x_i) dx_i} && (73)
\end{aligned}$$

8 Gates

A gate is a factor of the form

$$f_a(\mathbf{x}, y) = f_b(\mathbf{x})^{\delta(y=1)} f_c(\mathbf{x})^{\delta(y=0)} \tag{74}$$

We want to derive the messages $m_{a\to i}$, $m_{a\to y}$, and s_a in terms of f_b and f_c.

$$m_{a\to y}(y) \propto \left(\int_{\mathbf{x}} f_b(\mathbf{x})^{\alpha_a} \bar{q}^{\setminus a}(\mathbf{x}) d\mathbf{x} \right)^{\delta(y=1)/\alpha_a} \left(\int_{\mathbf{x}} f_c(\mathbf{x})^{\alpha_a} \bar{q}^{\setminus a}(\mathbf{x}) d\mathbf{x} \right)^{\delta(y=0)/\alpha_a} \tag{75}$$

Note that, regardless of the scaling of $m_{a\to i}$, we have the identity:

$$\frac{m_{a\to i}(x_i)^{\alpha_a} \bar{q}^{\setminus a}(x_i)}{\int_{x_i} m_{a\to i}(x_i)^{\alpha_a} \bar{q}^{\setminus a}(x_i) dx_i} = \frac{\mathrm{proj}\left[\int_{\mathbf{x}\setminus x_i} f_a(\mathbf{x})^{\alpha_a} \bar{q}^{\setminus a}(\mathbf{x}) d\mathbf{x} \right]}{\int_{\mathbf{x}} f_a(\mathbf{x})^{\alpha_a} \bar{q}^{\setminus a}(\mathbf{x}) d\mathbf{x}} \tag{76}$$

$$m_{a\to i}(x_i) \propto \left(\frac{\text{proj}\left[\bar{q}^{\setminus a}(y=1)\text{proj}\left[\int_{\mathbf{x}\setminus x_i} f_b(\mathbf{x})^{\alpha_a}\bar{q}^{\setminus a}(\mathbf{x})d\mathbf{x}\right] + \bar{q}^{\setminus a}(y=0)\text{proj}\left[\int_{\mathbf{x}\setminus x_i} f_c(\mathbf{x})^{\alpha_a}\bar{q}^{\setminus a}(\mathbf{x})d\mathbf{x}\right]\right]}{\bar{q}^{\setminus a}(x_i)} \right)^{1/\alpha_a} \tag{77}$$

$$= \left(\frac{\text{proj}\left[g_1 \frac{m_{b\to i}(x_i)^{\alpha_a}\bar{q}^{\setminus a}(x_i)}{\int_{x_i} m_{b\to i}(x_i)^{\alpha_a}\bar{q}^{\setminus a}(x_i)dx_i} + g_0 \frac{m_{c\to i}(x_i)^{\alpha_a}\bar{q}^{\setminus a}(x_i)}{\int_{x_i} m_{c\to i}(x_i)^{\alpha_a}\bar{q}^{\setminus a}(x_i)dx_i}\right]}{\bar{q}^{\setminus a}(x_i)} \right)^{1/\alpha_a} \tag{78}$$

$$\text{where } g_1 = \bar{q}^{\setminus a}(y=1)\int_{\mathbf{x}} f_b(\mathbf{x})^{\alpha_a}\bar{q}^{\setminus a}(\mathbf{x})d\mathbf{x} \tag{79}$$

$$g_0 = \bar{q}^{\setminus a}(y=0)\int_{\mathbf{x}} f_c(\mathbf{x})^{\alpha_a}\bar{q}^{\setminus a}(\mathbf{x})d\mathbf{x} \tag{80}$$

$$s_a = \left(\frac{\bar{q}^{\setminus a}(y=1)\int_{\mathbf{x}} f_b(\mathbf{x})^{\alpha_a}\bar{q}^{\setminus a}(\mathbf{x})d\mathbf{x} + \bar{q}^{\setminus a}(y=0)\int_{\mathbf{x}} f_c(\mathbf{x})^{\alpha_a}\bar{q}^{\setminus a}(\mathbf{x})d\mathbf{x}}{\left(\bar{q}^{\setminus a}(y=1)m_{a\to y}(y=1)^{\alpha_a} + \bar{q}^{\setminus a}(y=0)m_{a\to y}(y=0)^{\alpha_a}\right)\prod_{i\in a}\int_{x_i} m_{a\to i}(x_i)^{\alpha_a}\bar{q}^{\setminus a}(x_i)dx_i} \right)^{1/\alpha_a} \tag{81}$$

If we define s_b and s_c as follows, we can express the above formulas very simply:

$$s_b = \left(\frac{\int_{\mathbf{x}} f_b(\mathbf{x})^{\alpha_a}\bar{q}^{\setminus a}(\mathbf{x})d\mathbf{x}}{\prod_{i\in a}\int_{x_i} m_{a\to i}(x_i)^{\alpha_a}\bar{q}^{\setminus a}(x_i)dx_i} \right)^{1/\alpha_a} \tag{82}$$

$$s_c = \left(\frac{\int_{\mathbf{x}} f_c(\mathbf{x})^{\alpha_a}\bar{q}^{\setminus a}(\mathbf{x})d\mathbf{x}}{\prod_{i\in a}\int_{x_i} m_{a\to i}(x_i)^{\alpha_a}\bar{q}^{\setminus a}(x_i)dx_i} \right)^{1/\alpha_a} \tag{83}$$

$$m_{a\to y}(y) = \frac{s_b^{\delta(y=1)} s_c^{\delta(y=0)}}{s_b + s_c} \tag{84}$$

$$g_1 = \bar{q}^{\setminus a}(y=1)s_b^{\alpha_a} \tag{85}$$

$$g_0 = \bar{q}^{\setminus a}(y=0)s_c^{\alpha_a} \tag{86}$$

$$s_a = s_b + s_c \tag{87}$$

On directed graphs, this becomes:

$$s'_b = \left(\frac{\int_{\mathbf{x}} f_b(\mathbf{x})^{\alpha_a}\bar{q}^{\setminus a}(\mathbf{x})d\mathbf{x}}{\prod_{i\in \text{ch}(a)}\int_{x_i} m_{a\to i}(x_i)^{\alpha_a}\bar{q}^{\setminus a}(x_i)dx_i} \right)^{1/\alpha_a} \tag{88}$$

$$s'_c = \left(\frac{\int_{\mathbf{x}} f_c(\mathbf{x})^{\alpha_a}\bar{q}^{\setminus a}(\mathbf{x})d\mathbf{x}}{\prod_{i\in \text{ch}(a)}\int_{x_i} m_{a\to i}(x_i)^{\alpha_a}\bar{q}^{\setminus a}(x_i)dx_i} \right)^{1/\alpha_a} \tag{89}$$

$$s'_a = \bar{q}^{\setminus a}(y=1)s'_b + \bar{q}^{\setminus a}(y=0)s'_c \tag{90}$$

9 Variational Message Passing

Given any set of messages, properly scaled or not, the optimal normalizing constant for Variational Message Passing is (Minka, 2005):

$$\tilde{Z} = \left(\int_{\mathbf{x}} q(\mathbf{x}) d\mathbf{x} \right) \prod_{a=1}^{A} s_a \tag{91}$$

$$\text{where } s_a = \exp \left(\frac{\displaystyle\int_{\mathbf{x}} q(\mathbf{x}) \log \frac{f_a(\mathbf{x})}{\prod_i m_{a \to i}(x_i)} d\mathbf{x}}{\displaystyle\int_{\mathbf{x}} q(\mathbf{x}) d\mathbf{x}} \right) \tag{92}$$

To simplify this formula, let's define the following notation:

$$z_i = \int_{x_i} \prod_a m_{a \to i}(x_i) dx_i \tag{93}$$

$$\bar{q}(x_i) = \frac{1}{z_i} \prod_a m_{a \to i}(x_i) \tag{94}$$

Now we can simplify as follows:

$$\int_{\mathbf{x}} q(\mathbf{x}) d\mathbf{x} = \prod_i \int_{x_i} \prod_a m_{a \to i}(x_i) dx_i \tag{95}$$

$$= \prod_i z_i \tag{96}$$

$$s_a = \exp \left(\int_{\mathbf{x}} \left(\prod_i \bar{q}(x_i) \right) \log \frac{f_a(\mathbf{x})}{\prod_i m_{a \to i}(x_i)} d\mathbf{x} \right) \tag{97}$$

$$= \exp \left(\int_{\mathbf{x}_a} \left(\prod_{i \in a} \bar{q}(x_i) \right) \log \frac{f_a(\mathbf{x}_a)}{\prod_{i \in a} m_{a \to i}(x_i)} d\mathbf{x}_a \right) \tag{98}$$

Alternatively, (91) is equivalent to:

$$\tilde{Z} = \exp \left(\bar{q}(\mathbf{x}) \log \prod_a f_a(\mathbf{x}) - \bar{q}(\mathbf{x}) \log \bar{q}(\mathbf{x}) \right) \tag{99}$$

Therefore we can divide the work as follows: each factor computes $s'_a = \exp(\bar{q}(\mathbf{x}) \log \prod_a f_a(\mathbf{x}))$ and each variable computes $\exp(-\bar{q}(x_i) \log \bar{q}(x_i))$. Deterministic factors and their output variables send nothing.

9.1 Gates under VMP

$$m_{a \to i}(x_i) \propto \exp\left(\bar{q}(y=1) \log m_{b \to i}(x_i) + \bar{q}(y=0) \log m_{c \to i}(x_i)\right) \tag{100}$$

$$s'_b = \exp \left(\int_{\mathbf{x}} \bar{q}(\mathbf{x}) \log f_b(\mathbf{x}) \right) \tag{101}$$

$$s'_c = \exp \left(\int_{\mathbf{x}} \bar{q}(\mathbf{x}) \log f_c(\mathbf{x}) \right) \tag{102}$$

$$m_{a \to y}(y) = \frac{(s'_b)^{\delta(y=1)} (s'_c)^{\delta(y=0)}}{s'_b + s'_c} \tag{103}$$

$$s'_a = (s'_b)^{\bar{q}(y=1)} (s'_c)^{\bar{q}(y=0)} \tag{104}$$

10 Summary

Let every factor implement functions Approximate and Integrate with the following definitions:

$$\text{Approximate}(f_a, \bar{q}^{\backslash a}, \alpha_a) \propto \left(\frac{\text{proj}\left[f_a(\mathbf{x}_a)^{\alpha_a}\bar{q}^{\backslash a}(\mathbf{x}_a)\right]}{\bar{q}^{\backslash a}(\mathbf{x}_a)}\right)^{1/\alpha_a} \tag{105}$$

$$= \prod_i \left(\frac{\text{proj}\left[\bar{q}^{\backslash a}(x_i)\int_{\mathbf{x}_a\backslash x_i} f_a(\mathbf{x}_a)^{\alpha_a}\prod_{j\neq i}(\bar{q}^{\backslash a}(x_j)dx_j)\right]}{\bar{q}^{\backslash a}(x_i)}\right)^{1/\alpha_a} \tag{106}$$

$$\text{Integrate}(f_a, \bar{q}^{\backslash a}, \tilde{f}_a, \alpha_a) = \left(\frac{\int_{\mathbf{x}_a} f_a(\mathbf{x}_a)^{\alpha_a}\bar{q}^{\backslash a}(\mathbf{x}_a)d\mathbf{x}_a}{\int_{\mathbf{x}_a} \tilde{f}_a(\mathbf{x}_a)^{\alpha_a}\bar{q}^{\backslash a}(\mathbf{x}_a)d\mathbf{x}_a}\right)^{1/\alpha_a} = s_a \tag{107}$$

Note both routines are invariant to rescaling $\bar{q}^{\backslash a}$. For $\alpha_a = 0$:

$$\text{Approximate}(f_a, \bar{q}^{\backslash a}, \alpha_a) \propto \prod_i \exp\left(\int_{\mathbf{x}_a\backslash x_i} \log(f_a(\mathbf{x}_a))\prod_{j\neq i}(\bar{q}(x_j)dx_j)\right) \tag{108}$$

$$\text{Integrate}(f_a, \bar{q}^{\backslash a}, \tilde{f}_a, \alpha_a) = \exp\left(\int_{\mathbf{x}_a} \bar{q}(\mathbf{x}_a)\log\frac{f_a(\mathbf{x}_a)}{\tilde{f}_a(\mathbf{x}_a)}d\mathbf{x}_a\right) \tag{109}$$

For a power plate, we can define these functions recursively:

$$\text{Approximate}(f_a, \bar{q}^{\backslash a}, \alpha_a) \propto \text{Approximate}(f_c, \bar{q}^{\backslash a}, n\alpha_a)^n \tag{110}$$

$$\text{Integrate}(f_a, \bar{q}^{\backslash a}, \tilde{f}_a, \alpha_a) = \text{Integrate}(f_c, \bar{q}^{\backslash a}, \tilde{f}_c, n\alpha_a)^n\frac{\int_{\mathbf{x}_c}\tilde{f}_c(\mathbf{x}_c)^n d\mathbf{x}_c}{\int_{\mathbf{x}_a}\tilde{f}_a(\mathbf{x}_a)d\mathbf{x}_a} \tag{111}$$

$$\text{where } \tilde{f}_c(\mathbf{x}_c) \propto \tilde{f}_a(\mathbf{x}_a)^{1/n} \tag{112}$$

For a gate:

$$\text{Approximate}(f_a, \bar{q}^{\backslash a}, \alpha_a) \propto \text{Merge}(g_1, \text{Approximate}(f_b, \bar{q}^{\backslash a}(\mathbf{x}_b), \alpha_a), g_0, \text{Approximate}(f_c, \bar{q}^{\backslash a}(\mathbf{x}_c), \alpha_a), \alpha_a) \tag{113}$$

$$\times \frac{s_b^{\delta(y=1)}s_c^{\delta(y=0)}}{s_b + s_c} \tag{114}$$

$$\text{Integrate}(f_a, \bar{q}^{\backslash a}, \tilde{f}_a, \alpha_a) = s_b + s_c \tag{115}$$

$$\text{where } g_1 = \bar{q}^{\backslash a}(y=1)s_b^{\alpha_a} \tag{116}$$

$$g_0 = \bar{q}^{\backslash a}(y=0)s_c^{\alpha_a} \tag{117}$$

$$s_b = \text{Integrate}(f_b, \bar{q}^{\backslash a}(\mathbf{x}_b), \tilde{f}_a(\mathbf{x}_b), \alpha_a) \tag{118}$$

$$s_c = \text{Integrate}(f_c, \bar{q}^{\backslash a}(\mathbf{x}_c), \tilde{f}_a(\mathbf{x}_c), \alpha_a) \tag{119}$$

On directed graphs, we can instead use IntegrateD defined by:

$$\text{IntegrateD}(f_a, \bar{q}^{\backslash a}, \tilde{f}_a, \alpha_a) = s'_a = \left(\frac{\int_{\mathbf{x}_a} f_a(\mathbf{x}_a)^{\alpha_a}\prod_{i\in a}\bar{q}^{\backslash a}(x_i)d\mathbf{x}_a}{\prod_{\text{out } i}\int_{x_i} m_{a\to i}(x_i)^{\alpha_a}\bar{q}^{\backslash a}(x_i)d\mathbf{x}_i}\right)^{1/\alpha_a} \times \prod_{\text{out } i}\int_{x_i} m_{a\to i}(x_i)d\mathbf{x}_i \tag{120}$$

Note IntegrateD is invariant to rescaling $\bar{q}^{\backslash a}$ and $m_{a\to i}$ for outgoing edges but it requires $\bar{q}^{\backslash a}$ to be normalized for incoming edges.

11 Gate implementation in Infer.NET

11.1 EP directed

Replicate and UsesEqualDef both use the same formula for evidence contribution. Therefore no special handling is needed for deterministic factors.

The case variables collect IntegrateD$(f_b, \bar{q}^{\setminus a}(\mathbf{x}_b), \tilde{f}_b(\mathbf{x}_b), \alpha_a = 1)$ from the child factors. Nothing is collected from the Enter/Exit variables. The Exit operator contributes $\int_{\mathbf{x}} \tilde{f}_b(\mathbf{x}_b)\bar{q}^{\setminus a}(\mathbf{x}_b)d\mathbf{x}$ which cancels the denominator of IntegrateD. The missing factor in s_b is provided as an evidence message from Exit. Enter does not send an evidence message.

11.2 VMP

The case variables collect s'_b from the child factors. The Enter variables are marked deterministic and send nothing to the case variable. The Exit variables send $\exp(-\bar{q}(x_i)\log\bar{q}(x_i))$ to the case variable. These messages must be cancelled by the Exit operator.

12 Mixing EP and VMP

VMP can be mixed with EP by viewing it as Power EP with $\alpha = 1$ for EP factors and $\alpha = 0$ for VMP factors. However the behavior of variables and the mechanisms for collecting evidence are different for the two algorithms (as summarized in the Gates paper), so we need to make conversions. Note that the rules obey a consistency property, that if you split a variable into an EP copy and VMP copy then the results are unchanged.

Let $m^{e/v}_{a\to i}(x_i)$ be the ordinary EP/VMP message from factor f_a to variable x_i, computed from modified incoming messages. Let $m^{e/v}_{i\to a}(x_i)$ be the ordinary EP/VMP message from variable x_i to factor f_a. The modified message from variable to factor will be $m^{e\to v}_{i\to a}(x_i)$ or $m^{v\to e}_{i\to a}(x_i)$. The factor sends its usual evidence message, but the evidence will be multiplied by a factor s_{ia}.

Suppose x_i is an EP variable and f_a is a stochastic VMP factor. The modified message is:

$$m^{e\to v}_{i\to a}(x_i) = m^{e}_{i\to a}(x_i) m^{v}_{a\to i}(x_i) \tag{121}$$

$$s_{ia} = \exp\left(-\sum_{x_i} m^{e\to v}_{i\to a}(x_i) \log m^{v}_{a\to i}(x_i)\right) \tag{122}$$

Notice that if an EP variable is connected only to stochastic VMP factors, then it behaves equivalently (in messages and evidence contribution) to a VMP variable.

If x_i is an EP variable which parents a deterministic VMP factor, then the modified message is the same as above. An EP variable cannot be the child of a deterministic VMP factor. It must be treated as a derived VMP variable.

Suppose x_i is a stochastic VMP variable and f_a is an EP factor. The modified message is:

$$m^{v\to e}_{i\to a}(x_i) = m^{v}_{i\to a}(x_i) / m^{e}_{a\to i}(x_i) \tag{123}$$

$$s_{ia} = \exp\left(\sum_{x_i} m^{v}_{i\to a}(x_i) \log m^{e}_{a\to i}(x_i)\right) \tag{124}$$

Thus if a VMP variable is connected only to EP factors, it behaves equivalently (in messages and evidence contribution) to an EP variable.

If x_i is a derived VMP variable (a child of a deterministic VMP factor) and f_a is an EP factor, then the modified message is the same as above, using the definition of $m^{v}_{i\to a}(x_i)$ for derived VMP variables.

References

Minka, T. (2005). Divergence measures and message passing (Technical Report MSR-TR-2005-173). Microsoft Research Ltd.