

Application of the Information Bottleneck to LPAD Learning

Fabrizio Riguzzi¹ and Nicola Di Mauro²

¹ Dipartimento di Ingegneria - Università di Ferrara - Italy

² Dipartimento di Informatica - Università degli Studi di Bari "Aldo Moro" - Italy

1 Application of the Information Bottleneck to LPAD Learning

In order to apply the Information Bottleneck (IB) to LPADs, the network \mathcal{G}_{out} is the result of the translation of the LPAD for which we want to learn the parameters plus the addition of the Y variable. The set of hidden variables contains the vector of the choice variables \mathbf{CH} plus those atoms that are unobserved in the data, let us call them \mathbf{T} . With \mathbf{X} we indicate the set of atom variables that are observed in the data.

Suppose you want to learn the parameters of an LPAD using IB. Consider the LPAD L :

$$r_1 = x_1 : 0.4 \vee x_2 : 0.3.$$

$$r_2 = x_2 : 0.1 \vee x_3 : 0.2.$$

$$r_3 = x_4 : 0.6 \vee x_5 : 0.4 \leftarrow x_1.$$

$$r_4 = x_5 : 0.4 \leftarrow x_2, x_3.$$

$$r_5 = x_6 : 0.3 \vee x_7 : 0.2 \leftarrow x_2, x_5.$$

The Bayesian network equivalent to L is shown in Figure 1. For the \mathcal{G}_{in} network, we consider a naive

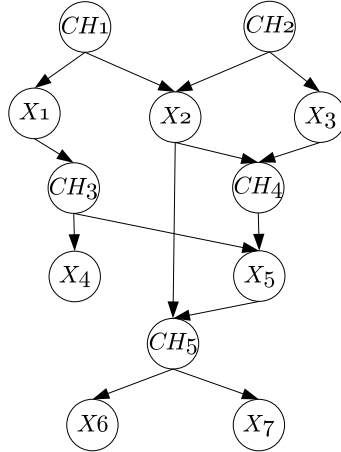


Fig. 1. Bayesian network

Bayes factorization:

$$Q(\mathbf{CH}, \mathbf{T}|Y) = \prod_i Q(CH_i|Y) \prod_j Q(T_j|Y).$$

Consider the LPAD L . Moreover, suppose that x_5 is unseen in the data. The networks \mathcal{G}_{in} and \mathcal{G}_{out} for this LPAD are shown in Figure 2. According to IB, the chosen Q distribution must be such that unobserved variables are independent of observed ones given Y . This requirement is satisfied by \mathcal{G}_{in} in Figure 2. As regards P , \mathbf{CH} and \mathbf{T} must be the only parents of Y . This requirement is also

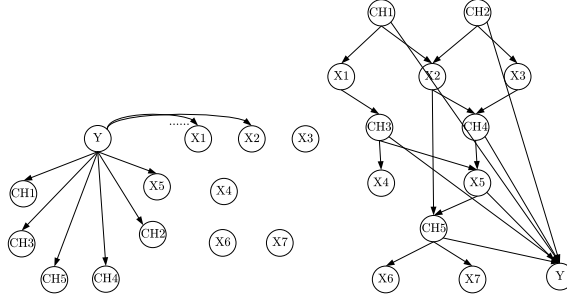


Fig. 2. $G_{in} = Q$ (left) and $G_{out} = P$ (right)

satisfied by \mathcal{G}_{out} : in fact, the observed variables are completely determined by knowing \mathbf{CH} and \mathbf{T} and so it the instance identity (assuming there are no duplicate examples, but this can be modeled by assigning them a different prior probability $Q(Y)$).

Note that for the network to be well defined the LPAD must be acyclic.

A ground normal logic program is *acyclic* [1] if the ground atoms can be assigned an integer level so that the level of the atom in the head of each rule is higher than the level of each atom in the body.

A disjunctive logic program is *acyclic* if the ground atoms can be assigned an integer level so that the level of each the atom in the head of each rule is the same and it is higher than the level of each atom in the body.

Let us compute the objective function \mathcal{L}_{EM} for such a case

$$\mathcal{L}_{EM} = \mathbf{I}_Q(\mathbf{CH}, \mathbf{T}; Y) - \gamma(\mathbb{E}_Q[\log P(\mathbf{X}, \mathbf{T}, \mathbf{CH})] - \mathbb{E}_Q[\log Q(\mathbf{CH}, \mathbf{T})]) \quad (1)$$

where

$$\begin{aligned} \mathbf{I}_Q(\mathbf{CH}, \mathbf{T}; Y) &= \frac{H(\mathbf{CH}, \mathbf{T}) - H(\mathbf{CH}, \mathbf{T}|Y)}{\log_2 e} = \\ &= -\mathbb{E}_Q[\log Q(\mathbf{CH}, \mathbf{T})] + \mathbb{E}_Q[\log Q(\mathbf{CH}, \mathbf{T}|Y)] = \\ &= -\mathbb{E}_Q[\log Q(\mathbf{CH}, \mathbf{T})] + \sum_i \mathbb{E}_Q[\log Q(CH_i|Y)] + \sum_i \mathbb{E}_Q[\log Q(T_i|Y)] \end{aligned}$$

and

$$\mathbb{E}_Q[\log Q(\mathbf{CH}, \mathbf{T})] \approx \sum_i \mathbb{E}_Q[\log Q(CH_i)] + \sum_i \mathbb{E}_Q[\log Q(T_i)]$$

following [2]. Thus

$$\begin{aligned} \mathcal{L}_{EM} &= \mathbb{E}_Q[\log Q(\mathbf{CH}, \mathbf{T}|Y)] - \gamma\mathbb{E}_Q[\log P(\mathbf{X}, \mathbf{T}, \mathbf{CH})] + (\gamma - 1)\mathbb{E}_Q[\log Q(\mathbf{CH}, \mathbf{T})] = \\ &= \sum_i \mathbb{E}_Q[\log Q(CH_i|Y)] + \sum_i \mathbb{E}_Q[\log Q(T_i|Y)] - \gamma\mathbb{E}_Q[\log P(\mathbf{X}, \mathbf{T}, \mathbf{CH})] + \\ &= (\gamma - 1) \sum_i \mathbb{E}_Q[\log Q(ch_i)] + (\gamma - 1) \sum_i \mathbb{E}_Q[\log Q(t_i)] \end{aligned}$$

2 Stationary Points of \mathcal{L}_{EM}

In this section we describe how to find the stationary points of \mathcal{L}_{EM} .

Proposition 1 (Stationary points of \mathcal{L}_{EM}). Let \mathcal{L}_{EM} be the function (1). $Q(ch_i|y)$ and $Q(t_i|y)$ are stationary points of \mathcal{L}_{EM} with respect to a fixed choice of P if and only if for all values ch_i, t_i and y :

$$Q(ch_i|y) = \frac{1}{Z_{\mathbf{CH}}(i, y, \gamma)} Q(ch_i)^{1-\gamma} e^{\gamma \mathbb{E}\mathbb{P}(ch_i, y)},$$

$$Q(t_i|y) = \frac{1}{Z_{\mathbf{T}}(i, y, \gamma)} Q(t_i)^{1-\gamma} e^{\gamma \mathbb{E}\mathbb{P}(t_i, y)}$$

where

$$Z_{\mathbf{CH}}(i, y, \gamma) = \sum_{ch'_i} Q(ch'_i)^{1-\gamma} e^{\gamma \mathbb{E}\mathbb{P}(ch'_i, y)},$$

$$Z_{\mathbf{T}}(i, y, \gamma) = \sum_{t'_i} Q(t'_i)^{1-\gamma} e^{\gamma \mathbb{E}\mathbb{P}(t'_i, y)}$$

are normalizing constants, and

$$\mathbb{E}\mathbb{P}(ch_i, y) = \mathbb{E}_{Q(\mathbf{CH}, \mathbf{T}|ch_i, y)}[\log P(\mathbf{x}[y], \mathbf{T}, \mathbf{CH})],$$

$$\mathbb{E}\mathbb{P}(t_i, y) = \mathbb{E}_{Q(\mathbf{CH}, \mathbf{T}|t_i, y)}[\log P(\mathbf{x}[y], \mathbf{T}, \mathbf{CH})].$$

Proof. We want to find the stationary points of $Q(ch_i|y)$ and $Q(t_i|y)$ under the constraint that

$$\sum_{ch_i} Q(ch_i|y) = 1$$

$$\sum_{t_i} Q(t_i|y) = 1$$

for any y .

Using Langrange multipliers, we want to optimize:

$$\mathcal{L} = \mathbb{E}_Q[\log Q(\mathbf{CH}, \mathbf{T}|Y)] - \gamma(\mathbb{E}_Q[\log P(\mathbf{X}, \mathbf{T}, \mathbf{CH})] + \mathbb{E}_Q[\log Q(\mathbf{CH}, \mathbf{T})])(\gamma - 1) +$$

$$+ \sum_i \sum_y \lambda_{i,y}^{\mathbf{CH}} \left(\sum_{ch'_i} Q(ch'_i|y) - 1 \right) + \sum_i \sum_y \lambda_{i,y}^{\mathbf{T}} \left(\sum_{t'_i} Q(t'_i|y) - 1 \right).$$

Computing the derivative of \mathcal{L} with respect to $Q(ch_{i0}|y_0)$ and $Q(t_{i0}|y_0)$ we get

$$\frac{\partial \mathbb{E}_Q[\log Q(\mathbf{CH}_k|Y)]}{\partial Q(ch_{i0}|y_0)} = \frac{\partial \sum_{ch_k, y} Q(ch_k, y) \log Q(ch_k|y)}{\partial Q(ch_{i0}|y_0)} =$$

$$\left(Q(y_0) \log Q(ch_k|y_0) + \frac{Q(ch_k, y_0)}{Q(ch_k|y_0)} \right) 1_{\{i0 = k, ch_{i0} = ch_k\}} =$$

$$Q(y_0)(\log Q(ch_k|y_0) + 1) 1_{\{i0 = k, ch_{i0} = ch_k\}}$$

$$\frac{\partial \mathbb{E}_Q[\log Q(\mathbf{T}_k|Y)]}{\partial Q(t_{i0}|y_0)} = \frac{\partial \sum_{t_k, y} Q(t_k, y) \log Q(t_k|y)}{\partial Q(t_{i0}|y_0)} =$$

$$\left(Q(y_0) \log Q(t_k|y_0) + \frac{Q(t_k, y_0)}{Q(t_k|y_0)} \right) 1_{\{i0 = k, t_{i0} = t_k\}} =$$

$$Q(y_0)(\log Q(t_k|y_0) + 1) 1_{\{i0 = k, t_{i0} = t_k\}}$$

$$\begin{aligned}
\frac{\partial \mathbb{E}_Q[\log P(\mathbf{X}, \mathbf{T}, \mathbf{CH})]}{\partial Q(ch_{i0}|y_0)} &= \frac{\partial \sum_{\mathbf{ch}, \mathbf{t}, y, \mathbf{x}} Q(\mathbf{ch}, \mathbf{t}, \mathbf{x}, y) \log P(\mathbf{x}, \mathbf{t}, \mathbf{ch})}{\partial Q(ch_{i0}|y_0)} = \\
&= \sum_{\mathbf{ch}, \mathbf{t}, y, \mathbf{x}} \frac{\partial Q(\mathbf{ch}, \mathbf{t}, \mathbf{x}, y)}{\partial Q(ch_{i0}|y_0)} \log P(\mathbf{x}, \mathbf{t}, \mathbf{ch}) = \\
&= \sum_{\mathbf{ch}, \mathbf{t}, y, \mathbf{x}} \frac{\partial Q(\mathbf{ch}, \mathbf{t}, \mathbf{x}|ch_i, y) Q(ch_i|y) Q(y)}{\partial Q(ch_{i0}|y_0)} \log P(\mathbf{x}, \mathbf{t}, \mathbf{ch}) = \\
&= \sum_{\mathbf{ch} \setminus \{ch_i\}, \mathbf{t}, \mathbf{x}} Q(\mathbf{ch}, \mathbf{t}, \mathbf{x}|ch_{i0}, y_0) Q(y_0) \log P(\mathbf{x}[y_0], \mathbf{t}, \mathbf{ch}) = \\
&= Q(y_0) \mathbb{E}_{Q(\mathbf{CH}, \mathbf{T}|ch_{i0}, y_0)}[\log P(\mathbf{x}[y_0], \mathbf{T}, \mathbf{CH})]
\end{aligned}$$

$$\frac{\partial \mathbb{E}_Q[\log P(\mathbf{X}, \mathbf{T}, \mathbf{CH})]}{\partial Q(t_{i0}|y_0)} = Q(y_0) E_{Q(\mathbf{CH}, \mathbf{T}|t_{i0}, y_0)}[\log P(\mathbf{x}[y_0], \mathbf{T}, \mathbf{CH})]$$

$$\begin{aligned}
\frac{\partial \mathbb{E}_Q[\log Q(\mathbf{CH}, \mathbf{T})]}{\partial Q(ch_{i0}|y_0)} &= \\
&= \sum_k \sum_{ch_k} \frac{\partial Q(ch_k) \log Q(ch_k)}{\partial Q(ch_{i0}|y_0)} + \sum_k \sum_{t_k} \frac{\partial Q(t_k) \log Q(t_k)}{\partial Q(ch_{i0}|y_0)} = \\
&= \sum_k \sum_{ch_k} \frac{\partial Q(ch_k)}{\partial Q(ch_{i0}|y_0)} \log Q(ch_k) + \sum_{ch_k} Q(ch_k) \frac{\partial \log Q(ch_k)}{\partial Q(ch_{i0}|y_0)} + \\
&= \sum_k \sum_{t_k} \frac{\partial Q(t_k)}{\partial Q(ch_{i0}|y_0)} \log Q(t_k) + \sum_{t_k} Q(t_k) \frac{\partial \log Q(t_k)}{\partial Q(ch_{i0}|y_0)} = \\
&= \sum_k \sum_{ch_k} \frac{\partial \sum_y Q(ch_k|y) Q(y)}{\partial Q(ch_{i0}|y_0)} \log Q(ch_k) + \sum_{ch_k} Q(ch_k) \frac{\partial \log Q(ch_k)}{\partial Q(ch_{i0}|y_0)} + \\
&= \sum_k \sum_{t_k} \frac{\partial \sum_y Q(t_k|y) Q(y)}{\partial Q(ch_{i0}|y_0)} \log Q(t_k) + \sum_{t_k} Q(t_k) \frac{\partial \log Q(t_k)}{\partial Q(ch_{i0}|y_0)} = \\
&= \sum_k \sum_{ch_k} \frac{\partial \sum_y Q(ch_k|y) Q(y)}{\partial Q(ch_{i0}|y_0)} \log Q(ch_k) + \sum_{ch_k} Q(ch_k) \frac{1}{Q(ch_k)} \frac{\partial \sum_y Q(ch_k|y) Q(y)}{\partial Q(ch_{i0}|y_0)} + \\
&= \sum_k \sum_{t_k} \frac{\partial \sum_y Q(t_k|y) Q(y)}{\partial Q(ch_{i0}|y_0)} \log Q(t_k) + \sum_{t_k} Q(t_k) \frac{1}{Q(t_k)} \frac{\partial \sum_y Q(t_k|y) Q(y)}{\partial Q(ch_{i0}|y_0)} = \\
&= Q(y_0) \log Q(ch_{i0}) + Q(ch_{i0}) \frac{Q(y_0)}{Q(ch_{i0})} = Q(y_0) (\log Q(ch_{i0}) + 1)
\end{aligned}$$

$$\frac{\partial \mathbb{E}_Q[\log Q(\mathbf{CH}, \mathbf{T})]}{\partial Q(t_{i0}|y_0)} = Q(y_0) (\log Q(t_{i0}) + 1)$$

Therefore

$$\begin{aligned}
\frac{\partial L}{\partial Q(ch_{i0}|y_0)} &= Q(y_0) (\log Q(ch_{i0}|y_0) + 1) - \gamma Q(y_0) \mathbb{E}_{Q(\mathbf{CH}, \mathbf{T}|ch_{i0}, y_0)}[\log P(\mathbf{x}[y_0], \mathbf{T}, \mathbf{CH})] + \\
&= (\gamma - 1) Q(y_0) (\log Q(ch_{i0}) + 1) + \lambda_{i0, y_0}^{\mathbf{CH}} = \\
&= Q(y_0) (\log Q(ch_{i0}|y_0) + 1) - \gamma \mathbb{E}_{Q(\mathbf{CH}, \mathbf{T}|ch_{i0}, y_0)}[\log P(\mathbf{x}[y_0], \mathbf{T}, \mathbf{CH})] +
\end{aligned}$$

$$\begin{aligned}
& (\gamma - 1)(\log Q(ch_{i0}) + 1) + \lambda_{i0,y_0}^{\mathbf{CH}} = \\
& Q(y_0)(\log Q(ch_{i0}|y_0) + 1 - \gamma \mathbb{E}_{Q(\mathbf{CH}, \mathbf{T}|ch_{i0}, y_0)}[\log P(\mathbf{x}[y_0], \mathbf{T}, \mathbf{CH})]) + \\
& \gamma(\log Q(ch_{i0}) + 1) - \log Q(ch_{i0}) - 1) + \lambda_{i0,y_0}^{\mathbf{CH}} = \\
& Q(y_0)(\log Q(ch_{i0}|y_0) - \gamma \mathbb{E}_{Q(\mathbf{CH}, \mathbf{T}|ch_{i0}, y_0)}[\log P(\mathbf{x}[y_0], \mathbf{T}, \mathbf{CH})]) + \\
& \gamma \log Q(ch_{i0}) + \gamma - \log Q(ch_{i0})) + \lambda_{i0,y_0}^{\mathbf{CH}}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial L}{\partial Q(t_{i0}|y_0)} &= Q(y_0)(\log Q(t_{i0}|y_0) - \gamma \mathbb{E}_{Q(\mathbf{CH}, \mathbf{T}|t_{i0}, y_0)}[\log P(\mathbf{x}[y_0], \mathbf{T}, \mathbf{CH})]) + \\
& \gamma \log Q(t_{i0}) + \gamma - \log Q(t_{i0})) + \lambda_{i0,y_0}^{\mathbf{T}}
\end{aligned}$$

Dividing by $Q(y_0)$ and equating to 0

$$\begin{aligned}
Q(ch_{i0}|y_0) &= e^{-\gamma - \lambda_{i0,y_0}^{\mathbf{CH}}/Q(y_0)} Q(ch_{i0})^{1-\gamma} e^{\gamma \mathbb{EP}(ch_{i0}, y_0)} \\
Q(t_{i0}|y_0) &= e^{-\gamma - \lambda_{i0,y_0}^{\mathbf{T}}/Q(y_0)} Q(t_{i0})^{1-\gamma} e^{\gamma \mathbb{EP}(t_{i0}, y_0)}
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{EP}(ch_i, y) &= \mathbb{E}_{Q(\mathbf{CH}, \mathbf{T}|ch_i, y)}[\log P(\mathbf{x}[y], \mathbf{T}, \mathbf{CH})] \\
\mathbb{EP}(t_i, y) &= \mathbb{E}_{Q(\mathbf{CH}, \mathbf{T}|t_i, y)}[\log P(\mathbf{x}[y], \mathbf{T}, \mathbf{CH})]
\end{aligned}$$

Normalizing we get

$$\begin{aligned}
Q(ch_i|y) &= \frac{1}{Z_{\mathbf{CH}}(i, y, \gamma)} Q(ch_i)^{1-\gamma} e^{\gamma \mathbb{EP}(ch_i, y)} \\
Q(t_i|y) &= \frac{1}{Z_{\mathbf{T}}(i, y, \gamma)} Q(t_i)^{1-\gamma} e^{\gamma \mathbb{EP}(t_i, y)}
\end{aligned}$$

where

$$\begin{aligned}
Z_{\mathbf{CH}}(i, y, \gamma) &= \sum_{ch'_i} Q(ch'_i)^{1-\gamma} e^{\gamma \mathbb{EP}(ch'_i, y)} \\
Z_{\mathbf{T}}(i, y, \gamma) &= \sum_{t'_i} Q(t'_i)^{1-\gamma} e^{\gamma \mathbb{EP}(t'_i, y)}
\end{aligned}$$

□

3 Derivatives of $G_{ch_i, y}(Q, \gamma)$ and $G_{t_j, y}(Q, \gamma)$

Let us now compute the update direction. The G functions are

$$\begin{aligned}
G_{ch_k, y}(Q, \gamma) &= -\log Q(ch_k|y) + (1 - \gamma) \log Q(ch_k) + \gamma \mathbb{EP}(ch_k, y) - \log Z_{\mathbf{CH}}(k, y, \gamma) \\
G_{t_k, y}(Q, \gamma) &= -\log Q(t_k|y) + (1 - \gamma) \log Q(t_k) + \gamma \mathbb{EP}(t_k, y) - \log Z_{\mathbf{T}}(k, y, \gamma)
\end{aligned}$$

We apply a computation similar to the one in appendix B in [2].

Let us first express the parameters of P : $\theta_{x_j|pa_{x_j}} = 1$ if $x_j \in \text{val}(\mathbf{pa}_{x_j})$ and 0 otherwise, where $\text{val}(\mathbf{pa}_{x_j})$ is the set of values assumed by \mathbf{pa}_{x_j} . For a non ground rule r , let $\theta_{hd_r=hd_r|body_r}$ or simply $\theta_{hd_r|body_r}$ be the probability that the head hd_r is selected given that the body has truth value $body_r$. Let $i(r)$ be the set of instances of r and, given a ground rule k , let $r(k)$ be the non ground rule of which k is an instance. Given the body \mathbf{pa}_{ch_k} of instantiated rule k , let $bt(\mathbf{pa}_{ch_k})$ be 1 if the observed

variables in \mathbf{pa}_{ch_k} do not make the body false and 0 otherwise. Let $tb(\mathbf{pa}_{ch_k})$ be a set of values for the unobserved variables that are parents of ch_k . The values are those that do not make the body false.

The values of $\theta_{hd_r|body_r}$ are

$$\theta_{HD_r=none|false} = 1.0$$

$$\theta_{HD_r=x|false} = 0.0$$

$$\begin{aligned} \theta_{hd_r|true} &= \frac{\sum_{s \in i(r)} \sum_y Q(y) Q(Ch_s = hd_r | y) bt(\mathbf{pa}_{ch_s}[y]) \prod_{t_i \in tb(\mathbf{pa}_{ch_s})} Q(t_i | y) + \alpha(r, hd_r, true)}{\sum_{s \in i(r)} \sum_y Q(y) bt(\mathbf{pa}_{ch_s}[y]) \prod_{t_i \in tb(\mathbf{pa}_{ch_s})} Q(t_i | y) + \alpha(r, true)} = \\ &= \frac{\sum_{s \in i(r)} \mathcal{N}(s, hd_r) + \alpha(r, hd_r, true)}{\mathcal{N}(r)} = \\ &= \frac{\mathcal{N}(r, hd_r)}{\mathcal{N}(r)} \end{aligned}$$

where

$$\mathcal{N}(s, hd_r) = \sum_y Q(y) Q(Ch_s = hd_r | y) bt(\mathbf{pa}_{ch_s}[y]) \prod_{t_i \in tb(\mathbf{pa}_{ch_s})} Q(t_i | y)$$

and

$$\mathcal{N}(r, hd_r) = \sum_{s \in i(r)} \mathcal{N}(s, hd_r)$$

The derivatives of the parameters are:

$$\frac{\partial \mathcal{N}(s, hd_r)}{\partial Q(ch_{i0} | y_0)} = Q(y_0) bt(\mathbf{pa}_{ch_s}[y_0]) \prod_{t_j \in tb(\mathbf{pa}_{ch_s})} Q(t_j | y_0) 1\{ch_{i0} = hd_r, i_0 = s, ch_{i0} = hd_r\}$$

$$\frac{\partial \mathcal{N}(r, hd_r)}{\partial Q(ch_{i0} | y_0)} = Q(y_0) bt(\mathbf{pa}_{ch_{i0}}[y_0]) \prod_{t_j \in tb(\mathbf{pa}_{ch_{i0}})} Q(t_j | y_0) 1\{ch_{i0} = hd_r, i_0 \in i(r)\}$$

$$\frac{\partial \mathcal{N}(r)}{\partial Q(ch_{i0} | y_0)} = 0$$

$$\frac{\partial \mathcal{N}(s, hd_r)}{\partial Q(t_{i0} | y_0)} = Q(y_0) Q(Ch_s = hd_r | y_0) bt(\mathbf{pa}_{ch_s}[y_0]) \prod_{t_i \in tb(\mathbf{pa}_{ch_s}) \setminus \{t_{i0}\}} Q(t_i | y_0) 1\{t_{i0} \in tb(\mathbf{pa}_{ch_s})\}$$

$$\frac{\partial \mathcal{N}(r, hd_r)}{\partial Q(t_{i0} | y_0)} = Q(y_0) \sum_{s \in i(r)} Q(Ch_s = hd_r | y_0) bt(\mathbf{pa}_{ch_s}[y_0]) \prod_{t_j \in tb(\mathbf{pa}_{ch_s}) \setminus \{t_{i0}\}} Q(t_j | y_0) 1\{t_{i0} \in tb(\mathbf{pa}_{ch_s})\}$$

$$\frac{\partial \mathcal{N}(r)}{\partial Q(t_{i0} | y_0)} = Q(y_0) \sum_{s \in i(r)} bt(\mathbf{pa}_{ch_s}[y_0]) \prod_{t_j \in tb(\mathbf{pa}_{ch_s}) \setminus \{t_{i0}\}} Q(t_j | y_0) 1\{t_{i0} \in tb(\mathbf{pa}_{ch_s})\}$$

$$\frac{\partial \theta_{x_j | \mathbf{pa}_{x_j}}}{\partial Q(ch_{i0} | y_0)} = 0$$

$$\frac{\partial \theta_{t_k | \mathbf{pa}_{t_k}}}{\partial Q(ch_{i0} | y_0)} = 0$$

$$\frac{\partial \theta_{x_j | \mathbf{pa}_{x_j}}}{\partial Q(t_{i0} | y_0)} = 0$$

$$\begin{aligned}\frac{\partial \theta_{t_k | \mathbf{pa}_{t_k}}}{\partial Q(t_{i0} | y_0)} &= 0 \\ \frac{\partial \theta_{hd_r | false}}{\partial Q(ch_{i0} | y_0)} &= 0 \\ \frac{\partial \theta_{hd_r | false}}{\partial Q(t_{i0} | y_0)} &= 0\end{aligned}$$

$$\begin{aligned}\frac{\partial \theta_{hd_r | true}}{\partial Q(ch_{i0} | y_0)} &= \\ \frac{\frac{\partial \mathcal{N}(r, hd_r)}{\partial Q(ch_{i0} | y_0)} \mathcal{N}(r) + \mathcal{N}(r, hd_r) \frac{\partial \mathcal{N}(r)}{\partial Q(ch_{i0} | y_0)}}{\mathcal{N}(r)^2} &= \\ Q(y_0) \frac{bt(\mathbf{pa}_{ch_{i0}}[y_0]) \prod_{t_j \in tb(\mathbf{pa}_{ch_s})} Q(t_j | y_0) 1\{ch_{i0} = hd_r, i_0 \in i(r)\}}{\mathcal{N}(r)} &\end{aligned}$$

$$\begin{aligned}\frac{\partial \theta_{hd_r | true}}{\partial Q(t_{i0} | y_0)} &= \\ \frac{\frac{\partial \mathcal{N}(r, hd_r)}{\partial Q(t_{i0} | y_0)} \mathcal{N}(r) + \mathcal{N}(r, hd_r) \frac{\partial \mathcal{N}(r)}{\partial Q(t_{i0} | y_0)}}{\mathcal{N}(r)^2} &= \\ \frac{Q(y_0) \sum_{s \in i(r)} Q(ch_s = hd_r | y_0) bt(\mathbf{pa}_{ch_s}[y_0]) \prod_{t_j \in tb(\mathbf{pa}_{ch_s}) \setminus \{t_{i0}\}} Q(t_j | y_0) 1\{t_{i0} \in tb(\mathbf{pa}_{ch_s})\}}{\mathcal{N}(r)} + & \\ \frac{Q(y_0) \sum_{s \in i(r)} bt(\mathbf{pa}_{ch_s}[y_0]) \prod_{t_j \in tb(\mathbf{pa}_{ch_s}) \setminus \{t_{i0}\}} Q(t_j | y_0) 1\{t_{i0} \in tb(\mathbf{pa}_{ch_s})\} \mathcal{N}(r, hd_r)}{\mathcal{N}(r)^2} &\end{aligned}$$

Let us now compute $\frac{\partial \log P(\mathbf{x}[y], \mathbf{t}, \mathbf{ch})}{\partial Q(ch_{i0} | y_0)}$. We can express $\log P(\mathbf{x}[y], \mathbf{t}, \mathbf{ch})$ with respect to ch_i as

$$\begin{aligned}\log P(\mathbf{x}[y], \mathbf{t}, \mathbf{ch}) &= \sum_{k \in t(i, y, \mathbf{x}[y], \mathbf{t})} \log \theta_{HD_r(i)=ch_k | true} + \sum_{k \notin t(i, y, \mathbf{x}[y], \mathbf{t})} \log \theta_{HD_r(i)=ch_k | \mathbf{pa}_{ch_k}[y]} \\ &+ \sum_j \log \theta_{x_j | \mathbf{pa}_{x_j}[y]} + \sum_k \log \theta_{t_k | \mathbf{pa}_{t_k}}\end{aligned}$$

where $t(i, y, \mathbf{x}[y], \mathbf{t})$ is the set of indexes of choice variables ch_k that are instances of rule $r(i)$ and such that the instantiated rule k has the body true with respect to $\mathbf{x}[y], \mathbf{t}$. Since the summations $\sum_{k \notin t(i, y, \mathbf{x}[y], \mathbf{t})}$, \sum_j and \sum_k have 0 derivative, it is possible to prove that

Proposition 2 (Derivatives of $\log P$ with respect to $Q(ch_{i0} | y_0)$).

$$\frac{\partial \log P(\mathbf{x}[y_0], \mathbf{t}, \mathbf{ch})}{\partial Q(ch_{i0} | y_0)} = Q(y_0) \mathcal{D}(y_0, ch_{t(i, y_0, \mathbf{x}[y_0], \mathbf{t})}, ch_{i0}, \mathbf{pa}_{ch_{t(i, y_0, \mathbf{x}[y_0], \mathbf{t})}})$$

where $ch_{t(i, y_0, \mathbf{x}[y_0], \mathbf{t})}$ is the set of choice variables ch_k with $k \in t(i, y_0, \mathbf{x}[y_0], \mathbf{t})$. and

$$\begin{aligned}\mathcal{D}(y_0, ch_{t(i, y_0, \mathbf{x}[y_0], \mathbf{t})}, ch_{i0}, \mathbf{pa}_{ch_{t(i, y_0, \mathbf{x}[y_0], \mathbf{t})}}) &= \\ \frac{1}{\mathcal{N}(r(i))} \sum_{k \in t(i, y_0, \mathbf{x}[y_0], \mathbf{t})} \frac{\prod_{t_j \in tb(\mathbf{pa}_{ch_{i0}})} Q(t_j | y_0) 1\{ch_{i0} = ch_k\}}{\theta_{HD_r(i)=ch_k | true}} &\end{aligned}$$

Proof.

$$\begin{aligned}
& \frac{\partial \log P(\mathbf{x}[y_0], \mathbf{t}, \mathbf{ch})}{\partial Q(ch_{i_0}|y_0)} = \\
& \sum_{k \in t(i, y_0, \mathbf{x}[y_0], \mathbf{t})} \frac{1}{\theta_{HD_{r(i)}=ch_k|true}} \frac{\partial \theta_{HD_{r(i)}=ch_k|true}}{\partial Q(ch_{i_0}|y_0)} = \\
& Q(y_0) \sum_{k \in t(i, y_0, \mathbf{x}[y_0], \mathbf{t})} \frac{1}{\theta_{HD_{r(i)}=ch_k|true}} \frac{\prod_{t_j \in tb(\mathbf{pa}_{ch_{i_0}})} Q(t_j|y_0) 1\{ch_{i_0} = ch_k\}}{\mathcal{N}(r(i))} = \\
& Q(y_0) \frac{1}{\mathcal{N}(r(i))} \sum_{k \in t(i, y_0, \mathbf{x}[y_0], \mathbf{t})} \frac{1}{\theta_{HD_{r(i)}=ch_k|true}} \prod_{t_j \in tb(\mathbf{pa}_{ch_{i_0}})} Q(t_j|y_0) 1\{ch_{i_0} = ch_k\} = \\
& Q(y_0) \frac{1}{\mathcal{N}(r(i))} \sum_{k \in t(i, y_0, \mathbf{x}[y_0], \mathbf{t})} \frac{\prod_{t_j \in tb(\mathbf{pa}_{ch_{i_0}})} Q(t_j|y_0) 1\{ch_{i_0} = ch_k\}}{\theta_{HD_{r(i)}=ch_k|true}} = \\
& Q(y_0) \mathcal{D}(y_0, ch_{t(i, y_0, \mathbf{x}[y_0], \mathbf{t})}, ch_{i_0}, \mathbf{pa}_{ch_{t(i, y_0, \mathbf{x}[y_0], \mathbf{t})}})
\end{aligned}$$

where $ch_{t(i, y_0, \mathbf{x}[y_0], \mathbf{t})}$ is the set of choice variables ch_k with $k \in t(i, y_0, \mathbf{x}[y_0], \mathbf{t})$. \square

Let us now compute $\frac{\partial \log P(\mathbf{x}[y], \mathbf{t}, \mathbf{ch})}{\partial Q(t_{i_0}|y_0)}$. We can express $\log P(\mathbf{x}[y], \mathbf{t}, \mathbf{ch})$ with respect to t_i as

$$\begin{aligned}
\log P(\mathbf{x}[y], \mathbf{t}, \mathbf{ch}) = & \sum_{k \in b(t_i, y, \mathbf{x}[y], \mathbf{t})} \log \theta_{HD_{r(k)}=ch_k|true} + \sum_{k \notin b(t_i, y, \mathbf{x}[y], \mathbf{t})} \log \theta_{HD_{r(k)}=ch_k|\mathbf{pa}_{ch_k}[y]} + \\
& \sum_j \log \theta_{x_j|\mathbf{pa}_{x_j}}[y] + \sum_k \log \theta_{t_k|\mathbf{pa}_{t_k}}
\end{aligned}$$

Let $b(t_i, y, \mathbf{x}[y], \mathbf{t})$ be the set of indexes of instantiations of rules for which t_i appears in the body with a matching truth value and such that the body is true with respect to $\mathbf{x}[y], \mathbf{t}$. The summations $\sum_{k \notin b(t_i, y, \mathbf{x}[y], \mathbf{t})}$, \sum_j and \sum_k have 0 derivative with respect to $Q(t_{i_0}|y_0)$.

Proposition 3 (Derivatives of $\log P$ with respect to $Q(t_{i_0}|y_0)$).

$$\frac{\partial \log P(\mathbf{x}[y_0], \mathbf{t}, \mathbf{ch})}{\partial Q(t_{i_0}|y_0)} = Q(y_0) \mathcal{F}(y_0, ech(t_{i_0}), et(t_{i_0}))$$

where

$$\begin{aligned}
\mathcal{F}(y_0, ech(t_{i_0}), et(t_{i_0})) = & \sum_{k \in b(t_{i_0}, y_0, \mathbf{x}[y_0], \mathbf{t})} \sum_{s \in i(r(k))} bt(\mathbf{pa}_{ch_s}[y_0]) \cdot \\
& \frac{\prod_{t'_j \in tb(\mathbf{pa}_{ch_s}) \setminus t_{i_0}} Q(t'_j|y_0) 1\{t_{i_0} \in tb(\mathbf{pa}_{ch_s})\}}{\mathcal{N}(r(k))} \left(\frac{Q(Ch'_s = ch_k|y_0)}{\theta_{HD_{r(k)}=ch_k|true}} + 1 \right)
\end{aligned}$$

and $ech(t_{i_0})$ is the set of ch_s where $s \in b(t_{i_0}, y_0, \mathbf{x}[y_0], \mathbf{t})$ and $et(t_{i_0})$ is the set of t_i such that t_i appears in the body of a rule $s \in b(t_{i_0}, y_0, \mathbf{x}[y_0], \mathbf{t})$ with a matching truth value.

Proof.

$$\frac{\partial \log P(\mathbf{x}[y_0], \mathbf{t}, \mathbf{ch})}{\partial Q(t_{i_0}|y_0)} =$$

$$\begin{aligned}
& \sum_{k \in b(t_{i_0}, y_0, \mathbf{x}[y_0], \mathbf{t})} \frac{1}{\theta_{HD_{r(k)}=ch_k|true}} \frac{\partial \theta_{HD_{r(k)}=ch_k|true}}{\partial Q(t_{i_0}|y_0)} = \\
& Q(y_0) \sum_{k \in b(t_{i_0}, y_0, \mathbf{x}[y_0], \mathbf{t})} \frac{1}{\theta_{HD_{r(k)}=ch_k|true}} \cdot \\
& \left(\frac{Q(y_0) \sum_{s \in i(r(k))} Q(Ch'_s = ch_k|y_0) bt(\mathbf{pa}_{ch_s}[y_0]) \prod_{t'_j \in tb(\mathbf{pa}_{ch_s}) \setminus \{t_{i_0}\}} Q(t'_j|y_0) 1\{t_{i_0} \in tb(\mathbf{pa}_{ch_s})\}}}{\mathcal{N}(r(k))} + \right. \\
& \left. \frac{Q(y_0) \sum_{s \in i(r(k))} bt(\mathbf{pa}_{ch_s}[y_0]) \prod_{t'_j \in tb(\mathbf{pa}_{ch_s}) \setminus \{t_{i_0}\}} Q(t'_j|y_0) 1\{t_{i_0} \in tb(\mathbf{pa}_{ch_s})\}} \mathcal{N}(r(k), ch_k)}{\mathcal{N}(r(k))^2} \right) = \\
& Q(y_0) \sum_{k \in b(t_{i_0}, y_0, \mathbf{x}[y_0], \mathbf{t})} \frac{1}{\theta_{HD_{r(k)}=ch_k|true}} \sum_{s \in i(r(k))} bt(\mathbf{pa}_{ch_s}[y_0]) \cdot \\
& \left(\frac{Q(Ch'_s = ch_k|y_0) \prod_{t'_j \in tb(\mathbf{pa}_{ch_s}) \setminus \{t_{i_0}\}} Q(t'_j|y_0) 1\{t_{i_0} \in tb(\mathbf{pa}_{ch_s})\}}{\mathcal{N}(r(k))} + \right. \\
& \left. \frac{\prod_{t'_j \in tb(\mathbf{pa}_{ch_s}) \setminus \{t_{i_0}\}} Q(t'_j|y_0) 1\{t_{i_0} \in tb(\mathbf{pa}_{ch_s})\}} \mathcal{N}(r(k), ch_k)}{\mathcal{N}(r(k))^2} \right) = \\
& Q(y_0) \sum_{k \in b(t_{i_0}, y_0, \mathbf{x}[y_0], \mathbf{t})} \frac{1}{\theta_{HD_{r(k)}=ch_k|true}} \sum_{s \in i(r(k))} bt(\mathbf{pa}_{ch_s}[y_0]) \cdot \\
& \frac{\prod_{t'_j \in tb(\mathbf{pa}_{ch_s}) \setminus \{t_{i_0}\}} Q(t'_j|y_0) 1\{t_{i_0} \in tb(\mathbf{pa}_{ch_s})\}}{\mathcal{N}(r(k))} \left(Q(Ch'_s = ch_k|y_0) + \frac{\mathcal{N}(r(k), ch_k)}{\mathcal{N}(r(k))} \right) = \\
& Q(y_0) \sum_{k \in b(t_{i_0}, y_0, \mathbf{x}[y_0], \mathbf{t})} \frac{1}{\theta_{HD_{r(k)}=ch_k|true}} \sum_{s \in i(r(k))} bt(\mathbf{pa}_{ch_s}[y_0]) \cdot \\
& \frac{\prod_{t'_j \in tb(\mathbf{pa}_{ch_s}) \setminus \{t_{i_0}\}} Q(t'_j|y_0) 1\{t_{i_0} \in tb(\mathbf{pa}_{ch_s})\}}{\mathcal{N}(r(k))} \left(Q(Ch'_s = ch_k|y_0) + \theta_{HD_{r(k)}=ch_k|true} \right) = \\
& Q(y_0) \sum_{k \in b(t_{i_0}, y_0, \mathbf{x}[y_0], \mathbf{t})} \sum_{s \in i(r(k))} bt(\mathbf{pa}_{ch_s}[y_0]) \cdot \\
& \frac{\prod_{t'_j \in tb(\mathbf{pa}_{ch_s}) \setminus \{t_{i_0}\}} Q(t'_j|y_0) 1\{t_{i_0} \in tb(\mathbf{pa}_{ch_s})\}}{\mathcal{N}(r(k))} \left(\frac{Q(Ch'_s = ch_k|y_0)}{\theta_{HD_{r(k)}=ch_k|true}} + 1 \right) = \\
& Q(y_0) \mathcal{F}(y_0, ech(t_{i_0}), et(t_{i_0}))
\end{aligned}$$

□

Proposition 4 (Derivatives of $\mathbb{E}\mathbb{P}(ch_i, y_0)$ with respect to $Q(ch_{i_0}|y_0)$).

$$\begin{aligned}
& \frac{\partial \mathbb{E}\mathbb{P}(ch_i, y_0)}{\partial Q(ch_{i_0}|y_0)} = \\
& Q(y_0) \sum_{ch_k \in ch_{t(i, y_0, \mathbf{x}[y_0])} \setminus \{ch_i\}, t \in tb(\mathbf{pa}_{ch_{t(i, y_0, \mathbf{x}[y_0])})} ch_k \in ch_{t(i, y_0, \mathbf{x}[y_0])} \setminus \{ch_i\}} \prod_{t_i \in tb(\mathbf{pa}_{ch_{t(i, y_0, \mathbf{x}[y_0])})} Q(ch_k|y_0) \\
& \prod_{t_i \in tb(\mathbf{pa}_{ch_{t(i, y_0, \mathbf{x}[y_0])})} Q(t_i|y_0) \mathcal{D}(y_0, ch_{t(i, y_0, \mathbf{x}[y_0], \mathbf{t})}, ch_{i_0}, \mathbf{pa}_{ch_{t(i, y_0, \mathbf{x}[y_0], \mathbf{t})}})
\end{aligned}$$

where $t(i, y_0, \mathbf{x}[y_0])$ is the set of indexes of choice variables ch_k that are instances of rule $r(i)$ and such that the instantiated rule has the body not false with respect to $\mathbf{x}[y_0]$.

Proof.

$$\begin{aligned}
\frac{\partial \mathbb{E}P(ch_i, y_0)}{\partial Q(ch_{i0}|y_0)} &= \frac{\partial \mathbb{E}_{Q(\mathbf{CH}, \mathbf{T}|ch_i, y_0)}[\log P(\mathbf{x}[y_0], \mathbf{T}, \mathbf{CH})]}{\partial Q(ch_{i0}|y_0)} \\
&= \frac{\partial \sum_{\mathbf{ch}, \mathbf{t}} Q(\mathbf{ch}, \mathbf{t}|ch_i, y_0) \log P(\mathbf{x}[y_0], \mathbf{t}, \mathbf{ch})}{\partial Q(ch_{i0}|y_0)} \\
&= \sum_{\mathbf{ch}} \frac{\partial Q(\mathbf{ch}, \mathbf{t}|ch_i, y_0)}{\partial Q(ch_{i0}|y_0)} \log P(\mathbf{x}[y_0], \mathbf{t}, \mathbf{ch}) + Q(\mathbf{ch}, \mathbf{t}|ch_i, y_0) \frac{\partial \log P(\mathbf{x}[y_0], \mathbf{t}, \mathbf{ch})}{\partial Q(ch_{i0}|y_0)}
\end{aligned}$$

The first term is 0 because $Q(\mathbf{ch}, \mathbf{t}|ch_i, y_0)$ is constant with respect to $Q(ch_{i0}|y_0)$. Thus

$$\begin{aligned}
\frac{\partial \mathbb{E}P(ch_i, y_0)}{\partial Q(ch_{i0}|y_0)} &= E_{Q(\mathbf{CH}, \mathbf{T}|ch_i, y_0)} \left[\frac{\partial \log P(\mathbf{x}[y_0], \mathbf{t}, \mathbf{ch})}{\partial Q(ch_{i0}|y_0)} \right] \{ch_i = ch_{i0}\} = \\
&E_{Q(\mathbf{CH}, \mathbf{T}|ch_i, y_0)} [Q(y_0) \mathcal{D}(y_0, ch_{t(i, y_0, \mathbf{x}[y_0], \mathbf{t})}), ch_{i0}, \mathbf{pa}_{ch_{t(i, y_0, \mathbf{x}[y_0], \mathbf{t})}})] \{ch_i = ch_{i0}\} = \\
&Q(y_0) \mathbb{E}_{Q(\mathbf{CH}, \mathbf{T}|ch_i, \mathbf{pa}_{ch_{i0}})} [\mathcal{D}(y_0, ch_{t(i, y_0, \mathbf{x}[y_0], \mathbf{t})}), ch_{i0}, \mathbf{pa}_{ch_{t(i, y_0, \mathbf{x}[y_0], \mathbf{t})}})] \{ch_i = ch_{i0}\}
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}_{Q(\mathbf{CH}, \mathbf{T}|ch_{i0}, y_0)} [\mathcal{D}(y_0, ch_{t(i, y_0, \mathbf{x}[y_0], \mathbf{t})}), ch_{i0}, \mathbf{pa}_{ch_{t(i, y_0, \mathbf{x}[y_0], \mathbf{t})}}] = \\
&\sum_{ch_k \in ch_{t(i, y_0, \mathbf{x}[y_0])} \setminus \{ch_i\}, t \in tb(\mathbf{pa}_{ch_{t(i, y_0, \mathbf{x}[y_0])}})} Q(ch_{t(i, y_0, \mathbf{x}[y_0])}, tb(\mathbf{pa}_{ch_{t(i, y_0, \mathbf{x}[y_0])}}) | y_0) \\
&\mathcal{D}(y_0, ch_{t(i, y_0)}, ch_{i0}, tb(\mathbf{pa}_{ch_{t(i, y_0)}})) = \\
&\sum_{ch_k \in ch_{t(i, y_0, \mathbf{x}[y_0])} \setminus \{ch_i\}, t \in tb(\mathbf{pa}_{ch_{t(i, y_0, \mathbf{x}[y_0])}})} \prod_{ch_k \in ch_{t(i, y_0, \mathbf{x}[y_0])} \setminus \{ch_i\}} Q(ch_k | y_0) \\
&\prod_{t_i \in tb(\mathbf{pa}_{ch_{t(i, y_0, \mathbf{x}[y_0])}})} Q(t_i | y_0) \mathcal{D}(y_0, ch_{t(i, y_0, \mathbf{x}[y_0], \mathbf{t})}), ch_{i0}, \mathbf{pa}_{ch_{t(i, y_0, \mathbf{x}[y_0], \mathbf{t})}})
\end{aligned}$$

where $t(i, y_0, \mathbf{x}[y_0])$ is the set of indexes of choice variables ch_k that are instances of rule $r(i)$ and such that the instantiated rule has the body not false with respect to $\mathbf{x}[y_0]$. \square

Proposition 5 (Derivatives of $\mathbb{E}P(t_i, y_0)$ with respect to $Q(t_{i0}|y_0)$).

$$\begin{aligned}
\frac{\partial \mathbb{E}P(t_i, y_0)}{\partial Q(t_{i0}|y_0)} &= \\
&\frac{Q(y_0)}{Q(t_{i0}|y_0)} \sum_{k, t_i \in body(k)} \prod_{t_j \in body(k) \setminus \{t_i\}} Q(t_j | y_0) \cdot \\
&\sum_{ch_k} Q(ch_k | y_0) \sum_{s \in i(r(k))} 1_{\{t_{j0} \in tb(\mathbf{pa}_{ch_s})\}} U(s, ch_k, y_0)
\end{aligned}$$

where

$$U(s, ch_k, y_0) = bt(\mathbf{pa}_{ch_s}[y_0]) \frac{\prod_{t'_j \in tb(\mathbf{pa}_{ch_s})} Q(t'_j | y_0)}{\mathcal{N}(r(k))} \left(\frac{Q(CH'_s = ch_k | y_0)}{\theta_{HD_{r(k)} = ch_k | true}} + 1 \right)$$

Proof.

$$\begin{aligned}
\frac{\partial \mathbb{E}P(t_i, y_0)}{\partial Q(t_{i0}|y_0)} &= \mathbb{E}_{Q(\mathbf{CH}, \mathbf{T}|t_i, y_0)} \left[\frac{\partial \log P(\mathbf{x}[y_0], \mathbf{t}, \mathbf{ch})}{\partial Q(t_{i0}|y_0)} \right] \{t_i = t_{i0}\} = \\
&\mathbb{E}_{Q(\mathbf{CH}, \mathbf{T}|t_i, y_0)} [Q(y_0) \mathcal{F}(y_0, ech(t_{i0}), et(t_{i0}))] \{t_i = t_{i0}\} = \\
&Q(y_0) \mathbb{E}_{Q(\mathbf{CH}, \mathbf{T}|t_i, y_0)} [\mathcal{F}(y_0, ech(t_{i0}), et(t_{i0}))] \{t_i = t_{i0}\}
\end{aligned}$$

$$\begin{aligned}
& \mathbb{E}_{Q(\mathbf{CH}, \mathbf{T}|t_{i0}, y_0)}[\mathcal{F}(y_0, ech(t_{i0}), et(t_{i0}))] = \\
& \sum_{\mathbf{ch}, \mathbf{t} \setminus \{t_{i0}\}} Q(\mathbf{ch}, \mathbf{t}|y_0) \mathcal{F}(y_0, ech(t_{i0}), et(t_{i0})) = \\
& \sum_{\mathbf{ch}, \mathbf{t} \setminus \{t_{i0}\}} Q(\mathbf{ch}, \mathbf{t}|y_0) Q(y_0) \sum_{k \in b(t_{i0}, y_0, \mathbf{x}[y_0], \mathbf{t})} \sum_{s \in i(r(k))} bt(\mathbf{pa}_{ch_s}[y_0]) \cdot \\
& \frac{\prod_{t'_j \in tb(\mathbf{pa}_{ch_s}) \setminus \{t_{i0}\}} Q(t'_j|y_0) 1\{t_{j0} \in tb(\mathbf{pa}_{ch_s})\}}{\mathcal{N}(r(k))} \left(\frac{Q(CH'_s = ch_k|y_0)}{\theta_{HD_{r(k)}=ch_k|true}} + 1 \right) = \\
& Q(y_0) \sum_{k, t_i \in body(k)} \prod_{t_j \in body(k) \setminus \{t_i\}} Q(t_j|y_0) \sum_{ch_k} Q(ch_k|y_0) \sum_{s \in i(r(k))} bt(\mathbf{pa}_{ch_s}[y_0]) \cdot \\
& \frac{\prod_{t'_j \in tb(\mathbf{pa}_{ch_s}) \setminus \{t_{i0}\}} Q(t'_j|y_0) 1\{t_{j0} \in tb(\mathbf{pa}_{ch_s})\}}{\mathcal{N}(r(k))} \left(\frac{Q(CH'_s = ch_k|y_0)}{\theta_{HD_{r(k)}=ch_k|true}} + 1 \right) = \\
& \frac{Q(y_0)}{Q(t_{i0}|y_0)} \sum_{k, t_i \in body(k)} \prod_{t_j \in body(k) \setminus \{t_i\}} Q(t_j|y_0) \cdot \\
& \sum_{ch_k} Q(ch_k|y_0) \sum_{s \in i(r(k))} 1\{t_{j0} \in tb(\mathbf{pa}_{ch_s})\} U(s, ch_k, y_0)
\end{aligned}$$

where

$$U(s, ch_k, y_0) = bt(\mathbf{pa}_{ch_s}[y_0]) \frac{\prod_{t'_j \in tb(\mathbf{pa}_{ch_s})} Q(t'_j|y_0)}{\mathcal{N}(r(k))} \left(\frac{Q(CH'_s = ch_k|y_0)}{\theta_{HD_{r(k)}=ch_k|true}} + 1 \right)$$

□

Theorem 1 (Derivatives of G with respect to $Q(ch_i|y)$ and $Q(t_i|y)$). Overall, the derivative of G with respect to ch_i and t_i are:

$$\begin{aligned}
& \frac{\partial G_{ch_i, y}(Q, \gamma)}{\partial Q(ch_i|y)} = \\
& -\frac{1}{Q(ch_i|y)} + Q(y)(1 - Q(ch_i|y)) \cdot \\
& \left(\frac{1 - \gamma}{Q(ch_i)} + \gamma \mathbb{E}_{Q(\mathbf{CH}, \mathbf{T}|ch_i, y_0)}[\mathcal{D}(y_0, ch_{t(i, y_0, \mathbf{x}[y_0], \mathbf{t}), ch_{i0}}, \mathbf{pa}_{ch_{t(i, y_0, \mathbf{x}[y_0], \mathbf{t})}})] \right)
\end{aligned}$$

$$\frac{\partial G_{t_i, y}(Q, \gamma)}{\partial Q(t_i|y)} = -\frac{1}{Q(t_i|y)} + Q(y)(1 - Q(t_i|y)) \left(\frac{1 - \gamma}{Q(t_i)} + \gamma \mathbb{E}_{Q(\mathbf{CH}, \mathbf{T}|t_i, y_0)}[\mathcal{F}(y_0, ech(t_{i0}), et(t_{i0}))] \right)$$

Proof.

$$\begin{aligned}
\frac{\partial \log Q(ch_i)}{\partial Q(ch_{i0}|y_0)} &= \frac{Q(y_0)}{Q(ch_i)} 1\{ch_i = ch_{i0}\} \\
\frac{\partial \log Q(t_i)}{\partial Q(t_{i0}|y_0)} &= \frac{Q(y_0)}{Q(t_i)} 1\{t_i = t_{i0}\}
\end{aligned}$$

In the following we need

$$\begin{aligned}
& \frac{\partial((1 - \gamma) \log Q(ch_i) + \gamma \mathbb{E}P(ch_i, y_0))}{\partial Q(ch_{i0}|y_0)} = \\
& Q(y_0) \left(\frac{1 - \gamma}{Q(ch_i)} 1\{ch_i = ch_{i0}\} + \gamma \mathbb{E}_{Q(\mathbf{CH}, \mathbf{T}|ch_i, y_0)}[\mathcal{D}(y_0, ch_{t(i, y), ch_{i0}}, tb(\mathbf{pa}_{ch_{t(i, y_0)}}))] \right)
\end{aligned}$$

$$\frac{\partial((1-\gamma)\log Q(t_i) + \gamma\mathbb{E}\mathbb{P}(t_i, y_0))}{\partial Q(t_{i0}|y_0)} =$$

$$Q(y_0) \left(\frac{1-\gamma}{Q(t_i)} \mathbf{1}\{t_i = t_{i0}\} + \gamma\mathbb{E}_{Q(\mathbf{CH}, \mathbf{T}|t_i, y_0)}[\mathcal{F}(y_0, ech(t_{i0}), et(t_{i0}))] \right)$$

Let us write $\log Z_{\mathbf{CH}}(i, y_0, \gamma)$

$$\log Z_{\mathbf{CH}}(i, y_0, \gamma) = \log \sum_{ch'_i} e^{(1-\gamma)\log Q(ch'_i) + \gamma\mathbb{E}\mathbb{P}(ch'_i, y_0)}$$

$$\frac{\partial \log Z_{\mathbf{CH}}(i, y_0, \gamma)}{\partial Q(ch_{i0}|y_0)} =$$

$$\frac{1}{Z_{\mathbf{CH}}(i, y_0, \gamma)} \sum_{ch'_i} \frac{\partial e^{(1-\gamma)\log Q(ch'_i) + \gamma\mathbb{E}\mathbb{P}(ch'_i, y_0)}}{\partial Q(ch_{i0}|y_0)} =$$

$$\frac{1}{Z_{\mathbf{CH}}(i, y_0, \gamma)} \sum_{ch'_i} e^{(1-\gamma)\log Q(ch'_i) + \gamma\mathbb{E}\mathbb{P}(ch'_i, y_0)} \frac{\partial((1-\gamma)\log Q(ch'_i) + \gamma\mathbb{E}\mathbb{P}(ch'_i, y_0))}{\partial Q(ch_{i0}|y_0)} =$$

$$\frac{1}{Z_{\mathbf{CH}}(i, y_0, \gamma)} \sum_{ch'_i} Q(ch'_i)^{1-\gamma} \exp(\mathbb{E}\mathbb{P}(ch'_i, y_0)^\gamma) \frac{\partial((1-\gamma)\log Q(ch'_i) + \gamma\mathbb{E}\mathbb{P}(ch'_i, y_0))}{\partial Q(ch_{i0}|y_0)} =$$

$$\sum_{ch'_i} Q(ch'_i|y_0) \frac{\partial((1-\gamma)\log Q(ch'_i) + \gamma\mathbb{E}\mathbb{P}(ch'_i, y_0))}{\partial Q(ch_{i0}|y_0)} =$$

$$Q(ch_{i0}|y_0) \frac{\partial((1-\gamma)\log Q(ch_{i0}) + \gamma\mathbb{E}\mathbb{P}(ch_{i0}, y_0))}{\partial Q(ch_{i0}|y_0)}$$

Let us write $\log Z_{\mathbf{T}}(i, y_0, \gamma)$

$$\log Z_{\mathbf{T}}(i, y_0, \gamma) = \log \sum_{t'_i} e^{(1-\gamma)\log Q(t'_i) + \gamma\mathbb{E}\mathbb{P}(t'_i, y_0)}$$

$$\frac{\partial \log Z_{\mathbf{T}}(i, y_0, \gamma)}{\partial Q(t_{i0}|y_0)} =$$

$$\frac{1}{Z_{\mathbf{T}}(i, y_0, \gamma)} \sum_{t'_i} \frac{\partial e^{(1-\gamma)\log Q(t'_i) + \gamma\mathbb{E}\mathbb{P}(t'_i, y_0)}}{\partial Q(t_{i0}|y_0)} =$$

$$\frac{1}{Z_{\mathbf{T}}(i, y_0, \gamma)} \sum_{t'_i} e^{(1-\gamma)\log Q(t'_i) + \gamma\mathbb{E}\mathbb{P}(t'_i, y_0)} \frac{\partial((1-\gamma)\log Q(t'_i) + \gamma\mathbb{E}\mathbb{P}(t'_i, y_0))}{\partial Q(t_{i0}|y_0)} =$$

$$\frac{1}{Z_{\mathbf{T}}(i, y_0, \gamma)} \sum_{t'_i} Q(t'_i)^{1-\gamma} \exp(\mathbb{E}\mathbb{P}(t'_i, y_0)^\gamma) \frac{\partial((1-\gamma)\log Q(t'_i) + \gamma\mathbb{E}\mathbb{P}(t'_i, y_0))}{\partial Q(t_{i0}|y_0)} =$$

$$\sum_{t'_i} Q(t'_i|y_0) \frac{\partial((1-\gamma)\log Q(t'_i) + \gamma\mathbb{E}\mathbb{P}(t'_i, y_0))}{\partial Q(t_{i0}|y_0)} =$$

$$Q(t_{i0}|y_0) \frac{\partial((1-\gamma)\log Q(t_{i0}) + \gamma\mathbb{E}\mathbb{P}(t_{i0}, y_0))}{\partial Q(t_{i0}|y_0)}$$

Thus, using the results obtained in Equations 2 and 2, the derivative of the G functions with respect to $Q(ch_i|y)$ and $Q(t_i|y)$ are

$$\begin{aligned}
\frac{\partial G_{ch_i, y}(Q, \gamma)}{\partial Q(ch_i|y)} &= \\
&-\frac{1}{Q(ch_i|y)} + \frac{\partial((1-\gamma)\log Q(ch_i) + \gamma\mathbb{E}\mathbb{P}(ch_i, y))}{\partial Q(ch_i|y)} - \\
Q(ch_i|y) \frac{\partial((1-\gamma)\log Q(ch_i) + \gamma\mathbb{E}\mathbb{P}(ch_i, y))}{\partial Q(ch_i|y)} &= \\
&-\frac{1}{Q(ch_i|y)} + (1-Q(ch_i|y)) \frac{\partial((1-\gamma)\log Q(ch_i) + \gamma\mathbb{E}\mathbb{P}(ch_i, y))}{\partial Q(ch_i|y)} = \\
&-\frac{1}{Q(ch_i|y)} + Q(y)(1-Q(ch_i|y)) \left(\frac{(1-\gamma)}{Q(ch_i)} + \right. \\
&\left. \gamma \mathbb{E}_{Q(\mathbf{CH}|ch_i, \mathbf{pa}_{ch_i})} [\mathcal{D}(y_0, ch_{t(i, y_0, \mathbf{x}[y_0], \mathbf{t}), ch_{i0}}, \mathbf{pa}_{ch_{t(i, y_0, \mathbf{x}[y_0], \mathbf{t})})}] \right)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial G_{t_i, y}(Q, \gamma)}{\partial Q(t_i|y)} &= \\
&-\frac{1}{Q(t_i|y)} + \frac{\partial((1-\gamma)\log Q(t_i) + \gamma\mathbb{E}\mathbb{P}(t_i, y))}{\partial Q(t_i|y)} - Q(t_i|y) \frac{\partial((1-\gamma)\log Q(t_i) + \gamma\mathbb{E}\mathbb{P}(t_i, y))}{\partial Q(t_i|y)} = \\
&-\frac{1}{Q(t_i|y)} + (1-Q(t_i|y)) \frac{\partial((1-\gamma)\log Q(t_i) + \gamma\mathbb{E}\mathbb{P}(t_i, y))}{\partial Q(t_i|y)} = \\
&-\frac{1}{Q(t_i|y)} + Q(y)(1-Q(t_i|y)) \left(\frac{(1-\gamma)}{Q(t_i)} + \gamma \mathbb{E}_{Q(\mathbf{CH}, \mathbf{T}|t_i, y)} [\mathcal{F}(y_0, ech(t_{i0}), et(t_{i0}))] \right)
\end{aligned}$$

Overall, the derivative of G with respect to ch_i and t_i are:

$$\begin{aligned}
\frac{\partial G_{ch_i, y}(Q, \gamma)}{\partial Q(ch_i|y)} &= \\
&-\frac{1}{Q(ch_i|y)} + Q(y)(1-Q(ch_i|y)) \left(\frac{1-\gamma}{Q(ch_i)} + \right. \\
&\left. \gamma \mathbb{E}_{Q(\mathbf{CH}, \mathbf{T}|ch_i, y_0)} [\mathcal{D}(y_0, ch_{t(i, y_0, \mathbf{x}[y_0], \mathbf{t}), ch_{i0}}, \mathbf{pa}_{ch_{t(i, y_0, \mathbf{x}[y_0], \mathbf{t})})}] \right)
\end{aligned}$$

$$\frac{\partial G_{t_i, y}(Q, \gamma)}{\partial Q(t_i|y)} = -\frac{1}{Q(t_i|y)} + Q(y)(1-Q(t_i|y)) \left(\frac{1-\gamma}{Q(t_i)} + \gamma \mathbb{E}_{Q(\mathbf{CH}, \mathbf{T}|t_i, y_0)} [\mathcal{F}(y_0, ech(t_{i0}), et(t_{i0}))] \right)$$

□

Theorem 2 (Derivatives of G with respect to γ).

$$\frac{\partial G_{ch_i, y}(Q, \gamma)}{\partial \gamma} = -\log Q(ch_i) + \mathbb{E}\mathbb{P}'(ch_i, y) - \mathbb{E}_{Q(ch'_i|y)} [\mathbb{E}\mathbb{P}'(ch'_i, y)] - \log Q(ch'_i)$$

where

$$\begin{aligned}
\mathbb{E}\mathbb{P}'(ch_i, y) &= \\
&\prod_{t_j \in \text{body}(i)} Q(t_j|y) \log \theta_{HD_{r(k)=ch_i} | \mathbf{pa}_{ch_k}} [y] 1\{\text{body}(i)[y] = \text{true}\} + \\
&\delta \left(\sum_{t_j \in \mathbf{pa}_{ch_i}^T} \prod_{t_j \in \mathbf{pa}_{ch_i}^T} Q(t_j|y) 1\{\text{body}(i)[y] = \text{false}, ch_i \neq \text{none}\} + \right. \\
&\left. R(i, \mathbf{ch}, y) + 1\{ch_i = x_j[y], \text{val}(ch_i)[y] = \text{false}\} + Q(T_j = \text{false}|y) 1\{ch_i = t_j, t_j = \text{false}\} \right)
\end{aligned}$$

$$\begin{aligned}
R(i, \mathbf{ch}, y) = & \sum_{j \in p(i), x_j[y]=\text{true}, ch_i \neq x_j[y]} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s \neq x_j[y]|y) + \\
& \sum_{j \in p(i), t_j=\text{true}, ch_i \neq t_j} \prod_{ch_s \in \mathbf{pa}_{t_j}, s \neq i} Q(ch_s \neq t_j|y) Q(T_j = \text{true}|y)
\end{aligned}$$

$body(i)$ is the body for rule i and $body(i)[y]$ the portion of body restricted to \mathbf{X} variables taking the values $\mathbf{x}[y]$, and

$$\frac{\partial G_{t_i, y}(Q, \gamma)}{\partial \gamma} = -\log Q(t_i) + \mathbb{E}\mathbb{P}'(t_i, y) - E_{Q(t'_i|y)}[\mathbb{E}\mathbb{P}'(t'_i, y) - \log Q(t'_i)]$$

where

$$\begin{aligned}
\mathbb{E}\mathbb{P}'(t_i, y) = & \sum_{k \in bb(t_i, y)} \sum_{ch_k \neq \text{none}} Q(ch_k|y) \prod_{t_j \in body(k), t_j \neq t_i} Q(t_j|y) \log \theta_{HD, r(k)=ch_k | \text{true}} + \\
& \delta(S(i, t, y) + \prod_{Ch_s \in \mathbf{pa}_{t_i}} Q(Ch_s \neq t_i|y) \{t_i = \text{true}\} \delta + \\
& \left(1 - \prod_{Ch_s \in \mathbf{pa}_{t_i}} Q(Ch_s \neq t_i|y)\right) 1\{t_i = \text{false}\} \delta
\end{aligned}$$

$$\begin{aligned}
S(i, \mathbf{t}, y) = & \sum_{k \in bb(t_i, y)} Q(Ch_k = \text{none}|y) \prod_{t_j \in body(k), t_j \neq t_i} Q(t_j|y) + \\
& \sum_{k \in bb(t_i, y)} Q(Ch_k \neq \text{none}|y) \left(1 - \prod_{t_j \in body(k), t_j \neq t_i} Q(t_j|y)\right) + \\
& \sum_{k, body(k)[y]=\text{true}, \bar{t}_i \in body(k)} Q(Ch_k \neq \text{none}|y) + \\
& \sum_{k, body(k)[y]=\text{false}, t_i \in body(k)} Q(Ch_k \neq \text{none}|y)
\end{aligned}$$

and $\delta \approx \log 0$ (e.g. $\delta = -/10$),

Proof.

$$\begin{aligned}
\frac{\partial \log Z_{\mathbf{CH}}(i, y, \gamma)}{\partial \gamma} = & \frac{1}{Z_{\mathbf{CH}}(i, y, \gamma)} \sum_{ch'_i} e^{(1-\gamma) \log Q(ch'_i) + \gamma \mathbb{E}\mathbb{P}(ch'_i, y)} \frac{\partial((1-\gamma) \log Q(ch'_i) + \gamma \mathbb{E}\mathbb{P}(ch'_i, y))}{\partial \gamma} = \\
& \frac{1}{Z_{\mathbf{CH}}(i, y, \gamma)} \sum_{ch'_i} Q(ch'_i)^{1-\gamma} e^{\gamma \mathbb{E}\mathbb{P}(ch'_i, y)} (-\log Q(ch'_i) + \mathbb{E}\mathbb{P}(ch'_i, y)) = \\
& \sum_{ch'_i} Q(ch'_i|y) (\mathbb{E}\mathbb{P}(ch'_i, y) - \log Q(ch'_i)) = \\
& \mathbb{E}_{Q(ch_i|y)} [\mathbb{E}\mathbb{P}(ch_i, y) - \log Q(ch_i)]
\end{aligned}$$

$$\frac{\partial G_{ch_i, y}(Q, \gamma)}{\partial \gamma} = -\log Q(ch_i) + \mathbb{E}\mathbb{P}(ch_i, y) - \mathbb{E}_{Q(ch'_i|y)}[\mathbb{E}\mathbb{P}(ch'_i, y) - \log Q(ch'_i)]$$

$$\begin{aligned} \frac{\partial \log Z_{\mathbf{T}}(i, y, \gamma)}{\partial \gamma} &= \\ \frac{1}{Z_{\mathbf{T}}(i, y, \gamma)} \sum_{t'_i} e^{(1-\gamma) \log Q(t'_i) + \gamma \mathbb{E}\mathbb{P}(t'_i, y)} \frac{\partial((1-\gamma) \log Q(t'_i) + \gamma \mathbb{E}\mathbb{P}(t'_i, y))}{\partial \gamma} &= \\ \frac{1}{Z_{\mathbf{T}}(i, y, \gamma)} \sum_{t'_i} Q(t'_i)^{1-\gamma} e^{\gamma \mathbb{E}\mathbb{P}(t'_i, y)} (-\log Q(t'_i) + \mathbb{E}\mathbb{P}(t'_i, y)) &= \\ \sum_{t'_i} Q(t'_i|y) (\mathbb{E}\mathbb{P}(t'_i, y) - \log Q(t'_i)) &= \\ \mathbb{E}_{Q(t_i|y)}[\mathbb{E}\mathbb{P}(t_i, y) - \log Q(t_i)] & \\ \frac{\partial G_{t_i, y}(Q, \gamma)}{\partial \gamma} &= -\log Q(t_i) + \mathbb{E}\mathbb{P}(t_i, y) - \mathbb{E}_{Q(t'_i|y)}[\mathbb{E}\mathbb{P}(t'_i, y) - \log Q(t'_i)] \end{aligned}$$

Let us now see how to compute $\mathbb{E}\mathbb{P}(ch_i, y)$

$$\begin{aligned} \mathbb{E}\mathbb{P}(ch_i, y) &= \mathbb{E}_{Q(\mathbf{CH}, \mathbf{T}|ch_i, y)}[\log P(\mathbf{x}[y], \mathbf{T}, \mathbf{CH})] = \\ & \sum_{\mathbf{ch}, \mathbf{ch}[i] \neq ch_i, \mathbf{t}} Q(\mathbf{ch}, \mathbf{t}|ch_i, y) \left(\sum_k \log \theta_{HD_{r(k)}=ch_k|\mathbf{pa}_{ch_k}}[y] + \sum_j \log \theta_{x_j|\mathbf{pa}_{x_j}}[y] + \sum_k \log \theta_{t_k|\mathbf{pa}_{t_k}} \right) = \\ & \sum_k \sum_{\mathbf{ch}, \mathbf{ch}[i] \neq ch_i, \mathbf{t}} Q(\mathbf{ch}, \mathbf{t}|ch_i, y) \log \theta_{HD_{r(k)}=ch_k|\mathbf{pa}_{ch_k}}[y] + \\ & \sum_j \sum_{\mathbf{ch}, \mathbf{ch}[i] \neq ch_i, \mathbf{t}} Q(\mathbf{ch}, \mathbf{t}|ch_i, y) \log \theta_{x_j|\mathbf{pa}_{x_j}}[y] + \\ & \sum_k \sum_{\mathbf{ch}, \mathbf{ch}[i] \neq ch_i, \mathbf{t}} Q(\mathbf{ch}, \mathbf{t}|ch_i, y) \log \theta_{t_k|\mathbf{pa}_{t_k}} = \\ & \sum_{\mathbf{t} \in \mathbf{pa}_{ch_i}} \prod_{t_j \in \mathbf{pa}_{ch_i}} Q(t_j|y) \log \theta_{HD_{r(i)}=ch_i|\mathbf{pa}_{ch_i}}[y] + \\ & \sum_{k \neq i} \sum_{ch_k, \mathbf{t} \in \mathbf{pa}_{ch_k}} Q(ch_k, \mathbf{t}|ch_i, y) \log \theta_{HD_{r(k)}=ch_k|\mathbf{pa}_{ch_k}}[y] + \\ & \sum_j \sum_{\mathbf{pa}_{x_j} \setminus \{ch_i\}} Q(\mathbf{pa}_{x_j}|ch_i, y) \log \theta_{x_j|\mathbf{pa}_{x_j}}[y] + \\ & \sum_k \sum_{\mathbf{pa}_{t_k} \setminus \{ch_i\}} Q(\mathbf{pa}_{t_k}|ch_i, y) \log \theta_{t_k|\mathbf{pa}_{t_k}} = \\ & \sum_{\mathbf{t} \in \mathbf{pa}_{ch_i}} \prod_{t_j \in \mathbf{pa}_{ch_i}} Q(t_j|y) \log \theta_{HD_{r(i)}=ch_i|\mathbf{pa}_{ch_i}}[y] + \\ & \sum_{k \neq i} \sum_{ch_k, \mathbf{t} \in \mathbf{pa}_{ch_k}} Q(ch_k, \mathbf{t}|ch_i, y) \log \theta_{HD_{r(k)}=ch_k|\mathbf{pa}_{ch_k}}[y] + \\ & \sum_j \sum_{\mathbf{pa}_{x_j} \setminus \{ch_i\}} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) \log \theta_{x_j|\mathbf{pa}_{x_j}}[y] + \sum_k \sum_{\mathbf{pa}_{t_k} \setminus \{ch_i\}} Q(ch_{t_k}|ch_i, y) \log \theta_{t_k|\mathbf{pa}_{t_k}} \end{aligned}$$

So $\mathbb{E}\mathbb{P}(ch_i, y)$ can be computed without inference. Let us write $\mathbb{E}\mathbb{P}(ch_i, y)$ as

$$\mathbb{E}\mathbb{P}(ch_i, y) = \mathbb{E}\mathbb{P}_1(ch_i, y) + \mathbb{E}\mathbb{P}_2(y)$$

where

$$\begin{aligned} \mathbb{E}\mathbb{P}_1(ch_i, y) = & \sum_{t \in \mathbf{pa}_{ch_i}} \prod_{t_j \in \mathbf{pa}_{ch_i}} Q(t_j|y) \log \theta_{HD_{r(k)}=ch_i|\mathbf{pa}_{ch_k}}[y] + \\ & \sum_{j \in p(i)} \sum_{\mathbf{pa}_{x_j} \setminus \{ch_i\}} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) \log \theta_{x_j|\mathbf{pa}_{x_j}}[y] + \\ & \sum_{j \in p(i)} \sum_{\mathbf{pa}_{t_j} \setminus \{ch_i\}, t_j} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) Q(t_j|y) \log \theta_{t_j|\mathbf{pa}_{x_j}}[y] \end{aligned}$$

with $p(i) = \{j | ch_i \in \mathbf{pa}_{x_j}\} = \text{values}(ch_i) \setminus \{none\}$ and $\text{values}(ch_i)$ is the set of values of ch_i , and

$$\begin{aligned} \mathbb{E}\mathbb{P}_2(y) = & \sum_{k \neq i} \sum_{ch_k, t \in \mathbf{pa}_{ch_k}} Q(ch_k, t | ch_i, y) \log \theta_{HD_{r(k)}=ch_k|\mathbf{pa}_{ch_k}}[y] + \\ & \sum_{j \notin p(i)} \sum_{\mathbf{pa}_{x_j} \setminus \{ch_i\}} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) \log \theta_{x_j|\mathbf{pa}_{x_j}}[y] + \sum_k \sum_{\mathbf{pa}_{t_k} \setminus \{ch_i\}} Q(ch_{t_k} | ch_i, y) \log \theta_{t_k|\mathbf{pa}_{t_k}} \end{aligned}$$

does not depend on ch_i , so

$$\begin{aligned} \mathbb{E}\mathbb{P}_1(ch_i, y) = & \prod_{t_j \in \text{body}(i)} Q(t_j|y) \log \theta_{HD_{r(i)}=ch_i|\mathbf{pa}_{ch_i}}[y] 1\{\text{body}(\mathbf{pa}_{ch_i}) = \text{true}\} + \\ & \sum_{t_j \in \mathbf{pa}_{ch_i}^T} \prod_{t_j \in \mathbf{pa}_{ch_i}^T} Q(t_j|y) 1\{\text{body}(\mathbf{pa}_{ch_i}) = \text{false}, ch_i \neq \text{none}\} \delta + \\ & \sum_{j \in p(i), x_j[y] = \text{true}} \sum_{\mathbf{pa}_{x_j} \setminus \{ch_i\}, x_j[y] \notin \text{val}(\mathbf{pa}_{x_j})} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) \delta + \\ & \sum_{j \in p(i), x_j[y] = \text{false}} \sum_{\mathbf{pa}_{x_j} \setminus \{ch_i\}, x_j[y] \in \text{val}(\mathbf{pa}_{x_j})} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) \delta + \\ & \sum_{j \in p(i), t_j = \text{true}} \sum_{\mathbf{pa}_{t_j} \setminus \{ch_i\}, t_j \notin \text{val}(\mathbf{pa}_{t_j})} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) Q(T_j = \text{true}|y) \delta + \\ & \sum_{j \in p(i), t_j = \text{false}} \sum_{\mathbf{pa}_{t_j} \setminus \{ch_i\}, t_j \in \text{val}(\mathbf{pa}_{t_j})} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) Q(T_j = \text{false}|y) \delta \end{aligned}$$

where $\text{body}(i)$ is the body for rule i and $\text{body}(i)[y]$ the portion of body restricted to \mathbf{X} variables taking the values $\mathbf{x}[y]$,

$$\begin{aligned} \mathbb{E}\mathbb{P}_1(ch_i, y) = & \prod_{t_j \in \text{body}(i)} Q(t_j|y) \log \theta_{HD_{r(i)}=ch_i|\mathbf{pa}_{ch_i}}[y] 1\{\text{body}(i)[y] = \text{true}\} + \end{aligned}$$

$$\begin{aligned}
& \delta \left(\sum_{t_j \in \mathbf{pa}_{ch_i}^T} \prod_{t_j \in \mathbf{pa}_{ch_i}^T} Q(t_j|y) 1\{body(body(i)[y] = false, ch_i \neq none)\} + \right. \\
& \sum_{j \in p(i), x_j[y]=true} \sum_{\mathbf{pa}_{x_j} \setminus \{ch_i\}, x_j[y] \notin val(\mathbf{pa}_{x_j})} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) + \\
& \sum_{j \in p(i), x_j[y]=false} \sum_{\mathbf{pa}_{x_j} \setminus \{ch_i\}, x_j[y] \in val(\mathbf{pa}_{x_j})} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) + \\
& \sum_{j \in p(i), t_j=true} \sum_{\mathbf{pa}_{t_j} \setminus \{ch_i\}, t_j \notin val(\mathbf{pa}_{t_j})} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) Q(T_j = true|y) + \\
& \left. \sum_{j \in p(i), t_j=false} \sum_{\mathbf{pa}_{t_j} \setminus \{ch_i\}, t_j \in val(\mathbf{pa}_{t_j})} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) Q(T_j = false|y) \right)
\end{aligned}$$

where δ is used to approximate $\log 0$ (e.g. $\delta = -10$).

$$\begin{aligned}
\mathbb{E}P_1(ch_i, y) = & \prod_{t_j \in body(r(i))} Q(t_j|y) \log \theta_{HD, r(i)=ch_i | \mathbf{pa}_{ch_i}}[y] 1\{body(i)[y] = true\} + \\
& \delta \left(\sum_{t_j \in \mathbf{pa}_{ch_i}^T} \prod_{t_j \in \mathbf{pa}_{ch_i}^T} Q(t_j|y) 1\{body(i)[y] = false, ch_i \neq none\} + \right. \\
& \sum_{j \in p(i), x_j[y]=true} \sum_{\mathbf{pa}_{x_j} \setminus \{ch_i\}, x_j[y] \notin val(\mathbf{pa}_{x_j}) \setminus \{ch_i\}} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) 1\{ch_i \neq x_j\} + \\
& \sum_{j \in p(i), x_j[y]=false} \sum_{\mathbf{pa}_{x_j} \setminus \{ch_i\}, x_j[y] \in val(\mathbf{pa}_{x_j}) \setminus \{ch_i\}} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) + \\
& \sum_{j \in p(i), x_j[y]=false} \sum_{\mathbf{pa}_{x_j} \setminus \{ch_i\}} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) 1\{ch_i = x_j[y]\} + \\
& \sum_{j \in p(i), t_j=true} \sum_{\mathbf{pa}_{t_j} \setminus \{ch_i\}, t_j \notin val(\mathbf{pa}_{t_j}) \setminus \{ch_i\}} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) Q(T_j = true|y) 1\{ch_i \neq t_j\} + \\
& \sum_{j \in p(i), t_j=false} \sum_{\mathbf{pa}_{t_j} \setminus \{ch_i\}, t_j \in val(\mathbf{pa}_{t_j}) \setminus \{ch_i\}} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) Q(T_j = false|y) + \\
& \left. \sum_{j \in p(i), t_j=false} \sum_{\mathbf{pa}_{t_j} \setminus \{ch_i\}} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) Q(T_j = false|y) 1\{ch_i = t_j\} \right) = \\
& \prod_{t_j \in body(r(i))} Q(t_j|y) \log \theta_{HD, r(i)=ch_i | \mathbf{pa}_{ch_i}}[y] 1\{body(i)[y] = true\} + \\
& \delta \left(\sum_{t_j \in \mathbf{pa}_{ch_i}^T} \prod_{t_j \in \mathbf{pa}_{ch_i}^T} Q(t_j|y) 1\{body(body(i)[y] = false, ch_i \neq none)\} + \right. \\
& \sum_{j \in p(i), x_j[y]=true} \sum_{\mathbf{pa}_{x_j} \setminus \{ch_i\}, x_j[y] \notin val(\mathbf{pa}_{x_j}) \setminus \{ch_i\}} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) 1\{ch_i \neq x_j\} + \\
& \sum_{j \in p(i), x_j[y]=false} \sum_{\mathbf{pa}_{x_j} \setminus \{ch_i\}, x_j[y] \in val(\mathbf{pa}_{x_j}) \setminus \{ch_i\}} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) +
\end{aligned}$$

$$\begin{aligned}
& \sum_{j \in p(i), x_j[y]=false} 1\{ch_i = x_j[y]\} + \\
& \sum_{j \in p(i), t_j=true} \sum_{\mathbf{pa}_{t_j} \setminus \{ch_i\}, t_j \notin val(\mathbf{pa}_{t_j}) \setminus \{ch_i\}} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y)Q(T_j = true|y)1\{ch_i \neq t_j\} + \\
& \sum_{j \in p(i), t_j=false} \sum_{\mathbf{pa}_{t_j} \setminus \{ch_i\}, t_j \in val(\mathbf{pa}_{t_j}) \setminus \{ch_i\}} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y)Q(T_j = false|y)) + \\
& \left. \sum_{j \in p(i), t_j=false} Q(T_j = false|y)1\{ch_i = t_j\} \right) = \\
& \prod_{t_j \in body(r(i))} Q(t_j|y) \log \theta_{HD, r(i)=ch_i | \mathbf{pa}_{ch_i}}[y] 1\{body(i)[y] = true\} + \\
& \delta \left(\sum_{t_j \in \mathbf{pa}_{ch_i}^T} \prod_{t_j \in \mathbf{pa}_{ch_i}^T} Q(t_j|y) 1\{body(i)[y] = false, ch_i \neq none\} + \right. \\
& \sum_{j \in p(i), x_j[y]=true} \sum_{\mathbf{pa}_{x_j} \setminus \{ch_i\}, x_j[y] \notin val(\mathbf{pa}_{x_j}) \setminus \{ch_i\}} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) 1\{ch_i \neq x_j\} + \\
& \sum_{j \in p(i), x_j[y]=false} \sum_{\mathbf{pa}_{x_j} \setminus \{ch_i\}, x_j[y] \in val(\mathbf{pa}_{x_j}) \setminus \{ch_i\}} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) + \\
& \sum_{j \in p(i), t_j=true} \sum_{\mathbf{pa}_{t_j} \setminus \{ch_i\}, t_j \notin val(\mathbf{pa}_{t_j}) \setminus \{ch_i\}} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y)Q(T_j = true|y)1\{ch_i \neq t_j\} + \\
& \sum_{j \in p(i), t_j=false} \sum_{\mathbf{pa}_{t_j} \setminus \{ch_i\}, t_j \in val(\mathbf{pa}_{t_j}) \setminus \{ch_i\}} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y)Q(T_j = false|y)) + \\
& 1\{ch_i = x_j[y], val(ch_i)[y] = false\} + \\
& \left. Q(T_j = false|y)1\{ch_i = t_j, val(ch_i) = t_j = false\} \right)
\end{aligned}$$

Let

$$\begin{aligned}
\mathbb{E}\mathbb{P}'(ch_i, y) = & \prod_{t_j \in body(r(i))} Q(t_j|y) \log \theta_{HD, r(k)=ch_i | \mathbf{pa}_{ch_k}}[y] 1\{body(i)[y] = true\} + \\
& \delta \left(\sum_{t_j \in \mathbf{pa}_{ch_i}^T} \prod_{t_j \in \mathbf{pa}_{ch_i}^T} Q(t_j|y) 1\{body(i)[y] = false, ch_i \neq none\} + \right. \\
& \sum_{j \in p(i), x_j[y]=true} \sum_{\mathbf{pa}_{x_j} \setminus \{ch_i\}, x_j[y] \notin val(\mathbf{pa}_{x_j}) \setminus \{ch_i\}} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) 1\{ch_i \neq x_j\} + \\
& \sum_{j \in p(i), t_j=true} \sum_{\mathbf{pa}_{t_j} \setminus \{ch_i\}, t_j \notin val(\mathbf{pa}_{t_j}) \setminus \{ch_i\}} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y)Q(T_j = true|y)1\{ch_i \neq t_j\} + \\
& 1\{ch_i = x_j[y], val(ch_i)[y] = false\} + \\
& \left. Q(T_j = false|y)1\{ch_i = t_j, val(ch_i) = t_j = false\} \right)
\end{aligned}$$

and

$$\begin{aligned} \mathbb{E}\mathbb{P}_3(y) = & \sum_{j \in p(i), x_j[y]=false} \sum_{\mathbf{pa}_{x_j} \setminus \{ch_i\}, x_j[y] \in val(\mathbf{pa}_{x_j}) \setminus \{ch_i\}} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) + \\ & \sum_{j \in p(i), t_j=false} \sum_{\mathbf{pa}_{t_j} \setminus \{ch_i\}, t_j \in val(\mathbf{pa}_{t_j}) \setminus \{ch_i\}} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) Q(T_j = false|y) \end{aligned}$$

where $\mathbb{E}\mathbb{P}_3(y)$ does not depend on ch_i , then $\mathbb{E}\mathbb{P}(ch_i, y) = \mathbb{E}\mathbb{P}'(ch_i, y) + \mathbb{E}\mathbb{P}_2(y) + \mathbb{E}\mathbb{P}_3(y) = \mathbb{E}\mathbb{P}'(ch_i, y) + \mathbb{E}\mathbb{P}''(y)$ and $\mathbb{E}\mathbb{P}''(y)$ does not depend on ch_i .

$$\begin{aligned} \mathbb{E}\mathbb{P}'(ch_i, y) = & \prod_{t_j \in body(r(i))} Q(t_j|y) \log \theta_{HD_{r(k)=ch_i} | \mathbf{pa}_{ch_k}}[y] 1\{body(i)[y] = true\} + \\ & \delta \left(\sum_{t_j \in \mathbf{pa}_{ch_i}^T} \prod_{t_j \in \mathbf{pa}_{ch_i}^T} Q(t_j|y) 1\{body(body(i)[y] = false, ch_i \neq none)\} + \right. \\ & \sum_{j \in p(i), x_j[y]=true} \sum_{\mathbf{pa}_{x_j} \setminus \{ch_i\}, x_j[y] \notin val(\mathbf{pa}_{x_j}) \setminus \{ch_i\}} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) 1\{ch_i \neq x_j\} + \\ & \sum_{j \in p(i), t_j=true} \sum_{\mathbf{pa}_{t_j} \setminus \{ch_i\}, t_j \notin val(\mathbf{pa}_{t_j}) \setminus \{ch_i\}} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s|y) Q(T_j = true|y) 1\{ch_i \neq t_j\} + \\ & \left. 1\{ch_i = x_j[y], val(ch_i)[y] = false\} + \right. \\ & \left. Q(T_j = false|y) 1\{ch_i = t_j, val(ch_i) = t_j = false\} \right) = \\ & \prod_{t_j \in body(r(i))} Q(t_j|y) \log \theta_{HD_{r(k)=ch_i} | \mathbf{pa}_{ch_k}}[y] 1\{body(i)[y] = true\} + \\ & \delta \left(\sum_{t_j \in \mathbf{pa}_{ch_i}^T} \prod_{t_j \in \mathbf{pa}_{ch_i}^T} Q(t_j|y) 1\{body(i)[y] = false, ch_i \neq none\} + \right. \\ & \sum_{j \in p(i), x_j[y]=true} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} \sum_{\mathbf{pa}_{x_j} \setminus \{ch_i\}, x_j[y] \notin val(\mathbf{pa}_{x_j}) \setminus \{ch_i\}} Q(ch_s|y) 1\{ch_i \neq x_j\} + \\ & \sum_{j \in p(i), t_j=true} \prod_{ch_s \in \mathbf{pa}_{t_j}, s \neq i} \sum_{\mathbf{pa}_{t_j} \setminus \{ch_i\}, t_j \notin val(\mathbf{pa}_{t_j}) \setminus \{ch_i\}} Q(ch_s|y) Q(T_j = true|y) 1\{ch_i \neq t_j\} + \\ & \left. 1\{ch_i = x_j[y], val(ch_i)[y] = false\} + \right. \\ & \left. Q(T_j = false|y) 1\{ch_i = t_j, val(ch_i) = t_j = false\} \right) = \\ & \prod_{t_j \in body(r(i))} Q(t_j|y) \log \theta_{HD_{r(k)=ch_i} | \mathbf{pa}_{ch_k}}[y] 1\{body(i)[y] = true\} + \\ & \delta \left(\sum_{t_j \in \mathbf{pa}_{ch_i}^T} \prod_{t_j \in \mathbf{pa}_{ch_i}^T} Q(t_j|y) 1\{body(i)[y] = false, ch_i \neq none\} + \right. \\ & \sum_{j \in p(i), x_j[y]=true, ch_i \neq x_j[y]} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s \neq x_j[y]|y) + \end{aligned}$$

$$\sum_{j \in p(i), t_j = \text{true}, ch_i \neq t_j} \prod_{ch_s \in \mathbf{pa}_{t_j}, s \neq i} Q(ch_s \neq t_j | y) Q(T_j = \text{true} | y) + 1\{ch_i = x_j[y], \text{val}(ch_i)[y] = \text{false}\} + Q(T_j = \text{false} | y) 1\{ch_i = t_j, \text{val}(ch_i) = t_j = \text{false}\})$$

Let

$$R(i, \mathbf{ch}, y) = \sum_{j \in p(i), x_j[y] = \text{true}, ch_i \neq x_j[y]} \prod_{ch_s \in \mathbf{pa}_{x_j}, s \neq i} Q(ch_s \neq x_j[y] | y) + \sum_{j \in p(i), t_j = \text{true}, ch_i \neq t_j} \prod_{ch_s \in \mathbf{pa}_{t_j}, s \neq i} Q(ch_s \neq t_j | y) Q(T_j = \text{true} | y)$$

then

$$\begin{aligned} \mathbb{E}\mathbb{P}'(ch_i, y) &= \prod_{t_j \in \text{body}(r(i))} Q(t_j | y) \log \theta_{HD_{r(k)} = ch_i | \mathbf{pa}_{ch_k}}[y] 1\{\text{body}(i)[y] = \text{true}\} + \\ &\delta \left(\sum_{t_j \in \mathbf{pa}_{ch_i}^T} \prod_{t_j \in \mathbf{pa}_{ch_i}^T} Q(t_j | y) 1\{\text{body}(\text{body}(i)[y]) = \text{false}, ch_i \neq \text{none}\} + \right. \\ &R(i, \mathbf{ch}, y) + 1\{ch_i = x_j[y], \text{val}(ch_i)[y] = \text{false}\} + \\ &\left. Q(T_j = \text{false} | y) 1\{ch_i = t_j, \text{val}(ch_i) = t_j = \text{false}\} \right) \end{aligned}$$

So

$$\begin{aligned} \frac{\partial G_{ch_i, y}(Q, \gamma)}{\partial \gamma} &= -\log Q(ch_i) + \mathbb{E}\mathbb{P}'(ch_i, y) + \mathbb{E}\mathbb{P}''(y) - \mathbb{E}_{Q(ch'_i | y)}[\mathbb{E}\mathbb{P}'(ch'_i, y) + \mathbb{E}\mathbb{P}''(y) - \log Q(ch'_i)] = \\ &-\log Q(ch_i) + \mathbb{E}\mathbb{P}'(ch_i, y) + \mathbb{E}\mathbb{P}''(y) - \mathbb{E}\mathbb{P}''(y) - \mathbb{E}_{Q(ch'_i | y)}[\mathbb{E}\mathbb{P}'(ch'_i, y) - \log Q(ch'_i)] = \\ &-\log Q(ch_i) + \mathbb{E}\mathbb{P}'(ch_i, y) - \mathbb{E}_{Q(ch'_i | y)}[\mathbb{E}\mathbb{P}'(ch'_i, y) - \log Q(ch'_i)] \end{aligned}$$

Let us now see how to compute $\mathbb{E}\mathbb{P}(t_i, y)$

$$\begin{aligned} \mathbb{E}\mathbb{P}(t_i, y) &= \mathbb{E}_{Q(\mathbf{CH}, \mathbf{T} | t_i, y)}[\log P(\mathbf{x}[y], \mathbf{T}, \mathbf{CH})] = \\ &\sum_{\mathbf{ch}, \mathbf{t} \setminus t_i} Q(\mathbf{ch}, \mathbf{t} | t_i, y) \left(\sum_{k \in b(t_i, y, \mathbf{x}[y], \mathbf{t})} \log \theta_{HD_{r(k)} = ch_k | \mathbf{pa}_{ch_k}} + \sum_{k \notin b(t_i, y, \mathbf{x}[y], \mathbf{t})} \log \theta_{HD_{r(i)} = ch_k | \mathbf{pa}_{ch_k}} + \right. \\ &\sum_j \log \theta_{x_j | \mathbf{pa}_{x_j}}[y] + \sum_{k \neq i} \log \theta_{t_k | \mathbf{pa}_{t_k}} + \log \theta_{t_i | \mathbf{pa}_{t_i}} \left. \right) = \\ &\sum_{k \in bb(t_i, y)} \sum_{\mathbf{ch}, \mathbf{t} \setminus t_i} Q(\mathbf{ch}, \mathbf{t} | t_i, y) \log \theta_{HD_{r(k)} = ch_k | \mathbf{pa}_{ch_k}} + \\ &\sum_{k \notin bb(t_i, y)} \sum_{\mathbf{ch}, \mathbf{t} \neq t_i} Q(\mathbf{ch}, \mathbf{t} | t_i, y) \log \theta_{HD_{r(i)} = ch_k | \mathbf{pa}_{ch_k}} + \\ &\sum_j \sum_{\mathbf{ch}, \mathbf{t} \neq t_i} Q(\mathbf{ch}, \mathbf{t} | t_i, y) \log \theta_{x_j | \mathbf{pa}_{x_j}} + \sum_{k \neq i} \sum_{\mathbf{ch}, \mathbf{t} \neq t_i} Q(\mathbf{ch}, \mathbf{t} | t_i, y) \log \theta_{t_k | \mathbf{pa}_{t_k}} + \end{aligned}$$

$$\begin{aligned}
& \sum_{\mathbf{ch}, \mathbf{t} \setminus t_i} Q(\mathbf{ch}, \mathbf{t} | t_i, y) \log \theta_{t_i | \mathbf{pa}_{t_i}} = \\
& \sum_{k \in bb(t_i, y)} \sum_{ch_k} Q(ch_k | y) \sum_{t \in body(k), t \neq t_i} \prod_{t_j \in body(k), t_j \neq t_i} Q(t_j | y) \log \theta_{HD_{r(k)}=ch_k | \mathbf{pa}_{ch_k}} + \\
& \sum_{k \notin bb(t_i, y)} \sum_{ch_k} Q(ch_k | y) \sum_{t \in body(k), t \neq t_i} \prod_{t_j \in body(k), t_j \neq t_i} Q(t_j | y) \log \theta_{HD_{r(i)}=ch_k | \mathbf{pa}_{ch_k}} + \\
& \sum_j \sum_{\mathbf{pa}_{x_j}} Q(\mathbf{pa}_{x_j} | y) \log \theta_{x_j | \mathbf{pa}_{x_j}} + \sum_{k \neq i} \sum_{\mathbf{pa}_{t_k}} Q(\mathbf{pa}_{t_k} | y) \log \theta_{t_k | \mathbf{pa}_{t_k}} + \sum_{\mathbf{pa}_{t_i}} Q(\mathbf{pa}_{t_i} | y) \log \theta_{t_i | \mathbf{pa}_{t_i}}
\end{aligned}$$

where $bb(t_i, y) = \{k | body(k)[y] = true, t_i \in body(k)\}$.

So $\mathbb{EP}(t_i, y)$ can be computed without inference.

$$\mathbb{EP}(t_i, y) = \mathbb{EP}_1(t_i, y) + \mathbb{EP}_2(y)$$

where $\mathbb{EP}_2(y)$ is

$$\mathbb{EP}_2(y) = \sum_j \sum_{\mathbf{pa}_{x_j}} Q(\mathbf{pa}_{x_j} | y) \log \theta_{x_j | \mathbf{pa}_{x_j}} + \sum_{k \neq i} \sum_{\mathbf{pa}_{t_k}} Q(\mathbf{pa}_{t_k} | y) \log \theta_{t_k | \mathbf{pa}_{t_k}}$$

and does not depend on t_i , and

$$\begin{aligned}
\mathbb{EP}_1(t_i, y) = & \sum_{k \in bb(t_i, y)} \sum_{ch_k} Q(ch_k | y) \sum_{t \in body(k), t \neq t_i} \prod_{t_j \in body(k), t_j \neq t_i} Q(t_j | y) \log \theta_{HD_{r(k)}=ch_k | \mathbf{pa}_{ch_k}} + \\
& \sum_{k \notin bb(t_i, y)} \sum_{ch_k} Q(ch_k | y) \sum_{t \in body(k), t \neq t_i} \prod_{t_j \in body(k), t_j \neq t_i} Q(t_j | y) \log \theta_{HD_{r(i)}=ch_k | \mathbf{pa}_{ch_k}} + \\
& \sum_{\mathbf{pa}_{t_i}} Q(\mathbf{pa}_{t_i} | y) \log \theta_{t_i | \mathbf{pa}_{t_i}} = \\
& \sum_{k \in bb(t_i, y)} \sum_{ch_k} Q(ch_k | y) \sum_{t \in body(k), t \neq t_i} \prod_{t_j \in body(k), t_j \neq t_i} Q(t_j | y) \log \theta_{HD_{r(k)}=ch_k | \mathbf{pa}_{ch_k}} + \\
& \sum_{k \notin bb(t_i, y), T_i \in body(k)} \sum_{ch_k} Q(ch_k | y) \sum_{t \in body(k), t \neq t_i} \prod_{t_j \in body(k), t_j \neq t_i} Q(t_j | y) \log \theta_{HD_{r(i)}=ch_k | \mathbf{pa}_{ch_k}} + \\
& \sum_{k \notin bb(t_i, y), T_i \notin body(k)} \sum_{ch_k} Q(ch_k | y) \sum_{t \in body(k), t \neq t_i} \prod_{t_j \in body(k), t_j \neq t_i} Q(t_j | y) \log \theta_{HD_{r(i)}=ch_k | \mathbf{pa}_{ch_k}} + \\
& \sum_{\mathbf{pa}_{t_i}} Q(\mathbf{pa}_{t_i} | y) \log \theta_{t_i | \mathbf{pa}_{t_i}}
\end{aligned}$$

$\mathbb{EP}_1(t_i, y)$ can be written as

$$\mathbb{EP}_1(t_i, y) = \mathbb{EP}'(t_i, y) + \mathbb{EP}_3(y)$$

where $\mathbb{EP}_3(y)$ is

$$\mathbb{EP}_3(y) = \sum_{k \notin bb(t_i, y), T_i \notin body(k)} \sum_{ch_k} Q(ch_k | y) \sum_{t \in body(k), t \neq t_i} \prod_{t_j \in body(k), t_j \neq t_i} Q(t_j | y) \log \theta_{HD_{r(i)}=ch_k | \mathbf{pa}_{ch_k}}$$

and it does not depend on t_i and

$$\mathbb{EP}'(t_i, y) =$$

$$\begin{aligned}
& \sum_{k \in bb(t_i, y)} \sum_{ch_k} Q(ch_k|y) \sum_{t \in body(k), t \neq t_i} \prod_{t_j \in body(k), t_j \neq t_i} Q(t_j|y) \log \theta_{HD_{r(k)}=ch_k | \mathbf{pa}_{ch_k}} + \\
& \sum_{k \notin bb(t_i, y), T_i \in body(k)} \sum_{ch_k} Q(ch_k|y) \sum_{t \in body(k), t \neq t_i} \prod_{t_j \in body(k), t_j \neq t_i} Q(t_j|y) \log \theta_{HD_{r(k)}=ch_k | \mathbf{pa}_{ch_k}} + \\
& \sum_{\mathbf{pa}_{t_i}} Q(\mathbf{pa}_{t_i}|y) \log \theta_{t_i | \mathbf{pa}_{t_i}} = \\
& \sum_{k \in bb(t_i, y)} \sum_{ch_k \neq none} Q(ch_k|y) \prod_{t_j \in body(k), t_j \neq t_i} Q(t_j|y) \log \theta_{HD_{r(k)}=ch_k | true} + \\
& \sum_{k \in bb(t_i, y)} Q(Ch_k = none|y) \prod_{t_j \in body(k), t_j \neq t_i} Q(t_j|y) \delta + \\
& \sum_{k \in bb(t_i, y)} \sum_{ch_k \neq none} Q(ch_k|y) (1 - \prod_{t_j \in body(k), t_j \neq t_i} Q(t_j|y)) \delta + \\
& \sum_{k \notin bb(t_i, y), T_i \in body(k)} \sum_{ch_k} Q(ch_k|y) \sum_{t \in body(k), t \neq t_i} \prod_{t_j \in body(k), t_j \neq t_i} Q(t_j|y) \log \theta_{HD_{r(k)}=ch_k | \mathbf{pa}_{ch_k}} + \\
& \sum_{\mathbf{pa}_{t_i}, T_i \notin \mathbf{pa}_{t_i}} \prod_{ch_s \in \mathbf{pa}_{t_i}} Q(ch_s|y) 1\{t_i = true\} \delta + \\
& \sum_{\mathbf{pa}_{t_i}, T_i \in \mathbf{pa}_{t_i}} \prod_{ch_s \in \mathbf{pa}_{t_i}} Q(ch_s|y) 1\{t_i = false\} \delta = \\
& \sum_{k \in bb(t_i, y)} \sum_{ch_k \neq none} Q(ch_k|y) \prod_{t_j \in body(k), t_j \neq t_i} Q(t_j|y) \log \theta_{HD_{r(k)}=ch_k | true} + \\
& \sum_{k \in bb(t_i, y)} Q(Ch_k = none|y) \prod_{t_j \in body(k), t_j \neq t_i} Q(t_j|y) \delta + \\
& \sum_{k \in bb(t_i, y)} Q(Ch_k \neq none|y) (1 - \prod_{t_j \in body(k), t_j \neq t_i} Q(t_j|y)) \delta + \\
& \sum_{k, body(k)[y]=true, \bar{t}_i \in body(k)} Q(Ch_k \neq none|y) \delta + \\
& \sum_{k, body(k)[y]=false, t_i \in body(k)} Q(Ch_k \neq none|y) \delta + \\
& \prod_{Ch_s \in \mathbf{pa}_{t_i}} Q(Ch_s \neq t_i|y) 1\{t_i = true\} \delta + \\
& \left(1 - \prod_{Ch_s \in \mathbf{pa}_{t_i}} Q(Ch_s \neq t_i|y) \right) 1\{t_i = false\} \delta = \\
& \sum_{k \in bb(t_i, y)} \sum_{ch_k \neq none} Q(ch_k|y) \prod_{t_j \in body(k), t_j \neq t_i} Q(t_j|y) \log \theta_{HD_{r(k)}=ch_k | true} + \\
& \delta \left(\sum_{k \in bb(t_i, y)} Q(Ch_k = none|y) \prod_{t_j \in body(k), t_j \neq t_i} Q(t_j|y) + \right. \\
& \left. \sum_{k \in bb(t_i, y)} Q(Ch_k \neq none|y) \left(1 - \prod_{t_j \in body(k), t_j \neq t_i} Q(t_j|y) \right) \right) +
\end{aligned}$$

$$\begin{aligned}
& \sum_{k, \text{body}(k)[y]=\text{true}, \bar{t}_i \in \text{body}(k)} Q(\text{Ch}_k \neq \text{none}|y) + \\
& \sum_{k, \text{body}(k)[y]=\text{false}, t_i \in \text{body}(k)} Q(\text{Ch}_k \neq \text{none}|y) + \\
& \prod_{\text{Ch}_s \in \text{pa}_{t_i}} Q(\text{Ch}_s \neq t_i|y) 1\{t_i = \text{true}\} + \\
& \left(1 - \prod_{\text{Ch}_s \in \text{pa}_{t_i}} Q(\text{Ch}_s \neq t_i|y) \right) 1\{t_i = \text{false}\}
\end{aligned}$$

Let

$$\begin{aligned}
S(i, \mathbf{t}, y) = & \sum_{k \in \text{bb}(t_i, y)} Q(\text{Ch}_k = \text{none}|y) \prod_{t_j \in \text{body}(k), t_j \neq t_i} Q(t_j|y) + \\
& \sum_{k \in \text{bb}(t_i, y)} Q(\text{Ch}_k \neq \text{none}|y) \left(1 - \prod_{t_j \in \text{body}(k), t_j \neq t_i} Q(t_j|y) \right) + \\
& \sum_{k, \text{body}(k)[y]=\text{true}, \bar{t}_i \in \text{body}(k)} Q(\text{Ch}_k \neq \text{none}|y) + \\
& \sum_{k, \text{body}(k)[y]=\text{false}, t_i \in \text{body}(k)} Q(\text{Ch}_k \neq \text{none}|y)
\end{aligned}$$

then

$$\begin{aligned}
\mathbb{E}\mathbb{P}'(t_i, y) = & \sum_{k \in \text{bb}(t_i, y)} \sum_{\text{ch}_k \neq \text{none}} Q(\text{ch}_k|y) \prod_{t_j \in \text{body}(k), t_j \neq t_i} Q(t_j|y) \log \theta_{HD_{r(k)}=\text{ch}_k|\text{true}} + \\
& \delta \left(S(i, \mathbf{t}, y) + \prod_{\text{Ch}_s \in \text{pa}_{t_i}} Q(\text{Ch}_s \neq t_i|y) 1\{t_i = \text{true}\} + \right. \\
& \left. \left(1 - \prod_{\text{Ch}_s \in \text{pa}_{t_i}} Q(\text{Ch}_s \neq t_i|y) \right) 1\{t_i = \text{false}\} \right)
\end{aligned}$$

So

$$\begin{aligned}
\frac{\partial G_{t_i, y}(Q, \gamma)}{\partial \gamma} = & -\log Q(t_i) + \mathbb{E}\mathbb{P}'(t_i, y) - \mathbb{E}_{Q(t'_i|y)}[\mathbb{E}\mathbb{P}'(t'_i, y) - \log Q(t'_i)]
\end{aligned}$$

$$\mathbb{E}_{Q(\text{ch}_i|y)}[\log Q(\text{ch}_i)] = \sum_{\text{ch}_i} Q(\text{ch}_i|y) \log Q(\text{ch}_i)$$

$$\mathbb{E}_{Q(t_i|y)}[\log Q(t_i)] = \sum_{t_i} Q(t_i|y) \log Q(t_i)$$

□

4 Choice of the Step Size

Now that we have the optimization direction, we have to decide the size of the step to take.

Theorem 3 (Expression of $\mathbf{I}_Q(\mathbf{CH}, \mathbf{T}; Y)$).

$$\begin{aligned} \mathbf{I}_Q(\mathbf{CH}, \mathbf{T}; Y) = & \\ & \sum_i \sum_{ch_i} \sum_y Q(y)Q(ch_i|y)(\log Q(ch_i|y) - \log Q(ch_i)) + \\ & \sum_i \sum_{t_i} \sum_y Q(y)Q(t_i|y)(\log Q(t_i|y) - \log Q(t_i)) \end{aligned}$$

Proof.

$$\begin{aligned} \mathbf{I}_Q(\mathbf{CH}, \mathbf{T}; Y) = & \\ & - \sum_i \mathbb{E}_Q[\log Q(CH_i)] - \sum_i \mathbb{E}_Q[\log Q(T_i)] + \sum_i \mathbb{E}_Q[\log Q(CH_i|Y)] + \sum_i \mathbb{E}_Q[\log Q(T_i|Y)] = \\ & - \sum_i \sum_{ch_i, y} Q(ch_i|y)Q(y) \log Q(ch_i) - \sum_i \sum_{t_i, y} Q(t_i|y)Q(y) \log Q(t_i) + \\ & \sum_i \sum_{ch_i, y} Q(ch_i|y)Q(y) \log Q(ch_i|y) + \sum_i \sum_{t_i, y} Q(t_i|y)Q(y) \log Q(t_i|y) = \\ & \sum_i \sum_{ch_i} \sum_y Q(y)Q(ch_i|y) \log Q(ch_i|y) - Q(ch_i|y)Q(y) \log Q(ch_i) + \\ & \sum_i \sum_{t_i} \sum_y Q(y)Q(t_i|y) \log Q(t_i|y) - Q(t_i|y)Q(y) \log Q(t_i) = \\ & \sum_i \sum_{ch_i} \sum_y Q(y)Q(ch_i|y)(\log Q(ch_i|y) - \log Q(ch_i)) + \\ & \sum_i \sum_{t_i} \sum_y Q(y)Q(t_i|y)(\log Q(t_i|y) - \log Q(t_i)) \end{aligned}$$

□

Theorem 4 (Derivatives of $\mathbf{I}_Q(\mathbf{CH}, \mathbf{T}; Y)$ with respect to $Q(ch_{i0}|y_0)$ and $Q(t_{i0}|y_0)$).

$$\frac{\partial \mathbf{I}_Q(\mathbf{CH}, \mathbf{T}; Y)}{\partial Q(ch_{i0}|y_0)} = Q(y_0)(\log Q(ch_{i0}|y_0) - \log Q(ch_{i0}))$$

$$\frac{\partial \mathbf{I}_Q(\mathbf{CH}, \mathbf{T}; Y)}{\partial Q(t_{i0}|y_0)} = Q(y_0)(\log Q(t_{i0}|y_0) - \log Q(t_{i0}))$$

Proof.

$$\begin{aligned} \frac{\partial \mathbf{I}_Q(\mathbf{CH}, \mathbf{T}; Y)}{\partial Q(ch_{i0}|y_0)} = & - \frac{\partial \mathbb{E}_Q[\log Q(CH_i)]}{\partial Q(ch_{i0}|y_0)} + \frac{\partial \mathbb{E}_Q[\log Q(CH_i|y_0)]}{\partial Q(ch_{i0}|y_0)} = \\ & -Q(y_0)(\log Q(ch_{i0}) + 1) + Q(y_0)(\log Q(ch_{i0}|y_0) + 1) = \\ & Q(y_0)(-\log Q(ch_{i0}) - 1 + \log Q(ch_{i0}|y_0) + 1) = \\ & Q(y_0)(\log Q(ch_{i0}|y_0) - \log Q(ch_{i0})) \end{aligned}$$

$$\begin{aligned}
\frac{\partial \mathbf{I}_Q(\mathbf{CH}, \mathbf{T}; Y)}{\partial Q(t_{i0}|y_0)} &= -\frac{\partial \mathbb{E}_Q[\log Q(T_i)]}{\partial Q(t_{i0}|y_0)} + \frac{\partial \mathbb{E}_Q[\log Q(T_i|y_0)]}{\partial Q(t_{i0}|y_0)} = \\
&= -Q(y_0)(\log Q(t_{i0}) + 1) + Q(y_0)(\log Q(t_{i0}|y_0) + 1) = \\
&= Q(y_0)(-\log Q(t_{i0}) - 1 + \log Q(t_{i0}|y_0) + 1) = \\
&= Q(y_0)(\log Q(t_{i0}|y_0) - \log Q(t_{i0}))
\end{aligned}$$

□

References

1. K. R. Apt and M. Bezem. Acyclic programs. *New Generation Computing*, 9(3/4):335–364, 1991.
2. G. Elidan and N. Friedman. Learning hidden variable networks: The information bottleneck approach. *Journal of Machine Learning Research*, 6:81–127, 2005.