5.2.5 に関連して、元論文 supplementary material の 1 を見てみる。

本の (5.72),(5.73) にて、ReLU の出力 z の平均と分散を求めているが、これで出力をガウス分布で近似した場合、期待値伝播法でガウス分布で近似された w の平均、分散は本の (5.59),(5.60) によって更新され、その際の Z は (5.65) にて求まることがわかっている。特に説明がなかったように思うが、元論文の supplementary material の 1 は (5.59),(5.60) に関して、$\frac{\partial ln\, Z_0}{\partial m_{i,j}^{(l)}}, \frac{\partial ln\, Z_0}{\partial v_{i,j}^{(l)}}$ をどう求めるかの説明になっている。

公式の実装の Theano 版では、network.py の generate_updates() にて、Theano の自動微分の機能を使って、求めている。(詳しくは見ていないが、)c 版では network.c の backward_PBP() にて、supplementary material の 1 の内容を使って、$\frac{\partial ln\, Z_0}{\partial m_{i,j}^{(l)}}, \frac{\partial ln\, Z_0}{\partial v_{i,j}^{(l)}}$ を計算し、更新を行っている。

(公式の実装では c 版のほうが 4-20 倍速いと言っているが、python と c の違いだけでなくて、微分の計算方法の違いも大きいように思う。)

まず、(5.68),(5.69) を振り返る。

$$\mathbb{E}[x_{mix}] = \sum_{k=1}^{K} \pi_k \mu_k \tag{1}$$

$$\mathbb{V}[x_{mix}] = \sum_{k=1}^{K} (\pi_k(\mu_k^2 + v_k)) - \mathbb{E}[x_{mix}]^2 \tag{2}$$

(2) は本に誤りがあるように思う。

切断分布について考える。

$$\pi_{low}{}_i^{(l)} = \Phi(\overline{\mu_i^{(l)}}) \tag{3}$$

$$\pi_{high}{}_i^{(l)} = \Phi(-\overline{\mu_i^{(l)}}) \tag{4}$$

ここで、

$$\overline{\mu_i^{(l)}} = -\frac{m_{a_i^{(l)}}}{\sqrt{v_{a_i^{(l)}}}} \tag{5}$$

となる。この部分は (5.70),(5.71) に対応している。

$$\mu_{low}{}_i^{(l)} = \mu_i^{(l)} - \sqrt{v_i^{(l)}} \frac{\phi(\overline{\mu_i^{(l)}})}{\Phi(\overline{\mu_i^{(l)}})} \tag{6}$$

$$\mu_{high}{}_i^{(l)} = \mu_i^{(l)} + \sqrt{v_i^{(l)}} \frac{\phi(\overline{\mu_i^{(l)}})}{\Phi(-\overline{\mu_i^{(l)}})} \tag{7}$$

(7) は (5.72) に対応しており、(6) は (5.72) の導出と同様、もしくは、(1) を用いて、導出できる。

$$v_{low}{}_i^{(l)} = v_i^{(l)} \left(1 - \overline{\mu_i^{(l)}} \frac{\phi(\overline{\mu_i^{(l)}})}{\Phi(\overline{\mu_i^{(l)}})} - \left(\frac{\phi(\overline{\mu_i^{(l)}})}{\Phi(\overline{\mu_i^{(l)}})} \right)^2 \right) \tag{8}$$

$$v_{high}{}_i^{(l)} = v_i^{(l)} \left(1 + \overline{\mu_i^{(l)}} \frac{\phi(\overline{\mu_i^{(l)}})}{\Phi(-\overline{\mu_i^{(l)}})} - \left(\frac{\phi(\overline{\mu_i^{(l)}})}{\Phi(-\overline{\mu_i^{(l)}})} \right)^2 \right) \tag{9}$$

(9) は (5.73) に対応しており、(8) は (5.73) の導出と同様、もしくは、(2) を用いて、導出できる。なお、(5.73) は一部誤りがあるように思う。

活性化関数 ReLU を通したあと、0 より小さい部分は 0 になるため、${m_{low}}_i^{(l)} = 0$,${v_{low}}_i^{(l)} = 0$ となる。

その時に、(1),(2) を利用して、$m_{z_i^{(l)}}, v_{z_i^{(l)}}$ を計算する。

$$m_{z_i^{(l)}} = \Phi(\overline{\mu_i^{(l)}})\times 0 + \Phi(-\overline{\mu_i^{(l)}})\left(m_{a_i^{(l)}} + \sqrt{v_{a_i^{(l)}}}\frac{\mathcal{N}(\overline{\mu_i^{(l)}}|0,1)}{\Phi(-\overline{\mu_i^{(l)}})}\right) = \Phi(-\overline{\mu_i^{(l)}})(m_{a_i^{(l)}} + \sqrt{v_{a_i^{(l)}}}\gamma_i^{(l)}) = \Phi(-\overline{\mu_i^{(l)}}){\mu_{high}}_i^{(l)} \tag{10}$$

$$\begin{aligned} v_{z_i^{(l)}} &= \Phi(-\overline{\mu_i^{(l)}})({\mu_{high}}_i^{(l)2} + v_{a_i^{(l)}}(1 + \overline{\mu_i^{(l)}}\gamma_i^{(l)} - (\gamma_i^{(l)})^2)) - (\Phi(-\overline{\mu_i^{(l)}}){\mu_{high}}_i^{(l)})^2 \\ &= \Phi(-\overline{\mu_i^{(l)}}){\mu_{high}}_i^{(l)2}(1 - \Phi(-\overline{\mu_i^{(l)}})) + \Phi(-\overline{\mu_i^{(l)}})v_{a_i^{(l)}}(1 + \overline{\mu_i^{(l)}}\gamma_i^{(l)} - (\gamma_i^{(l)})^2)) \\ &= m_{z_i^{(l)}}(m_{a_i^{(l)}} + \sqrt{v_{a_i^{(l)}}}\gamma_i^{(l)})\Phi(\overline{\mu_i^{(l)}}) + \Phi(-\overline{\mu_i^{(l)}})v_{a_i^{(l)}}(1 + \overline{\mu_i^{(l)}}\gamma_i^{(l)} - (\gamma_i^{(l)})^2)) \end{aligned} \tag{11}$$

ここで、

$$\gamma_i^{(l)} = \frac{\mathcal{N}(\overline{\mu_i^{(l)}}|0,1)}{\Phi(-\overline{\mu_i^{(l)}})} \tag{12}$$

元論文の supplementary material の 1 の (2) と比較して、$\overline{\mu_i^{(l)}}$ の符号が逆になっている。たちまち、定義が逆になっているので、値は等しくなる。

次に、元論文の supplementary material の 1 の (3),(4) を評価する。

そもそも、元論文 2 ページ目にあるように、$a^{(l)} = \frac{W^{(l)}z^{(l-1)}}{\sqrt{V_{l-1}}}$ で $\sqrt{V_{l-1}} = \sqrt{|I(l-1)|}$ となるので、(3) は W,Z が独立として、

$$\begin{aligned} m_{a_i^{(l)}} &= \int p(a_i^{(l)})a_i^{(l)}da_i^{(l)} = \frac{1}{\sqrt{|I_i|}}\int\int p(W^{(l)})p(z^{(l-1)})W^{(l)}z^{(l-1)}dWdz \\ &= \frac{1}{\sqrt{|I_i|}}\int\int\prod_j p(w_{i,j}^{(l)})\prod_j p(z_j^{(l-1)})\sum_j w_{i,j}^{(l)}z_j^{(l-1)}\prod_j dw_{i,j}^{(l)}\prod_j dz_j^{(l-1)} \\ &= \frac{1}{\sqrt{|I_i|}}\sum_j\left(\int p(w_{i,j}^{(l)})w_{i,j}^{(l)}dw_{i,j}^{(l)}\int p(z_j^{(l-1)})z_j^{(l-1)}dz_j^{(l-1)}\right) = \frac{1}{\sqrt{|I_i|}}\sum_j m_{w_{i,j}^{(l)}}m_{z_j^{(l-1)}} \end{aligned} \tag{13}$$

(確率変数の積の平均値を求めている。ここでは一種の変数変換を行っている。)

分散について、考慮する。

$$
\begin{aligned}
v_{a_i^{(l)}} &= \int p(a_i^{(l)})(a_i^{(l)} - m_{v_{a_i^{(l)}}})^2 da_i^{(l)} = \frac{1}{|I_i|}\int\int p(W_i^{(l)})p(z^{(l-1)})(W_i^{(l)} z^{(l-1)} - m_{W_i^{(l)}} m_{z^{(l-1)}})^2 dW dz \\
&= \frac{1}{|I_i|}\int\int \prod_j p(w_{i,j}^{(l)}) \prod_j p(z_j^{(l-1)}) (\sum_j (w_{i,j}^{(l)} z_j^{(l-1)} - m_{w_{i,j}^{(l)}} m_{z_j^{(l-1)}}))^2 \prod_j dw_{i,j}^{(l)} \prod_j dz_j^{(l-1)} \\
&= \frac{1}{|I_i|}(\sum_j \int\int p(w_{i,j}^{(l)})p(z_j^{(l-1)})(w_{i,j}^{(l)} z_j^{(l-1)} - m_{w_{i,j}^{(l)}} m_{z_j^{(l-1)}})^2 dw_{i,j}^{(l)} dz_j^{(l-1)} + \\
&\sum_j \sum_{k\neq j} \int\int\int\int p(w_{i,j}^{(l)})p(z_j^{(l-1)})p(w_{i,k}^{(l)})p(z_k^{(l-1)})(w_{i,j}^{(l)} z_j^{(l-1)} - m_{w_{i,j}^{(l)}} m_{z_j^{(l-1)}}) \\
&(w_{i,k}^{(l)} z_k^{(l-1)} - m_{w_{i,k}^{(l)}} m_{z_k^{(l-1)}}) dw_{i,j}^{(l)} dz_j^{(l-1)} dw_{i,k}^{(l)} dz_k^{(l-1)}) \\
&= \frac{1}{|I_i|}(\sum_j \left(\int p(w_{i,j}^{(l)}) {w_{i,j}^{(l)}}^2 dw_{i,j}^{(l)} \int p(z_j^{(l-1)}) {z_j^{(l-1)}}^2 dz_j^{(l-1)} - {m_{w_{i,j}^{(l)}}}^2 {m_{z_j^{(l-1)}}}^2\right) + \\
&\sum_j \sum_{k\neq j} (m_{w_{i,j}^{(l)}} m_{z_j^{(l-1)}} m_{w_{i,k}^{(l)}} m_{z_k^{(l-1)}} - m_{w_{i,j}^{(l)}} m_{z_j^{(l-1)}} m_{w_{i,k}^{(l)}} m_{z_k^{(l-1)}} - m_{w_{i,j}^{(l)}} m_{z_j^{(l-1)}} m_{w_{i,k}^{(l)}} m_{z_k^{(l-1)}} + m_{w_{i,j}^{(l)}} m_{z_j^{(l-1)}} m_{w_{i,k}^{(l)}} m_{z_k^{(l-1)}})) \\
&= \frac{1}{|I_i|}(\sum_j (({m_{w_{i,j}^{(l)}}}^2 + v_{w_{i,j}^{(l)}})({m_{z_j^{(l-1)}}}^2 + v_{z_j^{(l-1)}}) - {m_{w_{i,j}^{(l)}}}^2 {m_{z_j^{(l-1)}}}^2) + 0) \\
&= \frac{1}{|I_i|}\sum_j ({m_{w_{i,j}^{(l)}}}^2 v_{z_j^{(l-1)}} + {m_{z_j^{(l-1)}}}^2 v_{w_{i,j}^{(l)}} + v_{z_j^{(l-1)}} v_{w_{i,j}^{(l)}})
\end{aligned}
\tag{14}
$$

これにより、元論文 supplementary material(4) がわかる。

(13),(14) に関して、微分すると元論文 supplementary material の (5)-(8) になる。

$$\frac{\partial m_{a_i^{(l)}}}{\partial m_{z_j^{(l)}}} = \frac{m_{w_{i,j}^{(l)}}}{\sqrt{|I_i|}} \tag{15}$$

$$\frac{\partial m_{a_i^{(l)}}}{\partial v_{z_j^{(l)}}} = 0 \tag{16}$$

$$\frac{\partial v_{a_i^{(l)}}}{\partial m_{z_j^{(l)}}} = \frac{2 m_{z_j^{(l)}} v_{w_{i,j}^{(l)}}}{|I_i|} \tag{17}$$

$$\frac{\partial v_{a_i^{(l)}}}{\partial v_{z_j^{(l)}}} = \frac{{m_{w_{i,j}^{(l)}}}^2 + v_{w_{i,j}^{(l)}}}{|I_i|} \tag{18}$$

$$\frac{\partial m_{a_i^{(l)}}}{\partial m_{w_{i,j}^{(l)}}} = \frac{m_{z_j^{(l)}}}{\sqrt{|I_i|}} \tag{19}$$

$$\frac{\partial m_{a_i^{(l)}}}{\partial v_{w_{i,j}^{(l)}}} = 0 \tag{20}$$

$$\frac{\partial v_{a_i^{(l)}}}{\partial m_{w_{i,j}^{(l)}}} = \frac{2 m_{w_{i,j}^{(l)}} v_{z_j^{(l)}}}{|I_i|} \tag{21}$$

$$\frac{\partial v_{a_i^{(l)}}}{\partial v_{w_{i,j}^{(l)}}} = \frac{{m_{z_j^{(l)}}}^2 + v_{z_j^{(l)}}}{|I_i|} \tag{22}$$

(5) を微分して、元論文 supplementary material(9) を求める。

$$\frac{\partial \overline{\mu_i^{(l)}}}{\partial m_{a_i^{(l)}}} = \frac{1}{\sqrt{|I_i|}} \tag{23}$$

$$\frac{\partial \overline{\mu_i^{(l)}}}{\partial m_{v_i^{(l)}}} = \frac{m_{a_i^{(l)}}}{2{v_i^{(l)}}^{\frac{3}{2}}} = \frac{m_{a_i^{(l)}}}{2v_i^{(l)}\sqrt{v_i^{(l)}}} \tag{24}$$

(24) は $\overline{\mu_i^{(l)}}$ の定義のようにしないと成立しない気がする。

(12) を微分して、元論文 supplementary material(10) を求める。その際、以下に注意する。

$$\frac{\partial \Phi(-\overline{\mu_i^{(l)}})}{\partial \overline{\mu_i^{(l)}}} = -\mathcal{N}(-\overline{\mu_i^{(l)}}|0,1) = -\mathcal{N}(\overline{\mu_i^{(l)}}|0,1) \tag{25}$$

$$\frac{\partial \mathcal{N}(\overline{\mu_i^{(l)}}|0,1)}{\partial \overline{\mu_i^{(l)}}} = \frac{\partial \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{\overline{\mu_i^{(l)}}^2}{2})}{\partial \overline{\mu_i^{(l)}}} = -\overline{\mu_i^{(l)}} \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{\overline{\mu_i^{(l)}}^2}{2}) = -\overline{\mu_i^{(l)}} \mathcal{N}(\overline{\mu_i^{(l)}}|0,1) \tag{26}$$

$$\begin{aligned} \frac{\partial \gamma_i^{(l)}}{\partial \overline{\mu_i^{(l)}}} &= \frac{\partial \mathcal{N}(\overline{\mu_i^{(l)}}|0,1)}{\partial \overline{\mu_i^{(l)}}} \frac{1}{\Phi(-\overline{\mu_i^{(l)}})} - \mathcal{N}(\overline{\mu_i^{(l)}}|0,1) \frac{1}{{\Phi(-\overline{\mu_i^{(l)}})}^2} \frac{\partial \Phi(-\overline{\mu_i^{(l)}})}{\partial \overline{\mu_i^{(l)}}} \\ &= -\overline{\mu_i^{(l)}} \frac{\mathcal{N}(\overline{\mu_i^{(l)}}|0,1)}{\Phi(-\overline{\mu_i^{(l)}})} + \frac{{\mathcal{N}(\overline{\mu_i^{(l)}}|0,1)}^2}{{\Phi(-\overline{\mu_i^{(l)}})}^2} = -\overline{\mu_i^{(l)}} \gamma_i^{(l)} + {\gamma_i^{(l)}}^2 \end{aligned} \tag{27}$$

$$\frac{\partial \gamma_i^{(l)}}{\partial m_{a_i^{(l)}}} = \frac{\partial \gamma_i^{(l)}}{\partial \overline{\mu_i^{(l)}}} \frac{\partial \overline{\mu_i^{(l)}}}{\partial m_{a_i^{(l)}}} = -(\overline{\mu_i^{(l)}} \gamma_i^{(l)} - {\gamma_i^{(l)}}^2) \frac{\partial \overline{\mu_i^{(l)}}}{\partial m_{a_i^{(l)}}} \tag{28}$$

$$\frac{\partial \gamma_i^{(l)}}{\partial v_{a_i^{(l)}}} = \frac{\partial \gamma_i^{(l)}}{\partial \overline{\mu_i^{(l)}}} \frac{\partial \overline{\mu_i^{(l)}}}{\partial v_{a_i^{(l)}}} = -(\overline{\mu_i^{(l)}} \gamma_i^{(l)} - {\gamma_i^{(l)}}^2) \frac{\partial \overline{\mu_i^{(l)}}}{\partial v_{a_i^{(l)}}} \tag{29}$$

元論文 supplementary material と符号が異なるところがある。。。

元論文 supplementary material(11),(12) は (10) を微分する。

$$\begin{aligned} \frac{\partial m_{z_i^{(l)}}}{\partial m_{a_i^{(l)}}} &= \frac{\partial}{\partial m_{a_i^{(l)}}} \Phi(-\overline{\mu_i^{(l)}})(m_{a_i^{(l)}} + \sqrt{v_{a_i^{(l)}}}\gamma_i^{(l)}) = \frac{\partial \Phi(-\overline{\mu_i^{(l)}})}{\partial m_{a_i^{(l)}}}(m_{a_i^{(l)}} + \sqrt{v_{a_i^{(l)}}}\gamma_i^{(l)}) + \Phi(-\overline{\mu_i^{(l)}}) \frac{\partial}{\partial m_{a_i^{(l)}}}(m_{a_i^{(l)}} + \sqrt{v_{a_i^{(l)}}}\gamma_i^{(l)}) \\ &= -\mathcal{N}(\overline{\mu_i^{(l)}}|0,1) \frac{\partial \overline{\mu_i^{(l)}}}{\partial m_{a_i^{(l)}}}(m_{a_i^{(l)}} + \sqrt{v_{a_i^{(l)}}}\gamma_i^{(l)}) + \Phi(-\overline{\mu_i^{(l)}})(1 + \sqrt{v_{a_i^{(l)}}} \frac{\partial \gamma_i^{(l)}}{\partial m_{a_i^{(l)}}}) \end{aligned} \tag{30}$$

$$\begin{aligned} \frac{\partial m_{z_i^{(l)}}}{\partial v_{a_i^{(l)}}} &= \frac{\partial}{\partial v_{a_i^{(l)}}} \Phi(-\overline{\mu_i^{(l)}})(m_{a_i^{(l)}} + \sqrt{v_{a_i^{(l)}}}\gamma_i^{(l)}) = \frac{\partial \Phi(-\overline{\mu_i^{(l)}})}{\partial v_{a_i^{(l)}}}(m_{a_i^{(l)}} + \sqrt{v_{a_i^{(l)}}}\gamma_i^{(l)}) + \Phi(-\overline{\mu_i^{(l)}}) \frac{\partial}{\partial v_{a_i^{(l)}}}(m_{a_i^{(l)}} + \sqrt{v_{a_i^{(l)}}}\gamma_i^{(l)}) \\ &= -\mathcal{N}(\overline{\mu_i^{(l)}}|0,1) \frac{\partial \overline{\mu_i^{(l)}}}{\partial v_{a_i^{(l)}}}(m_{a_i^{(l)}} + \sqrt{v_{a_i^{(l)}}}\gamma_i^{(l)}) + \Phi(-\overline{\mu_i^{(l)}})(\frac{\gamma_i^{(l)}}{2\sqrt{v_{a_i^{(l)}}}} + \sqrt{v_{a_i^{(l)}}} \frac{\partial \gamma_i^{(l)}}{\partial v_{a_i^{(l)}}}) \end{aligned} \tag{31}$$

同様に、元論文 supplementary material(13),(14) は (11) を微分する。

$$\frac{\partial v_{z_i^{(l)}}}{\partial m_{a_i^{(l)}}} = \frac{\partial}{\partial m_{a_i^{(l)}}}(m_{z_i^{(l)}}(m_{a_i^{(l)}} + \sqrt{v_{a_i^{(l)}}}\gamma_i^{(l)})\Phi(\overline{\mu_i^{(l)}}) + \Phi(-\overline{\mu_i^{(l)}})v_{a_i^{(l)}}(1+\overline{\mu_i^{(l)}}\gamma_i^{(l)} - (\gamma_i^{(l)})^2))$$

$$= \frac{\partial m_{z_i^{(l)}}}{\partial m_{a_i^{(l)}}}(m_{a_i^{(l)}} + \sqrt{v_{a_i^{(l)}}}\gamma_i^{(l)})\Phi(\overline{\mu_i^{(l)}}) + m_{z_i^{(l)}}\left((1+\sqrt{v_{a_i^{(l)}}}\frac{\partial \gamma_i^{(l)}}{\partial m_{a_i^{(l)}}})\Phi(\overline{\mu_i^{(l)}}) + (m_{a_i^{(l)}} + \sqrt{v_{a_i^{(l)}}}\gamma_i^{(l)})\mathcal{N}(\overline{\mu_i^{(l)}}|0,1)\frac{\partial \overline{\mu_i^{(l)}}}{\partial m_{v_i^{(l)}}}\right)$$
$$-\mathcal{N}(\overline{\mu_i^{(l)}}|0,1)\frac{\partial \overline{\mu_i^{(l)}}}{\partial m_{v_i^{(l)}}}v_{a_i^{(l)}}(1+\overline{\mu_i^{(l)}}\gamma_i^{(l)} - (\gamma_i^{(l)})^2)) + \Phi(-\overline{\mu_i^{(l)}})v_{a_i^{(l)}}\left(\frac{\partial \overline{\mu_i^{(l)}}}{v_{a_i^{(l)}}}\gamma_i^{(l)} + \overline{\mu_i^{(l)}}\frac{\partial \gamma_i^{(l)}}{\partial v_{a_i^{(l)}}} - 2\gamma_i^{(l)}\frac{\partial \gamma_i^{(l)}}{\partial v_{a_i^{(l)}}})\right) \tag{32}$$

$$\frac{\partial v_{z_i^{(l)}}}{\partial v_{a_i^{(l)}}} = \frac{\partial}{\partial v_{a_i^{(l)}}}(m_{z_i^{(l)}}(m_{a_i^{(l)}} + \sqrt{v_{a_i^{(l)}}}\gamma_i^{(l)})\Phi(\overline{\mu_i^{(l)}}) + \Phi(-\overline{\mu_i^{(l)}})v_{a_i^{(l)}}(1+\overline{\mu_i^{(l)}}\gamma_i^{(l)} - (\gamma_i^{(l)})^2))$$

$$= \frac{\partial m_{z_i^{(l)}}}{\partial v_{a_i^{(l)}}}(m_{a_i^{(l)}} + \sqrt{v_{a_i^{(l)}}}\gamma_i^{(l)})\Phi(\overline{\mu_i^{(l)}}) + m_{z_i^{(l)}}\left((\frac{\gamma_i^{(l)}}{2\sqrt{v_{a_i^{(l)}}}} + \sqrt{v_{a_i^{(l)}}}\frac{\partial \gamma_i^{(l)}}{\partial v_{a_i^{(l)}}})\Phi(\overline{\mu_i^{(l)}}) + (m_{a_i^{(l)}} + \sqrt{v_{a_i^{(l)}}}\gamma_i^{(l)})\mathcal{N}(\overline{\mu_i^{(l)}}|0,1)\frac{\partial \overline{\mu_i^{(l)}}}{\partial v_{v_i^{(l)}}}\right)$$
$$-\mathcal{N}(\overline{\mu_i^{(l)}}|0,1)\frac{\partial \overline{\mu_i^{(l)}}}{\partial v_{a_i}^{(l)}}v_{a_i^{(l)}}(1+\overline{\mu_i^{(l)}}\gamma_i^{(l)} - (\gamma_i^{(l)})^2))+$$
$$\Phi(-\overline{\mu_i^{(l)}})\left((1+\overline{\mu_i^{(l)}}\gamma_i^{(l)} - (\gamma_i^{(l)})^2) + v_{a_i^{(l)}}(\frac{\partial \overline{\mu_i^{(l)}}}{v_{a_i^{(l)}}}\gamma_i^{(l)} + \overline{\mu_i^{(l)}}\frac{\partial \gamma_i^{(l)}}{\partial v_{a_i^{(l)}}} - 2\gamma_i^{(l)}\frac{\partial \gamma_i^{(l)}}{\partial v_{a_i^{(l)}}})\right) \tag{33}$$

最初の定義のところの符号が異なるので、ところどころ、符号が異なる。

元論文 supplementary material(15),(16) は定義になっている。

連鎖率を利用すると、一般に、元論文 supplementary material(17),(18) が成り立つ。

同様に、連鎖率を利用すると、一般に、元論文 supplementary material(19),(20),(21),(22) が成り立つ。

求めたいもの元論文 supplementary material(21),(22) であった。

(21),(22) の右辺を計算するには、元論文 supplementary material(7),(8),(15)-(18) を利用すれば良い。(7),(8) は更新前の値を利用すれば良い。出力層の方は (15),(16) を利用すれば良いが、それ以外の層では (17),(18) を利用する。出力層については (5.65) を利用することで、(15),(16) が検討でき、(15)-(18) に関して、(19),(20) も利用できる。(19),(20) に関して、(5),(6),(11)-(14) で計算できる。(9),(10) も利用すれば、更新前の値を利用すれば計算できる。