5.2.5.4 について確認してみる。

この部分については q と $p(y_n|x_n,W)=p(y_n|f(x_n;W))=p(y_n|x_n,W,\gamma_w,\gamma_w)$ の関係が、(4.55) の関係のようになっており、理解しやすい。つまり、

$$q_{new}(W,\gamma_w,\gamma_w)=q(W,\gamma_w,\gamma_w|y_i,...,y_1(,x_i,...,x_1))=\frac{1}{Z}p(y_n|f(x_i;W))q(W,\gamma_w,\gamma_w|y_{i-1},...,y_1(,x_{i-1},...,x_1)) \tag{1}$$

Z は正規化定数なので、(5.65) のように求められる。

$$\begin{aligned}Z(\alpha_{\gamma_y},\beta_{\gamma_y})&=\int\mathcal{N}(y_i|f(x_n;W),{\gamma_y}^{-1})q(W,\gamma_y,\gamma_w)dWd\gamma_yd\gamma_w\\ &=\int\int\int\mathcal{N}(y_i|f(x_n;W),{\gamma_y}^{-1})Gam(\gamma_y|\alpha_{\gamma_y},\beta_{\gamma_y})Gam(\gamma_w|\alpha_{\gamma_w},\beta_{\gamma_w})\prod_l\prod_i\prod_j\mathcal{N}({w_{i,j}}^{(l)}|{m_{i,j}}^{(l)},{v_{i,j}}^{(l)})dWd\gamma_yd\gamma_w\\ &=\int\int\mathcal{N}(y_i|f(x_n;W),{\gamma_y}^{-1})Gam(\gamma_y|\alpha_{\gamma_y},\beta_{\gamma_y})\int Gam(\gamma_w|\alpha_{\gamma_w},\beta_{\gamma_w})d\gamma_w\prod_l\prod_i\prod_j\mathcal{N}({w_{i,j}}^{(l)}|{m_{i,j}}^{(l)},{v_{i,j}}^{(l)})dWd\gamma_y\\ &=\int\int\mathcal{N}(y_i|f(x_n;W),{\gamma_y}^{-1})Gam(\gamma_y|\alpha_{\gamma_y},\beta_{\gamma_y})\prod_l\prod_i\prod_j\mathcal{N}({w_{i,j}}^{(l)}|{m_{i,j}}^{(l)},{v_{i,j}}^{(l)})dWd\gamma_y\end{aligned} \tag{2}$$

この式をもとに戻すと

$$\begin{aligned}Z(\alpha_{\gamma_y},\beta_{\gamma_y})&=\int\int\mathcal{N}(y_i|f(x_n;W),{\gamma_y}^{-1})Gam(\gamma_y|\alpha_{\gamma_y},\beta_{\gamma_y})\prod_l\prod_i\prod_j\mathcal{N}({w_{i,j}}^{(l)}|{m_{i,j}}^{(l)},{v_{i,j}}^{(l)})dWd\gamma_y\\ &=\int\int p(y_i|f(x_n;W),{\gamma_y}^{-1})q(\gamma_y)\prod_l\prod_i\prod_j q({w_{i,j}}^{(l)})dWd\gamma_y=\int\int p(y_i|f(x_n;W),{\gamma_y}^{-1})q(\gamma_y)q(W)dWd\gamma_y\end{aligned} \tag{3}$$

ここで、(5.48) 式を求めたときのように変数変換を行う。$z^L=f(x_i;W)$ として、

$$\begin{aligned}Z(\alpha_{\gamma_y},\beta_{\gamma_y})&=\int\int p(y_i|f(x_n;W),{\gamma_y}^{-1})q(\gamma_y)q(W)dWd\gamma_y\\ &=\int\int p(y_i|f(x_n;W),{\gamma_y}^{-1})q(\gamma_y)q(f(x_i;W)|x_i)\frac{\partial f}{\partial W}dWd\gamma_y=\int\int p(y_i|z^L,{\gamma_y}^{-1})q(\gamma_y)q(z^L|x_i)dz^Ld\gamma_y\end{aligned} \tag{4}$$

ここで、$q(z^L|x_i)=\mathcal{N}(z^L|m_{z^L},v_{z^L})$ と近似すると、(3.61)-(3.63) を考慮すると、

$$\begin{aligned}Z(\alpha_{\gamma_y},\beta_{\gamma_y})&=\int\int p(y_i|z^L,{\gamma_y}^{-1})q(\gamma_y)q(z^L|x_i)dz^Ld\gamma_y\approx\int\int p(y_i|z^L,{\gamma_y}^{-1})Gam(\gamma_y|\alpha_{\gamma_y},\beta_{\gamma_y})\mathcal{N}(z^L|m_{z^L},v_{z^L})dz^Ld\gamma_y\\ &=\int(\int p(y_i|z^L,{\gamma_y}^{-1})Gam(\gamma_y|\alpha_{\gamma_y},\beta_{\gamma_y})d\gamma_y)\mathcal{N}(z^L|m_{z^L},v_{z^L})dz^L=\int St(y_i|z^L,\frac{\alpha_{\gamma_y}}{\beta_{\gamma_y}},2\alpha_{\gamma_y})\mathcal{N}(z^L|m_{z^L},v_{z^L})dz^L\end{aligned} \tag{5}$$

このとき、t 分布の平均と分散は (3.64)(3.65) より、$\mathcal{E}=z^L,\mathcal{V}=\frac{2\alpha_{\gamma_y}}{\frac{\alpha_{\gamma_y}}{\beta_{\gamma_y}}(2\alpha_{\gamma_y}-2)}=\frac{\beta_{\gamma_y}}{(\alpha_{\gamma_y}-1)}$ となり、t 分布を平均と分散が一致する正規分布で近似すると以下のようになる。

$$\begin{aligned}Z(\alpha_{\gamma_y},\beta_{\gamma_y})&=\int St(y_i|z^L,\frac{\alpha_{\gamma_y}}{\beta_{\gamma_y}},2\alpha_{\gamma_y})\mathcal{N}(z^L|m_{z^L},v_{z^L})dz^L\\ &\approx\int\mathcal{N}(y_i|z^L,\frac{\beta_{\gamma_y}}{(\alpha_{\gamma_y}-1)})\mathcal{N}(z^L|m_{z^L},v_{z^L})dz^L=\mathcal{N}(y_i|m_{z^L},\frac{\beta_{\gamma_y}}{(\alpha_{\gamma_y}-1)}+v_{z^L})\end{aligned} \tag{6}$$

最後は、(A.24) を利用している。(ここは本の記述が間違っているように思う。[47] の (12) では上記の式のようになっている。)

本にあるように、m_{z^L}, v_{z^L} を前向き伝播によって近似的に求める。ベクトル $z^{(L)}$ の要素は対角ガウス分布に従っていると仮定するため、v_{z^L} は対角成分のみ 0 でない。よって、m_{z^L} はベクトルとみなすこともできる。また、各要素は共分散が 0 になり、独立になる。

活性について、考えると、本にあるように $a^{(l)} = \frac{w^{(l)}z^{(l-1)}}{\sqrt{H_{l-1}}}$ となる。なお、[47] の 2 章では $\sqrt{H_{l-1}+1}$ となっているが、P.23 の*3 のように定数項の扱いの違いと考えられる。$\sqrt{H_{l-1}}$ によって、入力の個数で活性の分散が変化しないようにしていると考えられる。

さて、活性の期待値を考える。

$$
\begin{aligned}
\mathbb{E}_{p(z^{(l-1)},W^{(l)})=p(z^{(l-1)})p(W^{(l)})}(a^{(l)}) = \int\int a^{(l)}p(z^{(l-1)})p(W^{(l)})dz^{(l-1)}dW^{(l)} \\
= \int\int \frac{W^{(l)}z^{(l-1)}}{\sqrt{H_{l-1}}}p(z^{(l-1)})p(W^{(l)})dz^{(l-1)}dW^{(l)} = \frac{1}{\sqrt{H_{l-1}}}\int W^{(l)}(\int z^{(l-1)}p(z^{(l-1)})dz^{(l-1)})p(W^{(l)})dW^{(l)} \\
= \frac{1}{\sqrt{H_{l-1}}}\int W^{(l)}p(W^{(l)})dW^{(l)}m_{z^{(l-1)}} = \frac{M^{(l)}m_{z^{(l-1)}}}{\sqrt{H_{l-1}}}
\end{aligned}
\tag{7}
$$

なお、${z^{(l-1)}}_i$ は上記のように独立なので、

$$
\begin{aligned}
\int z^{(l-1)}p(z^{(l-1)})dz^{(l-1)} = \int({z^{(l-1)}}_1, {z^{(l-1)}}_2, \cdots, {z^{(l-1)}}_n)\prod_i p(z_i^{(l-1)})\prod_i dz_i^{(l-1)} \\
= (\int {z^{(l-1)}}_1\prod_i p(z_i^{(l-1)})\prod_i dz_i^{(l-1)}, \int {z^{(l-1)}}_2\prod_i p(z_i^{(l-1)})\prod_i dz_i^{(l-1)}, \cdots, \int {z^{(l-1)}}_n\prod_i p(z_i^{(l-1)})\prod_i dz_i^{(l-1)}) \\
= (\int {z^{(l-1)}}_1 p(z_1^{(l-1)})dz_1^{(l-1)}, \int {z^{(l-1)}}_2 p(z_2^{(l-1)})dz_2^{(l-1)}, \cdots, \int {z^{(l-1)}}_n p(z_n^{(l-1)})dz_n^{(l-1)}) \\
= (m_{{z^{(l-1)}}_1}, m_{{z^{(l-1)}}_2}, \cdots, m_{{z^{(l-1)}}_n}) = m_{z^{(l-1)}}
\end{aligned}
\tag{8}
$$

W に関しても同様のことが言える。

分散についても同様に検討する。

$$
\begin{aligned}
v_{a^{(l)}} = (v_{{a^{(l)}}_1}, v_{{a^{(l)}}_2}, \cdots, v_{{a^{(l)}}_n}) \equiv (v_{{a^{(l)}}_i}) = (\mathbb{E}_{p(z^{(l-1)},W^{(l)})}({a_i^{(l)}}^2) - (\mathbb{E}_{p(z^{(l-1)},W^{(l)})}(a_i^{(l)}))^2) \\
= \frac{1}{H_{l-1}}(\int\int W^{(l)}z^{(l-1)} \odot W^{(l)}z^{(l-1)}p(z^{(l-1)})p(W^{(l)})dz^{(l-1)}dW^{(l)} - M^{(l)}m_{z^{(l-1)}} \odot M^{(l)}m_{z^{(l-1)}}) \\
= \frac{1}{H_{l-1}}(\int\int(\sum_j w_{i,j}^{(l)}z_j^{(l-1)})^2 p(z^{(l-1)})p(W^{(l)})dz^{(l-1)}dW^{(l)} - (\sum_j m_{i,j}^{(l)}m_{{z^{(l-1)}}_j})^2) \\
= \frac{1}{H_{l-1}}(\sum_j(\int(w_{i,j}^{(l)})^2 p(w_{i,j}^{(l)})dw_{i,j}^{(l)}\int(z_j^{(l-1)})^2 p(z_j^{(l-1)})dz_j^{(l-1)} - {m_{i,j}^{(l)}}^2 {m_{{z^{(l-1)}}_j}}^2))
\end{aligned}
\tag{9}
$$

最後の等号のところは各変数が一次のところは

$$
\begin{aligned}
&\int\int w_{i,j}^{(l)} z_j^{(l-1)} \quad w_{i,k}^{(l)} z_k^{(l-1)} p(z^{(l-1)}) p(W^{(l)}) dz^{(l-1)} dW^{(l)} \\
&= \int\int\int\int w_{i,j}^{(l)} z_j^{(l-1)} \quad w_{i,k}^{(l)} z_k^{(l-1)} p(w_{i,j}^{(l)}) p(w_{i,k}^{(l)}) p(z_j^{(l-1)}) p(z_k^{(l-1)}) dw_{i,j}^{(l)} dw_{i,k}^{(l)} dz_j^{(l-1)} dz_k^{(l-1)} \\
&= \int w_{i,j}^{(l)} p(w_{i,j}^{(l)}) dw_{i,j}^{(l)} \int w_{i,k}^{(l)} p(w_{i,k}^{(l)}) dw_{i,k}^{(l)} \int z_j^{(l-1)} \quad p(z_j^{(l-1)}) dz_j^{(l-1)} \int z_k^{(l-1)} p(z_k^{(l-1)}) dz_k^{(l-1)} \\
&= m_{i,j}^{(l)} m_{i,k}^{(l)} m_{z^{(l-1)}j} m_{z^{(l-1)}k}
\end{aligned}
\tag{10}
$$

となることに注意した。この式も考慮すると、

$$
\begin{aligned}
v_{a^{(l)}} &= \frac{1}{H_{l-1}} (\sum_j (\int (w_{i,j}^{(l)})^2 p(w_{i,j}^{(l)}) dw_{i,j}^{(l)} \int (z_j^{(l-1)})^2 p(z_j^{(l-1)}) dz_j^{(l-1)} - {m_{i,j}^{(l)}}^2 {m_{z^{(l-1)}j}}^2)) \\
&= \frac{1}{H_{l-1}} (\sum_j (\int (w_{i,j}^{(l)})^2 p(w_{i,j}^{(l)}) dw_{i,j}^{(l)} \int (z_j^{(l-1)})^2 p(z_j^{(l-1)}) dz_j^{(l-1)} \\
&- \int (w_{i,j}^{(l)})^2 p(w_{i,j}^{(l)}) dw_{i,j}^{(l)} {m_{z^{(l-1)}j}}^2 + \int (w_{i,j}^{(l)})^2 p(w_{i,j}^{(l)}) dw_{i,j}^{(l)} {m_{z^{(l-1)}j}}^2 - {m_{i,j}^{(l)}}^2 {m_{z^{(l-1)}j}}^2)) \\
&= \frac{1}{H_{l-1}} (\sum_j (\int (w_{i,j}^{(l)})^2 (p(w_{i,j}^{(l)}) dw_{i,j}^{(l)} (\int (z_j^{(l-1)})^2 p(z_j^{(l-1)}) dz_j^{(l-1)} - {m_{z^{(l-1)}j}}^2) \\
&+ (\int (w_{i,j}^{(l)})^2 p(w_{i,j}^{(l)}) dw_{i,j}^{(l)} - {m_{i,j}^{(l)}}^2) {m_{z^{(l-1)}j}}^2)) \\
&= \frac{1}{H_{l-1}} (\int \sum_j (w_{i,j}^{(l)})^2 p(w_{i,j}^{(l)} dw_{i,j}^{(l)} (\int \sum_j (z_j^{(l-1)})^2 p(z_j^{(l-1)}) dz_j^{(l-1)} - \sum_j {m_{z^{(l-1)}j}}^2) \\
&+ (\int \sum_j (w_{i,j}^{(l)})^2 p(w_{i,j}^{(l)}) dw_{i,j}^{(l)} - \sum_j {m_{i,j}^{(l)}}^2) \sum_j {m_{z^{(l-1)}j}}^2)) \\
&= \frac{1}{H_{l-1}} (\mathbb{E}(W \odot W) v_{z^{(l-1)}} + V^{(l)} m_{z^{(l-1)}} \odot m_{z^{(l-1)}}) \\
&= \frac{1}{H_{l-1}} ((\mathbb{E}((W - M) \odot (W - M)) + M \odot M) v_{z^{(l-1)}} + V^{(l)} m_{z^{(l-1)}} \odot m_{z^{(l-1)}}) \\
&= \frac{1}{H_{l-1}} ((V^{(l)} \odot V^{(l)} + M \odot M) v_{z^{(l-1)}} + V^{(l)} m_{z^{(l-1)}} \odot m_{z^{(l-1)}}) \\
&= \frac{1}{H_{l-1}} ((V^{(l)} \odot V^{(l)}) v_{z^{(l-1)}} + (M \odot M) v_{z^{(l-1)}} + V^{(l)} (m_{z^{(l-1)}} \odot m_{z^{(l-1)}}))
\end{aligned}
\tag{11}
$$

途中で行列の要素として示している部分と、行列として示している部分があるので注意のこと。