5.2.5 について、詳細を見ていく。適宜、文献 [7](日本語版の書籍を参考にしている),[47], [83] (https://tminka.github.io/papers/ep/minka-thesis.pdf) も参照する。

(5.54)-(5.58) までは仮定なので良いとする。

5.2.5.3 から検討が始まる。

まず、$m_{i,j}^{(l)} = 0, v_{i,j}^{(l)} = \infty, \alpha_{\gamma_y} = 1, \beta_{\gamma_y} = 0, \alpha_{\gamma_w} = 1, \beta_{\gamma_w} = 0$ とする。これはガウス分布、ガンマ分布の係数となっているので、

$$\mathcal{N}(w_{i,j}^{(l)}|m_{i,j}^{(l)}, v_{i,j}^{(l)}) = \frac{1}{2\pi\sqrt{v_{i,j}^{(l)}}} exp^{-\frac{1}{2v_{i,j}^{(l)}}(w_{i,j}^{(l)} - m_{i,j}^{(l)})^2} \tag{1}$$

$m_{i,j}^{(l)}$ が有限で $v_{i,j}^{(l)} = \infty$ なので、$exp^{-\frac{1}{2v_{i,j}^{(l)}}(w_{i,j}^{(l)} - m_{i,j}^{(l)})^2}$ が 1 になる。つまり、一様分布となる。

$$Gam(\gamma|\alpha, \beta) = C_G(\alpha, \beta)\gamma^{\alpha-1} exp^{-\beta\gamma} = C_G(1, 0)\gamma^0 exp^0 = C_G(1, 0) \tag{2}$$

となり、一様分布になる。

本文では仮定密度フィルタリングを利用して、パラメータを決めることとしているし、[47] でも同様の流れとなっている。しかし、[47] の supplemental material の 6 では EP 法を用いた方法が、記載されている。[83] の (3.62) 式の下にあるように、仮定密度フィルタリングは EP 法を 1 サイクル実施したときと等価になるため、それを考慮する。(本の仮定密度フィルタリングのところは $p(\theta|D) \propto p(D|\theta)p(\theta)$ をもととしているが、$p(W)$ について、$p(D|W)$ に見えないので、EP 法を踏まえた方法をやってみた。)

(5.57) は [7] の (10.188) のような形式で表されていることがわかる。そのため、EP 法が適用できる。

[47] の supplemental material の 6 で (29),(30) 式はそのまま記載されている。

さて、(4.81),(4.84) を考慮すると、$\theta = \{\{w_{i,j}^{(l)}\}, \gamma_w, \gamma_y\}, X = \{\}$ として、

$$p(W, \gamma_w, \gamma_y) = p(W|\gamma_w)p(\gamma_w)p(\gamma_y) = \left(\prod_l \prod_i \prod_j \mathcal{N}(w_{i,j}^{(l)}|0, {\gamma_w}^{-1})\right) Gam(\gamma_w|\alpha_{\gamma_{w0}}, \beta_{\gamma_{w0}}) Gam(\gamma_y|\alpha_{\gamma_{y0}}, \beta_{\gamma_{y0}}) \tag{3}$$

一方で、(30) はそのままとなり、

$$q(W, \gamma_w, \gamma_y) = \left(\prod_l \prod_i \prod_j \mathcal{N}(w_{i,j}^{(l)}|m_{i,j}^{(l)}, v_{i,j}^{(l)})\right) Gam(\gamma_w|\alpha_{\gamma_w}, \beta_{\gamma_w}) Gam(\gamma_y|\alpha_{\gamma_y}, \beta_{\gamma_y}) \tag{4}$$

このとき、(4.81) を考慮すると、

$$\begin{aligned} p(W, \gamma_w, \gamma_y) &= \prod_l \prod_i \prod_j f(w_{i,j}^{(l)}, \gamma_w) f_{\gamma_w}(\gamma_w) f_{\gamma_y}(\gamma_y) \\ f(w_{i,j}^{(l)}, \gamma_w) &= p(w_{i,j}^{(l)}|\gamma_w) = \mathcal{N}(w_{i,j}^{(l)}|0, {\gamma_w}^{-1}) \\ f_{\gamma_w}(\gamma_w) &= p(\gamma_w) = Gam(\gamma_w|\alpha_{\gamma_w}, \beta_{\gamma_w}) \\ f_{\gamma_y}(\gamma_y) &= p(\gamma_y) = Gam(\gamma_y|\alpha_{\gamma_y}, \beta_{\gamma_y}) \end{aligned} \tag{5}$$

同様に (4.84) を考慮して、[47],supplement material(32) を参考にして、以下のようにも仮定する。(平均などは異なるが、(4) も成り立っていることに注意する。)

$$\begin{aligned} q(W, \gamma_w, \gamma_y) &= \prod_l \prod_i \prod_j \tilde{f}(w_{i,j}^{(l)}, \gamma_w) \\ \tilde{f}(w_{i,j}^{(l)}, \gamma_w) &= \mathcal{N}(w_{i,j}^{(l)}|\tilde{m}_{i,j}^{(l)}, \tilde{v}_{i,j}^{(l)}) Gam(\gamma_w|\tilde{\alpha}_{i,j}^{(l)}, \tilde{\beta}_{i,j}^{(l)}) \end{aligned} \tag{6}$$

EP 法では (4.86) を考慮する。

$$
\begin{aligned}
q^{\backslash i,j,k}(w_{i,j}^{(l)},\gamma_w) &\equiv \prod_l \prod_i \prod_j \left(\mathcal{N}(w_{i,j}^{(l)}|m^{\backslash i,j,l},v^{\backslash i,j,l})Gam(\gamma_w|\alpha^{\backslash i,j,l},\beta^{\backslash i,j,l})\right) \\
= \frac{q^{old}(w_{i,j}^{(l)},\gamma_w)}{\tilde{f}(w_{i,j}^{(l)},\gamma_w)} &= \frac{\left(\prod_l \prod_i \prod_j \mathcal{N}(w_{i,j}^{(l)}|m_{i,j}^{old(l)},v_{i,j}^{old(l)})\right)Gam(\gamma_w|\alpha_{\gamma_w^{old}},\beta_{\gamma_w^{old}})Gam(\gamma_y|\alpha_{\gamma_{y0}},\beta_{\gamma_{y0}})}{\mathcal{N}(w_{i,j}^{(l)}|\tilde{m}_{i,j}^{(l)},\tilde{v}_{i,j}^{(l)})Gam(\gamma_w|\tilde{\alpha}_{i,j}^{(l)},\tilde{\beta}_{i,j}^{(l)})} \\
&\propto \frac{\mathcal{N}(w_{i,j}^{(l)}|m_{i,j}^{old(l)},v_{i,j}^{old(l)})}{\mathcal{N}(w_{i,j}^{(l)}|\tilde{m}_{i,j}^{(l)},\tilde{v}_{i,j}^{(l)})}\frac{{\gamma_w}^{\alpha_{\gamma_w^{old}}-1}exp(-\beta_{\gamma_w^{old}}\gamma_w)}{{\gamma_w}^{\tilde{\alpha}_{i,j}^{(l)}-1}exp(-\tilde{\beta}_{i,j}^{(l)}\gamma_w)} \\
= \mathcal{N}(w_{i,j}^{(l)}|({v_{i,j}^{old(l)}}^{-1}-(\tilde{v}_{i,j}^{(l)})^{-1})^{-1}&(m_{i,j}^{old(l)}{v_{i,j}^{old(l)}}^{-1}-\tilde{m}_{i,j}^{(l)}(\tilde{v}_{i,j}^{(l)})^{-1}),({v_{i,j}^{old(l)}}^{-1}-{(\tilde{v}_{i,j}^{(l)})}^{-1})^{-1}) \\
&{\gamma_w}^{(\alpha_{\gamma_w^{old}}-\tilde{\alpha}_{i,j}^{(l)}+1)-1}exp(-(\beta_{\gamma_w^{old}}-\tilde{\beta}_{i,j})^{(l)}\gamma_w)
\end{aligned}
\tag{7}
$$

これにより [47]supplement material の (36)(37) が求まる。

この $q^{\backslash i,j,k}$ を用いて、q^{new} のパラメータを求める。

(4.87) を考慮して、

$$
\begin{aligned}
r(W,\gamma_w,\gamma_y) &= \frac{1}{Z_0}f(w_{i,j}^{(l)},\gamma_w)q^{\backslash i,j,k}(w_{i,j}^{(l)},\gamma_w) \\
&= \mathcal{N}(w_{i,j}^{(l)}|0,{\gamma_w}^{-1})\mathcal{N}(w_{i,j}^{(l)}|m^{\backslash i,j,l},v^{\backslash i,j,l})Gam(\gamma_w|\alpha^{\backslash i,j,l},\beta^{\backslash i,j,l})
\end{aligned}
\tag{8}
$$

Z_0 を求めると、

$$
\begin{aligned}
Z_0 &= \int f(w_{I,J}^{(L)},\gamma_w)q^{\backslash I,J,L}(w_{I,J}^{(L)},\gamma_w)\prod_i \prod_j \prod_l dw_{i,j}^{(l)}d\gamma_w d\gamma_y \\
&= \int \mathcal{N}(w_{I,J}^{(L)}|0,{\gamma_w}^{-1})q^{\backslash I,J,L}(w_{I,J}^{(L)},\gamma_w)\prod_i \prod_j \prod_l dw_{i,j}^{(l)}d\gamma_w d\gamma_y \\
&= \int \mathcal{N}(w_{I,J}^{(L)}|0,{\gamma_w}^{-1})\mathcal{N}(w_{I,J}^{(L)}|m^{\backslash I,J,L},v^{\backslash I,J,L)})Gam(\gamma_w|\alpha^{\backslash I,J,L},\beta^{\backslash I,J,L})dw_{I,J}^{(L)}d\gamma_w
\end{aligned}
\tag{9}
$$

I,J,L でないところは $q^{\backslash I,J,L}$ が分布のため、積分すると 1 になる。(というか、正規化定数に含まれる)

一般に (4.60)-(4.62) は成立しているので、

$$
\mathbb{E}_r[w_{I,J}^{(L)}] = m_{i,j}^{new(l)} = m^{\backslash I,J,L} + v^{\backslash I,J,L}\frac{\partial}{\partial m^{\backslash I,J,L}}\ln Z_0 \tag{10}
$$

同様に (4.63),(4.64),(4.66) も成り立っているので、

$$
v_{i,j}^{new(l)} = v^{\backslash I,J,L} - {v^{\backslash I,J,L}}^2\left(\left(\frac{\partial}{\partial m^{\backslash I,J,L}}\ln Z_0\right)^2 - 2\frac{\partial}{\partial v^{\backslash I,J,L}}\ln Z_0\right) \tag{11}
$$

更に、(4.67) に対して、(4.69)-(4.71) が成立するので、(4.72),(4.73) すなわち、(5.62),(5.63) が成り立つ。

$$
\alpha_{\gamma_w}^{new} = \left(\frac{Z_2 Z_0(\alpha^{\backslash I,J,L}+1)}{{Z_1}^2\alpha^{\backslash I,J,L}} - 1\right)^{-1} \tag{12}
$$

$$
\beta_{\gamma_w}^{new} = \left(\frac{Z_2(\alpha^{\backslash I,J,L}+1)}{Z_1\beta^{\backslash I,J,L}} - \frac{Z_1\alpha^{\backslash I,J,L}}{Z_0\beta^{\backslash I,J,L}}\right)^{-1} \tag{13}
$$

最後に $\tilde{f}$ の更新を考える。

$$\tilde{f}^{new}(w_{i,j}^{(l)},\gamma_w) \equiv \mathcal{N}(w_{i,j}^{(l)}|\tilde{m}_{i,j}^{new(l)},\tilde{v}_{i,j}^{new(l)})Gam(\gamma_w|\tilde{\alpha}_{i,j}^{new(l)},\tilde{\beta}_{i,j}^{new(l)}) = Z_0\frac{q^{new}(w_{i,j}^{(l)},\gamma_w)}{q^{\backslash I,J,L}(w_{I,J}^{(L)},\gamma_w)}$$
$$= Z\frac{\mathcal{N}(w_{i,j}^{(l)}|m_{i,j}^{new(l)},v_{i,j}^{new(l)})Gam(\gamma_w|\alpha_{\gamma_w}^{new},\beta_{\gamma_w}^{new})}{\mathcal{N}(w_{i,j}^{(l)}|m^{\backslash i,j,l},v^{\backslash i,j,l})Gam(\gamma_w|\alpha^{\backslash i,j,l},\beta^{\backslash i,j,l})} = Z\frac{\mathcal{N}(w_{i,j}^{(l)}|m_{i,j}^{new(l)},v_{i,j}^{new(l)})Gam(\gamma_w|\alpha_{\gamma_w}^{new},\beta_{\gamma_w}^{new})}{\mathcal{N}(w_{i,j}^{(l)}|m^{\backslash i,j,l},v^{\backslash i,j,l})Gam(\gamma_w|\alpha^{\backslash i,j,l},\beta^{\backslash i,j,l})}$$
$$= \mathcal{N}(w_{i,j}^{(l)}|({v_{i,j}^{new(l)}}^{-1} - {v^{\backslash i,j,l}}^{-1})^{-1}(m_{i,j}^{new(l)}{v_{i,j}^{new(l)}}^{-1} - m^{\backslash i,j,l}{v^{\backslash i,j,l}}^{-1})^{-1}),({v_{i,j}^{new(l)}}^{-1} - {v^{\backslash i,j,l}}^{-1})^{-1})$$
$$\gamma_w{}^{(\alpha_{\gamma_w}^{new}-\alpha^{\backslash i,j,l}+1)-1}exp(-(\beta_{\gamma_w}^{new}-\tilde{\beta}^{\backslash i,j,l})\gamma_w) \tag{14}$$

Z は正規化定数になる。(同じ値とは限らない。)

これにより、[47],supplement material(38),(39) が成り立っていることがわかる。

これが、公式のコードと一致していることを確認する。

ところで、Z は (5.64) のように計算でき、以下のようになる。(別資料のように本文に誤りがある。[47] の (11) も参照のこと。)

$$Z_0 = \mathcal{N}(m^{\backslash I,J,L}|0,,v^{\backslash I,J,L)} + \frac{\beta_{\gamma_w}^{\backslash I,J,L)}}{\alpha_{\gamma_w}^{\backslash I,J,L)}-1}) \tag{15}$$

よって、$v = v^{\backslash I,J,L)} + \frac{\beta_{\gamma_w}^{\backslash I,J,L)}}{\alpha_{\gamma_w}^{\backslash I,J,L)}-1}$ として、

$$\begin{aligned} ln\,Z_0 &= ln\,\mathcal{N}(m^{\backslash I,J,L}|0,v^{\backslash I,J,L)} + \frac{\beta_{\gamma_w}^{\backslash I,J,L)}}{\alpha_{\gamma_w}^{\backslash I,J,L)}-1}) \\ &= ln\,\mathcal{N}(m^{\backslash I,J,L}|0,v) = -lnv - \frac{1}{2}{m^{\backslash I,J,L}}^2 v + C \end{aligned} \tag{16}$$

これを考えると Z_1, Z_2 とこれらの微分が求まる。

theano 版のコード (prior.py) を確認すると、以下のようになっている。

```
import numpy as np

class Prior:

    def __init__(self, layer_sizes, var_targets):

        # We refine the factor for the prior variance on the weights

        n_samples = 3.0
        v_observed = 1.0
# α_γ w0
        self.a_w = 2.0 * n_samples
# β_γ w0
        self.b_w = 2.0 * n_samples * v_observed
```

```
        # We refine the factor for the prior variance on the noise

        n_samples = 3.0
        a_sigma = 2.0 * n_samples
        b_sigma = 2.0 * n_samples * var_targets

        self.a_sigma_hat_nat = a_sigma - 1
        self.b_sigma_hat_nat = -b_sigma

        # We refine the gaussian prior on the weights

        self.rnd_m_w = []
        self.m_w_hat_nat = []
        self.v_w_hat_nat = []
        self.a_w_hat_nat = []
        self.b_w_hat_nat = []
        for size_out, size_in in zip(layer_sizes[ 1 : ], layer_sizes[ : -1 ]):
# 平均 m を平均 0，分散 1 で m をサンプリングする。
# 平均は 1/sqrt(size_in + 1) にしかなっていないので更に割る。
            self.rnd_m_w.append(1.0 / np.sqrt(size_in + 1) *
                np.random.randn(size_out, size_in + 1))
# \tilde{f}に関して\tilde{m}/\tilde{v}を 0 で初期化。
#これは\tilde{m}が (33) のように 0 なので、0 になる。
            self.m_w_hat_nat.append(np.zeros((size_out, size_in + 1)))
# \tilde{f}に関して 1/\tilde{v}を初期化 (33)
            self.v_w_hat_nat.append((self.a_w - 1) / self.b_w * \
                np.ones((size_out, size_in + 1)))
# \tilde{f}に関して (\tilde{α} - 1) を 0 で初期化
            self.a_w_hat_nat.append(np.zeros((size_out, size_in + 1)))
# \tilde{f}に関して\tilde{β}を 0 で初期化
            self.b_w_hat_nat.append(np.zeros((size_out, size_in + 1)))

    def get_initial_params(self):

        m_w = []
        v_w = []
        for i in range(len(self.rnd_m_w)):
            m_w.append(self.rnd_m_w[ i ])
            v_w.append(1.0 / self.v_w_hat_nat[ i ])
```

```
        return { 'm_w': m_w, 'v_w': v_w , 'a': self.a_sigma_hat_nat + 1,
            'b': -self.b_sigma_hat_nat }

    def get_params(self):

        m_w = []
        v_w = []
        for i in range(len(self.rnd_m_w)):
            m_w.append(self.m_w_hat_nat[ i ] / self.v_w_hat_nat[ i ])
            v_w.append(1.0 / self.v_w_hat_nat[ i ])

        return { 'm_w': m_w, 'v_w': v_w , 'a': self.a_sigma_hat_nat + 1,
            'b': -self.b_sigma_hat_nat }

#paramsにはq^{old}のm_{i, j}^{l}, v_{i, j}^{l}のデータが含まれる。
    def refine_prior(self, params):
# ここのiはレイヤの数→l
        for i in range(len(params[ 'm_w' ])):
# ここのjはレイヤの出力数→i
            for j in range(params[ 'm_w' ][ i ].shape[ 0 ]):
# ここのkはレイヤの出力数→j
                for k in range(params[ 'm_w' ][ i ].shape[ 1 ]):

                    # We obtain the parameters of the cavity distribution
# q^{old}に関してv_wが来るので1 / v_wで逆数を取る。
                    v_w_nat = 1.0 / params[ 'v_w' ][ i ][ j, k ]
# q^{old}に関してmv^{-1}を計算する。
                    m_w_nat = params[ 'm_w' ][ i ][ j, k ] / \
                        params[ 'v_w' ][ i ][ j, k ]
# (36)の1/v
                    v_w_cav_nat = v_w_nat - self.v_w_hat_nat[ i ][ j, k ]
# (36)のm/v
                    m_w_cav_nat = m_w_nat - self.m_w_hat_nat[ i ][ j, k ]
# (36)のv
                    v_w_cav = 1.0 / v_w_cav_nat
# (36)のmを求める。1/v_w_cav_nat = v_w_cavになる。
                    m_w_cav = m_w_cav_nat / v_w_cav_nat
# (37)に関して、-(\tilde{αγ w^{old} - 1)
                    a_w_nat = self.a_w - 1
# (37)に関して、-(\tilde{βγ w^{old})
```

```
                    b_w_nat = -self.b_w
# （37）に関して左辺のマイナス（nat がついているので α-1 になっている。）
                    a_w_cav_nat = a_w_nat - self.a_w_hat_nat[ i ][ j, k ]
# （37）に関して左辺のマイナス
                    b_w_cav_nat = b_w_nat - self.b_w_hat_nat[ i ][ j, k ]
# （37）の左辺を求める。
                    a_w_cav = a_w_cav_nat + 1
# （37）の左辺を求める。
                    b_w_cav = -b_w_cav_nat

# よく見ていないが値が悪ければ、更新されなくなりそうな。
                    if v_w_cav > 0 and b_w_cav > 0 and a_w_cav > 1 and \
                        v_w_cav < 1e6:

                        # We obtain the values of the new parameters of the
                        # posterior approximation

# Z0 の分散
                        v = v_w_cav + b_w_cav / (a_w_cav - 1)
# Z1 の分散
                        v1  = v_w_cav + b_w_cav / a_w_cav
# Z2 の分散
                        v2  = v_w_cav + b_w_cav / (a_w_cav + 1)
# log Z0。定数項は無視。Z0 は q^{\I,J,L}を見ている。
                        logZ = -0.5 * np.log(v) - 0.5 * m_w_cav**2 / v
# log Z1。定数項は無視。Z1 は q^{\I,J,L}を見ている。
                        logZ1 = -0.5 * np.log(v1) - 0.5 * m_w_cav**2 / v1
# log Z2。定数項は無視。Z0 は q^{\I,J,L}を見ている。
                        logZ2 = -0.5 * np.log(v2) - 0.5 * m_w_cav**2 / v2
# log Z0 を m_w_cav で偏微分
                        d_logZ_d_m_w_cav = -m_w_cav / v
# log Z0 を v_w_cav で偏微分。v に v_w_cav も含まれる。
                        d_logZ_d_v_w_cav = -0.5 / v + 0.5 * m_w_cav**2 / v**2
# 本の（5.59）*_cav を使っていることに注意
                        m_w_new = m_w_cav + v_w_cav * d_logZ_d_m_w_cav
# 本の（5.60）*_cav を使っていることに注意
                        v_w_new = v_w_cav - v_w_cav**2 * \
                            (d_logZ_d_m_w_cav**2 - 2 * d_logZ_d_v_w_cav)
# 本の（5.62）*_cav を使っていることに注意
                        a_w_new = 1.0 / (np.exp(logZ2 - 2 * logZ1 + logZ) * \
```

```
                            (a_w_cav + 1) / a_w_cav - 1.0)
# 本の (5.63) *_cav を使っていることに注意
                        b_w_new = 1.0 / (np.exp(logZ2 - logZ1) * \
                            (a_w_cav + 1) / (b_w_cav) - np.exp(logZ1 - \
                            logZ) * a_w_cav / b_w_cav)
# 新しい分散の逆数 (\tilde{f}の更新に備える。)
                        v_w_new_nat = 1.0 / v_w_new
# 新しい平均を v で割る。(\tilde{f}の更新に備える。)
                        m_w_new_nat = m_w_new / v_w_new
#α^{new}の-1(nat) を求める。
                        a_w_new_nat = a_w_new - 1
#β^{new}の nat は-にする。
                        b_w_new_nat = -b_w_new

                        # We update the parameters of the approximate factor,
                        # whih is given by the ratio of the new posterior
                        # approximation and the cavity distribution

# (38) m^{new}/v^{new}
                        self.m_w_hat_nat[ i ][ j, k ] = m_w_new_nat - \
                            m_w_cav_nat
# (38) 1/v^{new}
                        self.v_w_hat_nat[ i ][ j, k ] = v_w_new_nat - \
                            v_w_cav_nat
# (39) α^{new}-1
                        self.a_w_hat_nat[ i ][ j, k ] = a_w_new_nat - \
                            a_w_cav_nat
# (39) -β^{new}
                        self.b_w_hat_nat[ i ][ j, k ] = b_w_new_nat - \
                            b_w_cav_nat

                        # We update the posterior approximation

                        params[ 'm_w' ][ i ][ j, k ] = m_w_new
                        params[ 'v_w' ][ i ][ j, k ] = v_w_new

                        self.a_w = a_w_new
                        self.b_w = b_w_new

        return params
```

これまでの計算があっているように感じる。

なお、(4) のみを考慮して、γ_y, γ_w については $p(\gamma_y), p(\gamma_w)$ をそのままにすると述べられている。これについては、[83] の演習 10.37 で一般的に確認されている。(公式の解答もあり、そこまで難しくなさそう。)

よって、

$$\begin{aligned} \alpha_{\gamma_w} &= \alpha_{\gamma_{w0}} \\ \beta_{\gamma_w} &= \beta_{\gamma_{w0}} \\ \alpha_{\gamma_y} &= \alpha_{\gamma_{y0}} \\ \beta_{\gamma_y} &= \beta_{\gamma_{y0}} \end{aligned} \tag{17}$$

本、[47],supplement material6 にあるように、

$$\begin{aligned} m_{i,j}^{(l)} &= 0 \\ v_{i,j}^{(l)} &= \infty \\ \tilde{m}_{i,j}^{(l)} &= 0 \\ \tilde{v}_{i,j}^{(l)} &= \infty \\ \tilde{\alpha}_{\gamma_w} &= 1 \\ \tilde{\beta}_{\gamma_w} &= 0 \end{aligned} \tag{18}$$

こうしたときに、(33)(34) を考えると、最初の取り込みで、$q^{\backslash I,J,L}$ は γ_w のガンマ分布になるため、

$$\begin{aligned} m^{\backslash I,J,L} &= 0 \\ v^{\backslash I,J,L} &= \infty \\ {\alpha_{\gamma_w}}^{\backslash I,J,L} &= \alpha_{\gamma_{w0}} \\ {\beta_{\gamma_w}}^{\backslash I,J,L} &= \beta_{\gamma_{w0}} \end{aligned} \tag{19}$$

(9) で上記の値を代入すると

$$Z_0 = \int \mathcal{N}(w_{I,J}^{(L)}|0, {\gamma_w}^{-1}) Gam(\gamma_w|\alpha^{\backslash I,J,L}, \beta^{\backslash I,J,L}) dw_{I,J}^{(L)} d\gamma_w = \int p(w_{I,J}^{(L)}|\gamma_w) p(\gamma_w) dw_{I,J}^{(L)} d\gamma_w \tag{20}$$

となり、Z_0 は常に 1 になり、Z_1, Z_2 も 1 になる。(37) に対して、$\alpha^{\backslash I,J,L} = \alpha_0, \beta^{\backslash I,J,L} = \beta_0$ になる。Z たちが 1 なので、$\alpha^{new} = \alpha_0, \beta^{new} = \beta_0$ になり、(39) を使うと、$\tilde{\alpha}^{new} = \tilde{\alpha}^{old} = 1, \tilde{\beta}^{new} = \tilde{\beta}^{old} = 0$

一方、上記の式で、$w_{I,J}^{(L)}$ の関数だとするとスチューデントの t 分布の定義になる。(例えば、PRML の (2.158)-(2.166),(B.64)-(B.67), それに関する演習問題を参照のこと。)

直接平均を求めると、PRML の (B.65) から q^{new} の平均は 0 になり、q^{new} の分散は P.100 の ν, λ を使い、(B.66) に代入すると q^{new} の分散 v は

$$v = \frac{1}{\lambda}\frac{\nu}{\nu - 2} = \frac{\beta_0}{\alpha_0}\frac{2\alpha_0}{2\alpha_0 - 2} = \frac{\beta_0}{\alpha_0 - 1} \tag{21}$$

となり、(34) が求まる。(39) を考慮すると、(33) も求まる。