Efficient Model Evidence Computation in Tree-structured Factor Graphs

Supplementary Material

Hoang M.H. Nguyen Bart van Erp İsmail Şenöz Bert de Vries

Department of Electrical Engineering, *Eindhoven University of Technology*, Eindhoven, The Netherlands

Abstract

This supplementary document provides the message computation rules of sum-product messages in the paper "Efficient Model Evidence Computation in Tree-structured Factor Graphs".

I. Bernoulli node

The factor node function of the Bernoulli node is defined as

$$f(x,y) = \mathrm{Ber}(y \mid x) = x^{y}\,(1-x)^{1-y} \tag{1}$$

where $y \in \{0,1\}$ is a binary variable, and $x \in [0,1]$.

A. Forward message

The node function $f(x,y)$ represents a conditional distribution $p(y \mid x)$. Based on the general scale factor update rules of a conditional probability density function in [1, Ch.6], we obtain the scale factor update rule

$$\vec{\beta}_y = \vec{\beta}_x \tag{2}$$

where $\vec{\beta}_y$ and $\vec{\beta}_x$ are the scale factors of the messages $\vec{\mu}_y(y)$ and $\vec{\mu}_x(x)$, respectively.

B. Backward message

Now we compute the scale factor $\overleftarrow{\beta}_x$ of the backward message $\overleftarrow{\mu}_x(x)$. By definition, we have

$$\overleftarrow{\beta}_x = \int \overleftarrow{\mu}_x(x)\mathrm{d}x = \int \sum_{y} \overleftarrow{\mu}_y(y) f(y,x)\mathrm{d}x. \tag{3}$$

Assume that we have observed $y = \hat{y}$, then $\overleftarrow{\mu}_y(y) = \overleftarrow{\beta}_y \delta(y-\hat{y})$ and (3) is now equivalent to

$$\begin{aligned}
\overleftarrow{\beta}_x &= \overleftarrow{\beta}_y \int \mathrm{Ber}(\hat{y} \mid x)\mathrm{d}x && (4)\\
&= \overleftarrow{\beta}_y \int x^{\hat{y}}(1-x)^{1-\hat{y}}\mathrm{d}x && (5)\\
&= \overleftarrow{\beta}_y\, B(a,b) \underbrace{\int \frac{1}{B(a,b)}\, x^{a-1}(1-x)^{b-1}\mathrm{d}x}_{=1,\ \text{Beta distribution}} && (6)\\
&= \overleftarrow{\beta}_y\, B(a,b) && (7)
\end{aligned}$$

where $a = \hat{y}+1$, $b = 2-\hat{y}$ and $B(a,b)$ is the Beta function defined as

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \tag{8}$$

where $\Gamma(\cdot,\cdot)$ is the gamma function. From the definition of a, b and $B(a,b)$, we have

$$\begin{aligned}
\overleftarrow{\beta}_x &= \overleftarrow{\beta}_y\, B(\hat{y}+1, 2-\hat{y})\\
&= \overleftarrow{\beta}_y\, \frac{\Gamma(\hat{y}+1)\Gamma(2-\hat{y})}{\Gamma(3)}.
\end{aligned}$$

Note that $\hat{y} \in \{0,1\}$, and regardless of what value $\hat{y}$ takes, the following equality holds

$$\frac{\Gamma(\hat{y}+1)\Gamma(2-\hat{y})}{\Gamma(3)} = \frac{\Gamma(1)\Gamma(2)}{\Gamma(3)} = \frac{1}{2}. \tag{9}$$

Therefore, we achieve the following final equation of $\overleftarrow{\beta}_x$

$$\overleftarrow{\beta}_x = \frac{1}{2}\overleftarrow{\beta}_y \tag{10}$$

II. Categorical node

The factor node function of the Categorical node is

$$f(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{Cat}(\boldsymbol{y} \mid \boldsymbol{x}) = \prod_{i=1}^{K} x_i^{y_i} \tag{11}$$

where $\boldsymbol{y} = [y_1, \ldots, y_K]$ is a one-hot-coded vector, i.e. $y_i \in \{0, 1\}$ and $\sum_{i=1}^{K} y_i = 1$, and $\boldsymbol{x}$ is a vector such that $x_i \in [0, 1]$ and $\sum_{i=1}^{K} x_i = 1$. K is the number of categories.

A. Forward message

Similar to the Bernoulli node, we also regard the node function as a conditional distribution $p(\boldsymbol{y} \mid \boldsymbol{x})$, thus we have the following update rule

$$\vec{\beta}_{\boldsymbol{y}} = \vec{\beta}_{\boldsymbol{x}} \tag{12}$$

where $\vec{\beta}_{\boldsymbol{y}}$ and $\vec{\beta}_{\boldsymbol{x}}$ are the scale factors of the messages $\vec{\mu}_{\boldsymbol{y}}$ and $\vec{\mu}_{\boldsymbol{x}}$, respectively.

B. Backward message

Now we compute the scale factor $\overleftarrow{\beta}_{\boldsymbol{x}}$ of the backward message $\overleftarrow{\mu}_{\boldsymbol{x}}(\boldsymbol{x})$. By definition, we have:

$$\overleftarrow{\beta}_{\boldsymbol{x}} = \int \overleftarrow{\mu}_{\boldsymbol{x}} \mathrm{d}\boldsymbol{x} = \int \sum_{\boldsymbol{y}} \overleftarrow{\mu}(\boldsymbol{y}) f(\boldsymbol{y}, \boldsymbol{x}) \mathrm{d}\boldsymbol{x} \tag{13}$$

Assume that we have observed $\boldsymbol{y} = \hat{\boldsymbol{y}}$, then $\overleftarrow{\mu}_{\boldsymbol{y}}(\boldsymbol{y}) = \overleftarrow{\beta}_{\boldsymbol{y}} \delta(\boldsymbol{y} - \hat{\boldsymbol{y}})$ and (13) equals:

$$\begin{align}
\overleftarrow{\beta}_{\boldsymbol{x}} &= \overleftarrow{\beta}_{\boldsymbol{y}} \int \mathrm{Cat}(\hat{\boldsymbol{y}} | \boldsymbol{x}) \mathrm{d}\boldsymbol{x} \tag{14}\\
&= \overleftarrow{\beta}_{\boldsymbol{y}} \int \prod_{i=1}^{K} \boldsymbol{x}_i^{\hat{y}_i} \mathrm{d}\boldsymbol{x} \tag{15}\\
&= \overleftarrow{\beta}_{\boldsymbol{y}}\ B(\boldsymbol{\alpha}) \underbrace{\int \frac{1}{B(\boldsymbol{\alpha})} \prod_{i=1}^{K} \boldsymbol{x}_i^{\alpha_i - 1} \mathrm{d}\boldsymbol{x}}_{=1,\text{ Dirichlet distribution}} \tag{16}\\
&= \overleftarrow{\beta}_{\boldsymbol{y}}\ B(\boldsymbol{\alpha}) \tag{17}
\end{align}$$

where $\boldsymbol{\alpha} = \hat{\boldsymbol{y}} + \mathbf{1}$, and

$$B(\boldsymbol{\alpha}) = \frac{\prod_{i=1}^{K} \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{K} \alpha_i)}. \tag{18}$$

From the definition of $\boldsymbol{\alpha}$ and $B(\boldsymbol{\alpha})$, we have:

$$\begin{align}
\overleftarrow{\beta}_{\boldsymbol{x}} &= \overleftarrow{\beta}_{\boldsymbol{y}}\ B(\hat{\boldsymbol{y}} + \mathbf{1}) \tag{19}\\
&= \overleftarrow{\beta}_{\boldsymbol{y}} \frac{\prod_{i=1}^{K} \Gamma(\hat{y}_i + 1)}{\Gamma(\sum_{i=1}^{K} (\hat{y}_i + 1))} \tag{20}\\
&= \overleftarrow{\beta}_{\boldsymbol{y}} \frac{1}{\Gamma(K+1)} \tag{21}\\
&= \frac{\overleftarrow{\beta}_{\boldsymbol{y}}}{K!} \tag{22}
\end{align}$$

III. Transition node

The factor node function of the transition node is

$$f(\boldsymbol{y}, \boldsymbol{x}) = \mathrm{Cat}(\boldsymbol{y} \mid \mathbf{A}\boldsymbol{x}) = \prod_{j=1}^{J} \prod_{k=1}^{K} A_{jk}^{y_j x_k} \tag{23}$$

where $\boldsymbol{y}$ and $\boldsymbol{x}$ are one-hot-coded vectors. $\mathbf{A}$ is the transition matrix whose elements are constrained to $A_{jk} \in [0, 1]$ and whose columns are also constrained by $\sum_j A_{jk} = 1$. Since both $\boldsymbol{x}$ and $\boldsymbol{y}$ are one-hot coded, the node function (23) can be written in vector notation as follows

$$\mathrm{Cat}(\boldsymbol{y} \mid \mathbf{A}\boldsymbol{x}) = \boldsymbol{y}^\top \mathbf{A}\boldsymbol{x} \tag{24}$$

In the following derivation, we assume the matrix $\mathbf{A}$ is known.

A. Forward message

Again, we can recognize that the node function is a conditional distribution $p(\boldsymbol{y} \mid \boldsymbol{x})$, thus

$$\vec{\beta}_{\boldsymbol{y}} = \vec{\beta}_{\boldsymbol{x}} \tag{25}$$

where $\vec{\beta}_{\boldsymbol{y}}$ and $\vec{\beta}_{\boldsymbol{x}}$ are the scale factors of the messages $\vec{\mu}_{\boldsymbol{y}}(\boldsymbol{y})$ and $\vec{\mu}_{\boldsymbol{x}}(\boldsymbol{x})$, respectively.

B. Backward message

Assuming a categorical incoming message $\overleftarrow{\mu}_{\boldsymbol{y}}(\boldsymbol{y}) = \overleftarrow{\beta}_{\boldsymbol{y}} \mathrm{Cat}(\boldsymbol{y} \mid \boldsymbol{\pi}_{\boldsymbol{y}})$, now we compute the scale factor $\overleftarrow{\beta}_{\boldsymbol{x}}$ of the backward message $\overleftarrow{\mu}_{\boldsymbol{x}}$. By definition, we have

$$\begin{aligned}
\overleftarrow{\beta}_{\boldsymbol{x}} &= \sum_{\boldsymbol{x}} \overleftarrow{\mu}_{\boldsymbol{x}}(\boldsymbol{x}) = \sum_{\boldsymbol{x}} \sum_{\boldsymbol{y}} \overleftarrow{\mu}_{\boldsymbol{y}}(\boldsymbol{y}) \, \mathrm{Cat}(\boldsymbol{y} \mid \mathbf{A}\boldsymbol{x}) \\
&= \overleftarrow{\beta}_{\boldsymbol{y}} \sum_{\boldsymbol{x}} \sum_{\boldsymbol{y}} \mathrm{Cat}\left(\boldsymbol{y} | \boldsymbol{\pi}_{\boldsymbol{y}}\right) \boldsymbol{x}^\top \mathbf{A}^\top \boldsymbol{y} \\
&= \overleftarrow{\beta}_{\boldsymbol{y}} \sum_{\boldsymbol{x}} \boldsymbol{x}^\top \mathbf{A}^\top \underbrace{\mathbb{E}_{\mathrm{Cat}(\boldsymbol{y}|\boldsymbol{\pi}_{\boldsymbol{y}})}[\boldsymbol{y}]}_{\boldsymbol{\pi}_{\boldsymbol{y}}} \\
&= \overleftarrow{\beta}_{\boldsymbol{y}} \sum_{\boldsymbol{x}} \boldsymbol{x}^\top \mathbf{A}^\top \boldsymbol{\pi}_{\boldsymbol{y}} \\
&= \overleftarrow{\beta}_{\boldsymbol{y}} \, ||\mathbf{A}^\top \boldsymbol{\pi}_{\boldsymbol{y}}||_1 \sum_{\boldsymbol{x}} \boldsymbol{x}^\top \frac{\mathbf{A}^\top \boldsymbol{\pi}_{\boldsymbol{y}}}{||\mathbf{A}^\top \boldsymbol{\pi}_{\boldsymbol{y}}||_1}.
\end{aligned}$$

In the last equality, we can view $\boldsymbol{x}^\top \frac{\mathbf{A}^\top \boldsymbol{\pi}_{\boldsymbol{y}}}{||\mathbf{A}^\top \boldsymbol{\pi}_{\boldsymbol{y}}||_1}$ as a Categorical distribution over $\boldsymbol{x}$, and hence, the summation term is equal to unity. Therefore, the update rule for $\overleftarrow{\beta}_{\boldsymbol{x}}$ is:

$$\overleftarrow{\beta}_{\boldsymbol{x}} = \overleftarrow{\beta}_{\boldsymbol{y}} \, ||\mathbf{A}^\top \boldsymbol{\pi}_{\boldsymbol{y}}||_1,$$

where $||\cdot||_1$ denotes the ℓ^1-norm of a vector.

IV. Gaussian node

The node function is

$$f(\boldsymbol{x}, \boldsymbol{y}, \mathbf{V}) = \mathcal{N}(\boldsymbol{y} \mid \boldsymbol{x}, \mathbf{V}) = \frac{1}{|2\pi \mathbf{V}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{x})^\top \mathbf{V}^{-1} (\boldsymbol{y} - \boldsymbol{x})\right)$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are real-valued vectors, and $\mathbf{V}$ is the covariance matrix. In this derivation, we assume that $\mathbf{V}$ is known.

A. Forward message

Since the matrix $\mathbf{V}$ is known, we regard the Gaussian node as a conditional probability distribution function $p(\boldsymbol{y} \mid \boldsymbol{x})$. Thus

$$\vec{\beta}_{\boldsymbol{y}} = \vec{\beta}_{\boldsymbol{x}}, \tag{26}$$

where $\vec{\beta}_{\boldsymbol{y}}$ and $\vec{\beta}_{\boldsymbol{x}}$ are the scale factors of the messages $\vec{\mu}_{\boldsymbol{y}}$ and $\vec{\mu}_{\boldsymbol{x}}$, respectively.

B. Backward message

Due to the symmetry property of the quadratic term in the exponent of Gaussian distribution, we can also regard the node function as a conditional probability distribution function $p(\boldsymbol{x}|\boldsymbol{y})$, and thus we have

$$\overleftarrow{\beta}_{\boldsymbol{x}} = \overleftarrow{\beta}_{\boldsymbol{y}}.$$

V. Equality node

We first recall the sum-product rule for the equality node: if the incoming messages are $\vec{\mu}_x(x)$ and $\vec{\mu}_z(z)$, then the outgoing message on the edge y is $\vec{\mu}_y(y) = \vec{\mu}_x(y)\,\vec{\mu}_z(y)$.

A. Beta message

Recall the definition of Beta distribution:

$$\mathrm{Beta}(x \mid \alpha, \beta) = \frac{1}{B(\alpha,\beta)}\, x^{\alpha-1}\,(1-x)^{\beta-1},$$

where the random variable $x \in (0,1)$. Both $\alpha, \beta \in \mathbb{R}_{>0}$ are shape parameters. The normalizing constant is the inverse of the Beta function of α and β:

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

With the incoming messages

$$\begin{aligned}\vec{\mu}_x(x) &= \vec{\beta}_x\,\mathrm{Beta}(x \mid a_x, b_x),\\ \vec{\mu}_z(z) &= \vec{\beta}_z\,\mathrm{Beta}(z \mid a_z, b_z),\end{aligned}$$

the scale factor $\vec{\beta}_y$ of the message $\vec{\mu}_y$ is

$$\begin{aligned}\vec{\beta}_y &= \int_y \vec{\mu}_y(y)\mathrm{d}y = \int_y \vec{\mu}_x(y)\,\vec{\mu}_z(y)\mathrm{d}y = \vec{\beta}_x\vec{\beta}_z\int_y \vec{p}_x(y)\,\vec{p}_z(y)\mathrm{d}y\\ &= \vec{\beta}_x\vec{\beta}_z\frac{1}{B(a_x,b_x)\,B(a_z,b_z)}\int_y y^{a_x-1}(1-y)^{b_x-1}y^{a_z-1}(1-y)^{b_z-1}\mathrm{d}y\\ &= \vec{\beta}_x\vec{\beta}_z\frac{1}{B(a_x,b_x)\,B(a_z,b_z)}\int_y y^{a_x+a_z-2}(1-y)^{b_x+b_z-2}\mathrm{d}y\\ &= \vec{\beta}_x\vec{\beta}_z\frac{B(a_y,b_y)}{B(a_x,b_x)\,B(a_z,b_z)}\underbrace{\int_y \frac{1}{B(a_y,b_y)}y^{a_y-1}(1-y)^{b_y-1}\mathrm{d}y}_{=1,\text{Beta distribution}}\\ &= \vec{\beta}_x\vec{\beta}_z\frac{B(a_y,b_y)}{B(a_x,b_x)\,B(a_z,b_z)},\end{aligned}$$

where $\begin{cases} a_y &= a_x + a_z - 1\\ b_y &= b_x + b_z - 1\end{cases}$

B. Bernoulli message

Recall the definition of Bernoulli distribution:

$$\mathrm{Ber}(x \mid \pi) = \pi^x(1-\pi)^{1-x} \tag{27}$$

where the random variable $x \in \{0,1\}$, and $\pi \in (0,1)$.

With the incoming messages

$$\begin{aligned}\vec{\mu}_x(x) &= \vec{\beta}_x\,\mathrm{Ber}(x \mid \pi_x),\\ \vec{\mu}_z(z) &= \vec{\beta}_z\,\mathrm{Ber}(z \mid \pi_z),\end{aligned}$$

the scale factor $\vec{\beta}_y$ of the message $\vec{\mu}_y$ is

$$\begin{aligned}
\vec{\beta}_y &= \sum_{y=0}^{1} \vec{\mu}_y(y) = \sum_{y=0}^{1} \vec{\mu}_x(y)\vec{\mu}_z(y) = \vec{\beta}_x\vec{\beta}_z \sum_{y=0}^{1} \vec{p}_x(y)\vec{p}_z(y) \\
&= \vec{\beta}_x\vec{\beta}_z \sum_{y=0}^{1} \pi_x^y(1-\pi_x)^{1-y}\, \pi_z^y(1-\pi_z)^{1-y} \\
&= \vec{\beta}_x\vec{\beta}_z((1-\pi_x)(1-\pi_z)+\pi_x\pi_z)
\end{aligned}$$

C. Dirichlet message

Recall the definition of Dirichlet distribution:

$$\mathrm{Dir}(\boldsymbol{x} \mid \boldsymbol{\pi}) = \frac{1}{B(\boldsymbol{\pi})} \prod_{i=1}^{K} x_i^{\pi_i - 1}$$

where $K \in \mathbf{N}$ is the number of classes, $\boldsymbol{\pi} = [\pi_1, \ldots, \pi_K]^\top \in \mathbb{R}_{>0}^K$. The random variable $\boldsymbol{x} = [x_1, \ldots, x_K]^\top$ satisfies: $x_i \in (0,1)$ and $\sum_{i=1}^K x_i = 1$. The normalizing constant $B(\boldsymbol{\pi})$ is the multivariate beta function:

$$B(\boldsymbol{\pi}) = \frac{\prod_{i=1}^K \Gamma(\pi_i)}{\Gamma(\sum_{i=1}^K \pi_i)}$$

With the incoming messages

$$\begin{aligned}
\vec{\mu}_x(\boldsymbol{x}) &= \vec{\beta}_x\ \mathrm{Dir}(\boldsymbol{x} \mid \boldsymbol{\pi}_x), \\
\vec{\mu}_z(\boldsymbol{z}) &= \vec{\beta}_z\ \mathrm{Dir}(\boldsymbol{z} \mid \boldsymbol{\pi}_z),
\end{aligned}$$

the scale factor $\vec{\beta}_y$ of the message $\vec{\mu}_y$ is

$$\begin{aligned}
\vec{\beta}_y &= \int \vec{\mu}_y(\boldsymbol{y})\mathrm{d}\boldsymbol{y} = \int \vec{\mu}_x(\boldsymbol{y})\, \vec{\mu}_z(\boldsymbol{y})\mathrm{d}\boldsymbol{y} = \vec{\beta}_x\vec{\beta}_z \int \vec{p}_x(\boldsymbol{y})\vec{p}_z(\boldsymbol{y})\mathrm{d}\boldsymbol{y} \\
&= \vec{\beta}_x\vec{\beta}_z \frac{1}{B(\boldsymbol{\pi}_x)\, B(\boldsymbol{\pi}_z)} \int \prod_{i=1}^{K} \underbrace{y_i^{\pi_{x_i}-1}\, y_i^{\pi_{z_i}-1}}_{y_i^{\pi_{x_i}+\pi_{z_i}-2}} \mathrm{d}\boldsymbol{y} \\
&= \vec{\beta}_x\vec{\beta}_z \frac{B(\boldsymbol{\pi}_y)}{B(\boldsymbol{\pi}_x)\, B(\boldsymbol{\pi}_z)} \underbrace{\int \frac{1}{B(\boldsymbol{\pi}_y)} \prod_{i=1}^{K} y_i^{\pi_{y_i}-1}\mathrm{d}\boldsymbol{y}}_{=1,\text{Dirichlet distribution}} \\
&= \vec{\beta}_x\vec{\beta}_z \frac{B(\boldsymbol{\pi}_y)}{B(\boldsymbol{\pi}_x)\, B(\boldsymbol{\pi}_z)},
\end{aligned}$$

where $\boldsymbol{\pi}_y = \boldsymbol{\pi}_x + \boldsymbol{\pi}_z - \mathbf{1}$.

D. Categorical message

Recall the definition of Categorical distribution:

$$\mathrm{Cat}(\boldsymbol{x}|\boldsymbol{\pi}) = \prod_{k=1}^{K} \pi_k^{x_k}$$

where $\boldsymbol{x} = [x_1, \ldots, x_K]^\top$ is a one-hot coded vector, i.e. only one entry of $\boldsymbol{x}$ is 1 and the others are zero, and $\boldsymbol{\pi} = [\pi_1, \ldots, \pi_K]^\top$ satisfies $\pi_k \in [0,1]$ and $\sum_{k=1}^K \pi_k = 1$. Since $\boldsymbol{x}$ is one-hot coded, we rewrite the categorical distribution as follows:

$$\mathrm{Cat}(\boldsymbol{x}|\boldsymbol{\pi}) = \boldsymbol{\pi}^T\boldsymbol{x}.$$

With the incoming messages

$$\begin{aligned}
\vec{\mu}_x(\boldsymbol{x}) &= \vec{\beta}_x \mathrm{Cat}(\boldsymbol{x} \mid \boldsymbol{\pi}_x), \\
\vec{\mu}_z(\boldsymbol{z}) &= \vec{\beta}_z \mathrm{Cat}(\boldsymbol{z} \mid \boldsymbol{\pi}_z),
\end{aligned}$$

the scale factor $\vec{\beta}_y$ of the message $\vec{\mu}_y$ is

$$\begin{aligned}\vec{\beta}_y &= \sum_{\boldsymbol{y}} \vec{\mu}_y(\boldsymbol{y}) = \sum_{\boldsymbol{y}} \vec{\mu}_x(\boldsymbol{y})\vec{\mu}_z(\boldsymbol{y}) = \vec{\beta}_x\vec{\beta}_z \sum_{\boldsymbol{y}} \vec{p}_x(\boldsymbol{y})\vec{p}_z(\boldsymbol{y}) \\ &= \vec{\beta}_x\vec{\beta}_z \sum_{\boldsymbol{y}} \mathrm{Cat}(\boldsymbol{y} \mid \boldsymbol{\pi}_x)\boldsymbol{\pi}_z^\top \boldsymbol{y} = \vec{\beta}_x\vec{\beta}_z \boldsymbol{\pi}_z^\top \mathbb{E}_{\mathrm{Cat}(\boldsymbol{y}\mid\boldsymbol{\pi}_x)}[\boldsymbol{y}] \\ &= \vec{\beta}_x\vec{\beta}_z\, \boldsymbol{\pi}_x^\top \boldsymbol{\pi}_z\end{aligned}$$

E. Gamma message

Recall the definition of Gamma distribution:

$$\mathrm{Gam}(x \mid a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx), \tag{28}$$

where $a \in \mathbb{R}_{>0}$ and $b \in \mathbb{R}_{>0}$ are the shape and the rate parameters, respectively, and both are positive real numbers.

With the incoming messages

$$\begin{aligned}\vec{\mu}_x(x) &= \vec{\beta}_x \,\mathrm{Gam}(x \mid a_x, b_x), \\ \vec{\mu}_z(z) &= \vec{\beta}_z \,\mathrm{Gam}(z \mid a_z, b_z),\end{aligned}$$

the scale factor $\vec{\beta}_y$ of the message $\vec{\mu}_y$ is

$$\begin{aligned}\vec{\beta}_y &= \int \vec{\mu}_y(y)\mathrm{d}y = \int \vec{\mu}_x(y)\, \vec{\mu}_z(y)\mathrm{d}y = \vec{\beta}_x\vec{\beta}_z \int \vec{p}_x(y)\vec{p}_z(y)\mathrm{d}y \\ &= \vec{\beta}_x\vec{\beta}_z \frac{b_x^{a_x} b_z^{a_z}}{\Gamma(a_x)\Gamma(a_z)} \int y^{a_x-1} \exp\left(-b_x y\right) y^{a_z-1} \exp\left(-b_z y\right)\mathrm{d}y \\ &= \vec{\beta}_x\vec{\beta}_z \frac{b_x^{a_x} b_z^{a_z}}{\Gamma(a_x)\Gamma(a_z)} \int y^{a_x+a_z-2} \exp\left(-(b_x+b_z)y\right)\mathrm{d}y \\ &= \vec{\beta}_x\vec{\beta}_z \frac{b_x^{a_x} b_z^{a_z}}{\Gamma(a_x)\Gamma(a_z)} \frac{\Gamma(a_y)}{b_y^{a_y}} \underbrace{\int \frac{b_y^{a_y}}{\Gamma(a_y)} y^{a_y-1} \exp\left(-b_y y\right)\mathrm{d}y}_{=1,\ \text{Gamma distribution}} \\ &= \vec{\beta}_x\vec{\beta}_z \frac{\Gamma(a_y)}{\Gamma(a_x)\Gamma(a_z)} \frac{b_x^{a_x} b_z^{a_z}}{b_y^{a_y}},\end{aligned}$$

where $\begin{cases} a_y &= a_x + a_z - 1 \\ b_y &= b_x + b_z \end{cases}$

F. Wishart message

Recall the definition of Wishart distribution:

$$\mathrm{W}(\mathbf{X}|\mathbf{V}, n) = \frac{|\mathbf{X}|^{\frac{n-p-1}{2}} \exp\{-\frac{1}{2}\mathrm{tr}(\mathbf{V}^{-1}\mathbf{X})\}}{2^{\frac{np}{2}} |\mathbf{V}|^{\frac{n}{2}} \Gamma_p\left(\frac{n}{2}\right)},$$

where $\mathbf{V} \in \mathbb{R}^{p\times p}$ is the scale matrix, and $n > p-1$ is the degree of freedom.

With the incoming messages

$$\begin{aligned}\vec{\mu}_x(\mathbf{X}) &= \vec{\beta}_x \ \mathrm{W}(\mathbf{X} \mid \mathbf{V}_x, n_x), \\ \vec{\mu}_z(\mathbf{Z}) &= \vec{\beta}_z \ \mathrm{W}(\mathbf{Z} \mid \mathbf{V}_z, n_z),\end{aligned}$$

the scale factor of the message $\vec{\mu}_y(\mathbf{y})$ is

$$
\begin{aligned}
\vec{\beta}_y &= \int \vec{\mu}_y(\mathbf{y})\mathrm{d}\mathbf{y} = \int \vec{\mu}_x(\mathbf{y})\vec{\mu}_z(\mathbf{y})\mathrm{d}\mathbf{y} \\
&= \vec{\beta}_x\vec{\beta}_z \int \mathrm{W}(\mathbf{y} \mid \mathbf{V}_x, n_x)\mathrm{W}(\mathbf{y} \mid \mathbf{V}_z, n_z)\mathrm{d}\mathbf{y} \\
&= \frac{\vec{\beta}_x\vec{\beta}_z}{A} \int |\mathbf{y}|^{\frac{n_X+n_z-2p-2}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}((\mathbf{V}_x^{-1}+\mathbf{V}_z^{-1})\mathbf{y})\right)\mathrm{d}\mathbf{y} \\
&= \frac{\vec{\beta}_x\vec{\beta}_z}{A} B \underbrace{\int \frac{1}{B} |\mathbf{y}|^{\frac{n_y-p-1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}(\mathbf{V_y}^{-1}\mathbf{y})\right)\mathrm{d}\mathbf{y}}_{=1,\text{ Wishart distribution}} \\
&= \frac{\vec{\beta}_x\vec{\beta}_z}{A} B,
\end{aligned}
$$

where

$$
\begin{aligned}
A &= 2^{\frac{p(n_x+n_z)}{2}} |\mathbf{V}_x|^{\frac{n_x}{2}} |\mathbf{V}_z|^{\frac{n_z}{2}} \Gamma_p\left(\frac{n_x}{2}\right) \Gamma_p\left(\frac{n_z}{2}\right) \\
B &= 2^{\frac{p\,n_y}{2}} |\mathbf{V}_y|^{\frac{n_y}{2}} \Gamma_p\left(\frac{n_y}{2}\right) \\
n_y &= n_x + n_z - p - 1 \\
\mathbf{V}_y &= (\mathbf{V}_x^{-1} + \mathbf{V}_z^{-1})^{-1}
\end{aligned}
$$

References

[1] C. Reller, “State-Space Methods in Statistical Signal Processing: New Ideas and Applications,” Ph.D. dissertation, ETH Zurich, 2012.