Supplementary material (MLSP-2021): Message Passing-Based Inference in the Gamma Mixture Model

May 27, 2021

In this supplementary material the variational messages for the Gamma mixture node are derived.

1 Model specification

The likelihood function of the Gamma mixture node is specified by

$$f(x_k, \boldsymbol{s}_k, \boldsymbol{a}, \boldsymbol{b}) = p(x_k|\boldsymbol{s}_k, \boldsymbol{a}, \boldsymbol{b}) = \prod_{m=1}^{M} \Gamma(x_k|a_m, b_m)^{s_{km}}, \tag{1}$$

where $\Gamma(x_k|a_m, b_m)$ specifies the Gamma distribution for x_k with shape and rate parameters a_m and b_m, respectively. $\boldsymbol{a} \triangleq [a_1, \ldots, a_M]$ and $\boldsymbol{b} \triangleq [b_1, \ldots, b_M]$ are vectors of the parameters of the Gamma distributions such that $a_m, b_m \in \mathbb{R}_{>0}$ for every $m = 1, \ldots, M$. For each observation x_k we have a corresponding latent selector variable $\boldsymbol{s}_k$ comprising a 1-of-M binary vector with elements $s_{km} \in \{0, 1\}$, which are constrained by $\sum_m s_{km} = 1$.

We assume a mean-field factorization around the Gamma mixture node as

$$q(x_k, \boldsymbol{s}_k, \boldsymbol{a}, \boldsymbol{b}) = q(x_k)q(\boldsymbol{s}_k)q(\boldsymbol{a})q(\boldsymbol{b}) \tag{2}$$

where $q(\boldsymbol{a}) = \prod_{m=1}^{M} q(a_m)$ and $q(\boldsymbol{b}) = \prod_{m=1}^{M} q(b_m)$.

We assume the following functional forms for the approximate posterior marginals:

$$\begin{aligned}
q(x_k) &= \Gamma(x_k \mid \hat{\alpha}_k^{(x)}, \hat{\beta}_k^{(x)}) \quad \hat{\alpha}_k^{(x)},\ \hat{\beta}_k^{(x)} \in \mathbb{R}_{>0} \\
q(\boldsymbol{s}_k) &= \prod_{m=1}^{M} \hat{\pi}_m^{s_{km}} \text{ such that } \sum_{m=1}^{M} \hat{\pi}_m = 1 \\
q(a_m) &= \delta(a_m - \hat{a}_m) \text{ or } q(a_m) = \Gamma(a_m \mid \hat{\alpha}_m^{(a)}, \hat{\beta}_m^{(a)}) \quad \hat{\alpha}_k^{(a)},\ \hat{\beta}_k^{(a)} \in \mathbb{R}_{>0} \\
q(b_m) &= \Gamma(b_m \mid \hat{\alpha}_m^{(b)}, \hat{\beta}_m^{(b)}) \quad \hat{\alpha}_k^{(b)},\ \hat{\beta}_k^{(b)} \in \mathbb{R}_{>0}
\end{aligned}$$

Here the marginal $q(a_m)$ has two forms, depending on the inference methodology. For expectation-maximization the marginal follows a Dirac delta function and for moment matching the marginal follows a Gamma function. As a result, we will not express the expectations related to this distributions in terms of the corresponding parameters.

2 Mathematical identities

In this section we will derive some relatively common expectations to simplify derivations later on. We use $C \in \mathbb{R}$ to denote a constant. Consider the following expectation where $q(x) = \Gamma(x \mid \alpha, \beta)$:

$$\begin{aligned}
\mathrm{E}_{q(x)}\left[\ln \Gamma(x)\right] &= \mathrm{E}_{q(x)}\left[-\ln(x) + \ln \Gamma(x+1)\right] \qquad (\Gamma(x+1) = x\Gamma(x)) \\
&\approx -\mathrm{E}_{q(x)}\left[\ln(x)\right] + \mathrm{E}_{q(x)}\left[\ln\left(\sqrt{2\pi x}\left(\frac{x}{e}\right)^x\right)\right] \quad \textit{Stirling's approximation} \\
&= -\mathrm{E}_{q(x)}\left[\ln(x)\right] + \mathrm{E}_{q(x)}\left[\frac{1}{2}\ln(2\pi x) + x(\ln(x) - 1)\right] \\
&= \frac{1}{2}\ln(2\pi) - \frac{1}{2}\mathrm{E}_{q(x)}\left[\ln(x)\right] + \mathrm{E}_{q(x)}\left[x \ln(x)\right] - \mathrm{E}_{q(x)}\left[x\right] \\
&= \frac{1}{2}\ln(2\pi) - \frac{1}{2}(\psi(\alpha) - \ln(\beta)) - \frac{\alpha}{\beta} + \mathrm{E}_{q(x)}\left[x \ln(x)\right]
\end{aligned} \tag{3}$$

where Stirling's approximation approximates the Γ-function (especially well for $x \geq 1$). The term $\mathrm{E}_{q(x)}[x \ln(x)]$ can be determined as

$$\begin{aligned}
\mathrm{E}_{q(x)}[x \ln(x)] &= \int_0^\infty q(x) x \ln(x) \mathrm{d}x \\
&= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} x \ln(x) \mathrm{d}x \\
&= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\beta} \int_0^\infty \frac{\beta^{(\alpha+1)}}{\Gamma(\alpha)} x^{(\alpha+1)-1} e^{-\beta x} \ln(x) \mathrm{d}x \\
&= \frac{\alpha}{\beta} \mathrm{E}_{x \sim \Gamma(\alpha+1,\beta)}\left[\ln(x)\right] \\
&= \frac{\alpha}{\beta}(\psi(\alpha+1) - \ln(\beta))
\end{aligned} \tag{4}$$

Concluding

$$\mathrm{E}_{q(x)}\left[\ln \Gamma(x)\right] = \frac{1}{2}\ln(2\pi) - \frac{1}{2}(\psi(\alpha) - \ln(\beta)) + \frac{\alpha}{\beta}\left(-1 + \psi(\alpha+1) - \ln(\beta)\right) \tag{5}$$

Here $\psi(\cdot)$ denotes the digamma function.

3 Message $\vec{\nu}(x_k)$

The message $\ln \vec{\nu}(x_k)$ can be determined

$$
\begin{aligned}
\ln \vec{\nu}(x_k) &= \mathrm{E}_{q(\boldsymbol{s}_k)q(\boldsymbol{a})q(\boldsymbol{b})}\left[\ln\left(\prod_{m=1}^{M}\Gamma(x_k|a_m,b_m)^{s_{km}}\right)\right] + C \\
&= \mathrm{E}_{q(\boldsymbol{s}_k)q(\boldsymbol{a})q(\boldsymbol{b})}\left[\sum_{m=1}^{M} s_{km}\ln(\Gamma(x_k \mid a_m,b_m))\right] + C \\
&= \sum_{m=1}^{M}\mathrm{E}_{q(s_{km})}\left[s_{km}\right]\mathrm{E}_{q(a_m)q(b_m)}\left[\ln(\Gamma(x_k \mid a_m,b_m))\right] + C \\
&= \sum_{m=1}^{M}\hat{\pi}_m \mathrm{E}_{q(a_m)q(b_m)}\left[\ln\left(\frac{b_m^{a_m}}{\Gamma(a_m)}x_k^{a_m-1}e^{-b_m x_k}\right)\right] + C \\
&= \sum_{m=1}^{M}\hat{\pi}_m \mathrm{E}_{q(a_m)q(b_m)}\left[-\ln(\Gamma(a_m)) + a_m\ln(b_m) + (a_m-1)\ln(x_k) - b_m x_k\right] + C \\
&= \sum_{m=1}^{M}\hat{\pi}_m\left(\mathrm{E}_{q(a_m)q(b_m)}\left[(a_m-1)\ln(x_k) - b_m x_k\right]\right) \\
&= \sum_{m=1}^{M}\hat{\pi}_m\left(\left(\mathrm{E}_{q(a_m)}\left[a_m\right]-1\right)\ln(x_k) - \mathrm{E}_{q(b_m)}\left[b_m\right]x_k\right) + C \\
&= \sum_{m=1}^{M}\hat{\pi}_m\left(\left(\mathrm{E}_{q(a_m)}\left[a_m\right]-1\right)\ln(x_k) - \frac{\hat{\alpha}_m^{(b)}}{\hat{\beta}_m^{(b)}}x_k\right) + C \\
&= \left(\sum_{m=1}^{M}\hat{\pi}_m \mathrm{E}_{q(a_m)}\left[a_m\right]-1\right)\ln(x_k) - \left(\sum_{m=1}^{M}\hat{\pi}_m\frac{\hat{\alpha}_m^{(b)}}{\hat{\beta}_m^{(b)}}\right)x_k + C
\end{aligned}
\tag{6}
$$

From this the variational message $\vec{\nu}(x_k)$ can be determined as

$$
\boxed{\vec{\nu}(x_k) \ \propto \Gamma\left(x_k \,\middle|\, \sum_{m=1}^{M}\hat{\pi}_m \mathrm{E}_{q(a_m)}\left[a_m\right],\ \sum_{m=1}^{M}\hat{\pi}_m\frac{\hat{\alpha}_m^{(b)}}{\hat{\beta}_m^{(b)}}\right)}
\tag{7}
$$

4 Message $\overleftarrow{\nu}(s_k)$

The message $\ln \overleftarrow{\nu}(\boldsymbol{s}_k)$ can be determined as

$$
\begin{aligned}
\ln \overleftarrow{\nu}(\boldsymbol{s}_k) &= \mathrm{E}_{q(\boldsymbol{a})q(\boldsymbol{b})q(x_k)}\left[\ln\left(\prod_{m=1}^{M} \Gamma(x_k|a_m,b_m)^{s_{km}}\right)\right] + C \\
&= \mathrm{E}_{q(\boldsymbol{a})q(\boldsymbol{b})q(x_k)}\left[\sum_{m=1}^{M} s_{km} \ln\left(\Gamma(x_k|a_m,b_m)\right)\right] + C \\
&= \mathrm{E}_{q(\boldsymbol{a})q(\boldsymbol{b})q(x_k)}\left[\sum_{m=1}^{M} s_{km} \ln\left(\frac{b_m^{a_m} x_k^{a_m-1}}{\Gamma(a_m)} \exp\left(-b_m x_k\right)\right)\right] + C \\
&= \sum_{m=1}^{M} s_{km} \mathrm{E}_{q(\boldsymbol{a})q(\boldsymbol{b})q(x_k)}\left[a_m \ln b_m + (a_m-1)\ln x_k - \ln\Gamma(a_m) - b_m x_k\right] + C \\
&= \sum_{m=1}^{M} s_{km} \left(\mathrm{E}_{q(a_m)q(b_m)}[a_m \ln b_m] + \mathrm{E}_{q(a_m)q(x_k)}(a_m-1)\ln x_k\right) \\
&\quad + \sum_{m=1}^{M} s_{km} \left(-\mathrm{E}_{q(a_m)}\left[\ln\Gamma(a_m)\right] - \mathrm{E}_{q(x_k)q(b_m)}[b_m x_k]\right) + C \\
&= \sum_{m=1}^{M} s_{km} \ln \rho_{km} + C
\end{aligned}
$$

From this the variational message $\overleftarrow{\nu}(\boldsymbol{s}_k)$ can be determined as

$$
\boxed{\overleftarrow{\nu}(\boldsymbol{s}_k) \propto \exp \sum_{m=1}^{M} s_{km} \ln \rho_{km} = \prod_{m=1}^{M} \rho_{km}^{s_{km}}}
$$

where

$$
\rho_{km} = \exp\left\{\mathrm{E}_{q(a_m)q(b_m)}[a_m \ln b_m] + \mathrm{E}_{q(a_m)q(x_k)}(a_m-1)\ln x_k - \mathrm{E}_{q(a_m)}\left[\ln\Gamma(a_m)\right] - \mathrm{E}_{q(x_k)q(b_m)}[b_m x_k]\right\}
$$

In order to ensure that the message will be a proper distribution, the event probabilities have to sum to 1. Hence, all event probabilities are normalized and the message becomes:

$$
\overleftarrow{\nu}(\boldsymbol{s}_k) = \prod_{m=1}^{M} \left(\frac{\rho_{km}}{\sum_m \rho_{km}}\right)^{s_{km}} \tag{8}
$$

The individual expectations of ρ_{km} can be calculated as

$$
\mathrm{E}_{q(a_m)q(b_m)}\left[a_m \ln b_m\right] = \mathrm{E}_{q(a_m)}[a_m] \left(\psi(\hat{\alpha}_m^{(b)}) - \ln(\hat{\beta}_m^{(b)})\right) \tag{9a}
$$

$$
\mathrm{E}_{q(a_m)q(x_k)}[(a_m-1)\ln x_k] = \left(\mathrm{E}_{q(a_m)}[a_m] - 1\right) \left(\psi(\hat{\alpha}_m^{(x)}) - \ln(\hat{\beta}_m^{(x)})\right) \tag{9b}
$$

$$
\mathrm{E}_{q(x_k)q(b_m)}[b_m x_k] = \frac{\hat{\alpha}_m^{(b)} \hat{\alpha}_k^{(x)}}{\hat{\beta}_m^{(b)} \hat{\beta}_k^{(x)}} \tag{9c}
$$

The expectation $\mathrm{E}_{q(a_m)}[\ln \Gamma(a_m)]$ has been derived in Section 2.

5 Message $\breve{\nu}(b_m)$

The message $\breve{\nu}(b_m)$ can be determined as

$$\begin{aligned}
\ln \breve{\nu}(b_m) &= \mathrm{E}_{\backslash q(b_m)}\left[\ln\left(\prod_{m=1}^{M} \Gamma(x_k|a_m, b_m)^{s_{km}}\right)\right] + C \\
&= \mathrm{E}_{\backslash q(b_m)}\left[\sum_{m=1}^{M} s_{km} \ln(\Gamma(x_k \mid a_m, b_m))\right] + C \\
&= \mathrm{E}_{q(\boldsymbol{s}_k)}\left[s_{km}\right] \mathrm{E}_{q(a_m)q(x_k)}\left[\ln\left(\Gamma(x_k \mid a_m, b_m)\right)\right] + C \\
&= \hat{\pi}_m \mathrm{E}_{q(a_m)q(x_k)}\left[-\ln(\Gamma(a_m)) + a_m \ln(b_m) + (a_m - 1)\ln(x_k) - b_m x_k\right] + C \\
&= \hat{\pi}_m \left(\mathrm{E}_{q(a_m)q(x_k)}\left[a_m \ln(b_m) - b_m x_k\right]\right) + C \\
&= \hat{\pi}_m \left(\mathrm{E}_{q(a_m)}\left[a_m\right] \ln(b_m) - b_m \mathrm{E}_{q(x_k)}\left[x_k\right]\right) + C \\
&= \hat{\pi}_m \left(\mathrm{E}_{q(a_m)}\left[a_m\right] \ln(b_m) - \frac{\hat{\alpha}_k^{(x)}}{\hat{\beta}_k^{(x)}} b_m\right) + C \\
&= \left(\hat{\pi}_m \mathrm{E}_{q(a_m)}\left[a_m\right]\right) \ln(b_m) - \left(\hat{\pi}_m \frac{\hat{\alpha}_k^{(x)}}{\hat{\beta}_k^{(x)}}\right) b_m + C
\end{aligned} \tag{10}$$

$$\boxed{\breve{\nu}(b_m) \propto \Gamma\left(b_m \middle| 1 + \hat{\pi}_m \mathrm{E}_{q(a_m)}\left[a_m\right],\ \hat{\pi}_m \frac{\hat{\alpha}_k^{(x)}}{\hat{\beta}_k^{(x)}}\right)} \tag{11}$$

6 Message $\overleftarrow{\nu}(a_m)$

$$
\begin{aligned}
\ln \overleftarrow{\nu}(a_m) &= \mathrm{E}_{\backslash q(a_m)} \ln \left[\prod_{m=1}^{M} \Gamma(x_k|a_m, b_m)^{s_{km}} \right] + C \\
&= \mathrm{E}_{\backslash q(a_m)} \left[\sum_{m=1}^{M} s_{km} \ln \left(\frac{b_m^{a_m} x^{a_m - 1}}{\Gamma(a_m)} \exp\left(-b_m x\right) \right) \right] + C \\
&= \mathrm{E}_{\backslash q(a_m)} \left[\sum_{m=1}^{M} s_{km} \left(a_m \ln(b_m) + (a_m - 1)\ln(x_k) - \ln(\Gamma(a_m)) - b_m x_k \right) \right] + C \\
&= \hat{\pi}_m \left[a_m \mathrm{E}_{q(b_m)}[\ln(b_m)] + a_m \mathrm{E}_{q(x_k)}[\ln(x_k)] - \ln(\Gamma(a_m)) \right] + C \\
&= \hat{\pi}_m \left[a_m \underbrace{\left(\psi(\hat{\alpha}_m^{(b)}) - \ln(\hat{\beta}_m^{(b)}) + \psi(\alpha_k^{(x)}) - \ln(\beta_k^{(x)}) \right)}_{\zeta_{km}} - \ln(\Gamma(a_m)) \right] + C \\
&= \hat{\pi}_m \left(a_m \zeta_{km} - \ln \Gamma(a_m) \right) + C
\end{aligned}
$$

From this the variational message $\overleftarrow{\nu}(a_m)$ can be determined as

$$
\boxed{\overleftarrow{\nu}(a_m) \propto \exp\left(\hat{\pi}_m \left(a_m \zeta_{km} - \ln \Gamma(a_m)\right)\right)}
$$

7 Local variational free energy

The local variational free energy of the Gamma mixture node can be computed as follows:

$$
F[q] = \underbrace{-\mathrm{E}_{q(x_k)q(\boldsymbol{a})q(\boldsymbol{b})q(\boldsymbol{s}_k)} \Big[\ln(p(x_k|\boldsymbol{s}_k, \boldsymbol{a}, \boldsymbol{b})) \Big]}_{\text{Average energy}} + \underbrace{\mathrm{E}_{q(x_k)q(\boldsymbol{a})q(\boldsymbol{b})q(\boldsymbol{s}_k)} \Big[\ln(q(x_k)q(\boldsymbol{s}_k)q(\boldsymbol{a})q(\boldsymbol{b}) \Big])}_{\text{-Entropy}}
$$

Since the entropy of the incoming marginals can easily be computed, let us focus on the average energy term

$$
\begin{aligned}
&\mathrm{E}_{q(x_k)q(\boldsymbol{a})q(\boldsymbol{b})q(\boldsymbol{s}_k)} \Big[\ln(p(x_k|\boldsymbol{s}_k, \boldsymbol{a}, \boldsymbol{b})) \Big] = \mathrm{E}_{q(x_k)q(\boldsymbol{a})q(\boldsymbol{b})q(\boldsymbol{s}_k)} \left[\ln \left(\prod_{m=1}^{M} \Gamma(x_k|a_m, b_m)^{s_{km}} \right) \right] \\
&= \mathrm{E}_{q(x_k)q(\boldsymbol{a})q(\boldsymbol{b})q(\boldsymbol{s}_k)} \left[\sum_{m=1}^{M} s_{km} \ln\left(\Gamma(x_k|a_m, b_m)\right) \right] \\
&= \sum_{m=1}^{M} \hat{\pi}_m \mathrm{E}_{q(\boldsymbol{a})q(\boldsymbol{b})q(x_k)} \left[a_m \ln b_m + (a_m - 1)\ln(x_k) - \ln(\Gamma(a_m)) - b_m x_k \right]
\end{aligned}
$$

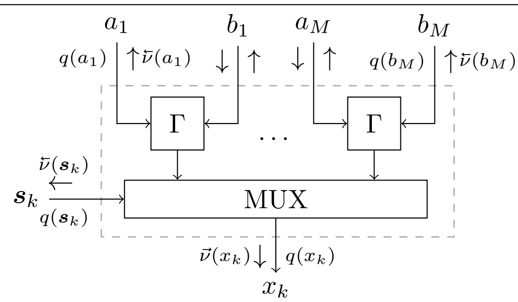
The required expectations are given in (3) and(9).

8 Summary

An overview of the derived messages is given in Table 1.

Table 1: Table containing (a) the Forney-style factor graph representation of the Gamma mixture node. The node indicated by MUX represents a multiplexer node, which selects the mixture component. (b) An overview of the chosen approximate posterior distributions. Here the $\hat{\cdot}$ accent refers to the parameters of these distributions. The choice of functional form for $q(a_m)$ will depend on the approximation method. (c) The derived messages for the Gamma mixture node. The definitions of ζ_{km} and ρ_{km} are presented in the the material below.

Factor graph



Marginals	**Functional form**
$q(a_m)$	$\delta(a_m - \hat{a}_m)$ or $\Gamma\left(a_m \mid \hat{\alpha}_m^{(a)}, \hat{\beta}_m^{(a)}\right)$ $\quad \hat{\alpha}_m^{(a)}, \hat{\beta}_m^{(a)} \in \mathbb{R}_{>0}$
$q(b_m)$	$\Gamma\left(b_m \mid \hat{\alpha}_m^{(b)},\ \hat{\beta}_m^{(b)}\right)$ $\quad \hat{\alpha}_m^{(b)}, \hat{\beta}_m^{(b)} \in \mathbb{R}_{>0}$
$q(\boldsymbol{s}_k)$	$\prod_{m=1}^{M} \hat{\pi}_m^{s_{km}}$ such that $\sum_{m=1}^{M} \hat{\pi}_m = 1$
$q(x_k)$	$\Gamma\left(x_k \mid \hat{\alpha}_k^{(x)},\ \hat{\beta}_k^{(x)}\right)$ $\quad \hat{\alpha}_k^{(x)}, \hat{\beta}_k^{(x)} \in \mathbb{R}_{>0}$
Messages	**Functional form**
$\overleftarrow{\nu}(a_m)$	$\exp\Big(\hat{\pi}_k \left(a_m \zeta_{km} - \log\Gamma(a_m)\right)\Big)$
$\overleftarrow{\nu}(b_m)$	$\Gamma\left(b_m \middle\vert 1 + \hat{\pi}_m \mathrm{E}\left[a_m\right],\ \hat{\pi}_m \frac{\hat{\alpha}_k^{(x)}}{\hat{\beta}_k^{(x)}}\right)$
$\overleftarrow{\nu}(\boldsymbol{s}_k)$	$\prod_{m=1}^{M} \rho_{km}^{s_{km}}$ such that $\sum_{m=1}^{M} \rho_{km} = 1$
$\overrightarrow{\nu}(x_k)$	$\Gamma\left(x_k \middle\vert \sum_{m=1}^{M} \hat{\pi}_m \mathrm{E}\left[a_m\right],\ \sum_{m=1}^{M} \hat{\pi}_m \frac{\hat{\alpha}_m^{(b)}}{\hat{\beta}_m^{(b)}}\right)$