1 Piman-Yor Diffusion Trees

1.1 Modelling of PYDT

$$
\begin{aligned}
\alpha &\sim \mathrm{Beta}(\alpha|a_\alpha, b_\alpha) && (1)\\
\theta &\sim \mathrm{G}(\theta|a_\theta, b_\theta) && (2)\\
p(\mathcal{T}|\alpha,\theta) &= \prod_{v\in\mathcal{I}} \frac{a(t_v)\prod_{k=3}^{K_v}[\theta+(k-1)\alpha]\prod_{l=1}^{K_v}\Gamma(n_l^v-\alpha)}{\Gamma(m(v)+\theta)\Gamma(1-\alpha)^{K_v-1}} && (3)\\
c &\sim \mathrm{G}(c|a_c, b_c) && (4)\\
1/\sigma^2 &\sim \mathrm{G}(1/\sigma^2|a_{\sigma^2}, b_{\sigma^2}) && (5)\\
p(t_v|c,\mathcal{T}) &= c(1-t_v)^{cJ^{\theta,\alpha}_{\mathbf{n}_v}-1} && (6)\\
\mathcal{N}(\mathbf{z}_v|\mathbf{z}_u, \sigma^2(t_v-t_u)\mathbf{I}) &= (2\pi\sigma^2(t_v-t_u))^{-\frac{D}{2}}\exp\Big(-\frac{1}{2}\frac{\parallel \mathbf{z}_v-\mathbf{z}_u\parallel^2}{\sigma^2(t_v-t_u)}\Big) && (7)\\
\mathcal{N}(\mathbf{z}_k|\mathbf{z}_v, \sigma^2(t_k-t_v)\mathbf{I}) &= (2\pi\sigma^2(t_k-t_v))^{-\frac{D}{2}}\exp\Big(-\frac{1}{2}\frac{\parallel \mathbf{z}_k-\mathbf{z}_v\parallel^2}{\sigma^2(t_k-t_v)}\Big) && (8)
\end{aligned}
$$

1.2 EM algorithm for PYDT

$$
\begin{aligned}
\ln P(\mathbf{X}|\boldsymbol{\theta}) &= \ln\Big\{\sum_{\mathbf{Z}} P(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})\Big\}\\
&= \ln\Big\{\sum_{\mathbf{Z}} q(\mathbf{Z})\frac{P(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})}{q(\mathbf{Z})}\Big\}\\
&= \sum_{\mathbf{Z}} q(\mathbf{Z})\ln\Big\{\frac{P(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})}{q(\mathbf{Z})}\Big\} + \sum_{\mathbf{Z}} q(\mathbf{Z})\ln\Big\{\frac{q(\mathbf{Z})}{P(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta})}\Big\} && (9)
\end{aligned}
$$

The main procedure of the algorithm is as follows.
1) Find $q(\mathbf{Z})$ which minimizes the KL divergence.
2) Take a gradient of the ELBO w.r.t. $\boldsymbol{\theta}$ using $q(\mathbf{Z})$ found in the step 1).
The step 2) can be written as below.

$$
\begin{aligned}
Q(\boldsymbol{\theta},\boldsymbol{\theta}') &= \sum_{\mathbf{Z}} q(\mathbf{Z})\ln\Big\{\frac{P(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})}{q(\mathbf{Z})}\Big\}\\
&= \sum_{\mathbf{Z}} q(\mathbf{Z})\ln P(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta}) - \sum_{\mathbf{Z}} q(\mathbf{Z})\ln q(\mathbf{Z}) && (10)\\
\frac{\partial Q(\boldsymbol{\theta},\boldsymbol{\theta}')}{\partial\boldsymbol{\theta}} &= \sum_{\mathbf{Z}} q(\mathbf{Z})\frac{\partial \ln P(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\\
&= \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}')\frac{\partial \ln P(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} && (11)
\end{aligned}
$$

An important point is that the EM algorithm is an algorithm designed for maximum likelihood estimation. Hence, if priors of parameters are combined in the model, EM algorithm becomes the maximum-a-posteriori EM algorithm (MAP-EM) (Gupta and Chen, 2011). Here is an example (Chen and John, 2010).

$$\begin{aligned}
\mu_j &= \mu + (j-1)\Delta\mu, j = 1, ..., k & (12)\\
\sigma_j^2 &= \sigma^2, j = 1, ..., k & (13)\\
p(y_j) &= \sum_{j=1}^{k} w_j \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \frac{(y_i - \mu - (j-1)\Delta\mu)^2}{2\sigma^2} \right) & (14)\\
\sigma^2 &\sim \text{Inv-Gamma}\left(\frac{\nu}{2}, \frac{\zeta^2}{2} \right) & (15)\\
\Delta\mu|\sigma^2 &\sim \mathcal{N}\left(\eta, \frac{\sigma^2}{\kappa} \right) & (16)\\
p(\theta) &\propto \left(\sigma^2\right)^{-\frac{\nu+3}{2}} \exp\left(- \frac{\zeta^2 + \kappa(\Delta\mu - \eta)^2}{2\sigma^2} \right) & (17)\\
\gamma_{ij}^{(m)} &\triangleq P(Z_i = j | y_i, \theta^{(m)})\\
&= \frac{w_j^{(m)} \phi(y_i | \mu_j^{(m)}, \sigma^{(m)})}{\sum_{l=1}^{k} w_l^{(m)} \phi(y_i | \mu_l^{(m)}, \sigma^{(m)})}, i = 1, ..., n \text{ and } j = 1, ..., k & (18)\\
Q(\theta|\theta^{(m)}) &= \sum_{i=1}^{n} \sum_{j=1}^{k} \gamma_{ij}^{(m)} \ln(w_j \phi(y_i | \mu + (j-1)\Delta\mu, \sigma)) & (19)\\
\theta^{(m+1)} &= \arg\max_{\theta} (Q(\theta|\theta^{(m)}) + \ln p(\theta)) & (20)
\end{aligned}$$

In case of PYDT, the MAP-EM formulation of the model is as follows.

$$
\begin{aligned}
p(\mathbf{X}, \mathbf{Z}|\mathbf{\Theta}, \mathcal{T}) &= \prod_{[uv]\in S(\mathcal{T})'} \mathcal{N}(\mathbf{z}_v|\mathbf{z}_u, \sigma^2(t_v - t_u)\mathbf{I}) \prod_{n=1}^{N} \mathcal{N}(\mathbf{x}_n|\mathbf{z}_{\mathrm{pa}(n)}, \sigma^2(1-t_{\mathrm{pa}(n)})\mathbf{I}) \qquad (21)\\
&\text{where } \mathbf{Z} = \{\mathbf{z}\}, \mathbf{\Theta} = \{\mathbf{t}, \sigma^2\}\\
p(\mathbf{\Theta}) &= \mathrm{G}(c|a_c, b_c)\mathrm{G}(1/\sigma^2|a_{\sigma^2}, b_{\sigma^2}) \prod_{v\in\mathcal{I}} p(t_v|c, \mathcal{T}) \qquad (22)\\
Q(\mathbf{\Theta}|\mathbf{\Theta}') &= \sum_{[uv]\in S(\mathcal{T})'} \left(-\frac{D}{2} \ln 2\pi\sigma^2(t_v - t_u) - \frac{\mathbb{E}_{p(\mathbf{z}|\mathbf{x},\mathbf{t},\sigma^2)}\big[\parallel \mathbf{z}_v - \mathbf{z}_u \parallel^2 \big]}{2\sigma^2(t_v - t_u)} \right)\\
&+ \sum_{n=1}^{N} \left(-\frac{D}{2} \ln 2\pi\sigma^2(1 - t_{\mathrm{pa}(n)}) - \frac{\mathbb{E}_{p(\mathbf{z}|\mathbf{x},\mathbf{t},\sigma^2)}\big[\parallel \mathbf{x}_v - \mathbf{z}_{\mathrm{pa}(n)} \parallel^2 \big]}{2\sigma^2(1 - t_{\mathrm{pa}(n)})} \right) \qquad (23)\\
\ln p(\mathbf{\Theta}) &= \Bigg(a_c \ln b_c + (a_c - 1)\ln c - b_c c - \ln\Gamma(a_c)\\
&+ a_{\sigma^2} \ln b_{\sigma^2} + (a_{\sigma^2} - 1)\ln \frac{1}{\sigma^2} - b_{\sigma^2}\frac{1}{\sigma^2} - \ln\Gamma(a_{\sigma^2})\\
&+ \sum_{v\in\mathcal{I}} (cJ^{\theta,\alpha}_{\mathbf{n}_v} - 1)\ln c(1 - t_v) \Bigg) \qquad (24)
\end{aligned}
$$

However, some distributions have conjugacy. Hence, those parameters can be marginalised out and that results in the collapsed version.

$$
\begin{aligned}
\int \mathrm{G}(c|a_c, b_c) \prod_{v\in\mathcal{I}} p(t_v|c, \mathcal{T})dc &= \frac{b_c^{a_c}}{\Gamma(a_c)}(1 - t_v)^{|\mathcal{I}|} \int c^{a_c - 1 + |\mathcal{I}|} \exp\Big(-\big(b_c - \sum_{v\in\mathcal{I}} J^{\theta,\alpha}_{\mathbf{n}_v} \ln(1 - t_v)\big)c\Big)dc\\
&= \frac{b_c^{a_c}}{\Gamma(a_c)}(1 - t_v)^{|\mathcal{I}|} \frac{\Gamma(a_c + |\mathcal{I}|)}{\big(b_c - \sum_{v\in\mathcal{I}} J^{\theta,\alpha}_{\mathbf{n}_v} \ln(1 - t_v)\big)^{a_c + |\mathcal{I}|}}\\
&= p(t_v|a_c, b_c, \mathcal{T}) \qquad (25)
\end{aligned}
$$

$$
\begin{aligned}
&\int \mathrm{G}(1/\sigma^2|a_{\sigma^2}, b_{\sigma^2}) \prod_{[uv]\in S(\mathcal{T})'} \mathcal{N}(\mathbf{z}_v|\mathbf{z}_u, \sigma^2(t_v - t_u)\mathbf{I}) \prod_{n=1}^{N} \mathcal{N}(\mathbf{x}_n|\mathbf{z}_{\mathrm{pa}(n)}, \sigma^2(1 - t_{\mathrm{pa}(n)})\mathbf{I})d(1/\sigma^2)\\
&= \frac{b_{\sigma^2}^{a_{\sigma^2}}}{\Gamma(a_{\sigma^2})} 2\pi^{-\frac{D}{2}(|\mathcal{I}|+N)} \int \left(\frac{1}{\sigma^2}\right)^{a_{\sigma^2} - 1 + \frac{D}{2}(|\mathcal{I}|+N)} \exp\left(-\left(b_{\sigma^2} + \frac{1}{2} \sum_{[uv]\in S(\mathcal{T})} \frac{\parallel \mathbf{z}_v - \mathbf{z}_u \parallel^2}{t_v - t_u} \right)\frac{1}{\sigma^2}\right)d(1/\sigma^2)\\
&= \frac{b_{\sigma^2}^{a_{\sigma^2}}}{\Gamma(a_{\sigma^2})} 2\pi^{-\frac{D}{2}(|\mathcal{I}|+N)} \frac{\Gamma\big(a_{\sigma^2} + \frac{D}{2}(|\mathcal{I}| + N)\big)}{\Big(b_{\sigma^2} + \frac{1}{2}\sum_{[uv]\in S(\mathcal{T})'} \frac{\|\mathbf{z}_v - \mathbf{z}_u\|^2}{t_v - t_u} + \frac{1}{2}\sum_{n=1}^{N} \frac{\|\mathbf{x}_n - \mathbf{z}_{\mathrm{pa}(n)}\|^2}{1 - t_{\mathrm{pa}(n)}}\Big)^{a_{\sigma^2} + \frac{D}{2}(|\mathcal{I}|+N)}}\\
&= p(\mathbf{X}, \mathbf{Z}|\mathbf{t}, a_{\sigma^2}, b_{\sigma^2}) \qquad (26)
\end{aligned}
$$

$$
\begin{aligned}
Q(\mathbf{\Theta}|\mathbf{\Theta}') &= a_{\sigma^2}\ln b_{\sigma^2} - \ln\Gamma(a_{\sigma^2}) + \ln\Gamma\Bigg(a_{\sigma^2} + \frac{D}{2}(|\mathcal{I}|+N)\Bigg) \\
&- \Bigg(a_{\sigma^2} + \frac{D}{2}(|\mathcal{I}|+N)\Bigg)\Bigg\langle \ln\Bigg(b_{\sigma^2} + \frac{1}{2}\sum_{[uv]\in S(\mathcal{T})'} \frac{\parallel \mathbf{z}_v - \mathbf{z}_u \parallel^2}{t_v - t_u} + \frac{1}{2}\sum_{n=1}^{N} \frac{\parallel \mathbf{x}_n - \mathbf{z}_{\mathrm{pa}(n)} \parallel^2}{1 - t_{\mathrm{pa}(n)}}\Bigg)\Bigg\rangle \\
&+ \text{Const.}
\end{aligned} \tag{27}
$$

$$
\begin{aligned}
\ln p(\mathbf{\Theta}) &= \sum_{v\in\mathcal{I}}\Bigg(a_c\ln b_c - \ln\Gamma(a_c) + |\mathcal{I}|\ln(1-t_v) \\
&\quad + \ln\Gamma(a_c + |\mathcal{I}|) - (a_c + \mathcal{I})\ln\Big(b_c - \sum_{v\in\mathcal{I}} J^{\theta,\alpha}_{\mathbf{n}_v}\ln(1-t_v)\Big)\Bigg) \\
&= |\mathcal{I}|\big(a_c\ln b_c - \ln\Gamma(a_c) + \ln\Gamma(a_c + |\mathcal{I}|)\big) \\
&\quad + \sum_{v\in\mathcal{I}}\Bigg(|\mathcal{I}|\ln(1-t_v) - (a_c + \mathcal{I})\ln\Big(b_c - \sum_{v\in\mathcal{I}} J^{\theta,\alpha}_{\mathbf{n}_v}\ln(1-t_v)\Big)\Bigg)
\end{aligned} \tag{28}
$$

$$\frac{\partial\big(Q(\mathbf{\Theta}|\mathbf{\Theta}') + \ln p(\mathbf{\Theta})\big)}{\partial t_v} \tag{29}$$

$$\frac{\partial\big(Q(\mathbf{\Theta}|\mathbf{\Theta}') + \ln p(\mathbf{\Theta})\big)}{\partial a_{\sigma^2}} \tag{30}$$

$$\frac{\partial\big(Q(\mathbf{\Theta}|\mathbf{\Theta}') + \ln p(\mathbf{\Theta})\big)}{\partial b_{\sigma^2}} \tag{31}$$

$$\frac{\partial\big(Q(\mathbf{\Theta}|\mathbf{\Theta}') + \ln p(\mathbf{\Theta})\big)}{\partial a_c} \tag{32}$$

$$\frac{\partial\big(Q(\mathbf{\Theta}|\mathbf{\Theta}') + \ln p(\mathbf{\Theta})\big)}{\partial b_c} \tag{33}$$

1.3 Posterior of $p(\mathbf{z}_v|\mathbf{z}_{\{u,k\}}, t_{\{u,v,k\}}, \sigma^2, \mathcal{T})$

The posterior distribution of $\mathbf{z}$ can be obtained by formulating the below joint distribution as a normal distribution. There is conjugacy among the distributions because the prior of mean is modelled as a normal distribution.

$$\mathcal{N}(\mathbf{z}_v|\mathbf{z}_u, \sigma^2(t_v - t_u)\mathbf{I})\prod_{k=1}^{K}\mathcal{N}(\mathbf{z}_k|\mathbf{z}_v, \sigma^2(t_k - t_v)\mathbf{I}) \tag{34}$$

$$p(\mathbf{z}_v|\mathbf{z}_{\{u,k\}}, t_{\{u,v,k\}}, \sigma^2, \mathcal{T}) = \mathcal{N}\left(\mathbf{z}_v \middle| r\left(\frac{\mathbf{z}_u}{t_v - t_u} + \sum_k \frac{\mathbf{z}_k}{t_k - t_v}\right), r\sigma^2\mathbf{I}\right) \tag{35}$$

$$\text{where } r = \left(\frac{1}{t_v - t_u} + \sum_k \frac{1}{t_k - t_v}\right)^{-1}$$

1.4 Posterior of $p(t_v|\mathcal{T})$

$$\begin{aligned}
p(t_v, \mathbf{z}_{\{u,v,k\}}, \sigma^2|c, \mathcal{T}) &= c(2\pi\sigma^2)^{-\frac{D(K+1)}{2}} \exp\Big\{(cJ^{\theta,\alpha}_{\mathbf{n}_v} - 1)\ln(1 - t_v) \\
&\quad -\frac{D}{2}\big(\ln(t_v - t_u) + \sum_k \ln(t_k - t_v)\big) \\
&\quad -\frac{\| \mathbf{z}_v - \mathbf{z}_u \|^2}{2\sigma^2}\frac{1}{t_v - t_u} - \sum_k \frac{\| \mathbf{z}_k - \mathbf{z}_v \|^2}{2\sigma^2}\frac{1}{t_k - t_v}\Big\} \\
&= C_{\sigma^2} \exp\big\{u(t_v)\big\} && (36) \\
\frac{du}{dt_v} &= -\frac{cJ^{\theta,\alpha}_{\mathbf{n}_v} - 1}{1 - t_v} - \frac{D}{2}\big(\frac{1}{t_v - t_u} - \sum_k \frac{1}{t_k - t_v}\big) \\
&\quad +\frac{\| \mathbf{z}_v - \mathbf{z}_u \|^2}{2\sigma^2}\frac{1}{(t_v - t_u)^2} - \sum_k \frac{\| \mathbf{z}_k - \mathbf{z}_v \|^2}{2\sigma^2}\frac{1}{(t_k - t_v)^2} \\
&= -\frac{cJ^{\theta,\alpha}_{\mathbf{n}_v} - 1}{1 - t_v} - \frac{1}{2(t_v - t_u)^2}\big(D(t_v - t_u) - \frac{\| \mathbf{z}_v - \mathbf{z}_u \|^2}{\sigma^2}\big) \\
&\quad +\sum_k \frac{1}{2(t_k - t_v)^2}\big(D(t_k - t_v) - \frac{\| \mathbf{z}_k - \mathbf{z}_v \|^2}{\sigma^2}\big) \\
&= A(t_v) && (37) \\
C_{\sigma^2} A(t_v)^{-1} \int \exp\big\{u(t_v)\big\} du &= Z, \text{ but this integral is intractable!} && (38) \\
&\Rightarrow \text{The range of integration w.r.t. } t_v \text{ is } 0 < t_u \le t_v \le \min(t_k) < 1, \\
&\quad \text{obtaining } Z \text{ by simply summing pdf by the interval } dt = 1e^{-2} \\
&\quad \text{is computationally tractable for modern hardwares.} \\
&\quad \text{A few } t \text{ points usually give sufficient accuracy.} \\
\mathbb{E}_{p(t_v|\mathbf{z}_{\{u,v,k\}}, \sigma^2, c, \mathcal{T})}[t_v] &= \frac{1}{Z}\int t_v p(t_v, \mathbf{z}_{\{u,v,k\}}, \sigma^2|c, \mathcal{T}) dt_v \\
&= \frac{1}{Z} C_{\sigma^2} A_\mu(t_v)^{-1} \int \exp\big\{u_\mu(t_v)\big\} du && (39)
\end{aligned}$$

1.5 Approximation of model evidence

Use the decomposition with variational distributions.

$$
\begin{aligned}
\ln P(\mathbf{X}|\boldsymbol{\theta}) &= \ln\Big\{\sum_{\mathbf{Z}} P(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})\Big\} \\
&= \ln\Big\{\sum_{\mathbf{Z}} q(\mathbf{Z})\frac{P(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})}{q(\mathbf{Z})}\Big\} \\
&= \sum_{\mathbf{Z}} q(\mathbf{Z}) \ln\Big\{\frac{P(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})}{q(\mathbf{Z})}\Big\} + \sum_{\mathbf{Z}} q(\mathbf{Z}) \ln\Big\{\frac{q(\mathbf{Z})}{P(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta})}\Big\} \qquad (40)
\end{aligned}
$$

$\mathbf{z}, c, \sigma^2$ have analytical posterior. In addition, the exact posterior of $\mathbf{t}$ can also be computed in a feasible way up to the error due to the finite numerical precision. Hence, we consider that the second term of (40), which is a Kullback-Leibler divergence of the calculated posterior and theoretical posterior in this context, vanishes. Incidentally, this is equivalent to calculate the ELBO of the model.

$$
\begin{aligned}
& \sum_{\mathbf{Z}} q(\mathbf{Z}) \ln\Big\{\frac{P(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})}{q(\mathbf{Z})}\Big\} \\
= \ & \mathbb{E}_{q(\mathbf{Z})}\big[\ln P(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta})\big] + H_q(\mathbf{Z}) \\
= \ & a_c \ln b_c + (a_c - 1)\mathbb{E}[\ln c] - b_c \mathbb{E}[c] - \ln\Gamma(a_c) \\
& + a_{\sigma^2} \ln b_{\sigma^2} + (a_{\sigma^2} - 1)\mathbb{E}\Big[\ln\frac{1}{\sigma^2}\Big] - b_{\sigma^2}\mathbb{E}\Big[\frac{1}{\sigma^2}\Big] - \ln\Gamma(a_{\sigma^2}) \\
& + \sum_{v\in\mathcal{I}} \Big\{ \big(\mathbb{E}[c] J^{\theta,\alpha}_{\mathbf{n}_v} - 1\big)\mathbb{E}[\ln c] + \big(\mathbb{E}[c] J^{\theta,\alpha}_{\mathbf{n}_v} - 1\big)\mathbb{E}\big[\ln(1-t_v)\big]\Big\} \\
& + \sum_{[uv]\in S(\mathcal{T})'} \Bigg\{ -\frac{D}{2}\bigg(\ln 2\pi - \mathbb{E}\Big[\ln\frac{1}{\sigma^2}\Big] + \mathbb{E}\big[\ln(t_v - t_u)\big]\bigg) \\
& - \mathbb{E}\Big[\frac{1}{\sigma^2}\Big]\mathbb{E}\Big[\frac{1}{t_v - t_u}\Big] \frac{\sum_{d=1}^{D} \mathbb{E}[z_{v,d}^2] - 2\mathbb{E}[z_{v,d}]\mathbb{E}[z_{u,d}] + \mathbb{E}[z_{u,d}^2]]}{2} \Bigg\} \\
& + \sum_{n=1}^{N} \Bigg\{ -\frac{D}{2}\bigg(\ln 2\pi - \mathbb{E}\Big[\ln\frac{1}{\sigma^2}\Big] + \mathbb{E}\Big[\ln(1 - t_{\mathrm{pa}(n)})\Big]\bigg) \\
& - \mathbb{E}\Big[\frac{1}{\sigma^2}\Big]\mathbb{E}\Big[\frac{1}{1 - t_{\mathrm{pa}(n)}}\Big] \frac{\sum_{d=1}^{D} x_{v,d}^2 - 2x_{n,d}\mathbb{E}[z_{\mathrm{pa}(n),d}] + \mathbb{E}[z_{\mathrm{pa}(n),d}^2]]}{2} \Bigg\} \\
& + H_q(\mathbf{Z}) \qquad (41)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[\ln c] &\approx \ln \mathbb{E}[c] - \frac{\mathbb{V}[c]}{2\mathbb{E}[c]^2} \\
&= \ln\big(a_c + |\mathcal{I}|\big) - \ln\Big(b_c - \sum_{v\in\mathcal{I}} J^{\theta,\alpha}_{\mathbf{n}_v} \ln(1-t_v)\Big) + \frac{1}{2\big(a_c + |\mathcal{I}|\big)} && (42) \\
\mathbb{E}\left[\ln\frac{1}{\sigma^2}\right] &\approx \ln \mathbb{E}\left[\frac{1}{\sigma^2}\right] - \frac{\mathbb{V}\left[\frac{1}{\sigma^2}\right]}{2\mathbb{E}\left[\frac{1}{\sigma^2}\right]^2} \\
&= \ln\Big(a_{\sigma^2} + \frac{D(|\mathcal{I}|+N)}{2}\Big) - \ln\Big(b_{\sigma^2} + \frac{1}{2}\sum_{[uv]\in S(\mathcal{T})} \frac{\sum_{d=1}^{D}(z_{v,d}-z_{u,d})^2}{t_v - t_u}\Big) \\
&\quad + \frac{1}{2\big(a_{\sigma^2} + \frac{D(|\mathcal{I}|+N)}{2}\big)} \\
\mathbb{E}\big[\ln(1-t_v)\big] &\approx \ln\big(1 - \mathbb{E}[t_v]\big) - \frac{\mathbb{V}[t_v]}{2\big(1-\mathbb{E}[t_v]\big)^2} && (43) \\
\mathbb{E}\big[\ln(t_v - t_u)\big] &\approx \ln\left(\mathbb{E}[t_v] - \mathbb{E}[t_u]\right) \\
&\quad + \frac{1}{2}\Big\{ - \frac{\mathbb{V}[t_v]}{(\mathbb{E}[t_v]-\mathbb{E}[t_u])^2} + \frac{2\big(\mathbb{E}[t_v t_u] - \mathbb{E}[t_v]\mathbb{E}[t_u]\big)}{(\mathbb{E}[t_v]-\mathbb{E}[t_u])^2} - \frac{\mathbb{V}[t_u]}{(\mathbb{E}[t_v]-\mathbb{E}[t_u])^2} \Big\} \\
&= \ln\left(\mathbb{E}[t_v] - \mathbb{E}[t_u]\right) + \frac{1}{2}\Big\{ - \frac{\mathbb{V}[t_v] + \mathbb{V}[t_u] - 2\big(\mathbb{E}[t_v t_u] - \mathbb{E}[t_v]\mathbb{E}[t_u]\big)}{(\mathbb{E}[t_v]-\mathbb{E}[t_u])^2} \Big\} && (44) \\
\mathbb{E}\left[\frac{1}{t_v - t_u}\right] &\approx \frac{1}{\mathbb{E}[t_v] - \mathbb{E}[t_u]} + \frac{1}{2}\Big\{ \frac{2\mathbb{V}[t_v] + 2\mathbb{V}[t_u] - 4\big(\mathbb{E}[t_v t_u] - \mathbb{E}[t_v]\mathbb{E}[t_u]\big)}{\big(\mathbb{E}[t_v]-\mathbb{E}[t_u]\big)^3} \Big\} && (45)
\end{aligned}
$$

Here, the expectations of $\mathbf{z}$ terms with the posterior distributions create a loopy graph. Hence, we approximate the expectation by directly substituting the samples of parent and children nodes to the calculation of each node v instead of recursively taking expectations of parent and children nodes which again introduces the node v.

Although the above equations show 2nd order approximation, we basically use 0th order approximation which is equal to 1st order approximation.