1 Piman-Yor Diffusion Trees

1.1 Modelling of PYDT

$$
\begin{aligned}
\alpha &\sim \mathrm{Beta}(\alpha|a_\alpha, b_\alpha) && (1)\\
\theta &\sim \mathrm{G}(\theta|a_\theta, b_\theta) && (2)\\
c &\sim \mathrm{G}(c|a_c, b_c)\\
&= \frac{b_c^{a_c} c^{a_c-1} e^{-b_c c}}{\Gamma(a_c)} && (3)\\
1/\sigma^2 &\sim \mathrm{G}(1/\sigma^2|a_{\sigma^2}, b_{\sigma^2}) && (4)\\
p(t_v|c,\mathcal{T}) &= c(1-t_v)^{cJ^{\theta,\alpha}_{\mathbf{n}_v}-1} && (5)\\
p(\mathbf{z}_v|\mathbf{z}_u, \sigma^2(t_v-t_u)\mathbf{I}) &= (2\pi\sigma^2(t_v-t_u))^{-\frac{D}{2}} \exp\Big(-\frac{1}{2}\frac{\parallel \mathbf{z}_v - \mathbf{z}_u \parallel^2}{\sigma^2(t_v-t_u)}\Big) && (6)\\
p(\mathbf{z}_k|\mathbf{z}_v, \sigma^2(t_k-t_v)\mathbf{I}) &= (2\pi\sigma^2(t_k-t_v))^{-\frac{D}{2}} \exp\Big(-\frac{1}{2}\frac{\parallel \mathbf{z}_k - \mathbf{z}_v \parallel^2}{\sigma^2(t_k-t_v)}\Big) && (7)\\
p(\mathbf{X},\mathbf{Z}|\mathbf{\Theta},\mathcal{T}) &= \mathrm{G}(c|a_c,b_c)\mathrm{G}(1/\sigma^2|a_{\sigma^2}, b_{\sigma^2})\\
&\quad \prod_{v\in\mathcal{I}}^{|\mathcal{I}|} p(t_v|c,\mathcal{T}) p(\mathbf{z}_v|\mathbf{z}_u, \sigma^2(t_v-t_u)\mathbf{I})\\
&\quad \prod_{n=1}^{N} p(\mathbf{x}_n|\mathbf{z}_{pa(n)}, \sigma^2(1-t_{pa(n)})\mathbf{I}) && (8)\\
&\quad \text{where } \mathbf{Z} = \{c, \sigma^2, \mathbf{z}, \mathbf{t}\}, \mathbf{\Theta} = \{a_c, b_c, a_{\sigma^2}, b_{\sigma^2}\}
\end{aligned}
$$

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.

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

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

$$
\begin{aligned}
p(t_v, \mathbf{z}_{\{u,v,k\}}, \sigma^2 | c, \mathcal{T}) &= c(2\pi\sigma^2)^{-\frac{D(K+1)}{2}} \exp \Big\{ (cJ^{\theta,\alpha}_{\mathbf{n}_v} - 1) \ln(1 - t_v) \\
&\quad - \frac{D}{2} \big(\ln(t_v - t_u) + \sum_k \ln(t_k - t_v) \big) \\
&\quad - \frac{\parallel \mathbf{z}_v - \mathbf{z}_u \parallel^2}{2\sigma^2} \frac{1}{t_v - t_u} - \sum_k \frac{\parallel \mathbf{z}_k - \mathbf{z}_v \parallel^2}{2\sigma^2} \frac{1}{t_k - t_v} \Big\} \\
&= C_{\sigma^2} \exp \big\{ u(t_v) \big\} \qquad (12) \\
\frac{du}{dt_v} &= -\frac{cJ^{\theta,\alpha}_{\mathbf{n}_v} - 1}{1 - t_v} - \frac{D}{2} \big(\frac{1}{t_v - t_u} - \sum_k \frac{1}{t_k - t_v} \big) \\
&\quad + \frac{\parallel \mathbf{z}_v - \mathbf{z}_u \parallel^2}{2\sigma^2} \frac{1}{(t_v - t_u)^2} - \sum_k \frac{\parallel \mathbf{z}_k - \mathbf{z}_v \parallel^2}{2\sigma^2} \frac{1}{(t_k - t_v)^2} \\
&= -\frac{cJ^{\theta,\alpha}_{\mathbf{n}_v} - 1}{1 - t_v} - \frac{1}{2(t_v - t_u)^2} \big(D(t_v - t_u) - \frac{\parallel \mathbf{z}_v - \mathbf{z}_u \parallel^2}{\sigma^2} \big) \\
&\quad + \sum_k \frac{1}{2(t_k - t_v)^2} \big(D(t_k - t_v) - \frac{\parallel \mathbf{z}_k - \mathbf{z}_v \parallel^2}{\sigma^2} \big) \\
&= A(t_v) \qquad (13) \\
C_{\sigma^2} A(t_v)^{-1} \int \exp \big\{ u(t_v) \big\} du &= Z, \text{ but this integral is intractable!} \qquad (14) \\
&\Rightarrow \text{The range of integration w.r.t. } t_v \text{ is } 0 < t_u \leq t_v \leq \min(t_k) < 1, \\
&\quad \text{obtaining } Z \text{ by simply summing pdf by the interval } dt = 1e^{-2} \\
&\quad \text{is computationally tractable for modern hardwares.} \\
&\quad \text{A few } t \text{ points usually give sufficient accuracy for } Z. \\
\mathbb{E}_{p(t_v|\mathbf{z}_{\{u,v,k\}}, \sigma^2, c, \mathcal{T})}[t_v] &= \frac{1}{Z} \int t_v p(t_v, \mathbf{z}_{\{u,v,k\}}, \sigma^2 | c, \mathcal{T}) dt_v \\
&= \frac{1}{Z} C_{\sigma^2} A_\mu(t_v)^{-1} \int \exp \big\{ u_\mu(t_v) \big\} du \qquad (15)
\end{aligned}
$$