

1 Piman-Yor Diffusion Trees

1.1 Modelling of PYDT

$$\alpha \sim \text{Beta}(\alpha|a_\alpha, b_\alpha) \quad (1)$$

$$\theta \sim \text{G}(\theta|a_\theta, b_\theta) \quad (2)$$

$$p(\mathcal{T}|\alpha, \theta) = \prod_{v \in \mathcal{I}} \frac{a(t_v) \prod_{k=3}^{K_v} [\theta + (k-1)\alpha] \prod_{l=1}^{K_v} \Gamma(n_l^v - \alpha)}{\Gamma(m(v) + \theta) \Gamma(1 - \alpha)^{K_v - 1}} \quad (3)$$

$$c \sim \text{G}(c|a_c, b_c) \quad (4)$$

$$1/\sigma^2 \sim \text{G}(1/\sigma^2|a_{\sigma^2}, b_{\sigma^2}) \quad (5)$$

$$p(t_v|c, \mathcal{T}) = c(1 - t_v)^{cJ_{\mathbf{n}_v}^{\theta, \alpha} - 1} \quad (6)$$

$$\mathcal{N}(\mathbf{z}_v|\mathbf{z}_u, \sigma^2(t_v - t_u)\mathbf{I}) = (2\pi\sigma^2(t_v - t_u))^{-\frac{D}{2}} \exp\left(-\frac{1}{2} \frac{\|\mathbf{z}_v - \mathbf{z}_u\|^2}{\sigma^2(t_v - t_u)}\right) \quad (7)$$

$$\mathcal{N}(\mathbf{z}_k|\mathbf{z}_v, \sigma^2(t_k - t_v)\mathbf{I}) = (2\pi\sigma^2(t_k - t_v))^{-\frac{D}{2}} \exp\left(-\frac{1}{2} \frac{\|\mathbf{z}_k - \mathbf{z}_v\|^2}{\sigma^2(t_k - t_v)}\right) \quad (8)$$

1.2 EM algorithm for PYDT

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

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.

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}') &= \sum_{\mathbf{Z}} q(\mathbf{Z}) \ln \left\{ \frac{P(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})}{q(\mathbf{Z})} \right\} \\ &= \sum_{\mathbf{Z}} q(\mathbf{Z}) \ln P(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}) - \sum_{\mathbf{Z}} q(\mathbf{Z}) \ln q(\mathbf{Z}) \quad (10) \end{aligned}$$

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

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).

$$\mu_j = \mu + (j - 1)\Delta\mu, j = 1, \dots, k \quad (12)$$

$$\sigma_j^2 = \sigma^2, j = 1, \dots, k \quad (13)$$

$$p(y_j) = \sum_{j=1}^k w_j \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \mu - (j - 1)\Delta\mu)^2}{2\sigma^2}\right) \quad (14)$$

$$\sigma^2 \sim \text{Inv-Gamma}\left(\frac{\nu}{2}, \frac{\zeta^2}{2}\right) \quad (15)$$

$$\Delta\mu|\sigma^2 \sim \mathcal{N}\left(\eta, \frac{\sigma^2}{\kappa}\right) \quad (16)$$

$$p(\theta) \propto (\sigma^2)^{-\frac{\nu+3}{2}} \exp\left(-\frac{\zeta^2 + \kappa(\Delta\mu - \eta)^2}{2\sigma^2}\right) \quad (17)$$

$$\begin{aligned} \gamma_{ij}^{(m)} &\triangleq P(Z_i = j|y_i, \theta^{(m)}) \\ &= \frac{w_j^{(m)} \phi(y_i|\mu_j^{(m)}, \sigma^{(m)})}{\sum_{l=1}^k w_l^{(m)} \phi(y_i|\mu_l^{(m)}, \sigma^{(m)})}, i = 1, \dots, n \text{ and } j = 1, \dots, k \end{aligned} \quad (18)$$

$$Q(\theta|\theta^{(m)}) = \sum_{i=1}^n \sum_{j=1}^k \gamma_{ij}^{(m)} \ln(w_j \phi(y_i|\mu + (j - 1)\Delta\mu, \sigma)) \quad (19)$$

$$\theta^{(m+1)} = \arg \max_{\theta} (Q(\theta|\theta^{(m)}) + \ln p(\theta)) \quad (20)$$

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

$$p(\mathbf{X}, \mathbf{Z} | \boldsymbol{\Theta}, \mathcal{T}) = \prod_{[uv] \in S(\mathcal{T})'} \mathcal{N}(\mathbf{z}_v | \mathbf{z}_u, \sigma^2(t_v - t_u)) \mathbf{I} \prod_{n=1}^N \mathcal{N}(\mathbf{x}_n | \mathbf{z}_{\text{pa}(n)}, \sigma^2(1 - t_{\text{pa}(n)})) \mathbf{I} \quad (21)$$

where $\mathbf{Z} = \{\mathbf{z}\}$, $\boldsymbol{\Theta} = \{\mathbf{t}, \sigma^2\}$

$$p(\boldsymbol{\Theta}) = \text{G}(c | a_c, b_c) \text{G}(1/\sigma^2 | a_{\sigma^2}, b_{\sigma^2}) \prod_{v \in \mathcal{I}} p(t_v | c, \mathcal{T}) \quad (22)$$

$$\begin{aligned} Q(\boldsymbol{\Theta} | \boldsymbol{\Theta}') &= \sum_{[uv] \in S(\mathcal{T})'} \left(-\frac{D}{2} \ln 2\pi\sigma^2(t_v - t_u) - \frac{\mathbb{E}_{p(\mathbf{z} | \mathbf{x}, \mathbf{t}, \sigma^2)}[\|\mathbf{z}_v - \mathbf{z}_u\|^2]}{2\sigma^2(t_v - t_u)} \right) \\ &+ \sum_{n=1}^N \left(-\frac{D}{2} \ln 2\pi\sigma^2(1 - t_{\text{pa}(n)}) - \frac{\mathbb{E}_{p(\mathbf{z} | \mathbf{x}, \mathbf{t}, \sigma^2)}[\|\mathbf{x}_n - \mathbf{z}_{\text{pa}(n)}\|^2]}{2\sigma^2(1 - t_{\text{pa}(n)})} \right) \end{aligned} \quad (23)$$

$$\begin{aligned} \ln p(\boldsymbol{\Theta}) &= \left(a_c \ln b_c + (a_c - 1) \ln c - b_c c - \ln \Gamma(a_{\sigma^2}) \right. \\ &+ a_{\sigma^2} \ln b_{\sigma^2} + (a_{\sigma^2} - 1) \ln \frac{1}{\sigma^2} - b_{\sigma^2} \frac{1}{\sigma^2} - \ln \Gamma(a_{\sigma^2}) \\ &\left. + \sum_{v \in \mathcal{I}} (c J_{\mathbf{n}_v}^{\theta, \alpha} - 1) \ln c(1 - t_v) \right) \end{aligned} \quad (24)$$

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

$$\begin{aligned} \int \text{G}(c | a_c, b_c) \prod_{v \in \mathcal{I}} p(t_v | c, \mathcal{T}) dc &= \frac{b_c^{a_c}}{\Gamma(a_c)} (1 - t_v)^{|\mathcal{I}|} \int c^{a_c - 1 + |\mathcal{I}|} \exp\left(-\left(b_c - \sum_{v \in \mathcal{I}} J_{\mathbf{n}_v}^{\theta, \alpha} \ln(1 - t_v)\right)c\right) dc \\ &= \frac{b_c^{a_c}}{\Gamma(a_c)} (1 - t_v)^{|\mathcal{I}|} \frac{\Gamma(a_c + |\mathcal{I}|)}{\left(b_c - \sum_{v \in \mathcal{I}} J_{\mathbf{n}_v}^{\theta, \alpha} \ln(1 - t_v)\right)^{a_c + |\mathcal{I}|}} \\ &= p(t_v | a_c, b_c, \mathcal{T}) \end{aligned} \quad (25)$$

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

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

$$\begin{aligned}
\ln p(\boldsymbol{\Theta}) &= \sum_{v \in \mathcal{I}} \left(a_c \ln b_c - \ln \Gamma(a_c) + |\mathcal{I}| \ln(1 - t_v) \right. \\
&\quad \left. + \ln \Gamma(a_c + |\mathcal{I}|) - (a_c + |\mathcal{I}|) \ln \left(b_c - \sum_{v \in \mathcal{I}} J_{\mathbf{n}_v}^{\theta, \alpha} \ln(1 - t_v) \right) \right) \\
&= |\mathcal{I}| (a_c \ln b_c - \ln \Gamma(a_c) + \ln \Gamma(a_c + |\mathcal{I}|)) \\
&\quad + \sum_{v \in \mathcal{I}} \left(|\mathcal{I}| \ln(1 - t_v) - (a_c + |\mathcal{I}|) \ln \left(b_c - \sum_{v \in \mathcal{I}} J_{\mathbf{n}_v}^{\theta, \alpha} \ln(1 - t_v) \right) \right)
\end{aligned} \tag{28}$$

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

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

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

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

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

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

$$\begin{aligned}
p(t_v, \mathbf{z}_{\{u,v,k\}}, \sigma^2 | c, \mathcal{T}) &= c(2\pi\sigma^2)^{-\frac{D(K+1)}{2}} \exp \left\{ (cJ_{\mathbf{n}_v}^{\theta, \alpha} - 1) \ln(1 - t_v) \right. \\
&\quad - \frac{D}{2} (\ln(t_v - t_u) + \sum_k \ln(t_k - t_v)) \\
&\quad \left. - \frac{\|\mathbf{z}_v - \mathbf{z}_u\|^2}{2\sigma^2} \frac{1}{t_v - t_u} - \sum_k \frac{\|\mathbf{z}_k - \mathbf{z}_v\|^2}{2\sigma^2} \frac{1}{t_k - t_v} \right\} \\
&= C_{\sigma^2} \exp \{u(t_v)\} \tag{34}
\end{aligned}$$

$$\begin{aligned}
\frac{du}{dt_v} &= -\frac{cJ_{\mathbf{n}_v}^{\theta, \alpha} - 1}{1 - t_v} - \frac{D}{2} \left(\frac{1}{t_v - t_u} - \sum_k \frac{1}{t_k - t_v} \right) \\
&\quad + \frac{\|\mathbf{z}_v - \mathbf{z}_u\|^2}{2\sigma^2} \frac{1}{(t_v - t_u)^2} - \sum_k \frac{\|\mathbf{z}_k - \mathbf{z}_v\|^2}{2\sigma^2} \frac{1}{(t_k - t_v)^2} \\
&= -\frac{cJ_{\mathbf{n}_v}^{\theta, \alpha} - 1}{1 - t_v} - \frac{1}{2(t_v - t_u)^2} \left(D(t_v - t_u) - \frac{\|\mathbf{z}_v - \mathbf{z}_u\|^2}{\sigma^2} \right) \\
&\quad + \sum_k \frac{1}{2(t_k - t_v)^2} \left(D(t_k - t_v) - \frac{\|\mathbf{z}_k - \mathbf{z}_v\|^2}{\sigma^2} \right) \\
&= A(t_v) \tag{35}
\end{aligned}$$

$$C_{\sigma^2} A(t_v)^{-1} \int \exp \{u(t_v)\} du = Z, \text{ but this integral is intractable!} \tag{36}$$

\Rightarrow The range of integration w.r.t. t_v is $0 < t_u \leq t_v \leq \min(t_k) < 1$, obtaining Z by simply summing pdf by the interval $dt = 1e^{-2}$ is computationally tractable for modern hardwares.

A few t points usually give sufficient accuracy.

$$\begin{aligned}
\mathbb{E}_{p(t_v | \mathbf{z}_{\{u,v,k\}}, \sigma^2, c, \mathcal{T})} [t_v] &= \frac{1}{Z} \int t_v p(t_v, \mathbf{z}_{\{u,v,k\}}, \sigma^2 | c, \mathcal{T}) dt_v \\
&= \frac{1}{Z} C_{\sigma^2} A_{\mu}(t_v)^{-1} \int \exp \{u_{\mu}(t_v)\} du \tag{37}
\end{aligned}$$