1 HDP&HMM

Model set-up:

\[ \pi_0 \sim \text{GEM}(\gamma) \]
\[ \pi_k \sim \text{Dir}(\alpha \pi_0) \]
\[ \phi_k \sim H \]
\[ Z_t \sim \text{Cat}(\pi_{z_{t-1}}) \]
\[ Y_t \sim f(\phi_{Z_t}) \]

Here we also include the initial distribution of \( Z_1 \) in the model by setting \( \theta \sim \text{Dir}(\alpha \pi_0), Z_1 \sim \text{Cat}(\theta) \).

The result is

\[ \pi_k \sim \text{Dir}(\alpha \pi_{0,1}, \ldots, \alpha \pi_{0,k+K}, \ldots) \]
\[ \pi_0 \sim \text{Dir}(\frac{\gamma_k}{K}, \frac{\gamma_k}{K}, \ldots, \frac{\gamma_k}{K}) \].

Figure 1: Plate Diagram.
2 variation inference

2.1 Model set-up

The model is

\[
\begin{align*}
\text{parameters} & \quad \theta \\
\text{data} & \quad y \\
\mathbb{P}(\theta|y) & \\
q(\theta) & \\
\end{align*}
\]

We want to find a good \( q(\theta) \) to approximate the posterior \( \mathbb{P}(\theta|y) \). The measure the distance between \( q(\theta) \) and \( \mathbb{P}(\theta|y) \), we use the KL distance, which is

\[
KL(q(\theta)||\mathbb{P}(\theta|y)) = \int_\theta q(\theta) \log \frac{q(\theta)}{\mathbb{P}(\theta|y)} d\theta \\
= -\mathcal{H}(q) - E[\log \mathbb{P}(y, \theta)] - \log \mathbb{P}(y) \\
\geq 0
\]

Thus we get a lower bound for \( y \)'s marginal density

\[
\log \mathbb{P}(y) \geq E[\log \mathbb{P}(y, \theta)] + \mathcal{H}(q) \equiv \mathcal{L}
\]

2.2 Degenerate case

Given the density of Gamma distribution, it’s notable that the density tends to \(+\infty\) when \( x \) tends to 0, which point estimation doesn’t apply here. Variation inference method is often used as an alternative way to deal with such degenerate case.

For Gamma distribution, we could avoid the degenerate problem by taking log, which is shown in the following figures.
2.3 Estimation for Exponential Family

Suppose the conditional posterior follows an exponential family

\[
P(\theta_k|y, \theta_{\backslash k}) = h_k(\theta_k) \exp\{\eta_k(y, \theta_k) \cdot t_k(\theta_{\backslash k}) - A_k(\eta_k(y, \theta_{\backslash k}))\}
\]

For \( q(\theta) \), we assume \( \theta_k \) to be independent, each from an exponential family

\[
q(\theta) = \Pi_k q(\theta_k).
\]

\[
q(\theta_k) = h_k(\theta_k) \exp\{\gamma_k - t_k(\theta_k) - A_k(\gamma_k)\}
\]

Then we get \( \mathcal{L} \),

\[
\mathcal{L} = \mathbb{E}[\log P(y, \theta)] - \sum_k \mathbb{E}[\log q(\theta_k)]
\]

\[
= \mathbb{E}[\log P(\theta_k|y, \theta_{\backslash k})] + \mathbb{E}[\log P(y, \theta_{\backslash k})] - \mathbb{E}[\log q(\theta_k)] - \sum_{l \neq k} \mathbb{E}[\log q(\theta_l)]
\]

\[
= \text{const} + \mathbb{E}[t_k(\theta_k)](\mathbb{E}[\eta_k(y, \theta_{\backslash k})] - \gamma_k) + A_k(\gamma_k)
\]

Take derivative with respect to \( \gamma_k \)

\[
\nabla_{\gamma_k} \mathcal{L}(\gamma) = \nabla_{\gamma_k}^2 A_k(\gamma_k)(\mathbb{E}[\eta_k(y, \theta_{\backslash k})] - \gamma_k) - \nabla_{\gamma_k} A_k(\gamma_k) + \nabla_{\gamma_k} A_k(\gamma_k)
\]

\[
= 0
\]

Finally we get the estimate for \( \gamma_k \)

\[
\Rightarrow \gamma_k^* = \mathbb{E}[\eta_k(y, \theta_k)].
\]

2.4 Disadvantage

- local optima
- uncertainty estimation(understate variance)
3  An example

In this example, we have such a model

\[ \theta \sim \text{Dir}(\frac{\alpha}{k}, \frac{\alpha}{k}, \ldots, \frac{\alpha}{k}) \]
\[ \phi_k \sim \text{Dir}(\beta, \beta, \ldots, \beta) \]
\[ Z_n \sim \text{Cat}(\theta) \]
\[ y_n \sim \text{Multi}(\phi_{Z_n, M_n}) \]

So the conditional posteriors are exponential family

\[ P(\theta | Z) = \text{Dir}(\theta; \frac{\alpha}{k} + \sum_n \mathbb{1}[Z_n = 1], \ldots, \frac{\alpha}{k} + \sum_n \mathbb{1}[Z_n = k]) \]
\[ P(\phi_k | z, y) \sim \text{Dir}(\phi_k; \beta + \sum_n \mathbb{1}[Z_n = 1]y_n, \ldots, \beta + \sum_n \mathbb{1}[Z_n = k]y_{n,k}) \]

\[ P(Z_n = k | \theta, \phi, y) \propto \theta_k \text{Multi}(y_n, \phi_k, M_n). \]

And \( \mathcal{L} \) take a form like this

\[ \mathcal{L} = \mathbb{E}[\log P(y, z, \theta, \phi)] - \mathbb{E}[\log q(z, \theta, \phi)]. \]

Here we set \( q \) for the parameter \( \theta, \phi \)'s and \( Z \)'s to be

\[ q(\theta) = \text{Dir}(\theta; \gamma) \]
\[ q(\phi_k) = \text{Dir}(\phi_k, \lambda_k) \]
\[ q(Z_n = k) = \xi_{n,k} \]
\[ \gamma_k = \frac{\alpha}{k} + \sum_n \xi_{n,k} \]
\[ \lambda_{k,v} = \beta + \sum_n \xi_{n,k} y_{n,v} \]
So it’s easy to get the estimate for the parameters of $\theta$ and $\phi$’s.

$$
\mathbb{E}[\log P(Z_n|y_n, \theta, \phi)] - \mathbb{E}[\log q(Z_n)]
= \sum_k \xi_{n,k} (\mathbb{E}[\log \theta_k] + \sum_v y_{n,v} \mathbb{E}[\log \phi_{k,v}] - \log \xi_{n,k}) - \Lambda(\xi_{n,k} - 1)
$$

Finally, we estimate the parameter of $Z$’s

$$
\frac{\partial \mathcal{L}}{\partial \xi_{n,k}} = \mathbb{E}[\log \theta_k] + \sum_v y_{n,v} \mathbb{E}[\log \phi_{k,v}] - 1 - \log \xi_{n,k} + \Lambda
$$

$$
\log \xi_{n,k}^* = \Lambda - 1 + \mathbb{E}[\log \theta_k] + \sum_v y_{n,v} \mathbb{E}[\log \phi_{k,v}]
$$

$$
\Rightarrow \xi_{n,k}^* \propto \exp\{\mathbb{E}[\log \theta_k] + \sum_v y_{n,v} \mathbb{E}[\log \phi_{k,v}]\}
$$