A Detailed Derivation of Dirichlet Process Mixtures

Zihao Hu
zihaohu@sjtu.edu.cn

For the stick-breaking representation of the Dirichlet process mixture, the KL-divergence reads

$$\begin{aligned}\mathrm{KL}(q||p) = &\sum_{i=1}^{T}\Big\{\mathbb{E}_{q_{v_i}}\Big[\ln\frac{q_{v_i}(v_i;\phi_i^v)}{p_{v_i}(v_i|\alpha)}\Big]\Big\} + \sum_{i=1}^{T}\Big\{\mathbb{E}_{q_{\eta_i}}\Big[\ln\frac{q_{\eta_i}(\eta_i;\phi_i^\eta)}{p_{\eta_i}(\eta_i|\lambda)}\Big]\Big\} \\ &+ \sum_{n=1}^{N}\Big\{\mathbb{E}_{q}\Big[\ln\frac{q_{z_n}(z_n)}{p_z(z_n|v)p_x(x_n|\eta_{x_n})}\Big]\Big\}.\end{aligned} \tag{1}$$

where

$$\begin{aligned}p_{v_i}(v_i;\alpha) &= \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}v_i^{\alpha_1-1}(1-v_i)^{\alpha_2-1}\\ p_{\eta_i}(\eta_i;\lambda) &= h(\eta_i)\exp\{\lambda_1\eta_i+\lambda_2(-a(\eta_i))\}\\ p_z(z_n|v) &= \prod_{i=1}^{T}(1-v_i)^{1[z_n>i]}v_i^{1[z_n=i]}\\ p_x(x_n|\eta_i) &= h(x_n)\exp\{\eta_i x_n - a(\eta_i)\}\\ q_{v_i}(v_i;\phi_i^v) &= \frac{\Gamma(\phi_{i,1}^v+\phi_{i,2}^v)}{\Gamma(\phi_{i,1}^v)+\Gamma(\phi_{i,2}^v)}v_i^{\phi_{i,1}^v-1}(1-v_i)^{\phi_{i,2}^v-1}\\ q_{\eta_i}(\eta_i;\phi_i^\eta) &= h(\eta_i)\exp\{\phi_{i,1}^v\eta_i+\phi_{i,2}^v(-a\eta_i)-a(\phi_i^\eta)\}.\end{aligned} \tag{2}$$

We define G as the function that contains all items about ϕ_i^v in $\mathrm{KL}(q||p)$, then

$$\begin{aligned}G(\phi_i^v) = &\ln\frac{\Gamma(\phi_{i,1}^v+\phi_{i,2}^v)}{\Gamma(\phi_{i,1}^v)+\Gamma(\phi_{i,2}^v)} + (\phi_{i,1}^v-1)[\psi(\phi_{i,1}^v)-\psi(\phi_{i,1}^v+\phi_{i,2}^v)]\\ &+(\phi_{i,2}^v-1)[\psi(\phi_{i,2}^v)-\psi(\phi_{i,1}^v+\phi_{i,2}^v)] - \ln\frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\\ &-(\alpha_1-1)[\psi(\phi_{i,1}^v)-\psi(\phi_{i,1}^v+\phi_{i,2}^v)] - (\alpha_2-1)[\psi(\phi_{i,2}^v)-\psi(\phi_{i,1}^v+\phi_{i,2}^v)]\\ &-\sum_{n=1}^{N}\Big\{q(z_n>i)[\psi(\phi_{i,2}^v)-\psi(\phi_{i,1}^v+\phi_{i,2}^v)] + q(z_n=i)[\psi(\phi_{i,1}^v)-\psi(\phi_{i,1}^v+\phi_{i,2}^v)]\Big\}.\end{aligned} \tag{3}$$

Taking derivatives with $\phi^v_{i,1}$, we obtain

$$\begin{aligned}\frac{\partial G}{\phi^v_{i,1}} &= \psi(\phi^v_{i,1}+\phi^v_{i,2}) - \psi(\phi^v_{i,1}) + \psi(\phi^v_{i,1}) - \psi(\phi^v_{i,1}+\phi^v_{i,2}) \\ &+ \Big(\phi^v_{i,1} - \alpha_1 - \sum_{n=1}^{N} q(z_n = i)\Big)\Big(\psi'(\phi^v_{i,1}) - \psi'(\phi^v_{i,1}+\phi^v_{i,2})\Big) \\ &+ \Big(\phi^v_{i,2} - \alpha_2 - \sum_{n=1}^{N} q(z_n > i))(-\psi'(\phi^v_{i,1}+\phi^v_{i,2})\Big),\end{aligned} \tag{4}$$

where $\psi(x)$ is the digamma function, and $\psi'(x)$ is the trigamma function. The case for $\phi^v_{i,2}$ is similar

$$\begin{aligned}\frac{\partial G}{\phi^v_{i,2}} &= \Big(\phi^v_{i,1} - \alpha_1 - \sum_{n=1}^{N} q(z_n = i)\Big)\Big(-\psi'(\phi^v_{i,1}+\phi^v_{i,2})\Big) \\ &+ \Big(\phi^v_{i,2} - \alpha_2 - \sum_{n=1}^{N} q(z_n > i)\Big)\Big(\psi'(\phi^v_{i,2}) - \psi'(\phi^v_{i,1}+\phi^v_{i,2})\Big).\end{aligned} \tag{5}$$

Letting the derivative of ϕ^v_i be zero yields

$$\begin{aligned}\phi^v_{i,1} &= \alpha_1 + \sum_{n=1}^{N} q(z_n = i) \\ \phi^v_{i,2} &= \alpha_2 + \sum_{n=1}^{N} q(z_n > i) = \alpha_2 + \sum_{n=1}^{N}\sum_{j=i+1}^{\infty} q(z_n = j)\end{aligned} \tag{6}$$

Considering the case for deriving q_{η_i}

$$\begin{aligned}\ln q^\star_{\eta_i}(\eta_i;\phi^n_i) &= \mathbb{E}_{q\neq q_{\eta_i}}\Big[\ln h(\eta_i) + \lambda_1\eta_i + \lambda_2(-a(\eta_i)) + \\ &+ \sum_{n=1}^{N} q(z_n = i)[\eta_i x_n - a(\eta_i)]\Big] \\ &= \Big(\lambda_1 + \sum_{n=1}^{N} q(z_n = i)x_n\Big)\eta_i - \Big(\lambda_2 + \sum_{n=1}^{N} q(z_n = i)\Big)a(\eta_i) + C\end{aligned} \tag{7}$$

hence,

$$\begin{aligned}\phi^n_{i,1} &= \lambda_1 + \sum_{n=1}^{N} q(z_n = i)x_n \\ \phi^n_{i,2} &= \lambda_2 + \sum_{n=1}^{N} q(z_n = i).\end{aligned} \tag{8}$$

Finally, for z

$$\ln q^\star(z_n = i) = \mathbb{E}[\ln p_z(z_n = i|v) + \ln p_x(x_n|\eta_i)] + C \triangleq S_{n,i}. \tag{9}$$

After normalizing, we obtain

$$q^\star(z_n = i) = \frac{\exp\{S_{n,i}\}}{\sum_{i=1}^{\infty}\exp\{S_{n,i}\}}. \tag{10}$$