P2. $\tau \sim IG(\alpha, \beta)$

prior $P(\tau) = \frac{\beta^\alpha}{\Gamma(\alpha)} \tau^{-\alpha-1} e^{-\frac{\beta}{\tau}}$ $(\alpha, \beta > 0)$

likelihood, $x \sim N(\mu, \sigma^2)$ $\underline{\tau = \sigma^2}$

$$P(x|\tau) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \Big|_{\sigma^2=\tau} = \frac{1}{\sqrt{2\pi\tau}} e^{-\frac{(x-\mu)^2}{2\tau}}$$

posterior. $P(\tau|x) = \frac{P(\tau)P(x|\tau)}{P(x)} \propto \tau^{-\alpha-1} \frac{1}{\sqrt{\tau}} \cdot e^{-(\beta + \frac{(x-\mu)^2}{2}) \frac{1}{\tau}}$

$$\propto \tau^{-\alpha-1-\frac{1}{2}} e^{-\frac{1}{\tau}(\beta + \frac{(x-\mu)^2}{2})}$$

$\beta' = \beta + \frac{(x-\mu)^2}{2}$. $\alpha' = \alpha + \frac{1}{2}$.

$P(\tau|x) \propto \tau^{-\alpha'-1} e^{-\frac{\beta'}{\tau}} \sim IG(\alpha', \beta')$

$\Downarrow$

$IG(\alpha + \frac{1}{2}, \beta + \frac{(x-\mu)^2}{2})$

P3. $Dir(x|a) = c\prod_{k=1}^{K} x_k^{a_k-1}$ $\sum_{k=1}^{K} x_k = 1$

a) Proof. (by induction)

1° $K=2$. $Dir(x|a) = c\prod_{k=1}^{2} x_k^{a_k-1}$ $\sum_{k=1}^{2} x_k = 1 \Rightarrow x_1+x_2=1$

$= c\,(x_1)^{a_1-1}(1-x_1)^{a_2-1}$ $c = \frac{\Gamma(a_1+a_2)}{\Gamma(a_1)\Gamma(a_2)}$

$\sim Beta(a_1,a_2)$

Therefore. $c = \frac{\Gamma(a_1+a_2)}{\Gamma(a_1)\Gamma(a_2)}$ is true for $k=2$.

2° $K\geq 3$. suppose it is true for $k=K-1$ s.t.:

$Dir(x|a) = c_{k-1}\prod_{k=1}^{K-1} x_k^{a_k-1}$ $\sum_{k=1}^{K-1} x_k = 1$ $c_{k-1} = \frac{\Gamma(a_1+a_2+\cdots a_{K-1})}{\Gamma(a_1)-\Gamma(a_{K-1})}$

Now, we need to show that it is true for $k=K$

Since. when $k=K-1$ $\frac{1}{c_{K-1}} = \frac{1}{\int_{x_1\cdots x_{K-1}} \prod_{k=1}^{K-1} x_k^{a_k-1}\,dx_1\cdots x_{K-1}}$

when $k=K$,

$$\frac{1}{c_K} = \int_{x_1\cdots x_K} \prod_{k=1}^{K} x_k^{a_k-1}\,dx_1\cdots x_K = \int_0^1 x_K^{a_K-1}dx_K \int \prod_{k=1}^{K-1}\underbrace{\left(\frac{x_k}{1-x_K}\right)}_{y_k}^{a_k-1}(1-x_K)^{a_k-1}\,dx_1 dx_{K-1}$$

$$= \int_0^1 x_K^{a_K-1}dx_K \int (1-x_K)^{K-2}\prod_{k=1}^{K-1} y_k^{a_k-1}(1-x_K)^{a_k-1}\,dy_1\cdots dy_{K-1}$$

$$= \int_0^1 x_K^{a_K-1}dx_K \int (1-x_K)^{K-2}\prod_{k=1}^{K-1} x_k^{a_k-1}(1-x_K)^{a_k-1}\,dx_1\cdots dx_{K-1}$$

$$= \frac{1}{c_{K-1}}\int_0^1 x_K^{a_K-1}(1-x_K)^{K-2}(1-x_K)^{\sum_{k=1}^{K-1}(a_k-1)}\,dx_1\cdots dx_K$$

$\sum_{k=1}^{K-1} a_k-(K-1)$

$$= \frac{1}{c_{K-1}}\underbrace{\int_0^1 x_K^{a_K-1}(1-x_K)^{\sum_{k=1}^{K-1}a_k-1}\,dx_1\cdots dx_K}_{Beta(a_K,\sum_{k=1}^{K-1}a_k)}$$

$$= \frac{\Gamma(a_1)\cdots\Gamma(a_{K-1})}{\Gamma(a_1+a_2+\cdots a_{K-1})}\frac{\Gamma(a_K)\Gamma(\sum_{k=1}^{K-1}a_k)}{\Gamma(\sum_{k=1}^{K}a_k)} = \frac{\Gamma(a_1)\cdots\Gamma(a_K)}{\Gamma(a_1+\cdots a_K)}$$

which completes the proof. for $\forall K\in\{2,3\ldots\}$

b). $\because$ $Dir(x|a) = c\prod_{k=1}^{K} x_k^{a_k-1}$

$$\frac{1}{c_k} = \int \prod_{k=1}^{K} x_k^{a_k-1} = \int_0^1 x_1^{a_1-1}dx_1 \int_0^1 x_2^{a_2-1}dx_2 \cdots \int_0^1 x_k^{a_k-1}dx_k$$

$$= \int_0^1 x_k^{a_k-1}dx_k \cdot \int_0^1 \prod_{k=1}^{k-2} x_k^{a_k-1}(1-x_1-\cdots-x_{k-2}-x_k)\,dx_1\cdots dx_{k-2}$$

(without x_{k-1})

$$= \int_0^1 x_1^{a_1-1}dx_1 \cdots \int_0^1 (1-\sum_{k=1}^{k-2} x_k)^{a_{k-1}+a_k-1} \left(\frac{x_{k-1}}{1-\sum_{k=1}^{k-2} x_k}\right)^{a_{k-1}-1} \left(1-\frac{x_{k-1}}{1-\sum_{k=1}^{k-2} x_k}\right)^{a_k-1} d\left(\frac{x_{k-1}}{1-\sum_{k=1}^{k-2} x_k}\right)$$

$$= Beta(a_{k-1}, a_k)\, Beta(a_{k-2}, a_{k-1}+a_k)\, Beta(a_{k-3}, a_{k-2}+a_{k-1}+a_k) \cdots Beta(a_1, a_2+a_3+\cdots a_k)$$

$$= \frac{\Gamma(a_k)\Gamma(a_{k-1})}{\Gamma(a_{k-1}+a_k)} \cdot \frac{\Gamma(\cancel{a_{k-1}+a_k})\Gamma(a_{k-2})}{\Gamma(a_{k-2}+\cancel{a_{k-1}}+a_k)} \cdot \frac{\Gamma(\cancel{a_{k-2}}+a_{k-1}+a_k)\Gamma(a_{k-3})}{\Gamma(a_{k-3}+\cancel{a_{k-2}}+\cdots a_k)} + \cdots \frac{\Gamma(\cancel{a_1+\cdots a_k})\Gamma(a_k)}{\Gamma(a_1+a_2+\cdots a_k)}$$

$$= \frac{\Gamma(a_1)\Gamma(a_2)\cdots\Gamma(a_k)}{\Gamma(a_1+\cdots a_k)}$$

$$\Rightarrow c_k = \frac{\Gamma(a_1+\cdots a_k)}{\Gamma(a_1)\Gamma(a_2)\cdots\Gamma(a_k)}$$

P4.

a).
$$\theta_i \sim \mathrm{Bern}(\lambda)$$
$$\varepsilon_i \sim N(0, \sigma_{\theta_i}^2)$$
$$X_{2,i} = \alpha_{\theta_i} X_{1,i} + \beta_{\theta_i} + \varepsilon_i$$

$$P(X_{2,i}; \xi) = P(X_{2,i} \mid \theta = 0; \xi) \cdot P(\theta = 0; \xi) + P(X_{2,i} \mid \theta = 1, \xi) P(\theta = 1; \xi)$$

$$X_{2,i} \mid \theta = 1, \xi \sim N(\underbrace{\alpha_1 X_{1,i}}_{\text{const}} + \beta_1, \sigma_1^2)$$

$$X_{2,i} \mid \theta = 0, \xi \sim N(\alpha_0 X_{1,i} + \beta_0, \sigma_0^2)$$

$$P(\theta = 1; \xi) = \lambda \qquad P(\theta = 0; \xi) = 1 - \lambda)$$

Thus. $P(X_{2,i}, \xi) = \lambda N(X_{2,i} : \alpha_1 X_{1,i} + \beta_1, \sigma_1^2) +$
$(1-\lambda) N(X_{2,i}; \alpha_0 X_{1,i} + \beta_0, \sigma_0^2)$

b). $P(X_{2,i}, \theta_i; \xi) = P(X_{2,i} \mid \theta_i; \xi) \cdot P(\theta_i; \xi)$

$$X_{2,i} \mid \theta_i; \xi \sim N(\alpha_{\theta_i} X_{1,i} + \beta_{\theta_i}, \sigma_{\theta_i}^2)$$
$$\theta_i \sim \mathrm{Bern}(\lambda)$$

$$\Rightarrow P(X_{2,i}, \theta_i; \xi) = (\lambda \cdot N(X_{2,i}; \alpha_1 X_{1,i} + \beta_1, \sigma_1^2))^{\theta_i} \cdot ((1-\lambda) N(X_{2,i}; \alpha_0 X_{1,i} + \beta_0, \sigma_0^2))^{1-\theta_i}$$

c). $\log P((X_{2,i}, \theta_i)_{i=1}^{n}; \xi)$

$$= \sum_{i=1}^{n} \log P(X_{2,i}, \theta_i; \xi)$$

$$= \sum_{i=1}^{n} \Big(\theta_i (\log \lambda + \log N(X_{2,i}, \alpha_1 X_{1,i} + \beta_1, \sigma_1^2)) + (1-\theta_i)(\log(1-\lambda) + \log N(X_{2,i}, \alpha_0 X_{1,i} + \beta_0, \sigma_0^2)) \Big)$$

d)

$$\underbrace{P(\theta_i \mid x_{2,i}, \zeta^{(l)})}_{p_i^{(l)}} = \frac{P(x_{2,i} \mid \theta_i ; \zeta^{(l)}) P(\theta_i ; \zeta^{(l)})}{P(x_{2,i} ; \zeta^{(l)})}$$

$$= \frac{N(\alpha_1^{(l)} x_{1,i} + \beta_1^{(l)}, \sigma_1^{(l)2}) \lambda_i^{(l)}}{\lambda^{(l)} N(\alpha_1^{(l)} x_{1,i} + \beta_1^{(l)}, \sigma_1^{(l)2}) + (1-\lambda) N(\alpha_0 x_{1,i} + \beta_0, \sigma_0^{(l)2})}$$

e). $$\log P(\{x_{2,i}, \theta_i\}_{i=1}^n ; \zeta) = \sum_{i=1}^n p_i^{(l)} \Big(\log \lambda - \log(1-\lambda) - \tfrac{1}{2}\log 2\pi\sigma_1 + \tfrac{1}{2}\log 2\pi\sigma_0 - \frac{1}{2\sigma_1^2}(x_{2,i} \alpha_1 x_{1,i} - \beta_0)^2 + \frac{1}{2\sigma_0^2}(x_{2,i} - \alpha_0 x_{1,i} - \beta_0)\Big)$$

$$+ \sum_{i=1}^n \Big(\log(1-\lambda) - \tfrac{1}{2}\log 2\pi\sigma_0 - \frac{1}{2\sigma_0^2}(x_{2,i} \alpha_0 x_{1,i} - \beta_0) \Big)$$

① $$\lambda_i^{(l+1)} = \arg\min \underbrace{\sum_{i=1}^n p_i^{(l)} (\log \tfrac{\lambda}{1-\lambda}) + \sum_{i=1}^n \log(1-\lambda)}_{f(\lambda_i)}$$

$$\nabla f(\lambda_i) = 0.$$

$$\sum_{i=1}^n p_i^{(l)} \left(\frac{1}{\lambda} + \frac{1}{1-\lambda}\right) + \left(\frac{-n}{1-\lambda}\right) = 0$$

$$\sum_{i=1}^n p_i^{(l)} \frac{1}{\lambda} = \frac{1}{1-\lambda} \left(-\sum_{i=1}^n p_i^{(l)} + n\right)$$

$$\lambda_i^{(l+1)} = \frac{\sum_{i=1}^n p_i^{(l)}}{n}$$

② $$\alpha_1 = \arg\min \underbrace{\sum_{i=1}^n \frac{p_i^{(l)} (x_{2,i} - \alpha_1 x_{1,i} - \beta_1)^2}{2\sigma_1^2}}_{f(\alpha_1)}$$

$$\nabla f(\alpha_1) = 0$$

$$\Rightarrow \sum_{i=1}^n p_i^{(l)} (x_{2,i} - \alpha_1 x_{1,i} - \beta_1) x_{1,i} = 0 \quad ①$$

Similarly for β_1,

$$\nabla f(\beta_1) = 0 \quad \sum_{i=1}^N p_i^{(l)} (x_{2,i} - \alpha_1 x_{1,i} - \beta_1) = 0 \quad ②$$

Based on ①②

$$\sum_{i=1}^{n} P_i^{(\ell)} X_{2,i} X_{1,i} = \sum_{i=1}^{n} P_i^{(\ell)} X_{1,i} \alpha_1 + \sum_{i=1}^{n} P_i^{(\ell)} X_{1,i} \beta_1$$

$$= \left[\sum_{i=1}^{n} P_i^{(\ell)} X_{1,i}^2 \quad \sum_{i=1}^{n} P_i^{(\ell)} X_{1,i} \right] \begin{bmatrix} \alpha_1^{(\ell+1)} \\ \beta_1^{(\ell+1)} \end{bmatrix}$$

$$\sum_{i=1}^{n} P_i^{(\ell)} X_{2,i} = \sum_{i=1}^{n} P_i^{(\ell)} X_{1,i} \alpha_1 + \sum_{i=1}^{n} P_i^{(\ell)} \beta_1$$

$$= \left[\sum_{i=1}^{n} P_i^{(\ell)} X_{1,i} \quad \sum_{i=1}^{n} P_i^{(\ell)} \right] \begin{bmatrix} \alpha_1^{(\ell+1)} \\ \beta_1^{(\ell+1)} \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \sum_{i=1}^{n} P_i^{(\ell)} X_{2,i} X_{1,i} \\ \sum_{i=1}^{n} P_i^{(\ell)} X_{2,i} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} P_i^{(\ell)} X_{1,i}^2 & \sum_{i=1}^{n} P_i^{(\ell)} X_{1,i} \\ \sum_{i=1}^{n} P_i^{(\ell)} X_{1,i} & \sum_{i=1}^{n} P_i^{(\ell)} \end{bmatrix} \begin{bmatrix} \alpha_1^{(\ell+1)} \\ \beta_1^{(\ell+1)} \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \alpha_1^{(\ell+1)} \\ \beta_1^{(\ell+1)} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} P_i^{(\ell)} X_{1,i}^2 & \sum_{i=1}^{n} P_i^{(\ell)} X_{1,i} \\ \sum_{i=1}^{n} P_i^{(\ell)} X_{1,i} & \sum_{i=1}^{n} P_i^{(\ell)} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} P_i^{(\ell)} X_{2,i} X_{1,i} \\ \sum_{i=1}^{n} P_i^{(\ell)} X_{2,i} \end{bmatrix}$$

③

Similar to (α_0, β_0)

$$\alpha_0^{(\ell+1)} = \arg\min \underbrace{\sum_{i=1}^{n} \frac{(1-P_i^{(\ell)})(X_{2,i} - \alpha_0 X_{1,i} - \beta_0)^2}{2\sigma_0^2}}_{f(\alpha_0)}$$

$$\nabla f(\alpha_0^{(\ell+1)}) = \sum_{i=1}^{n} (1-P_i^{(\ell)})(X_{2,i} - \alpha_0 X_{1,i} - \beta_0) X_{1,i} = 0.$$

Similarly.

$$\nabla f(\beta_0^{(\ell+1)}) = 0. \quad \sum_{i=1}^{n} (1-P_i^{(\ell)})(X_{2,i} - \alpha_0 X_{1,i} - \beta_0) = 0.$$

$$\begin{bmatrix} \alpha_0^{(\ell+1)} \\ \beta_0^{(\ell+1)} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} (1-P_i^{(\ell)}) X_{1,i}^2 & \sum_{i=1}^{n} (1-P_i^{(\ell)}) X_{1,i} \\ \sum_{i=1}^{n} (1-P_i^{(\ell)}) X_{1,i} & \sum_{i=1}^{n} (1-P_i^{(\ell)}) \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} (1-P_i^{(\ell)}) X_{2,i} X_{1,i} \\ \sum_{i=1}^{n} (1-P_i^{(\ell)}) X_{2,i} \end{bmatrix}$$

$$\sigma_1^{(l+1)} = \arg\max \sum_{i=1}^{n} p_i^{(l)} \left(-\frac{1}{2}\log 2\pi\sigma_1 - \frac{1}{2\sigma_1^2}(x_{2,i} - \alpha_1 x_{1,i} - \beta_1)^2 \right)$$

$$-\frac{1}{2}\log 2\pi - \frac{1}{2}\log \sigma_1$$

$$f(\sigma_1)$$

$$\nabla f(\sigma_1)^{(l+1)} = \sum_{i=1}^{n} p_i^{(l)} \left(-\frac{1}{2\sigma_1} + \frac{1}{2\sigma_1^2}(x_{2,i} - \alpha_1 x_{1,i} - \beta_1)^2 \right) = 0$$

$$\sum_{i=1}^{n} p_i^{(l)} = \sum_{i=1}^{n} p_i^{(l)} (x_{2,i} - \alpha_1 x_{1,i} - \beta_1)^2 \frac{1}{\sigma_1^{2(l+1)}}$$

$$\sigma_1^{2(l+1)} = \frac{\sum_{i=1}^{n} p_i^{(l)} (x_{2,i} - \alpha_1 x_{1,i} - \beta_1)^2}{\sum_{i=1}^{n} p_i^{(l)}}$$

④ Similarly,

$$\sigma_0^{2(l+1)} = \frac{\sum_{i=1}^{n} (1 - p_i^{(l)}) (x_{2,i} - \alpha_0 x_{1,i} - \beta_0)^2}{\sum_{i=1}^{n} (1 - p_i^{(l)})}$$

12.12.

P5 $D_{KL}(P\|Q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} \, dx$

$f(x) = \log x$ is a concave function and use Jensen's inequality

$f(\int g(x) p(x) dx) \leq \int f(g(x)) p(x) dx$ for convex function

$f(x) = -\log x$

$g(x) = \frac{q(x)}{p(x)}$

$\int -\log g(x) \, p(x) \, d(x) \geq -f(\int g(x) p(x) \, dx$

$\geq -f(\int \frac{q(x)}{\cancel{p(x)}} \cancel{p(x)} \, dx)$

$\geq -f(1) = -\log 1 = 0$

$\Rightarrow D_{KL}(P\|Q) \geq 0$