BCM

$$p(f^* \mid \mathcal{D}^{(j)}, \mathcal{D}^{(k)}) \propto p(\mathcal{D}^{(j)}, \mathcal{D}^{(k)} \mid f^*)\, p(f^*)$$

Assumption: $\mathcal{D}^{(j)} \perp\!\!\!\perp \mathcal{D}^{(k)} \mid f^*$ (BCM)

Then: $$p(f^* \mid \mathcal{D}^{(j)}, \mathcal{D}^{(k)}) \overset{BCM}{\propto} p(\mathcal{D}^{(j)} \mid f^*)\, p(\mathcal{D}^{(k)} \mid f^*)\, p(f^*)$$

$$= \frac{p(\mathcal{D}^{(j)}, f^*)\, p(\mathcal{D}^{(k)}, f^*)}{p(f^*)}$$

$$\propto \frac{p_j(f^* \mid \mathcal{D}^{(j)})\, p_k(f^* \mid \mathcal{D}^{(k)})}{p(f^*)}$$

expert j

expert k

"Normalization" by GP prior

For M experts, we apply the above repeatedly and obtain

$$p(f^* \mid \mathcal{D}^{(1)}, \ldots, \mathcal{D}^{(M)}) \propto \frac{\prod_{i=1}^{M} p_i(f^* \mid \mathcal{D}^{(i)})}{p^{M-1}(f^*)}$$

## PoE Computational graph

① $p(f^* | \mathcal{D}) = \prod_{k=1}^{M} p_k(f^* | \mathcal{D}^{(k)})$

②

$$\prod_{k=1}^{M} p_k(f^* | \mathcal{D}^{(k)}) \qquad p_1 p_2 p_3 p_4$$

$$= \prod_{k=1}^{L} \prod_{i=1}^{L_k} p_{k_i}(f^* | \mathcal{D}^{(k,i)}), \qquad = (p_1 p_2)(p_3 p_4) \qquad L=2,\ L_k=2$$

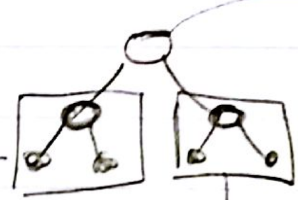
$\sum_{k=1}^{L} L_k = M \rightarrow$ ensures that we have M GP experts

## gPoE Computational graph

① $p(f^* | \mathcal{D}) = \prod_{k=1}^{M} p_k^{\beta_k}(f^* | \mathcal{D}^{(k)}), \quad \sum_{k=1}^{M} \beta_k = 1$



② $= \prod_{k=1}^{L} \prod_{i=1}^{L_k} p_{k_i}^{\beta_{k_i}}(f^* | \mathcal{D}^{(k,i)}), \quad \sum \beta_{k_i} = 1$





$$p_1^{\beta_1} p_2^{\beta_2} p_3^{\beta_3} p_4^{\beta_4}$$

$$= \underbrace{\underbrace{\left(p_1^{\beta_1} p_2^{\beta_2}\right)}\underbrace{\left(p_3^{\beta_3} p_4^{\beta_4}\right)}}_{\text{root node}}$$

## BCM Computational graph

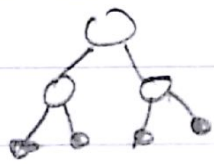
① $p(f^* | \mathcal{D}) = \dfrac{\prod_{k=1}^{M} p_k(f^* | \mathcal{D}^{(k)})}{p^{M-1}(f^*)}$ "normalized PoE"

② $= \dfrac{\prod_{k=1}^{L} \prod_{i=1}^{L_k} p_{k_i}(f^* | \mathcal{D}^{(k,i)})}{\underbrace{p^{M-1}(f^*)}}$ — see PoE

normalization can happen at the root of the computational graph

## Robust BCM Computational graph

$$p(f^*|D) = \frac{\prod_{k=1}^{M} p_k^{\beta_k}(f^*|D^{(k)})}{p^{\sum \beta_k - 1}(f^*)}$$



$$= \frac{\prod_{k=1}^{L} \prod_{i=1}^{L_k} p_{ki}^{\beta_{ki}}(f^*|D^{(ki)})}{\underbrace{p^{\sum_{k,i} \beta_{ki} - 1}(f^*)}} \quad \text{— gPoE}$$

To normalize this model, the root node has to know the sum of the individual weights $\sum_{k,i} \beta_{ki}$.