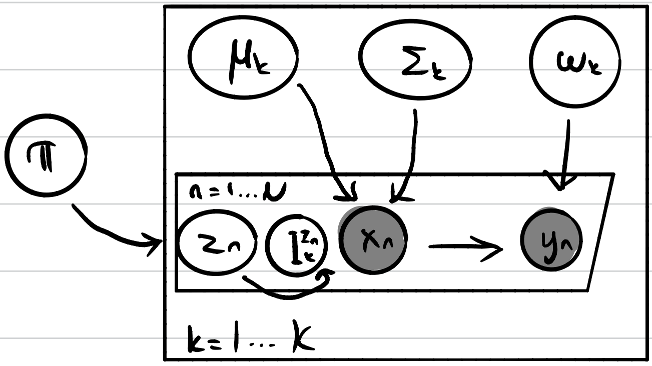
1.



So the indicator checks if it is the correct cluster or something similar. I am not sure how to add a power of into the diagram so indicator is just it's own node.

$$p(\{z_n\}, \{x_n\}, \{\alpha_m\}, W, U, V, \pi)$$
$$= p(\pi)p(U)p(V) \prod_{m=1}^{M} \left(p(\alpha_m \mid U, V)\, p(W^m \mid \alpha_m)\right) \prod_{i=1}^{N} \left(p(x_i^{(m)} \mid z_i, W^m, \pi)\right)$$

2. $p(z_n) = \text{Categorical}(\pi) = \pi_k$

$$p(x_n \mid z_n = k, \mu) = \prod_{d=1}^{D} \text{Bernoulli}(\mu_{kd}) = \prod_{d=1}^{D} \mu_{kd}^{x_{nd}} (1-\mu_{kd})^{1-x_{nd}}$$

Observed data log-likelihood:

$$\sum_{n=1}^{N} \log p(x_n \mid \mu) = \sum_{n=1}^{N} \log \sum_{k=1}^{K} p(x_n, z_n = k \mid \mu) = \sum_{n=1}^{N} \log \sum_{k=1}^{K} p(z_n \mid \mu)\, p(x_n \mid z_n = k, \mu)$$
$$= \sum_{n=1}^{N} \log \sum_{k=1}^{K} \pi_k \left(\prod_{d=1}^{D} \mu_{kd}^{x_{nd}} (1-\mu_{kd})^{1-x_{nd}} \right)$$

Complete data likelihood for one point

$$p(x_n, z_n \mid \mu, \pi) = \prod_{k=1}^{K} \left(p(x_n \mid z_n = k, \mu, \pi)\, p(z_n = k \mid \mu, \pi) \right)^{I(z_n = k)}$$

Complete data log-likelihood

$$\log P(X, Z, \mu, \pi) = \sum_{n=1}^{N}\sum_{k=1}^{K} p(z_n=k \mid \pi, \mu)\log(\pi_k \, p(x_n \mid z_n=k, \mu))$$
$$= \sum_{n=1}^{N}\sum_{k=1}^{K} p(z_n=k \mid \pi, \mu)\left(\log(\pi_k) + \sum_{d=1}^{D} \log \mu_{kd}^{x_{nd}} (1-\mu_{kd})^{1-x_{nd}}\right)$$
$$= \sum_{n=1}^{N}\sum_{k=1}^{K} p(z_n=k \mid \pi, \mu)\left(\log(\pi_k) + \sum_{d=1}^{D} x_{nd}\log \mu_{kd} + (1-x_{nd})\log(1-\mu_{kd})\right)$$

Expected complete data log likelihood

$f(\mu, \pi)$

$$E_{p(z_n=k \mid \mu, \pi)}(z_{nk}) \left(\log(\pi_k) + \sum_{d=1}^{D}\left(x_{nd}\log \mu_{kd} + (1-x_{nd})\log(1-\mu_{kd})\right)\right)$$

$$p(z_{nk} \mid x_n, \mu, \pi) = \frac{p(x_n, z_{nk} \mid \mu, \pi)}{\sum_k p(x_n, z_{nk} \mid \mu, \pi)}$$
$$= \frac{\pi_k \prod_{d=1}^{D} \mu_{kd}^{x_{nd}} (1-\mu_{kd})^{(1-x_{nd})}}{\sum_k \pi_k \prod_{d=1}^{D} \mu_{kd}^{x_{nd}} (1-\mu_{kd})^{(1-x_{nd})}} = r_{nk}$$

Derivate expectation wrt μ_{kd}

$$f'(\mu, \pi) = \sum_{n=1}^{N} r_{nk}\left(0 + \sum_{d=1}^{D}\left(\frac{x_{nd}}{\mu_{kd}} + \frac{(1-x_{nd})}{(1-\mu_{nd})}\right)\right)$$
$$= \sum_{n=1}^{N} r_{nk}\left(\sum_{d=1}^{D}\frac{x_{nd}}{\mu_{kd}} + \frac{(1-x_{nd})}{(1-\mu_{nd})}\right)$$

To optimize π_k:

$$\pi_k = \frac{1}{N}\sum_i r_{ik} = \frac{r_k}{N}$$

To optimize μ_{kd}

$$\sum_{n=1}^{N} r_{nk} \left(\sum_{d=1}^{D} \frac{x_{nd}}{\mu_{kd}} + \frac{(1-x_{nd})}{(1-\mu_{nd})} \right) = 0$$

$$\sum_{n=1}^{N} r_{nk} \sum_{d=1}^{D} \frac{x_{nd}}{\mu_{kd}} = \sum_{n=1}^{N} r_{nk} \sum_{d=1}^{D} \frac{1-x_{nd}}{1-\mu_{kd}}$$

$$\sum_{n=1}^{N} r_{nk} \sum_{d=1}^{D} x_{nd} = \mu_{kd} \sum_{n=1}^{N} r_{nk} \sum_{d=1}^{D} \frac{1-x_{nd}}{1-\mu_{kd}}$$

$$(1-\mu_{kd}) \sum_{n=1}^{N} r_{nk} \sum_{d=1}^{D} x_{nd} = \mu_{nd} \sum_{n=1}^{N} r_{nk} \sum_{d=1}^{D} (1-x_{nd})$$

$$\sum_{n}^{N} r_{nk} \sum_{d=1}^{D} x_{nd} - \mu_{kd} \sum_{n} r_{nk} \sum_{d=1}^{D} x_{nd} = \mu_{nd} \sum_{n} r_{nk} \left(\sum_{d} -x_{nd} + 1 \right)$$

$$1-\mu_{kd} = \frac{\mu_{nd} \sum_{n} r_{nk} \sum_{d} (-x_{nd} + 1)}{\sum_{n} r_{nk} \sum_{d} x_{nd}}$$

$$1-\mu_{kd} = \frac{\mu_{nd} \sum_{n} r_{nk} (1-x_{nd}}{\sum_{n} r_{nk} x_{nd}}$$

$$\frac{1-\mu_{kd}}{\mu_{kd}} = \frac{\sum_{n} (r_{nk} - r_{nk} x_{nd})}{\sum_{n} r_{nk} x_{nd}}$$

$$\frac{1-\mu_{kd}}{\mu_{kd}} = \frac{\sum_{n} r_{nk}}{\sum_{n} r_{nk} x_{nd}} - 1$$

$$1-\mu_{kd} = \mu_{kd} \frac{\sum r_{nk}}{\sum r_{nk} x_{nd}} - \mu_{kd}$$

$$\mu_{kd} = \frac{\sum_{n} r_{nk} x_{nd}}{\sum_{n} r_{nk}}$$