# Machine Learning 2 - Homework Week 4


Minh Ngo
MSc Artificial Intelligence
University of Amsterdam
nlminhtl@gmail.com


May 3, 2016

Collaborators: Riaan Zoetmulder

## Problem 3

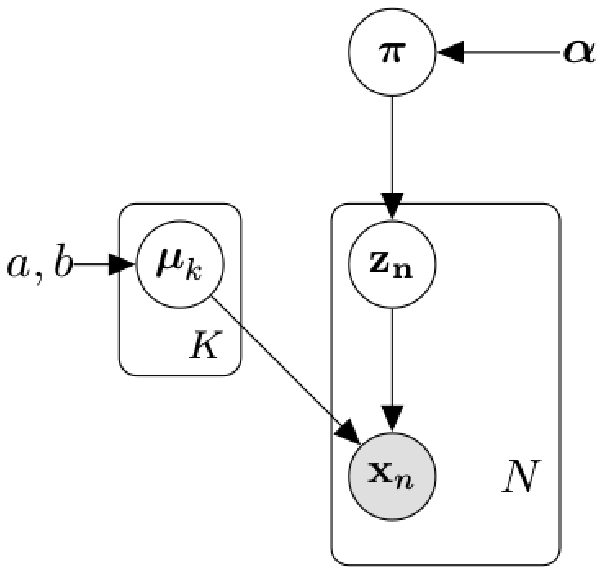


Figure 1: Mixtures of Bernoulli distribution

First we can write down a posterior probability:

$$p(\mu|\pi|\{x_n\}_{n=1}^N) = p(\pi|\alpha)\prod_{k=1}^K p(\mu_k|\alpha_k,\beta_k)\prod_{n=1}^N p(z_{nk}|\pi_k)p(x_n|\mu_k)^{z_{nk}} \tag{1}$$

To simplify the notation we will consider $\{x_n\}_{n=1}^N$ as $\{x_n\}$ in the later derivation.

Given the fact that:

$$p(z_{nk}|\pi_k) = \pi_k^{z_{nk}} \tag{2}$$

We can simplify log posterior:

$$\begin{aligned}
\ln p(\mu|\pi|\{x_n\}) &= \ln p(\pi|\alpha) + \sum_{k=1}^K \left[\ln p(\mu_k|\alpha_k,\beta_k) + \sum_{n=1}^N \left(\ln p(z_{nk}|\pi_k) + z_{nk}\ln p(x_n|\mu_k)\right)\right] \\
&= \ln p(\pi|\alpha) + \sum_{k=1}^K \ln p(\mu_k|\alpha_k,\beta_k) + \sum_{k=1}^K\sum_{n=1}^N \left(\ln p(z_{nk}|\pi_k) + z_{nk}\ln p(x_n|\mu_k)\right) \\
&= \ln p(\pi|\alpha) + \sum_{k=1}^K \ln p(\mu_k|\alpha_k,\beta_k) + \sum_{k=1}^K\sum_{n=1}^N z_{nk}\left(\ln \pi_k + \ln p(x_n|\mu_k)\right)
\end{aligned} \tag{3}$$

From Bishop (9.55, 9.56):

$$E[\ln p(\mu|\pi|\{x_n\})] = \ln p(\pi|\alpha) + \sum_{k=1}^K \ln p(\mu_k|\alpha_k,\beta_k) + \sum_{k=1}^K\sum_{n=1}^N \gamma(z_{nk})\left(\ln \pi_k + \ln p(x_n|\mu_k)\right) \tag{4}$$

where:

$$\gamma(z_{nk}) = \frac{\pi_k p(x_n|\mu_k)}{\sum_{j=1}^K \pi_j p(x_n|\mu_j)} \tag{5}$$

From Bishop (B.16, B.23, B.24):

$$p(\pi|\alpha) \propto \prod_{k=1}^K \pi_k^{\alpha_k - 1} \tag{6}$$

From Bishop (B.6):

$$p(\mu_k|\alpha_k, \beta_k) \propto \prod_{i=1}^{D} \mu_{ki}^{\alpha_k - 1}(1-\mu_{ki})^{\beta_k - 1} \tag{7}$$

From Bishop (B.1):

$$p(x_n|\mu_k) = \prod_{i=1}^{D} \mu_{ki}^{x_{ni}}(1-\mu_{ki})^{1-x_{ni}} \tag{8}$$

Consequently:

$$\ln p(\pi|\alpha) = \sum_{k=1}^{K}(\alpha_k - 1)\ln \pi_k + C_1 \tag{9}$$

$$\ln p(\mu_k|\alpha_k, \beta_k) = \sum_{i=1}^{D} \Big[(\alpha_k - 1)\ln \mu_{ki} + (\beta_k - 1)\ln(1-\mu_{k_i})\Big] + C_2 \tag{10}$$

$$\ln p(x_n|\mu_k) = \sum_{i=1}^{D} \Big[x_{ni} \ln \mu_{ki} + (1 - x_{ni})\ln(1-\mu_{ki})\Big] \tag{11}$$

M Step:

$$\frac{\partial E[..]}{\partial \mu_{ki}} = \frac{\partial \ln p(\mu_k|\alpha_k, \beta_k)}{\partial \mu_{ki}} + \sum_{n=1}^{N} \gamma(z_{nk}) \frac{\partial \ln p(x_n|\mu_k)}{\partial \mu_{ki}} = 0 \tag{12}$$

Other terms are gone since they don't depend on $\mu_{ki}$. Consequently:

$$\frac{\alpha_k - 1}{\mu_{ki}} + \frac{\beta_k - 1}{1-\mu_{ki}} + \sum_{n=1}^{N} \gamma(z_{nk}) \left[\frac{x_{ni}}{\mu_{ki}} - \frac{1-x_{ni}}{1-\mu_{ki}}\right] = 0 \tag{13}$$

$$\frac{\alpha_k - 1}{\mu_{ki}} - \frac{\beta_k - 1}{1-\mu_{ki}} + \sum_{n=1}^{N} \gamma(z_{nk}) \frac{x_{ni} - \mu_{ki}}{\mu_{ki}(1-\mu_{ki})} = 0 \tag{14}$$

$$(\alpha_k - 1) - \mu_{ki}(\beta_k + \alpha_k - 2) + \sum_{n=1}^{N} \gamma(z_{nk}) x_{ni} - \mu_{ki} \sum_{n=1}^{N} \gamma(z_{nk}) = 0 \tag{15}$$

Setting $N_k = \sum_{n=1}^{N} \gamma(z_{nk})$ we get:

$$\alpha_k - 1 + \sum_{n=1}^{N} \gamma(z_{nk} x_{ni}) = \mu_{ki}(\beta_k + \alpha_k - 2 + N_k) \tag{16}$$

Finally:

$$\mu_{ki} = \frac{\alpha_k - 1 + \sum_{n=1}^{N} \gamma(z_{nk}) x_{ni}}{\beta_k + \alpha_k - 2 + N_k} \tag{17}$$

To derive the E step we are taking derivative of the following Lagrangian:

$$\mathcal{L}(.., \lambda) = E[..] + \lambda(\sum_{k=1}^{K} \pi_k - 1) \tag{18}$$

$$\begin{aligned} \frac{\partial \mathcal{L}(..)}{\partial \pi_k} &= \frac{\alpha_k - 1}{\pi_k} + \frac{1}{\pi_k} \sum_{n=1}^{N} \gamma(z_{nk}) + \lambda \\ &= \frac{\alpha_k - 1}{\pi_k} + \frac{N_k}{\pi_k} + \lambda = 0 \end{aligned} \tag{19}$$

Consequently:

$$\lambda \pi_k = -\alpha_k + 1 - N_k \tag{20}$$

$$\lambda \sum_{d=1}^{K} \pi_d = \sum_{d=1}^{K} (-\alpha_d + 1 - N_d) \tag{21}$$

$$\lambda = K - N - \sum_{d=1}^{K} \alpha_d \tag{22}$$

$$\pi_k = \frac{\alpha_k - 1 + N_k}{N - K + \sum_{d=1}^{K} \alpha_d} \tag{23}$$