Problem 2.1

Similarities:

- Both models fall under the category of topic models, which aim to discover the underlying topics in a collection of documents
- Both are based on the assumption that each document is a mixture of different topics, and each topic is associated with a probability distribution over words

Differences:

- Multinomial mixture models modes the distribution of words within a document using a multinomial distribution. Each topic is associated with a probability distribution over the entire vocabulary. Mixture of unigrams models assume that the words in a document are generated independently from a multinomial distribution over the vocabulary for each topic. In other words, it treats each word in isolation and does not capture dependencies between adjacent words.
- Multinomial mixture models do not assume independence between words in a document. The multinomial distribution allows for capturing the co-occurrence patterns of words within topics. Mixture of unigrams models assume word independence within a document, meaning the occurrence of one word does not affect the likelihood of another word appearing.
  - This lack of independence makes multinomial mixture models have more parameters (mixture proportions) and be more complex in general.

Problem 2.2

(a) $p(w,d,z;\theta,\beta) = p(w|z;\beta)p(z|d;\theta)p(d) = \beta_{zw}\theta_{dz}\pi_d$

(b) $p(z|w,d;\theta,\beta) = \frac{p(w,d,z;\theta,\beta)}{p(w,d;\theta,\beta)} = \frac{p(w,d,z;\theta,\beta)}{\sum_{z^*} p(w,d,z=z^*;\theta,\beta)} = \frac{p(w|z;\beta)p(z|d;\theta)p(d)}{\sum_{z^*} p(w|z=z^*;\beta)p(z=z^*|d;\theta)p(d)} = \frac{p(w|z;\beta)p(z|d;\theta)}{\sum_{z^*} p(w|z=z^*;\beta)p(z=z^*|d;\theta)} = \frac{\beta_{zw}\theta_{dz}}{\sum_{z^*\in Z}\beta_{z^*w}\theta_{dz^*}}$

(c) $p(w|z,d;\theta,\beta) = \frac{p(w,z,d;\theta,\beta)}{p(z,d;\theta,\beta)} = \frac{p(w|z;\beta)p(z|d;\theta)p(d)}{p(z|d;\theta)p(d)} = \frac{\beta_{zw}\theta_{dz}}{\theta_{dz}} = \beta_{zw}$

Problem 2.3

$$\mathcal{L}(\Theta) = \sum_{d=1}^{D}\sum_{n=1}^{N_d} \log \sum_{k} P(z_n = k|d,\theta_d)P(w_n|\beta_k)$$

$$= \sum_{d=1}^{M}\sum_{n=1}^{N_d}\sum_{k=1}^{K} q_d(z_n = k)(\log P(z_n = k, w_n|d,\theta_d,\beta_k) - \log q_d(z_n = k))$$

$$-\sum_{n=1}^{N_d}\sum_{d=1}^{M} q(z_n = k|d,w_n)(\log q_d(z_n = k) - \log P(z_n = k,\theta|w_n))$$

$$
\begin{aligned}
&= \sum_{d=1}^{M} \sum_{n=1}^{N_d} \left[E_{q_d}[\log P(z_n, w_n | d, \Theta)] - E_{q_d}[\log q_d(\theta, z_n)] \right] \\
&\qquad\qquad - \sum_{d=1}^{M} E_{q_d}[\log q_d(z_n) - \log P(\theta, z_n | w_n)] \\
&= L(\Theta, q) - D(q(\theta, z) || P(Z | d, \Theta)
\end{aligned}
$$