Matt Dickenson `mcd31`
STA561/CS571 — Fall 2013 **Homework 7** Due: 11 November, 2013

*Homework Notes:* I did not work with anyone else on this homework or refer to resources other than the course notes, textbook, and course Piazza page.

**Problem 1**

**A** We use the forward KL divergence, minimized with respect to $q(x)$:

$$
\begin{aligned}
KL(p||q) &= \sum_x p(x,y) \log \frac{p(x,y)}{q(x,y)} \\
&= \sum_x p(x,y) \log \frac{p(x,y)}{q(x)q(y)} \\
\log q(x) &= \log \int_y q(x,y) dy \\
&= \log \int_y q(x)q(y) dy \\
&= \log \int_y q(x)q(y) \frac{p(x,y)}{p(x,y)} dy \\
&= \log \int_y p(x,y)q(x)q(y)\ over p(x,y) dy \\
&= \log p(x) \frac{q(x)\mathbb{E}[q(y)]}{p(x)} \\
&= \log p(x)q(x)\mathbb{E}[q(y)] - p(x) \\
KL(\tilde{p}||q) &= -\mathbb{E}[\log q(x)q(y)] + \mathbb{E}[\log p(x)] - \log p(x)
\end{aligned}
$$

We can disregard the last term of the last line as a constant.

$$
\begin{aligned}
\frac{\partial}{\partial q(x)} KL(\tilde{p}||q) &= 0 \\
-\frac{\mathbb{E}[q(y)]}{q(x)} + \frac{1}{\mathbb{E}[p(x)]} &= 0 \\
q(x) &= \mathbb{E}[p(x)q(y)]
\end{aligned}
$$

**B** The variational updates applied from this approach will be appropriate under two conditions: we expect that $q(x)$ will not be close to zero (which would lead to very large divergence) and that the true distribution is unimodal. These constraints are valid in some situations, but in general it is safer to use the backwards KL divergence as we did in class.

Matt Dickenson `mcd31`


### Problem 2

**A** We expect values of $\pi_j$ (with $K = 3$ for this example) close to $(1, 0, 0), (0, 1, 0)$, or $(0, 0, 1)$ with high probability, but assign very low probbaility to $\pi_j = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.

**B** In this case, $\pi_j$ is uniform.

**C** When $\alpha = 10$, we assign high probability to $\pi_j = (\frac{1}{k}, \frac{1}{k}, ...)$.

**D** Figure 1 presents 1,000 samples from three different parameterizations of the Beta distribution: $(0.2, 0.2), (1, 1)$, and $(10, 10)$. These samples support the answers above. For the first panel (Beta$(0.2, 0.2)$), samples of 0 and 1 are highly probable. In the second panel (Beta$(1, 1)$), the samples appear to be uniformly distributed over the unit interval. Samples in the third panel (Beta$(10, 10)$) favor values near the center of the distribution ($\frac{1}{k} = \frac{1}{2}$) rather than at the extremes of 0 or 1.

**E** I suspect that inference in these models is easier with values of $\alpha$ similar to $[\alpha_1, ..., \alpha_k] = 0.2$. Inference with this setting is easier because each document is assigned to a single topic with high probability, rather than a mixture of topics. That is, assigning $z_{i,j}$ is easier when $[\alpha_1, ..., \alpha_k] < 1$.

### Problem 3

One major challenge in moving forward with my project has been deciding how to make daily event data commensurable with dyadic data about military disputes that includes a start and end date. The current solution, which I will like employ for the duration of the project, is to aggregate the event data into dyad-months and put the dispute data into the same format. I am currently working on selecting features to use in training the HMM model.

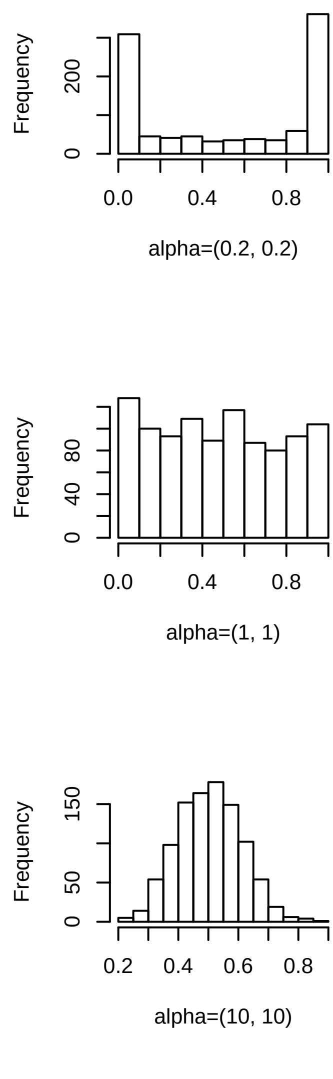


Figure 1: 1,000 Samples from Beta Distributions with Three Different Values of $\alpha$