Assignment 1

This assignment is due by the start of class on Monday, September 19.

1. Please consult the solutions of problems 1.17 and 2.5 from the textbook that are available on the text's web page. This provides useful properties of the Beta and Dirichlet distributions.

   Use these properties to solve problems 2.6 and 2.10 in the textbook. For 2.10 solve only $\mathrm{var}[\mu_j]$.

2. Bayesian Unigram model for text learning.

   We have a simple probabilistic model for text generation using a vocabulary of K words $w_1, \ldots, w_K$, specified by a discrete distribution over the words with parameters $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)$, so that $Pr[\text{next word is } w_j] = \mu_j$. The discrete distribution is specified by Eq (B.54) in the Appendix.

   To generate a document with N word tokens using this model we sample each token independently from $\boldsymbol{\mu}$. As shown in Section 2.2 of the textbook we can use a Dirichlet prior for this problem to yield a Dirichlet posterior $Pr(\boldsymbol{\mu}|\text{Data})$, given in Eq (2.41), where the prior is specified by the vector of counts $\boldsymbol{\alpha}$.

   Our data is one document of length N, given by the token sequence Data$= x_1, \ldots, x_N$, where each x_i is some word w_j in the vocabulary.

   (i) Calculate the predictive distribution for this problem. Specifically, show

   $$Pr[\text{next word is } w_j|\text{Data}] = \int_{\boldsymbol{\mu}} Pr(\boldsymbol{\mu}|\text{Data})Pr([\text{next word is } w_j|\boldsymbol{\mu}) \, d\boldsymbol{\mu} = \frac{m_k + \alpha_k}{N + \alpha_0}.$$

   (ii) Calculate the evidence function. Specifically, show

   $$Pr[\text{Data}|\boldsymbol{\alpha}] = \int_{\boldsymbol{\mu}} Pr(\boldsymbol{\mu}|\boldsymbol{\alpha})Pr(\text{Data}|\boldsymbol{\mu}) \, d\boldsymbol{\mu} = \frac{\Gamma(\alpha_0) \prod_{k=1}^{K} \Gamma(\alpha_k + m_k)}{\Gamma(\alpha_0 + N) \prod_{k=1}^{K} \Gamma(\alpha_k)}.$$

   As we discuss later in the course, the evidence function can be used to select a suitable value for $\boldsymbol{\alpha}$. For now we just focus on the calculation.