Assignment 1

This assignment is due by the start of class on Thursday, September 17.

1. Solve problems 2.6 and 2.10 in the textbook.

   **Notes:** (1) For 2.6: please consult the solutions of problems 1.17 and 2.5 that are available on the text’s web page. (2) For 2.10 solve only $E[\mu_j]$.

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 $\mu = (\mu_1, \ldots, \mu_K)$, 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 $\mu$. As shown in Section 2.2 of the textbook we can use a Dirichlet prior for this problem to yield a Dirichlet posterior $Pr(\mu|\text{Data})$, given in Eq (2.41), where the prior is specified by the vector of counts $\alpha$.

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

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

   \[
   Pr[\text{next word is } w_j|\text{Data}] = \int_{\mu} Pr(\mu|\text{Data})Pr(\text{next word is } w_j|\mu) \, d\mu.
   \]

   (ii) Calculate the evidence function

   \[
   Pr[\text{Data}|\alpha] = \int_{\mu} Pr(\mu|\alpha)Pr(\text{Data}|\mu) \, d\mu.
   \]

   As we discuss later in the course, the evidence function can be used to select a suitable value for $\alpha$. For now we just focus on the calculation. Your answer should be written in terms of the Gamma function, $\Gamma(\cdot)$. 