**Problem 1:**

We’ll show that s.t .   
As we saw in class we know that:  
   
Let’s denote a LaGrange multiplier function, with the constraints:  
So we’ll partially derive:  
 and   
So:   
   
And:   
together:  
   
We’ll recall,

**Problem 2:  
(i)** We’ll notice that if we choose we’ll get:  
Because we get an expression which doesn’t involve , and is a close expression, we can say is sampled from a uniform distribution.

**(ii)**  Uniform On the other hand , that is sampled from uniform distribution is (as proven above):   
Meaning, a uniform is a vector where entry is and , while a that is sampled from uniform distribution is a vector that can receive any value with equal probability.

**(iii)** In the context of mixture models and estimating π , using the prior instead of a pure likelihood-based approach can prevent overfitting or misleading results. In particular, if the data set is too small or noisy, the prior can provide a useful source of information about the distribution, which can help better generalize new data.

For example, if we sampled 𝑁 (a relatively small number) samples from a mixture model and received data in which no sample comes from the entry with the largest weight – we’ll denote as c, then a likelihood-based calculation will give . On the other hand, if we can assume in advance that the data is sampled from some distribution where gets a relatively high value with a high probability, then the model will give a more accurate value for .