**1.a**

Estimate for words that occur in a query is

Different forms of estimation come from different .

For Jelinek-Mercer smoothing:

,

and that is a constant:

**1.b**

For Dirichlet priori smoothing:

and that depends on document length:

**1.c**

In this two-stage smoothing strategy, the document language model is first smoothed by Dirichlet priori to explain unseen words and then smoothed further by Jelinek Mercer to explain noise in query

and that

Particularly, when , the model reduces to Dirichlet priori smoothing and ; when , the model is only smoothed by Jelinek-Mercer and .

**6.a**

In general, the verbose queries, whether long or short, exhibit more variation across the whole range than their counterpart keyword queries. Particularly, the **short verbose queries** have the most variation compared with other types of queries as it has the smallest MAP.

Because smoothing is responsible for explaining the common words in a query. Insufficient smoothing will result in worse performance of verbose queries than that of keyword performance.

**6.b**

From the shape of lines, we find that the two types of keyword queries behave similarly, while the two types of verbose queries have the similar shape. However, the shape of keyword queries is different from the shape of verbose queries.

**6.c**

Still the shape of long keyword query and the shape of short keyword query are similar. The shape of keyword query and the shape of verbose query are different.

With the help of Jelinek-Mercer, Dirichlet smoothing becomes much less sensitive to the type/length of queries and can achieve a higher precision on long queries.

**6.d**

I would adopt two-stage smoothing method with proposed parameter estimation due to its effectiveness, robustness and no ad hoc parameter tuning involved.

From the observation, we find that Jelinek-Mercer tends to perform better for verbose queries while Dirichlet prior is better at keyword queries

Using two-stage smoothing, we can decouple the two roles of smoothing and let Jelinek-Mercer take care of query modeling role and Dirichlet prior focus on the document collection.

I would set the default at around 0.7 and set at around 2000.