Compatibility calculation

Start from the Kullback-Leibler distance,  
  
where is the entropy in the distribution of . The second term depends on , but this is zero in all cases, except for the symbol present in at this site. Let us split the second term into its two parts. First, the components where the symbol in is the same as the one in and index this with . Then and is the probability of this symbol in the main distribution. Then

|  |  |
| --- | --- |
|  | (1) |

Consider the middle term where the summation runs over the cases where the symbol in the bulk of the data is not the one in our sequence of interest. Note also that for all , or we can say, for some very small number .  
since the probabilities sum to 1, we can say , so

and substitute this into (1)

Now the second term has no information about the data set, so we can remove and say,   
we also know that , which makes the term disappear, so one is left with