The log-likelihood function

$L((f_1, \ldots, f_k), (p_1, \ldots, p_k)) = \sum_{i=1}^{k} f_i \log p_i$

that you (hopefully) obtained in final project is also know as cross entropy. It is extremely popular in machine learning, namely, in classification problems. It allows you to measure how good probability distribution $(p_1, \ldots, p_k)$ fits the actual absolute frequencies obtained from the data $(f_1, \ldots, f_k)$. Assume that frequencies $(f_1, \ldots, f_k)$ are fixed. What is the best distribution $(p_1, \ldots, p_k)$ from likelihood's perspective? Intuitively, it seems that we have to put relative frequencies

$r_i = \frac{f_i}{\sum_{j=1}^{k} f_j}$

as p_i to get best fit. In fact, it is true. To prove it, let us use Jensen's inequality for logarithms. It is stated as follows:

For any values $\alpha_1, \ldots, \alpha_k$, such that $\sum_{j=1}^{k} \alpha_j = 1$ and $\alpha_j \geq 0$ for all $j = 1, \ldots, k$, and any positive values $x_1, \ldots, x_k$, the following inequality holds:

$\log \sum_{j=1}^{k} \alpha_j x_j \geq \sum_{j=1}^{k} \alpha_j \log(x_j)$.

Use this inequality to prove that

$\sum_{j=1}^{k} r_j \log p_j - \sum_{j=1}^{k} r_j \log r_j \leq 0,$

then prove that to obtain maximum log-likelihood (and therefore maximum likelihood) for fixed $(f_1, \ldots, f_k)$ we have to put $p_i = r_i$, $i = 1, \ldots, k$.

Hint. Use properties of logarithm to transform left-hand part of the last inequality to right-rand part of the previous inequality.

1) First let us prove that

$$\sum_{j=1}^{k} r_j \log p_j - \sum_{j=1}^{k} r_j \log r_j \leq 0$$

Using properties of logarithm we can rewrite the above like

$$\sum_{j=1}^{k} r_j (\log p_j - \log r_j) \leq 0$$

$$\sum_{j=1}^{k} r_j \log \frac{p_j}{r_j} \leq 0$$

Now, supposing that $\alpha_j = r_j$ and $x_j = \frac{p_j}{r_j}$, we can rewrite Jensen's inequality the following way:

$$\log \sum_{j=1}^{k} r_j \frac{p_j}{r_j} \geq \sum_{j=1}^{k} r_j \log \frac{p_j}{r_j}$$

$$\sum_{j=1}^{k} r_j \log \frac{p_j}{r_j} \leq \log \sum_{j=1}^{k} p_j$$

Since $\sum_{j=1}^{k} p_j = 1$, we obtain the following:

$$\sum_{j=1}^{k} r_j \log \frac{p_j}{r_j} \leq \log 1 \,,$$

which means that

$$\sum_{j=1}^{k} r_j \log \frac{p_j}{r_j} \leq 0$$

2) From the inequality proved above, and considering that $\sum_{j=1}^{k} r_j \log p_j$ represents the log-likelihood function, we can see that $\sum_{j=1}^{k} r_j \log \frac{p_j}{r_j}$ reaches its maximum at 0, which can be obtained only when $\log \frac{p_j}{r_j} = 0$, which means when $\frac{p_j}{r_j} = 1$ or just $p_j = r_j$.

Alternatively, we can consider the initial inequality with unchanged log-likelohood function:

$$\sum_{j=1}^{k} r_j \log p_j - \sum_{j=1}^{k} r_j \log r_j \leq 0$$

Knowing that the inequality holds, we can see that maximum value of the left-hand expression is 0, when

$$\sum_{j=1}^{k} r_j \log p_j = \sum_{j=1}^{k} r_j \log r_j$$

Obviously, this equality holds only when $p_j = r_j$.