

$$H(p) = - \sum_{i=1}^K p_i \log(p_i)$$

Probability:

$$P(X = x_i) = p_i, \text{ for } i = 1, 2, \dots, K$$

Constrant:

$$\sum_{i=1}^K p_i = 1 \rightarrow \sum_{i=1}^K p_i - 1 = 0$$

Lagrange multipliers:

$$\begin{aligned} \mathcal{L} &= H(p) + \lambda \left(\sum_{i=1}^K p_i - 1 \right) \\ &= - \sum_{i=1}^K p_i \log(p_i) + \lambda \left(\sum_{i=1}^K p_i - 1 \right) \end{aligned}$$

To find the highest entropy, set $\frac{\partial \mathcal{L}}{\partial p_i} = 0$

$$\frac{\partial \mathcal{L}}{\partial p_i} = \frac{\partial \left(- \sum_{i=1}^K p_i \log(p_i) + \lambda (\sum_{i=1}^K p_i - 1) \right)}{\partial p_i} = -\log(p_i) - 1 + \lambda = 0$$

$$\log(p_i) = \lambda - 1$$

$$\text{Thus, } p_1 = p_2 = \dots = p_i = \frac{1}{K}$$

Therefore, when $p^ = \frac{1}{K}$, the categorical distribution has the highest entropy,*

$$H(p^*) = \log K.$$