



Maximum Likelihood

1 Why

We provide a principle for distribution selection based on maximizing the probability of the dataset.

2 Definition

We have a set of outcomes A and a dataset (a^1, \dots, a^n) . We want a distribution $p : A \rightarrow \mathbf{R}$. We define $\bar{p} : A^n \rightarrow \mathbf{R}$ by $\bar{p}(a^1, \dots, a^n) = \prod_{i=1}^n p(a^i)$.

The *principle of maximum likelihood* says to solve:

find p , a distribution
to maximize $\bar{p}(a^1, \dots, a^n)$



