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 $(a^1, \ldots, a^n)$. We want a distribution $p : A \to \mathbf{R}$. We define $\bar{p} : A^n \to \mathbf{R}$ by $\bar{p}(a^1, \ldots, a^n) = \prod_{i=1}^n p(a^i)$.

The **principle of maximum likelihood** says to solve:

$$\begin{aligned} \text{find} \quad & p, \text{ a distribution} \\ \text{to maximize} \quad & \bar{p}(a^1, \ldots, a^n) \end{aligned}$$