

## MAXIMUM LIKELIHOOD DISTRIBUTIONS

## Why

We want to summarize a dataset with a distribution.

## Overview

The likelihood (or distribution likelihood) of a probability distribution  $p:A\to \mathbf{R}$  on a dataset  $a^1,\ldots,a^n\in A$  is  $\prod_{i=1}^n p(a^i)$ . A maximum likelihood distribution  $p^*:A\to \mathbf{R}$  is one which maximizes the likelihood over all distributions on A.

We call the correspondence between datasets and distributions the *maximum likelihood algorithm*. We say that we are selecting the distribution according to the *maximum likelihood principle*. In general, we call any function from datasets to distributions a *distribution selector*.

