Why

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

Definition

Let $(x^1, \ldots, x^n)$ be a dataset in R^d.

The *principle of maximum likelihood* says to solve:

$$\textbf{find} \quad f \text{ a density}$$
$$\textbf{to maximize} \quad \prod_{k=1}^{n} f(x^k)$$

We call a solution of the above problem a *maximum likelihood density*. We call the objective of the above problem *density likelihood*.