Maximum-Likelihood Estimation at First Glance

Detlef Prescher
University of Heidelberg
prescher@cl.uni-heidelberg.de

April 30, 2007

Why Maximum-Likelihood Estimation?

Maximum-likelihood estimation (Fisher 1912), typically yields an excellent estimate if the given corpus is large:

- maximum-likelihood estimators fulfill the so-called **invariance principle**,
- under certain conditions which are typically satisfied in practical problems, they are **consistent estimators**,
- unlike the relative-frequency estimator, maximum-likelihood estimators typically **do not over-fit the given corpus** in practice.

Maximum-Likelihood Estimation is probably the most widely used estimation method

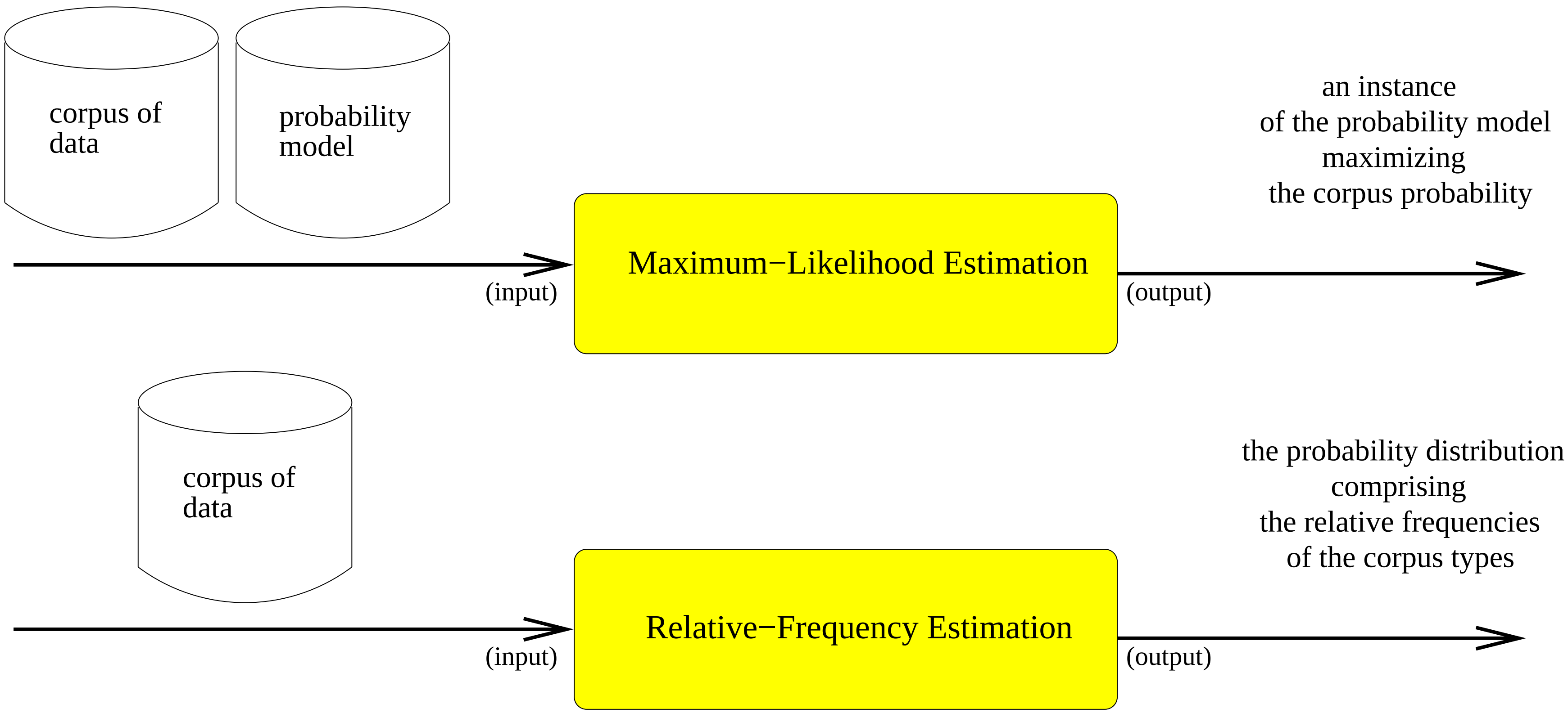
More on Maximum-Likelihood Estimation...

Common wisdom is: Maximum-likelihood estimation equals relative-frequency estimation!

In this lecture, you learn:

Common wisdom is rarely wise, merely common.

A First Simple Argument against Common Wisdom



Maximum-likelihood estimation is a fully-fledged estimation method with a probability model as input, whereas relative-frequency estimation acts on a corpus only...

Reminder: Probability Models

Definition 1. *A non-empty set $\mathcal{M}$ of probability distributions on a set $\mathcal{X}$ of types is called a* ***probability model on*** *$\mathcal{X}$. The elements of $\mathcal{M}$ are called* ***instances of the model*** *$\mathcal{M}$. The* ***unrestricted probability model*** *is the set $\mathcal{M}(\mathcal{X})$ of all probability distributions on the set of types*

$$\mathcal{M}(\mathcal{X}) = \left\{ p\colon \mathcal{X} \to [0,1] \;\middle|\; \sum_{x \in \mathcal{X}} p(x) = 1 \right\}$$

A probability model $\mathcal{M}$ is called ***restricted*** *in all other cases*

$$\mathcal{M} \subseteq \mathcal{M}(\mathcal{X}) \quad \text{and} \quad \mathcal{M} \neq \mathcal{M}(\mathcal{X})$$

Reminder: Maximum-Likelihood Estimates

Definition 2. *Let f be a non-empty and finite corpus on a countable set $\mathcal{X}$ of types. Let $\mathcal{M}$ be a probability model on $\mathcal{X}$. The **probability of the corpus** allocated by an instance p of the model $\mathcal{M}$ is defined as*

$$L(f;p) = \prod_{x \in \mathcal{X}} p(x)^{f(x)}$$

*An instance $\hat{p}$ of the model $\mathcal{M}$ is called a **maximum-likelihood estimate of $\mathcal{M}$ on f**, if and only if the corpus f is allocated a maximum probability by $\hat{p}$*

$$L(f;\hat{p}) = \max_{p \in \mathcal{M}} L(f;p)$$

Problems of Maximum-Likelihood Estimation

Maximum-likelihood estimates are the solutions of a quite complex optimization problem. So, some nasty questions about maximum-likelihood estimation arise:

Existence Is there for any probability model and any corpus a maximum-likelihood estimate of the model on the corpus?

Uniqueness Is there for any probability model and any corpus a unique maximum-likelihood estimate of the model on the corpus?

Computability For which probability models and corpora can maximum-likelihood estimates be efficiently computed?

A Very Special Case

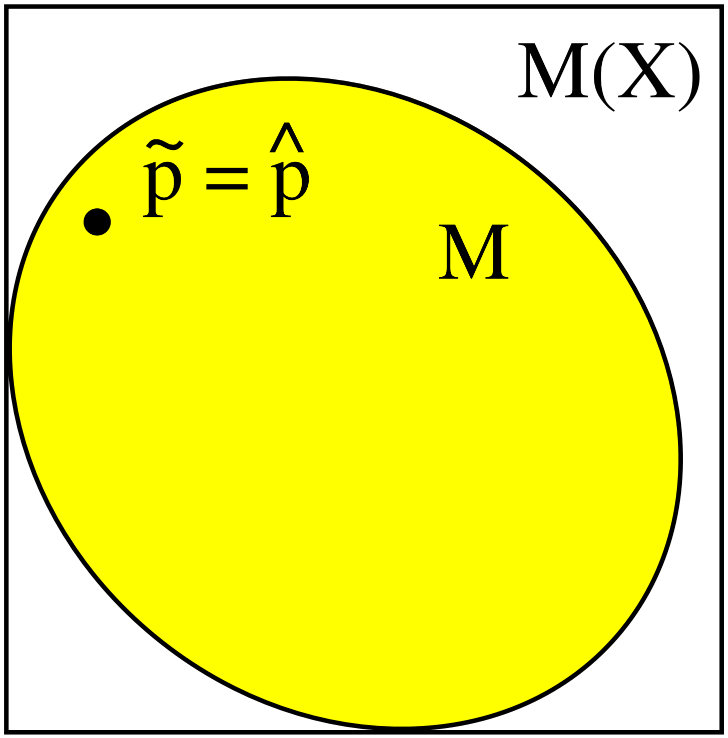
Fortunately, existence, uniqueness and computability of maximum-likelihood estimates are not always a problem:

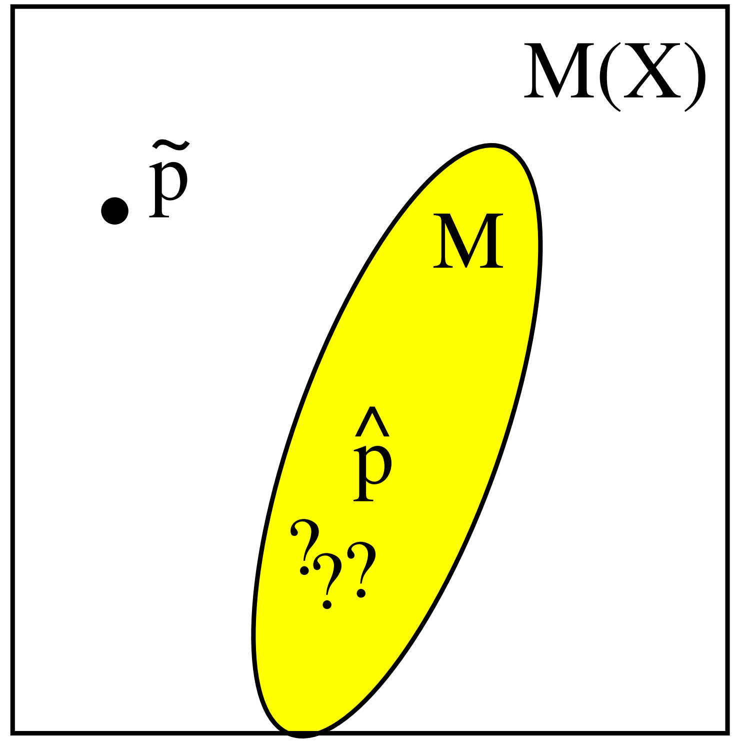
Theorem 1. *Let f be a non-empty and finite corpus on a countable set $\mathcal{X}$ of types. Then:*

(i) The relative-frequency estimate $\tilde{p}$ is a unique maximum-likelihood estimate of the unrestricted probability model $\mathcal{M}(\mathcal{X})$ on f.

(ii) The relative-frequency estimate $\tilde{p}$ is a maximum-likelihood estimate of a (restricted or unrestricted) probability model $\mathcal{M}$ on f, if and only if $\tilde{p}$ is an instance of the model $\mathcal{M}$. In this case, $\tilde{p}$ is a unique maximum-likelihood estimate of $\mathcal{M}$ on f.

Maximum-Likelihood and Relative-Frequency Estimation





In practice, Maximum-Likelihood Estimation and Relative-Frequency Estimation typically yield different estimates.