Estimation Theory at First Glance

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

April 23, 2007

Corpora

Definition 1. *Let $\mathcal{X}$ be a countable set. Then, each function $f : \mathcal{X} \rightarrow \mathcal{N}$ is called a **corpus**, each $x \in \mathcal{X}$ is called a **type**, and each value of f is called a **type frequency**. The **corpus size** is defined as*

$$|f| = \sum_{x \in \mathcal{X}} f(x)$$

(Here, $\mathcal{N}$ is the set of all natural numbers, including 0.)

Occurence Frequencies

Definition 2. *Let* $x_1, \ldots, x_n$ *be a finite sequence of type instances from* $\mathcal{X}$. *Each* x_i *of this sequence is called a* ***token***. *The* ***occurrence frequency*** *of a type* x *in the sequence is defined as the following count*

$$f(x) = |\{\, i \mid x_i = x\}\,|$$

Probability Distributions

Definition 3. *Let $\mathcal{X}$ be a countable set of types. A real-valued function $p\colon \mathcal{X} \to \mathcal{R}$ is called a* ***probability distribution on*** $\mathcal{X}$*, if p has two properties: First, p´s values are non-negative numbers*

$$p(x) \geq 0 \quad \text{for all } x \in \mathcal{X}$$

and second, p´s values sum to 1

$$\sum_{x \in \mathcal{X}} p(x) = 1$$

The Three Axioms of Probability Theory

Define

$$p(A) := \sum_{x \in A} p(x) \text{ for all } A \subseteq \mathcal{X}$$

to satisfy the three axioms of standard Probabiliy Theory.

Axiom 1. $p(A) \geq 0$ *for any event* $A \subseteq \mathcal{X}$.

Axiom 2. $p(\mathcal{X}) = 1$.

Axiom 3. $p(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} p(A_i)$ *for any infinite sequence of disjoint events* $A_1, A_2, A_3, ...$

Relative-Frequency Estimates

Definition 4. *Let f be a non-empty and finite corpus. The probability distribution*

$$\tilde{p} \colon \mathcal{X} \to [0,1] \quad \textit{where} \quad \tilde{p}(x) = \frac{f(x)}{|f|}$$

is called the ***relative-frequency estimate*** *on f.*

Notes:

$\tilde{p}$ is well defined, since both $|f| > 0$ and $|f| < \infty$.

$\tilde{p}$'s values sum to one: $\sum_{x \in \mathcal{X}} \tilde{p}(x) = \sum_{x \in \mathcal{X}} |f|^{-1} \cdot f(x) = |f|^{-1} \cdot \sum_{x \in \mathcal{X}} f(x) = |f|^{-1} \cdot |f| = 1$.

Probability Models

Definition 5. *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*

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

Maximum-Likelihood Estimates

Definition 6. *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 $\hat{p}$*

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

Overview of Maximum-Likelihood and Relative-Frequency