

# Towards a Notion of Credal Maximum Likelihood – Imprecise Probability-based Mixed Models and Unobserved Heterogeneity

Thomas Augustin

Department of Statistics  
University of Munich (LMU)  
thomas@stat.uni-muenchen.de

## Abstract

The poster presents some ideas to contribute to a theory of credal maximum likelihood. We consider parametric sampling models with unobserved heterogeneity, where one parameter may vary in an unknown way within an interval. These models are a natural and powerful imprecise probability alternative to mixed models in traditional statistics, where – according to the dogma of ideal precision – a certain precise distribution for the parameter is assumed in order to be able to integrate out the unobserved heterogeneity.

To formulate the basic idea more precisely, let a sample  $y_1, \dots, y_n$  of  $n$  independent observations be given, where, for every  $i$ , the distribution of the underlying variable  $Y_i$  belongs to a parametric family, indexed by a real-valued parameter  $\vartheta_i$ , and the distribution of  $Y_i$  can be described by a density or a probability mass function  $f_{\vartheta_i}(y_i)$ . In generalization of the classical i.i.d. maximum likelihood estimate  $\hat{\vartheta}$ , which solves de facto,

$$\prod_{i=1}^n f_{\vartheta_i}(y_i) \rightarrow \max \quad (1)$$

in  $\vartheta_1, \dots, \vartheta_n$  under the assumption

$$\vartheta_i \equiv \vartheta, i = 1, \dots, n,$$

we consider, for some given constant  $\delta \geq 0$ , the solution of (1) in  $\vartheta_1, \dots, \vartheta_n$  and  $L\vartheta_\delta, U\vartheta_\delta$  subject to the constraints

$$\vartheta_i \in [L\vartheta_\delta, U\vartheta_\delta], \quad i = 1, \dots, n, \quad (2)$$

$$U\vartheta_\delta - L\vartheta_\delta \leq \delta. \quad (3)$$

The interval  $[\widehat{L\vartheta}_\delta, \widehat{U\vartheta}_\delta]$  derived from the optimal values of  $L\vartheta_\delta$  and  $U\vartheta_\delta$  is called *credal maximum likelihood estimate (of level  $\delta$ )*.

We illustrate our approach with some of the most popular parametric models. In particular, we present powerful algorithms to calculate the credal maximum likelihood estimate, investigate its behaviour and discuss criteria for the choice of delta, in the cases of

- the Bernoulli distribution
- the Poisson distribution
- the normal distribution with unknown mean and known variance.

Then we will generalize our approach to cover a multidimensional parameter and to deal with covariates, which we illustrate in the case of

- linear regression models.

**Keywords.** Credal sampling models, imprecise sampling models, maximum likelihood, frequentist statistical inference, unobserved heterogeneity, regression models, mixed models.