

Proposal for Chapter in the Book “Geometric Structures of Information”:

Chapter Title: New Data-Set-Size Asymptotics of Scaled Bregman Distances

Estimated Length: 30-40 pages

Wolfgang Stummer¹ and Anna-Lena Kißlinger²

¹ Department of Mathematics, University of Erlangen–Nürnberg, Cauerstrasse 11, 91058 Erlangen, Germany,
as well as Affiliated Faculty Member of the School of Business and Economics, University of Erlangen–Nürnberg, Lange Gasse 20, 90403 Nürnberg, Germany, Corresponding Author.

² Chair of Statistics and Econometrics, University of Erlangen–Nürnberg, Lange Gasse 20, 90403 Nürnberg, Germany.

Abstract. Divergences (distances) which measure the similarity respectively proximity between two probability distributions have turned out to be very useful for several different tasks in information geometry, statistics, machine learning, etc. Some prominent examples are the Kullback-Leibler information KL (relative entropy), the Csiszar-Ali-Silvey ϕ -divergences CASD, the “classical” (i.e., unscaled) Bregman distances OBD and the more flexible scaled Bregman distances SBD (of Stummer and Vajda) which cover all the KL, CASD and OBD at once. We present new results on the problem of robustly finding the distribution – within a preselected class of parametric models – which is closest (in SBD) to the appropriate data-derived distribution. In particular, we derive the associated uncertainty-natured asymptotics for large data sets as well as corresponding tests for goodness of fit respectively goodness of approximation. These investigations will be visualized by numerous figures and diagrams.