Goodness-of-fit tests based on (h, Φ)-divergences and entropy differences

Jean-François Bercher, V Girardin, J Lequesne, Ph Regnault

Goodness-of-fit tests based on (h, φ)-divergences and entropy differences.

J.-F. Bercher*, V. Girardin⋆, J. Lequesne⋆ and Ph. Regnault†

* Laboratoire d'Informatique Gaspard-Monge, UMR8049, ESIEE, Cité Descartes BP99, 93162 Noisy-le-Grand cedex, France.

jf.bercher@esiee.fr

⋆ Laboratoire de Mathématiques N. Oresme, UMR6139, Université de Caen – Basse Normandie, Campus II, BP 5186, 14032 Caen, France.

girardin@math.unicaen.fr and justine.lequesne@unicaen.fr

† Laboratoire de Mathématiques de Reims, EA4535, Université de Reims Champagne-Ardenne, UFR SEN, BP 1039, 51687 Reims, France.

philippe.regnault@univ-reims.fr

Abstract

We consider fitting uncategorical data to a parametric family of distributions by means of tests based on (h, φ)-divergence estimates.

The class of (h, φ)-divergences, introduced in Salicrú *et al.* (1993), includes the well-known classes of ϕ-divergences, of Bregman divergences and of distortion measures. The most classic are Kullback-Leibler, Rényi and Tsallis divergences. Most of (h, φ)-divergences are associated to (h, φ)-entropy functionals, e.g., Kullback-Leibler divergence to Shannon entropy. Distributions maximizing (h, φ)-entropies under moment constraints are involved in numerous applications and are also of theoretic interest. Besides the family of exponential distributions maximizing Shannon entropy, see, e.g., Bercher (2014) for an overview of various information inequalities involving the so-called q-Gaussian distributions, i.e., distributions maximizing Rényi (or Tsallis) entropy under variance constraints.

For distributions maximizing Shannon or Rényi entropy under moment constraints, the related divergence is well known to reduce to an entropy difference. Then estimating divergence reduces to estimating entropy; see Girardin and Lequesne (2013a, 2013b).

A commonly used non-parametric procedure for estimating entropy is the nearest neighbors method; see Vasicek (1976) for Shannon entropy and Leonenko *et al.* (2008) for Rényi entropy. Vasicek (1976) deduced a test of normality whose statistics involves Shannon entropy difference, thus opening the way to numerous authors who adapted or extended the procedure to obtain goodness-of-fit tests for various sub-families of exponential distributions.

Recently, Girardin and Lequesne (2013b) considered goodness-of-fit tests for q-Gaussian distributions (among which the non-standard Student distribution arises as a meaningful example) based on Rényi's divergence and entropy differences. Further, we will show how this methodology may extend to families of distributions maximizing other (h, φ)-entropies.

References:

J.-F. Bercher. (2014). On generalized Cramér-Rao inequalities and an extension of the Shannon-Fisher-Gauss setting. In *New Perspectives on Stochastic Modeling and Data Analysis*, J. Bozeman, V. Girardin and C. H. Skiadas (Eds), ISAST, 19-35.

V. Girardin and J. Lequesne. (2013a). Relative Entropy Versus Entropy Difference in Goodness-of-Fit Tests. Application to Pareto fitting. *Prépublication du LMNO-UCBN*, Caen, France.

V. Girardin and J. Lequesne. (2013b). Entropy based goodness-of-fit tests. *Proceedings 2d Marrakesh International Conference on Probability and Statistics*.

N. Leonenko, L. Pronzato and V. Savani. (2008). A class of Rényi information estimators for multidimensional densities. *Ann. Stat.* 36(5) 2153-2182.

M. Salicrú, M. L. Menéndez, D. Morales and L. Pardo. (1993). Asymptotic distribution of (h, φ)-entropies. *Comm. Stat. (Theory and Methods) 22(7) 2015-2031.*

O. Vasicek. (1976). A test of normality based on sample entropy. *J. Roy. Statist. Soc. Ser. B* 38(1) 54-59.

Key Words: (h, ϕ)-divergence, (h, ϕ)-entropy, Kullback-Leibler divergence, Shannon entropy, Rényi entropy, maximum entropy distribution, q-Gaussian distribution, goodness-of-fit testing, nearest neighbors estimation.