KYBERNETIKA — VOLUME *43* (2007), NUMBER 5, PAGES 731-746

Optimality Conditions for Maximizers of the Information Divergence from an Exponential Family

František Matúš

Abstract: The information divergence of a probability measure P from an exponential family $\mathcal{E}$ over a finite set is defined as infimum of the divergences of P from Q subject to $Q \in \mathcal{E}$. All directional derivatives of the divergence from $\mathcal{E}$ are explicitly found. To this end, behaviour of the conjugate of a log-Laplace transform on the boundary of its domain is analysed. The first order conditions for P to be a maximizer of the divergence from $\mathcal{E}$ are presented, including new ones when P is not projectable to $\mathcal{E}$.

Keywords: Kullback–Leibler divergence; relative entropy; exponential family; information projection; log-Laplace transform; cumulant generating function; directional derivatives; first order optimality conditions; convex functions; polytopes;

AMS Subject Classification: 94A17; 62B10; 60A10; 52A20;

References

[1] N. Ay: An information-geometric approach to a theory of pragmatic structuring. Ann. Probab. 30 (2002), 416–436.

[2] N. Ay: Locality of Global Stochastic Interaction in Directed Acyclic Networks. Neural Computation 14 (2002), 2959–2980.

[3] N. Ay and A. Knauf: Maximizing multi-information. Kybernetika 45 (2006), 517–538.

[4] N. Ay and T. Wennekers: Dynamical properties of strongly interacting Markov chains. Neural Networks 16 (2003), 1483–1497.

[5] O. Barndorff-Nielsen: Information and Exponential Families in Statistical Theory. Wiley, New York 1978.

[6] L. D. Brown: Fundamentals of Statistical Exponential Families. (Lecture Notes – Monograph Series 9.) Institute of Mathematical Statistics, Hayward, CA 1986.

[7] N. N. Chentsov: Statistical Decision Rules and Optimal Inference. Translations of Mathematical Monographs, American Mathematical Society, Providence, R.I. 1982. (Russian original: Nauka, Moscow 1972.)

[8] I. Csiszár and F. Matúš: Information projections revisited. IEEE Trans. Inform. Theory 49 (2003), 1474–1490.

[9] I. Csiszár and F. Matúš: Closures of exponential families. Ann. Probab. 33 (2005), 582–600.

[10] I. Csiszár and F. Matúš: Generalized maximum likelihood estimates for exponential families. To appear in Probab. Theory Related Fields (2008).

[11] S. Della Pietra, V. Della Pietra, and J. Lafferty: Inducing features of random fields. IEEE Trans. Pattern Anal. Mach. Intell. 19 (1997), 380–393.

[12] G. Letac: Lectures on Natural Exponential Families and their Variance Functions. (Monografias de Matemática 50.) Instituto de Matemática Pura e Aplicada, Rio de Janeiro 1992.

[13] F. Matúš: Maximization of information divergences from binary i.i.d. sequences. In: Proc. IPMU 2004, Perugia 2004, Vol. 2, pp. 1303–1306.

[14] F. Matúš and N. Ay: On maximization of the information divergence from an exponential family. In: Proc. WUPES'03 (J. Vejnarová, ed.), University of Economics, Prague 2003, pp. 199–204.

[15] R. T. Rockafellar: Convex Analysis. Princeton University Press, Priceton, N.J. 1970.

[16] T. Wennekers and N. Ay: Finite state automata resulting from temporal information maximization. Theory in Biosciences 122 (2003), 5–18.