

Minimum Description Length Principle and Distribution Complexity of Spherical Distributions

Bono Nonchev¹

¹*Faculty of Mathematics and Informatics, University of Sofia, Bulgaria*

Abstract

The application of the MDL principle to discern from which distribution a sample originates is discussed with the focus is on the general class of spherical distributions. Their trivial generalization the elliptical distributions are widely used in financial theory and have properties that enable us to calculate a closed form solution of their distribution complexity.

The MDL principle and its codelength/model interpretation is discussed first, as well as its application in model selection. Then the NML model is introduced as a suitable choice and its equivalent formulation as the model complexity is explored. After that the distribution complexity is presented as a solution of the problem of infinite model complexity, with the rest of the paper exposing the main result - the calculation of the distribution complexity for spherical distributions.

The analytical formulas for the distribution complexity are explicitly shown in three cases - the Gaussian distribution, the Student-T distribution and the Laplace distribution. Thoughts on their interpretation of the change of complexity with the size of the sample are presented with a somewhat surprising characterization of the NML model for the spherical distributions that has potential impact on robust estimation.

Keywords: MDL, Model Selection, Complexity, Distribution Selection, Spherical distributions, Student-T distribution, Laplace distribution

AMS subject classifications: 94A17, 62B10, 62F03

Acknowledgements: This work was supported by the European Social Fund through the Human Resource Development Operational Programme under contract BG051PO001-3.3.06-0052 (2012/2014).

Bibliography

- [1] Peter Grünwald, Jay Injae Myung, and Mark Pitt. *Advances in Minimum Description Length: Theory and Applications*. The MIT Press, April 2005.
- [2] Robert Kass and Adrian Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995.
- [3] Andrey Nikolaevich Kolmogorov. On Tables of Random Numbers. *Sankhya: The Indian Journal of Statistics, Series A*, 25(4):369–376, December 1963.
- [4] Guoqi Qian and Hans Künsch. On Model Selection in Robust Linear Regression. 1996.
- [5] Jorma Rissanen. MDL Denoising. *IEEE Transactions on Information Theory*, 46(7):2537–2543, 2000.

- [6] Jorma Rissanen. *Information and Complexity in Statistical Modeling (Information Science and Statistics)*. Springer, January 2007. ISBN 0387366105.
- [7] Yuri Shtarkov. Universal Sequential Coding of Single Messages, *Problems of Information Transmission*, (3):175–186, July .