

Full Title of Article

Author Name1

ABC@SAMPLE.COM

Address 1

Author Name2

XYZ@SAMPLE.COM

Address 2

Editor: Cheng Soon Ong and Tu Bao Ho

Abstract

This paper aims at bringing out the merits of using the Laplace distribution over the Normal distribution. The choice of the best distribution is objectively made using the minimum message length (MML) principle. The message length for transmitting data using a Laplace distribution is derived and its parameters are estimated. This method of transmission is compared with that of transmitting the data using the Normal distribution. This is explored in the context of superposition of protein structures. The optimal superposition of protein structures (described using a Normal distribution) minimizing the L2 norm is computed using the Kearsley's method and the superposition minimizing the L1 norm (described using a Laplace model) is approximated using Monte Carlo simulation. These two are compared with respect to their model complexity and the overall fit to the data using MML.

Keywords: Laplace, Normal, MML, Kearsley, Monte Carlo simulation

1. Introduction

Normal distribution is widely used in modelling a set of data whose true distribution is unknown. In many problems, the objective function is formulated as a sum of squares, (the L2 norm) and this function is minimized or maximized depending on the application. Normal distribution has a huge impact on the cost function because of the squared nature of the individual terms. If there are outliers in the dataset, the final inference might be skewed to accommodate the outliers in the model description. The Laplace distribution, however, is robust to outliers as the objective function involves the sum of the absolute values of the difference of the individual terms (L1 norm). The choice of selecting a Laplace over Normal is investigated in this paper. This selection is made by formulating the objective function using Minimum Message Length (MML). The distribution which results in the best compression of data is chosen to be the best model.

The general procedure to formulate the message length expression for transmitting data using some statistical model is outlined in [Wallace and Freeman \(1987\)](#). The MML method of estimating parameters for a number of distributions has been well established ([Wallace, 2005](#)). If the data is generated from a Normal distribution, its approximation using a Normal distribution would result in a better compression

Examples of protein structures are considered and they are superposed using Kearsley's transformation. This results in an orientation of the proteins which minimizes the total least squares of the corresponding coordinates. A superposition which minimizes the total absolute value of the difference of the coordinates is computed using a Monte Carlo simulation. The two orientations

Acknowledgments

Acknowledgements should go at the end, before appendices and references.

References

- C. S. Wallace. *Statistical and Inductive Inference using Minimum Message Length*. Information Science and Statistics. SpringerVerlag, 2005.
- C. S. Wallace and P. R. Freeman. Estimation and inference by compact coding. *Journal of the Royal Statistical Society. Series B (Methodological)*, 49(3):pp. 240–265, 1987.

Appendix A. First Appendix

This is the first appendix.

Appendix B. Second Appendix

This is the second appendix.