Research Highlight #149

Improving Distance Measurements Using Information Geometry

Keith A. Earle
Department of Physics, University at Albany-SUNY

Introduction and Methods: Improving error estimates is crucial for inferring accurate structural information via distance measurements. Most methods of estimating errors tacitly assume that parameter space is Euclidean. In practice, this is not the case. It is straightforward to show that parameter space, even for a simple Lorentzian line, has negative curvature[1]. The methods of Information Geometry facilitate precise, mathematically consistent error analysis. This is especially important for challenging signals that may consist of multiple, poorly resolved features at low signal to noise ratio (SNR). As an example of a model system on which these ideas could be tested, Earle and Hock[2] used Information Geometry allied with methods of Bayesian analysis, in particular the Nested Sampling Algorithm[3], to assess the effects of varying SNR and unmodelled signal, *e.g.*, a broad background, on parameter inference and error estimates for an inhomogeneously broadened Pake doublet spectrum.

Results: The effects of parameter space curvature were most significant for cases of low SNR, corresponding to signal detection at the sensitivity limit. Error estimates also change significantly in the presence of unmodelled signal, *e.g.*, a broad, inhomogeneous background. The Fisher information[4] was the most useful metric for summarizing correlations among fit parameters, particularly in the presence of unmodelled signals. Plots of the Fisher information landscape as the model parameters were varied allowed a clear visualization of the effects of unmodelled signals on the best-fit model parameters and parameter uncertainties.

Implications: When reporting parameter uncertainties, the most common procedure is to use the uncertainties returned by some non-linear, least-squares fitting routine without accounting for curvature corrections. For systems under study near the sensitivity limit, it is imperative to account for curvature corrections in order to achieve reliable results. In the case studied here, the inhomogeneous broadening parameterizes the width of the distance distribution. If the width of this distribution is under- or over-estimated, then incorrect inferences about the structure of conformers may be drawn with further implications for unraveling mechanistic details. As an example of a situation where accurate widths for distance distributions play a crucial role, inhibitors often change distance distributions of ‘flaps’, or other nearby protein structures. The work highlighted here emphasizes that for systems with low SNR it is particularly important to carefully assess the contribution to error estimates arising from the curvature of the parameter space. If the width of the distance distribution is inferred incorrectly, then promising drug candidates may be overlooked, or ineffective drugs may be chosen erroneously for clinical trial.

Discussion: The model system studied here had an analytical expression for the homogeneous line shape, and the inhomogeneity could be modelled well by Gaussian convolution. The concepts of Information Geometry and methods of Bayesian data analysis are generic, however, and can be applied whenever a computable model is available. Earle and coworkers have demonstrated proof of principle for the rotational diffusion problem, for example[1]. These methods will also be useful for rigorous analysis of NARS spectra in order to achieve maximally unbiased parameter and error estimates. Rigorous analysis of time-series, *e.g.*, pulse spectra, are also amenable to the Information Geometric, Bayesian analysis approach discussed here. The particular implementation of the nested sampling algorithm developed for this work is also being extended to the case of simultaneous, multi-frequency spectral analysis.

---

[1] Earle, K. A., L, Mainali, I. Dev Sahu, and D. J. Schneider. “Magnetic Resonance Spectra and Statistical Geometry.” Applied Magnetic Resonance, 37:(1-4)865-880 (2010).
[2] Hock, K., and K. A. Earle. "Information Theory Applied to Parameter Inference of Pake Doublet Spectra." Applied Magnetic Resonance, 45(9): 859-879 (2014).
[3] Sivia, D. S., and J. Skilling. “Data Analysis: A Bayesian Tutorial,” 2nd Edition, (Oxford Science Publications, 2006)
[4] Earle, K. A., and D, J. Schneider. “Parameter Estimation as a Problem in Statistical Thermodynamics.” AIP Conf Proc. 1305(1):357–364 (2011).