The geometry of model space

An underlying objective for many of us working with information geometry is understanding what geometric space is most appropriate for embedding statistical models in and, in particular, to think what the 'local neighbourhood' of a model is. One important motivation for this question include building a framework in which sensitivity analysis of modelling assumptions can be undertaken and finding the limits to knowledge of the statistical method. For example a important set of ideas on the geometry of model sensitivity which raises important questions on the way that inference is done in practice is found in [6], [5], [16], and [7].

While the information geometric properties of regular parametric families of distributions are well understood it has been surprisingly difficult to extend this concept to the 'space of all distributions', see for example [1], [21], [11], [20], [2], [4], [10], [13], [14] and [12]. Key assumptions in all these approaches are that all distributions have a common support and that a manifold structure is appropriate.

One very general proposal for finding this general space proposes that simplicial structures are, by their nature, more appropriate than manifold based ones. Specifically, they are considerably more tractable while automatically accommodating distributions with different supports. The hierarchical structure of a simplicial complex [15] is determined by the support and moment structure of the set of distributions. Furthermore, they arise naturally under suitable discretisation of the sample space. While this is clearly not the most general case (an obvious equivalence relation being thereby induced), it does provide an excellent foundation on which to construct a theory of computational information geometry. Indeed, in many practical applications, it can be argued (see, for example, [22]) that, since continuous data can only be measured to finite accuracy, this discretisation is sufficient for a complete analysis.

Another approach to looking at models which are close to, say exponential families and the related exponential dispersion models, is through mixing and in particular local mixing, see [17], [18] and [3]. The resulting highly interpretable and flexible models have enough structure to retain the attractive inferential properties of exponential families. In particular results on identification, parameter orthogonality and log-concavity of the likelihood can be proved. Of interest are the differences between mixing with small exponential dispersion families and more general mixtures. There are strong links with Amari's -1-fibre bundles and their associated geometry.

It will be of interest to explore a related theme on model selection through ideas of sparsity, [23], [19], [8] and [9]. A good deal of the foundation of the power of the idea of sparsity has an underlying geometric basis, where the geometry is typically based on l_1-norms and on ideas of convexity rather than the typically l_2 Riemannian geometry of information geometry. One aim of this session is to explore the relationships between the geometry of sparsity, convex geometry of mixtures and information geometry.

References

[1] S-I Amari. *Lecture notes in statistics-28: Differential-geometrical methods in statistics.* Springer-Verlag Inc, 1990.

[2] Shun-Ichi Amari and Hiroshi Nagaoka. *Methods of information geometry.* American Mathematical Society, 2000.

[3] K. Anaya-Izquierdo and P. Marriott. Local mixtures of exponential families. *Bernoulli*, 13(3):623–640, 2007.

[4] Alberto Cena. Geometric structures on the non-parametric statistical manifold. *PhD Thesis, University of Milan*, 2003.

[5] J. Copas and S. Eguchi. Local model uncertainty and incomplete data bias (with discussion). *J. Royal Statistical Society, B*, 67:459–512, 2005.

[6] J. Copas and S. Eguchi. Local sensitivity approximation for selectivity bias. *J. Royal Statistical Society, B*, 63:871–895, 2005.

[7] J. Copas and S. Eguchi. Likelihood for statistically equivalent models. *J. Royal Statistical Society, B*, 72:193–217, 2010.

[8] R. A. DeVore and V. N. Temlyakov. Nonlinear approximation in finite-dimensional spaces,. *J. of Complexity*, 13:489–508, 1997.

[9] D.L. Donoho and J. Tanner. Counting the faces of randomly-projected hyper-cubes and orthants, with applications,. *Discrete and Computational Geometry*, 43(3):522541, 2010.

[10] K. Fukumizu. Infinite dimensional exponential families by reproducing kernel hilbert spaces. *Proceedings of the 2nd International Symposium on Information Geometry and its Applications*, pages 324–333, 2005.

[11] Paolo Gibilisco and Giovanni Pistone. Connections on non-parametric statistical manifolds by orlicz space geometry. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 1(2):325–347, 1998.

[12] M. R. Grasselli. Dual connections in nonparametric classical information geometry. *to appear in the Annals of the Institute for Statistical Mathematics*, 2008.

[13] Henryk Gzyl and Lzaro Recht. A geometry on the space of probabilities ii. projective spaces and exponential families. *Revista Matematica Iberoamericana*, 22(3):833–849, 2006.

[14] Henryk Gzyl and Lzaro Recht. Intrinsic geometry on the class of probability densities and exponential families. *Publicacions Matematiques*, 51(2):309–332, 2007.

[15] A.T. Lundell and S. Weingram. *The Topology of CW Complexes*. Van Nostrand Reinhold: New York, 1969.

[16] J. Copas M. Henmi and S. Eguchi. Confidence intervals and p-values for meta analysis with publication bias. *Biometrics*, 63:475–482, 2007.

[17] Paul Marriott. On the local geometry of mixture models. *Biometrika*, 89(1):77–93, 2002.

[18] Paul Marriott. On the geometry of measurement error models. *Biometrika*, 90(3):567–576, 2003.

[19] A. Pinkus. Sparse representations and approximation theory. *Journal of Approximation Theory*, 163:388412, 2011.

[20] Giovanni Pistone and Maria Piera Rogantin. The exponential statistical manifold: Mean parameters, orthogonality and space transformations. *Bernoulli*, 5:721–760, 1999.

[21] Giovanni Pistone and Carlo Sempi. An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. *The Annals of Statistics*, 23:1543–1561, 1995.

[22] E.J.G. Pitman. *Some basic theory for statistical inference*. Chapman and Hall, London, 1979.

[23] S.J. Szarek. Convexity, complexity, and high dimensions. *Proceedings of the International Congress of Mathematicians*, 2006.