Geometries of Inference

Miķelis Bickis
Department of Mathematics and Statistics, University of Saskatchewan
bickis@snoopy.usask.ca

Abstract

Inferential processes, having to do with the closeness of models to data, lend themselves to geometric ideas. There are several geometries that are relevant to probability models, and one must avoid the temptation to attribute features of one geometry to another, but to keep in mind the images appropriate to the task at hand.

A probability measure can be represented as an expectation, i.e., a linear functional on random variables. The set of all probability measures thus inherits a linear structure, and can be viewed as a convex subset of the linear space (a simplex in the finite-dimensional case). Walley's lower prevision [4] can be represented as the infimum of a convex subset of this larger set. To define a geometry, a linear structure also needs a distance. The appealing Euclidean norm does not adequately describe the distance concepts that are appropriate to inferential problems.

Kullback-Leibler divergence [2], while lacking the properties of a norm (or even a metric), is an inferentially meaningful measure of distance between probability measures since it is the expectation of a log-likelihood ratio. It is appealing to quantify the imprecision of a lower prevision by the information diamger—i.e., the supremum of Kullback-Leibler divergences—of the set of probability measures. This diameter, however, would be infinite if the measures in the set have different null events.

Walley's imprecise Dirichlet model [5] and the imprecise exponential family models of Quaeghebeur and de Cooman [3] are based on a convex set of hyperparameters for prior distributions of the model parameters, which are then modified by Bayesian updating. Upper and lower previsions of future observations can then be described geometrically in terms of tangent planes to the hyperparameter set. This interpretation is complicated for other predictands, or for models outside the class discussed by Diaconis and Ylvisaker [1].

The various issues are illustrated graphically by reference to 2×2 contingency tables.

Keywords. Exponential family, information geometry,

References

[1] Diaconis, P. and Ylvisaker, D. 1979. Conjugate priors for exponential families. *Ann. Statist.* **7** , 269–281.

[2] Kullback, S., 1959. *Information Theory and Statistics.* John Wiley and Sons, Inc., New York;

[3] Quaeghebeur, E., and G. de Cooman, 2005. Imprecise probability models for inference in exponential families. *Proceedings of the Fourth International Symposium on Imprecise Probabilities and their Applications*, G. Coman, R. Nau, & T. Seidenfeld, eds., SIPTA.

[4] Walley, Peter 1991 *Statistical Reasoning with Imprecise Probabilities*, Chapman & Hall, London.

[5] Walley, Peter., 1996. Inferences from multinomial data: learning about a bag of marbles. *JRSS* **B 58**, 3–57.