Abstracts of the Ninth Workshop on Simulation (2018) 1-2

Design of Experiments, Bayesian Quadrature and Sensitivity Analysis

Luc Pronzato [1]

A standard objective in computer experiments is to predict/interpolate the behaviour of an unknown function f on a compact domain from a few evaluations inside the domain. When little is known about the function, space-filling design is advisable: typically, points of evaluation spread out across the available space are obtained by minimizing a geometrical (for instance, minimax-distance) or a discrepancy criterion measuring distance to uniformity. We shall focus our attention to sequential constructions where design points are added one at a time. This work is in collaboration with Anatoly Zhigljavsky (Cardiff University) and is motivated by recent results [4] indicating that the sequence of design points generated by a vertex-direction algorithm applied to the minimization of a convex functional of a design measure can have better space filling properties than points generated by the greedy minimization of the corresponding supermodular set function.

We shall make a survey of some recent results on energy functionals, that can be used to measure distance to uniformity, based on [1, 5, 6]. In particular, we shall investigate connections between design for integration of f with respect to a measure μ (quadrature design), construction of the (continuous) BLUE for the location model, and minimization of energy (kernel discrepancy) for signed measures. Integrally strictly positive definite kernels define strictly convex energy functionals, with an equivalence between the notions of potential and directional derivative for smooth kernels, showing the strong relation between discrepancy minimization and more traditional design of optimal experiments. In particular, kernel herding algorithms are special instances of vertex-direction methods used in optimal design, and can be applied to the construction of point sequences with suitable space-filling properties.

Finally, when using a Gaussian-process prior for f, a suitable Karhunen-Loève decomposition of the process yields a Bayesian linear model. The machinery of optimal design theory can then be used to construct optimal design measures that minimize the integrated mean-squared prediction error (a Bayesian A-optimal design problem [2]), or the variance of estimators of Sobol' indices for sensitivity analysis (which corresponds to Bayesian L-optimal design [3]), or the variance of an unbiased estimator of the integral of f with respect to μ (Bayesian c-optimal design).

[1] Laboratoire I3S, CNRS, Université Côte d'Azur, Sophia Antipolis, France, E-mail: Luc.Pronzato@cnrs.fr

References

[1] S.B. Damelin, F.J. Hickernell, D.L. Ragozin, and X. Zeng. On energy, discrepancy and group invariant measures on measurable subsets of Euclidean space. *J. Fourier Anal. Appl.*, 16:813–839, 2010.

[2] B. Gauthier and L. Pronzato. Convex relaxation for IMSE optimal design in random field models. *Computational Statistics and Data Analysis*, 113:375–394, 2017.

[3] L. Pronzato. Sensitivity analysis via Karhunen-Loève expansion of a random field model: estimation of Sobol' indices and experimental design. *Reliability Engineering and System Safety*, 2018. to appear, hal-01545604v2.

[4] L. Pronzato and A.A. Zhigljavsky. Measures minimizing regularized dispersion. *Submitted*, 2017.

[5] S. Sejdinovic, B. Sriperumbudur, A. Gretton, and K. Fukumizu. Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *The Annals of Statistics*, 41(5):2263–2291, 2013.

[6] B.K. Sriperumbudur, A. Gretton, K. Fukumizu, B. Schölkopf, and G.R.G. Lanckriet. Hilbert space embeddings and metrics on probability measures. *Journal of Machine Learning Research*, 11(Apr):1517–1561, 2010.