GSD2016

Support points - A new way to compact distributions

Simon Mak, ISyE, Georgia Institute of Technology

In this presentation, we present a new method for compacting any continuous distribution F into a set of optimal representative points called support points. Support points can have many useful applications in statistics, engineering and operations research. Two applications will be highlighted here: (1) as a way to optimally allocate experimental runs for stochastic simulations, and (2) as an improved alternative to MCMC thinning in Bayesian computation. For numerical integration, we will also show the improved performance of support points over Monte Carlo (MC) and quasi-Monte Carlo (qMC) methods using several simulation studies.

On a technical level, support points differ from existing qMC methods on two fronts. The latter often assumes integration on the unit hypercube to obtain a closed-form expression for discrepancy, and employs number-theoretic methods to generate sampling points (see, e.g., Dick-Kuo-Sloan, 2013). Support points, on the other hand, are obtained from the energy goodness-of-fit statistic (Székely and Rizzo, 2013), which can be efficiently optimized using subsampling methods, despite having no closed-form expression. As a result, these points can be generated for any distribution from which a random sample can be obtained. We will present a Koksma-Hlawka-like bound connecting integration error to the energy statistic, and establish an existence result for error convergence rate using support points. A quick blockwise coordinate-descent algorithm is then proposed for generating support points, with time-complexity scaling linearly in both sample space dimension and number of desired points.

This work is joint with V. Roshan Joseph (Georgia Institute of Technology).