From measures to machine learning: principled computation in big spaces

The common feature of the frontier of modern science is the study of complex phenomena with large datasets that require substantial computation. How can that best be done, in a principled and effective manner?

The foundation of quantification is the sum rule: if a system has parts that can be arbitrarily re-ordered and re-packaged, then quantification of the system must obey the ordinary sum rule of addition. The foundation of inference is the product rule: if systems are independent, the outcomes must be the same whether they are analysed separately or together. Those rules, embodying quantity and probability, are the unique calculus of science. Computation should respect that basic structure. At first, it seems that we need to add up all the possibilities if we are to apply the sum rule properly. But in large problems, we can only compute a limited number of possibilities. Successful computation uses those limited evaluations as statistical proxies for similar possibilities that might have been computed but were not.

Technically, we do not attempt direct Riemannian integration

\[ \text{Quantity} = \int \text{Property} \, d(\text{Possibility}). \]

Rather, we program the equivalent Lebesgue integral

\[ \text{Quantity} = \int (\text{number of possibilities}) \, d(\text{Property}). \]

This “nested-sampling algorithm” re-orders a multi-dimensional problem into a simple one-dimensional sum, ordered by Property value. Applied to probability, the algorithm generates both halves of Bayesian inference, evidence value and posterior distribution. Simplicity yields generality and power, and nested sampling is capable of computing multimodal and multi-phase applications.