

Sampling-based statistical analysis in large-scale physics experiments

Advisor	Balázs Kégl
Laboratory	Linear Accelerator Laboratory (LAL) and Computer Science Laboratory (LRI) University Paris-Sud/CNRS/INRIA
Research group	Machine Learning and Applied Statistics (AppStat/LAL) and Machine Learning and Optimization (TAO/LRI)
Address	University Paris-Sud, 91898 Orsay Cedex
Mail	balazs.kegl@gmail.com
Web	http://users.web.lal.in2p3.fr/kegl
Keywords	astroparticle physics, signal processing, statistics, simulation
Required background	statistics or signal processing or machine learning or physics

Summary

The context of this project is to investigate the application of sampling-based statistical analysis techniques to large-scale physics experiments. These techniques are based on Monte Carlo Markov chains (MCMC) where simulation is driven by both prior knowledge and observed data. Design choices can be made on higher levels in the model, making the approach appealing to experimental scientists whose goal is to test different models and parametrizations on the observed data. If used in a Bayesian setup, the output of these techniques is a posterior distribution over the model parameters that can be used in a fine-scale analysis.

The main testbed for the project is the [Pierre Auger experiment](#). The objective of this experiment is to study the properties of ultra-high energy cosmic ray particles by observing the generated atmospheric "particle showers", using a large observatory of 1600 water tanks and four fluorescence detectors built on the Pampas of Argentina. The principal objective of the project is to implement a recently developed adaptive MCMC algorithm [1, 2] and adapt it to the inference problem for the surface detector signal in the Auger experiment.

The ideal candidate has a strong background either in computer science, statistics, or physics, and an open mind to acquire the necessary knowledge in the other two disciplines. The project has statistical (e.g., model selection, convergence, optimal sampling in large hierarchical models), computational (e.g., parallelization, grid computation), and physical aspects, and the project can be modulated between these three angles based on the candidate's profile. The project can be followed by a Ph.D. thesis.

References

- [1] H. Haario, E. Saksman, and J. Tamminen, "An adaptive Metropolis algorithm," *Bernoulli*, vol. 7, pp. 223–242, 2001.
- [2] R. Bardenet, O. Cappé, G. Fort, and B. Kégl, "Adaptive Metropolis with online relabeling," in *submitted*, 2011.