Markov Chain Monte Carlo Simulations

Credit: 3 hours

Prerequisites: Calculus I and II. Either willingness to learn Fortran “on the fly” or good knowledge of another programing language like C or C++.

With the emergence of high performance computing Markov Chain Monte Carlo (MCMC) simulations have become a major enabling technique in applied sciences. The goal of this course is to provide a self-contained presentation. This includes not only Monte Carlo methods themselves, but also its statistical foundations and the statistics needed to analyze MCMC data, which are autocorrelated. The course is addressed to applied scientists, who need ready to work recipes and running computer code will be provided for all subjects covered. Illustrative examples may be drawn from statistical physics or from topics of other participating departments, for instance from simulations of biomolecules. Students are encouraged to bring in own examples, related to their planned or actual research work.

1. Sampling, statistics and computer code
   - probability distributions, sampling and random numbers
   - organization of the computer code
   - Gaussian distribution
   - confidence intervals
   - order statistics and heapsort
   - functions and expectation values of random variables
   - sums of independent random variables and characteristic functions
   - linear transformation of random variables
   - the central limit theorem and binning
2. Error analysis for independent random variables
   - Gaussian confidence intervals and error bars
   - estimators of the variance and bias
   - statistical error bar routines
   - the Gaussian difference test
   - the χ^2 and sample variance distribution
   - Gosset's student distribution and student difference test
   - the error of the error bar and the variance ratio test (F-test)
   - When are distributions consistent? (χ^2 and Kolmogorov tests)
   - The jackknife approach
   - determination of parameters (linear and Levenberg-Marquardt fitting)
3. Markov Chain Monte Carlo
   - preliminaries: The canonical ensemble and the two-dimensional Ising model
   - lattice labeling
   - sampling and re-weighting
   - importance sampling
   - the Metropolis algorithm
   - the heath bath algorithm
   - illustrations of Monte Carlo simulations I : discrete systems
   - illustrations of Monte Carlo simulations II: continuous systems
4. Error analysis for Markov chain data
   - autocorrelations
   - integrated autocorrelation time and binning
   - self-consistent versus reasonable error analysis

- analysis of MCMC data (examples)
- comparison of MCMC algorithms
- fitting of MCMC data

5. Advanced Topics (the instructor will choose a subset)
   - simulations of generalized ensembles
   - free energy and entropy calculations
   - event driven simulations
   - cluster algorithms
   - molecular dynamics
   - hybrid Monte Carlo
   - kinetic Monte Carlo
   - large scale simulations
   - parallel computing
   - **to be anounced (may be chosen by students)**

Text:

Markov Chain Monte Carlo Simulations and Their Statistical Analysis, Bernd A. Berg, World Scientific 2004.

Objectives:

The students should aquire a thorough understanding of the introduced methods. They should become able to implement the methods in their own computer programs and to present the thus obtained results in class.

Assessment:

The course grade will be based to 70% on assignments (mainly computational) and to 30% on a final project, which each student is expected to present in class. Regular class attendance is expected, some assignments will be given during class hours and the others as homework. The course grade dividing lines are: $A > 85\%,\ A^- > 80\%,\ B^+ > 75\%,\ B > 64\%,\ B^- > 60\%,\ C^+ > 56\%,\ C > 44\%,\ C^- > 40\%,\ D > 24\%$ and F for less or equal 24%.