Specified Topic 3

Learning goals: Metropolis Algorithm, Markov Chain Monte Carlo, Intro to Bayesian Methods

Applying the Metropolis Algorithm

Methods now classified as Markov Chain Monte Carlo have a long history in statistical mechanics. The first real implementation was the Metropolis Algorithm by Metropolis et al in 1953.

There are a multitude of problems for which this type of algorithm is now applied.

The basic idea is to sample from a probability distribution based on constructing a Markov chain whose equilibrium distribution is the targeted probability distribution. The method scales well with the dimension of the problem and is particularly suited to high multi-dimensional problems.

Metropolis Algorithm outline. Consider a problem where we want to sample from an equilibrium distribution, $\pi(\vec{x})$, where the sample space, $\vec{x}$ has n-dimensions.

- Generate a random starting vector $\vec{x}_0$.
- Consider a transition from this state to a candidate state $\vec{x}_1$ characterized by a **symmetric** proposal pdf. A standard choice would be a multi-dimensional Gaussian centered on $\vec{x}_0$.
- Calculate the acceptance ratio, $\alpha = \pi(\vec{x}_i)/\pi(\vec{x}_{i-1})$
- If $\alpha \geq 1$ accept $\vec{x}_i$. Else if $\alpha < 1$ accept with probability α.
- If $\vec{x}_i$ is rejected in the previous step, try a new candidate state starting again from state $\vec{x}_{i-1}$.
- If $\vec{x}_i$ is accepted, use this as the starting point for new candidates.

After a while (known as burn-in), the accepted $\vec{x}$ vectors, should be consistent with being sampled from the desired target distribution, $\pi(\vec{x})$.

This general technique is applicable to many different problems. Two of the main-stays are the simulated annealing method of finding the lowest energy state of a system - basically a minimization problem, and the application to Bayesian inference.

In Bayesian inference one usually wants to be able to sample the posterior density which is the pdf of the model given the data, expressed in terms of the product of the likelihood (probability of the data given the model) and the prior. If nothing is known about the values of the model parameters, then frequently (pun) one uses a non-informative prior like a uniform distribution which then makes the problem one of simply sampling from the likelihood distribution. One does not need to worry about normalizing the posterior density, as the deliverable is simply a sequence of random $\vec{x}$ vectors drawn from the distribution of model parameters in proportion to the likelihood.

$$p(\text{model}|\text{data}) = p(\text{data}|\text{model})p(\text{model}) / \int p(\text{data}|\text{model})p(\text{model})d\text{model}$$

So

$$p(\text{model}|\text{data}) \sim p(\text{data}|\text{model})$$

for our case of a flat prior.

One can then for example turn around fitting problems of data into Bayesian type problems of finding the posterior density from being able to sample the likelihood function. As an example, find the posterior densities of the constant and quadratic terms of the parabolic fit to the data of the Fits examples, by formulating the likelihood function, (in this case very closely related to the chi-squared function). You should also plot parameter 1 vs parameter 2 and measure the correlation coefficient. As is the case with likelihood fitting, it is numerically sensible to calculate $\log p$. See Press 15.8 for more implementation details.

There are a number of applications where this kind of method can be applied. Well known ones using the simulated annealing approach are the Thomson problem (that of finding the minimum energy configuration of N electrons on a unit sphere) and the application to the traveling salesman problem.

So, this project consists of

- Applying the Metropolis Algorithm to the HW3 data analysis or similar data as described above.
- Applying Metropolis or a related algorithm to a problem of your choice. Some possibilities are the Thomson problem, the traveling salesman problem and the Ising model.