MCMC Notes

Shaobo

March 5, 2015

1 Markov Chain

A Markov chain is a special type of stochastic process. The standard definition of a stochastic process is an ordered collection of random variables:

where t is frequently (but not necessarily) a time index. If we think of X_t as a state X at time t, and invoke the following dependence condition on each state:

then the stochastic process is known as a Markov chain. This conditioning specifies that the future depends on the current state, but not past states. Thus, the Markov chain wanders about the state space,remembering only where it has just been in the last time step. The collection of transition probabilities is sometimes called a transition matrix when dealing with discrete states, or more generally, a transition kernel.

In the context of Markov chain Monte Carlo, it is useful to think of the Markovian property as mild non-independence MCMC allows us to indirectly generate independent samples from a particular posterior distribution.

2 Gibbs Sampling

1. If a posterior has k parameters to be estimated, we may condition each parameter on current values of the other $k-1$ parameters, and sample from the resultant distributional form (usually easier), and repeat this operation on the other parameters in turn. This procedure generates samples from the posterior distribution. Note that we have now combined Markov chains (conditional independence) and Monte Carlo techniques (estimation by simulation) to yield Markov Chain Monte Carlo.

3 PYMC

PyMC is a python module that implements Bayesian statistical models and fitting algorithms, including Markov chain Monte Carlo. Its flexibility and extensibility make it applicable to a large suite of problems. Along with core sampling functionality, PyMC includes methods for summarizing output, plotting, goodness-of-fit and convergence diagnostics.

PyMC provides functionalities to make Bayesian analysis as painless as possible. Here is a short list of some of its features:

- Fits Bayesian statistical models with Markov chain Monte Carlo and other algorithms.
- Includes a large suite of well-documented statistical distributions.
- Uses NumPy for numerics wherever possible.
- Includes a module for modeling Gaussian processes.
- Sampling loops can be paused and tuned manually, or saved and restarted later.
- Creates summaries including tables and plots.
- Traces can be saved to the disk as plain text, Python pickles, SQLite or MySQL database, or hdf5 archives.
- Several convergence diagnostics are available.
- Extensible: easily incorporates custom step methods and unusual probability distributions.
- MCMC loops can be embedded in larger programs, and results can be analyzed with the full power of Python.