Advanced Sampling Algortihms

April 3, 2014

In this presentation we discussed about how we can do the exact sampling from a probability distribution π over a set of states S. A simple way is to construct an ergodic Markov Chain with stationary distribution π. We are using a 2D Ising model to sample from the distribution. A 2D Ising model has set of sites which interact with their immediate neighbours. Each site can have a spin of $+1$ or -1. Each combination of spins constitutes a state $s_i \in S$. If the size of grid is L then $|S| = 2^{L^2}$. We can run a Markov chain by using Monte Carlo method over the 2D Ising Model. We start with some initial state. We run the chain for long enough time. Then the markov chain returns a state $s \in S$ which is distributed by **(approximately)** π .

But the issue here is how long we need to wait. Also we don't want approximate samples. Propp Wilson Algorithm is a method to overcome these two issues. Propp Wilson detects the convergence of Markov chain to stationary distribution and the output states $s \in S$ are distributed according to π. The main idea of Propp Wilson algorithm is coupling from the past. Let us say we a sequence of increasing numbers $(N_1, N_2, \ldots, N_m, \ldots, N_M)$ We run the chains from all $|S|$ states as initial states from time $t = -N_m$. At time $t = 0$ we check if all chains are giving the same state s as output. If the output is the same then it means that the chains have converged to a stationary distribution π and we can sample exactly from the distribution π. If there was no convergence that the chains are restarted from $-N_{m+1}$. This process is repeated until all the chains converge at time $t = 0$.

The idea of Propp Wilson was good but we had the drawback of running chains from all possible states in S. The size of S can be very high as in case of 2D Ising Model. Let us consider that the state space S follows an ordering and the update function of Markov Chain retains the ordering such that all chains remain in between the chains starting from the lowest state and the highest state. In this case we only need to run the boundary chains for the Propp Wilson algorithm. This idea is known as Sandwiching. We use sandwiching to reduce the computational complexity of Propp Wilson. Instead of all states in S we run the chains from highest state s^{max} which has spin at all sites as '$+1$' and lowest state s^{min} which has all states with spin '-1'. When these two chain converge on or before $t = 0$ we can use the converged chain to sample from the distribution π.