

Gibbs sampling and the Boltzmann machine

C.W. Seitz

January 3, 2022

1 Monte Carlo Methods

The Monte Carlo method was invented by scientists working on the atomic bomb in the 1940s, who named it for the city in Monaco famed for its casinos and games of chance. Its core idea is to use random samples of parameters or inputs to explore the behavior of a complex system or process. The Monte Carlo methods are basically a class of computational algorithms that rely on repeated random sampling to obtain certain numerical results, and can be used to solve problems that have a probabilistic interpretation

2 Markov Chains

A sequence of random variables (X_0, X_1, \dots, X_n) follows a Markov chain if the probability of arriving at a state X_n is dependent only on the previous state X_{n-1} . That is,

$$X_n \sim P(X_n | X_0, \dots, X_{n-1}) = P(X_n | X_{n-1}) \quad (1)$$

We say that the chain is time-homogeneous if the transition-density is independent of time. The probability of arriving at any particular state X_n is then given by integrating over all possible paths to that state

$$P(X_n) = \int T(X_n | X_{n-1}) P(X_{n-1}) dx \quad (2)$$

which is known as the Chapman-Kolmogorov equation.

2.1 Ergodicity

Ergodicity expresses the idea that a stochastic process will eventually visit every state in the phase space χ that the system moves in. This implies that the statistics of the system can be deduced from a sufficiently long trajectory through χ . Equivalently, a sufficiently large collection of random samples from a process can represent the average statistical properties of the entire process. This simultaneously implies that the statistics of the system's behavior is independent of the initial conditions, since we have not specified *which* trajectory is

under consideration. As a side note, the concept of ergodicity was crucial for thermodynamics as it indicates that all possible microstates are equiprobable over a long period of time. Ergodic Markov chains are useful algorithmic tools in that, regardless of their initial state, they eventually reach a unique stationary distribution $\pi(x)$

2.2 Detailed balance

Detailed balance is a useful condition because it allows us to prove the existence of a stationary state. The condition states that the probability flux $X_i \rightarrow X_j$ exactly balances the flux $X_j \rightarrow X_i$ so that the net flux is zero. If π is such a stationary distribution, then the detailed balance condition reads:

$$\pi_i T(X_i|X_j) = \pi_j T(X_j|X_i) \quad (3)$$

3 Gibbs Sampling

An arbitrary joint probability distribution can be expressed as $p(x, y) = p(x|y)p(y) = p(x|y) \int p(x, y) dx$ where $p(y)$ is called the marginal distribution of y . Suppose that we would like to sample from the distribution $p(x, y)$, but for some reason, we cannot.

4 The Hopfield Network and Boltzmann Machine