In this paper, we have shown that stochastic gradient descent approaches extend in an elegant way to a sampler for posterior distributions when computing the likelihood from the full data is intensive. We are able to establish the stationary distribution of the procedure and gave several examples showing the accuracy and uses of our result. By taking a Bayesian route, we are able to show that the SGD algorithm is a sampler of a posterior under a non-informative prior. This enables us to establish connection between SGD and maximize \emph{*a posteriori}* (MAP) procedure, which is an analogy to gradient descent solving a maximum likelihood estimator in the frequentist paradigm.

We

In addition, our results give a relatively clean and simple way to derive finite sample rates of convergence for statistical estimators with dependent data without requiring the full machinery of empirical process theory (e.g. [Yu94]).

A natural extension of this work, which we hope to be able to accomplish, is to relax the assumptions on the uniformity of the mixing times in Assumption B, which would allow a wider range of applications of our results.

Extensive simulation of the methodology is also needed to fully understand the novel method.