Basis Expansion Monte Carlo

Eric Kernfeld *

University of Washington, Seattle, WA, USA

November 8, 2014

Abstract

Most Monte Carlo inference methods are left to run until the markov chain reaches its steady-state. This leaves the user with a large chain of samples from the distribution of interest. We introduce Basis Expansion Monte Carlo, which runs the sampler piecewise starting at different places in the parameter space in order extract information more quickly. To make inference about the steady state, we gradually update an approximation of the hidden linear operator that underlies any Metropolis-Hastings or Gibbs sampler. We use the steady-state of the approximate operator in place of the true steady-state. Results show ...

1 Introduction

In many statistical models, it is impossible to find a closed form for the distribution of interest (we will call this π). One work-around, originating in computational physics, relies on the fact that for points x_1 and x_2 in the parameter space, $\pi(x_1)/\pi(x_2)$ may still be calculable, though $\pi(x_1)$ and $\pi(x_2)$ are not.

More and references about history, background, and/or tutorials on monte carlo methods

In this setting, we will consider two classes of algorithms. The first, the Metropolis-Hastings scheme, consists of the following procedure.

Algorithm 1: Metropolis-Hastings algorithm

Set $x_0 = 0, i = 0$
Repeat ad nauseum:
  Increment i
  Draw x from a proposal distribution $q(x|x_{i-1})$
  Set $\alpha = min(1, \frac{\pi(x)q(x_{i-1}|x)}{\pi(x_{i-1})q(x|x_{i-1})})$
  Draw u from a uniform density on $[0, 1]$. Set $x_i = x$ with probability α, i.e. if $u < \alpha$, and $x_i = x_{i-1}$ otherwise.

Suppose this MCMC algorithm produces a chain $x_1, x_2, x_3, ...$ of samples. Because the algorithm is stochastic, these samples can be viewed as realizations of random variables $X_1, X_2, X_3, ...$ with marginal density functions f_1, f_2, f_3, etc. We can write the conditional density of X_2 given X_1 as $f_{2|1}(x_2, x_1) = (1 - \alpha(x_1|x_2))\delta_{x_1}(x_2) + \alpha(x_1|x_2)q(x_2|x_1)$. Then, we have that $f_2(x_2) = \int f_{2|1}(x_2, x_1)f_1(x_1)dx_1$; in more generality, $f_i(x_i) = \int f_{i|i-1}(x_i, x_{i-1})f_{i-1}(x_{i-1})dx_{i-1}$. Noting that $f_{i|i-1}$ doesn't depend on i, we can replace it with a function K so that $f_i(x_i) = \int K(x_i, x_{i-1})f_{i-1}(x_{i-1})dx_{i-1}$

In most cases, MCMC schemes have followed the orthodoxy that the chain must continue until equilibrium. There ought to be more efficient strategies. To be briefly technical, if an MCMC algorithm produces a chain $x_1, x_2, x_3, ...$ of samples, these samples can be viewed as realizations of random variables $X_1, X_2, X_3, ...$ with probability density functions f_1, f_2, f_3, etc. By the Markov property, the first "M" in "MCMC", each

*Electronic address: ekernf01@u.washington.edu; Corresponding author

density determines the next. So, there is some unknown function L with $f_i = Lf_{i-1}$; for the equilibrium f_{eq} that we seek, $Lf_{eq} = f_{eq}$. I intend to approximate L, then compute f_{eq} from that. My current plan is to pick a length-B list of functions $\{q_i\}_{i=1}^{B}$ as a basis for the space of distributions. Then, for each new MCMC sampler, I'll estimate a function $\alpha()$ and a matrix M so that $L \approx L_\alpha + L_M$, where $(L_\alpha f)(x) = \alpha(x)f(x)$ and $(L_M f)(x) = \sum_{i,j=1}^{B} q_j(x)M_{ij} \int q_i(x)f(x)dx$. The first term mimics the rejection probability–the chance that the crawler will stay put–while the next term tracks movement. It is easy to tell when the sampler rejects and when it doesn't, so we can bump α up or down accordingly. Meanwhile, when the crawler moves, we know it's time to update L_M. Terms cancel so that M_{ij} is an inner product between $L_M q_i$ and q_j. So, it can be written as the expectation $E_{L_M q_i}[q_j(x)]$. We can sample x from basis function q_i, run the sampler to turn it into a sample from Lq_i, then evaluate q_j. An average of many such samples will converge to the expectation, rendering an estimate for M_{ij}. This approximation can also be adapted to Gibbs sampling, a ubiquitous MCMC variant. Because of its mathematical form, I call the scheme Basis Expansion Monte Carlo (BEMC).