

We have discussed in class how a structured exploration of the target probability distribution can present significant computation advantages over the random walk approach. HMC presents a method for introducing a momentum variable that assists in exploring the unnormalized distribution more systematically. For numerical implementation of Hamilton's differential equations we use the leapfrog method. It is called the leapfrog as the first (numerical approximation) step for the momentum variable is a half step. The position step then uses this value from the momentum to perform a full step. The two variables (offset by the initial $\frac{1}{2}$) then essentially 'jump' passed each other for each successive update (in essence leapfrogging passed one another).

(5.6) and (5.7) show the advantage of using HMC over a random walk metropolis algorithm for approximating a high dimension distribution with varied standard deviations. (5.6) shows the **highly** correlated samples coming from the Metropolis algorithm where HMC appears to have uncorrelated and randomly sampled chains. Moreover, in (5.7) we see that Metropolis is unable to get **effective** samples on some of the parameters (especially those with the larger standard deviations). This is seen by the large sample standard deviation. The sample standard deviation does not seem to dramatically increase for HMC suggesting there are more effective samples.

We are unable to use HMC to directly approximate the LDA posterior. HMC uses local gradient information to perform leapfrog steps. In LDA, the word-topic and document-topic distributions are multinomials (i.e. they are discrete and the local gradients will have discontinuities). We can marginalize over these variables to adjust the problem for inference using HMC. Alternatively, Gibbs sampling might present a more appropriate approach in this case.