

CS880: Approximations Algorithms

Scribe: Dave Andrzejewski

Lecturer: Shuchi Chawla

Topic: Metropolis method, volume estimation

Date: 4/26/07

The previous lecture discussed some of the key concepts of Markov Chain Monte Carlo (MCMC) methods, including the stationary distribution π^* and the mixing time τ_ϵ . This lecture introduces the Metropolis method for constructing Markov chains in order to sample from some distribution. The use of sampling methods for volume estimation is also introduced.

26.1 Metropolis method

26.1.1 MCMC review

Recall from last time the key properties of a random walk Markov chain.

- Ω = the state space
- $n = |\Omega|$
- P = the transition matrix, $P_{ij} = Pr[i \rightarrow j]$
- π^* = the stationary distribution such that $\pi^*P = \pi^*$
- τ_ϵ = the mixing time, after which the ℓ_1 -norm of the difference between the chain distribution and the stationary distribution is guaranteed to be $< \epsilon$.

Also recall this important theorem concerning the existence and uniqueness of the stationary distribution π^* .

Theorem 26.1.1 *An aperiodic irreducible finite Markov chain is ergodic and has a unique stationary distribution.*

We can easily guarantee the aperiodicity of our chain by simply adding self-loops to all vertices. This will increase the mixing time by no more than a factor of 2.

26.1.2 Metropolis filter

But how do we actually construct a Markov chain with a stationary distribution equal to our target distribution? Also, we want this method to have a good (that is, small) mixing time. The Metropolis method allows us achieve these goals by defining our Markov chain as a random walk over a suitably defined graph.

We define the approach as follows. Say we wish to sample values $i \in \Omega$ from a distribution $Q(i)$. Then we define an undirected d -regular graph G on Ω , picking this graph in such a way that it has high conductance. Then from node v , pick the next node u uniformly from the d neighbors. Then:

- If $Q(u) \geq Q(v)$, move to node u
- Else move to node u with probability $\frac{Q(u)}{Q(v)}$, stay with probability $(1 - \frac{Q(u)}{Q(v)})$.

First we examine the graph itself. Since it is fully connected and undirected, it is irreducible. Since all nodes have self-edges, it is aperiodic. Therefore this random walk is guaranteed to have a unique stationary distribution π^* . Now we must show that this stationary distribution is equal to our target distribution Q .

Claim 26.1.2 $\pi^* = Q$

Proof: Say that our initial $\pi = Q$, then take one step. Consider any node v , and calculate the probability of arriving at node v after this one step. If it is equal to $Q(v)$, then we have shown that $QP = Q$, and therefore $\pi^* = Q$.

We need to calculate the probability of starting at distribution Q , taking one step, and then ending up in state v . This can be decomposed into three cases: we move from a neighbor u into v where $Q(u) \geq Q(v)$, we move from a neighbor u into v where $Q(u) < Q(v)$, or we are already in v and we choose a neighbor u such that $Q(u) < Q(v)$ but we end up staying at v . Let n be the number of neighbors u such that $Q(u) \geq Q(v)$.

$$Q'(v) = \sum_{\substack{u|(u,v) \in G, \\ Q(u) \geq Q(v)}} \frac{1}{d} Q(u) \frac{Q(v)}{Q(u)} + \sum_{\substack{u|(u,v) \in G, \\ Q(u) < Q(v)}} \frac{1}{d} Q(u) + \sum_{\substack{u|(u,v) \in G, \\ Q(u) < Q(v)}} \frac{1}{d} Q(v) \left(1 - \frac{Q(u)}{Q(v)}\right) \quad (26.1.1)$$

$$= \frac{n}{d} Q(v) + \frac{d-n}{d} Q(u) + \frac{d-n}{d} Q(v) - \frac{d-n}{d} Q(u) \quad (26.1.2)$$

$$= Q(v) \quad (26.1.3)$$

■

This shows that a random walk using the Metropolis method is guaranteed to converge to our target distribution Q . It is worth noting that our scheme of uniformly choosing a neighbor is a special case of the general Metropolis-Hastings sampler [3]. In the more general case, a proposal distribution is used to select the next candidate state conditioned on the current state. This proposal distribution need not be uniform over neighbors, and in fact need not even be symmetric.

26.1.3 Volume estimation

An interesting application of sampling techniques is the problem of estimating the volume of a convex shape $K \in \mathbb{R}^n$ using an inclusion oracle which reveals whether a given point is contained in the shape or not. We are also given two balls, one completely enclosing K and one completely enclosed by K . Call these $K \subseteq B(0, R)$ and $K \supseteq B(0, r)$. This technique that we use has interesting parallels to the concept of self-reducibility.

What is the probability that a uniformly chosen point in the larger ball will be in K ? We can use the smaller ball to bound this probability as $\geq \frac{\text{vol}(B(0,r))}{\text{vol}(B(0,R))}$. However, for large n we will suffer the 'curse of dimensionality' [4], and this lower bound will be very small, in particular $(\frac{r}{R})^n$.

The key insight of our approach is that we can use a sample from a *series* of regions K_i in order to make the probability of finding a point in the region as large as possible.

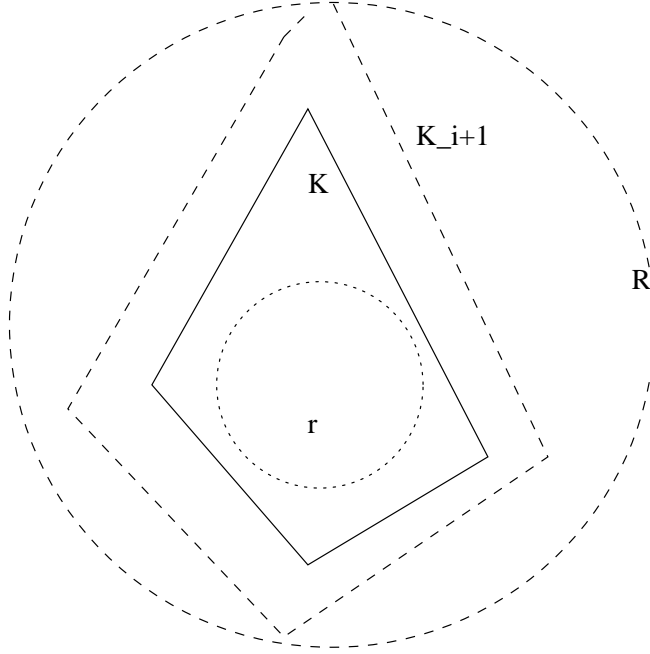


Figure 26.1.1: An example of our volume estimation problem setup.

$$K = K_0 \subseteq K_1 \subseteq \dots \subseteq K_\ell = B(0, R) \quad (26.1.4)$$

These regions are constructed such that $\frac{\text{vol}(K_{i+1})}{\text{vol}(K_i)}$ is small for all i . Furthermore we choose a small $\ell \approx n \log(\frac{R}{r})$, so we do not need to iterate too many times. We can estimate the volume of K using estimates of the ratios of adjacent K_i in the following way.

$$\text{vol}(K) = \frac{\text{vol}(K_0)}{\text{vol}(K_1)} \frac{\text{vol}(K_1)}{\text{vol}(K_2)} \dots \frac{\text{vol}(K_{\ell-1})}{\text{vol}(K_\ell)} \text{vol}(B(0, R)) \quad (26.1.5)$$

So now we simply need to figure out how to sample from K_i . We define K_i as a re-scaling of K , subject to containment by $B(0, R)$.

$$K_i = (1 + 1/n)^i K \cap B(0, R) \quad (26.1.6)$$

This gives us a nice bound on the volume ratio of adjacent K_i .

$$\frac{\text{vol}(K_{i+1})}{\text{vol}(K_i)} \leq (1 + 1/n)^n = e \quad (26.1.7)$$

Note that $(1 + \frac{1}{n})^{n \log R/r} K$ contains $B(0, r)$, therefore e is indeed $O(n \log R/r)$.

To sample from K_i , we then simply sample uniformly from K and then re-scale. But how to sample from K itself? To approach this problem, we employ the MCMC methods we have been discussing.

We define our random walk, known as the Ball-walk, as follows. From any point $u \in K$, sample a point randomly from the ball centered at u with radius δ , $B(u, \delta)$, and move to the new point if it is inside K . If the point is outside K , stay at u .

Note that the graph defined by this rule allows us to reach any point from any other point, and also allows self-loops. Therefore it is irreducible and aperiodic, and must have a unique stationary distribution π^* . The resulting Markov chain is time-reversible. Therefore, the stationary distribution is uniform.

For the practicality of this scheme, it is important to choose a good value for δ in order to get good samples from K . Taken to the extreme, a huge δ value would result in constantly picking points outside K , and therefore remaining at the current point. Likewise, a very small δ would result in taking very small steps, making it very slow to explore all of K . Also, if something is known about the geometry of K , it may be helpful to rescale the proposal ball to an ellipse, for example. This is accomplished by putting the body in an “isotropic” position via an affine transformation, so as to remove all sharp corners.

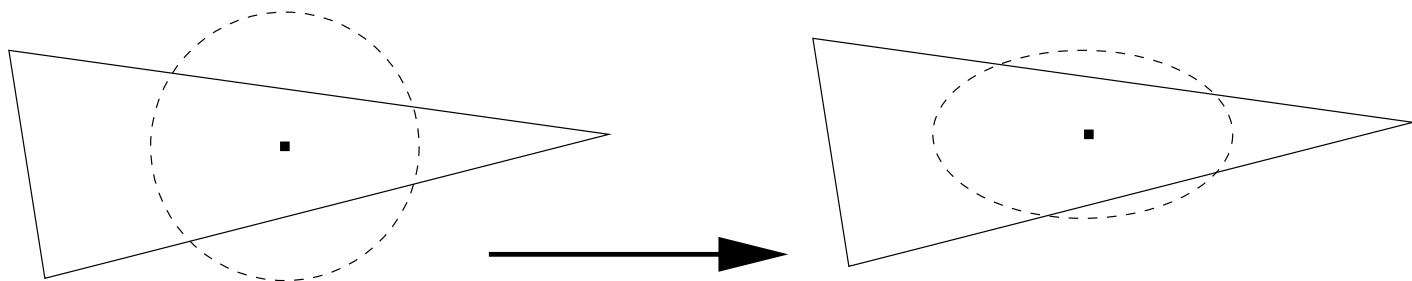


Figure 26.1.2: Rescaling the proposal ball to an ellipse based on the geometry of K .

The first approach based on this technique was polynomial in n , but with an unfortunate order $O(n^{23})$ [1]. Newer approaches, dubbed ‘hit and run’, first choose a direction, and then sample uniformly from the line segment along that direction contained in K . This approach drastically improves mixing time, achieving $\tilde{O}(n^4)$ [2].

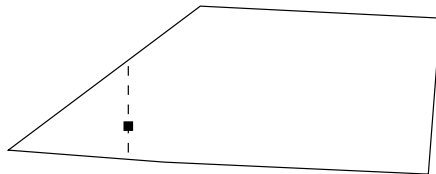


Figure 26.1.3: The ‘hit and run’ technique.

The inapproximability result is that one cannot estimate volume within a constant factor in $\Omega(n^2)$ time.

References

- [1] Martin Dyer, Alan Frieze, Ravi Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. JACM 1991.
- [2] Laszlo Lovasz, Santosh Vempala. Simulated Annealing in Convex Bodies and an $O(n^4)$ Volume Algorithm FOCS 2003.
- [3] D. MacKay. Information Theory, Inference, and Learning Algorithms. Cambridge University Press, 2003.
- [4] Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer-Verlag, 2001.
- [5] V. Vazirani. Approximation Algorithms. Springer, 2001.