Markov Chain Monte Carlo

Marcus Gallagher

ENGG7302
Advanced Computational Techniques in Engineering
(Main reference: (Andrieu et. al 2003))

Outline

Sampling

- In many applications (where we need to simulate a probabilistic model) we are interested in sampling from a stochastic system.
- The fundamental problem here can be seen as:
  - Drawing samples from some probability distribution.
  - Evaluating the expectation of some function with respect to some probability distribution.

Sampling

- General idea: obtain a set of samples drawn independently from $p(x)$ and approximate the expectation:

$$E[f] = \int f(x)p(x)dx$$

$$\hat{f} = \frac{1}{K}\sum_{k=1}^{L} f(x_k)$$

Basic Sampling Techniques

- Pseudo-random number generators
  - Using a source of uniformly distributed random numbers, transform to non-uniformly distributed numbers (of known functional form - e.g. Gaussian).
- Methods for sampling from more complex distributions (e.g. with unknown functional form):
  - Rejection sampling
  - Importance sampling, etc.
- Unfortunately these techniques have some important severe limitations particularly in high-dimensional spaces.

Markov chain Monte Carlo (MCMC)

- MCMC is a framework that allows sampling from a large class of distributions which scales well with dimensionality.
- Works by simulating a Markov chain, which is set up to converge the desired distribution.
- To build our sampler, we require that the chain be ergodic (irreducible and aperiodic) and have the target distribution as the equilibrium (invariant) distribution.

- Recall that the equilibrium distribution is the normalized left eigenvector of the transition probability matrix P (with corresponding eigenvalue 1).
- It turns out that all other eigenvalues will have absolute value less than one.
- Furthermore, the second largest eigenvalue determines the rate of convergence of the chain (smaller is better).

- So far, we have only discussed discrete Markov chains.
- It turns out to be a bit easier to appreciate MCMC if we consider Markov chains with a continuous state space (though it all still works in discrete space!).
- This also shows how generally applicable the techniques are.
- (Point of detail - don't worry too much about this!): In this case, the transition matrix becomes an integral kernel K and π_X becomes the corresponding eigenfunction

$$\int \pi_{x_i} K(x_{i+1}|x_i) dx_i = \pi_{x_{i+1}}$$

Metropolis-Hastings Algorithm

- The Metropolis-Hastings algorithm is the most popular MCMC method and has been considered among the top 10 algorithms that have had the greatest influence on the development and practice of science and engineering in the 20th century.
- The idea is to simulate a Markov chain by (at each time step) sampling from a specified proposal distribution $q(x^*|x)$ (that we can sample from easily!)
- This proposal distribution $q(x^*|x)$ will depend on the current point x, so the sequence of samples forms a Markov chain.
- To produce samples from some probability distribution of interest, we want that distribution to be the equilibrium distribution of the chain, π.

- The candidate value x^* from the proposal distribution $q()$ is then accepted with probability:

$$A(x, x^*) = \min\left(1, \frac{\pi_{x^*} q(x|x^*))}{\pi_x q(x^*|x))}\right)$$

- Otherwise the sampler rejects the move to x^* and remains at x.

Metropolis-Hastings Algorithm

1. Initialize $x(0)$.
2. For $i = 0$ to $N - 1$
   - Generate a random number with uniform distribution over $(0, 1)$:
     $$u \sim \mathcal{U}_{[0,1]}$$
   - Sample $x^* \sim q(x^*|x(i))$
   - If $u < A(x(i), x^*) = \min\left(1, \frac{\pi_{x^*} q(x|x^*))}{\pi_x q(x^*|x))}\right)$
     $$x(i + 1) = x^*$$
   - Else
     $$x(i + 1) = x(i)$$

Note that if the step from x to x^* causes an increase in the value of π_x, then the candidate point is certain to be kept.

- It can be shown that samples generated by the MH algorithm will mimic samples from the target distribution asymptotically.
  - Because of the possibility of rejection, the chain is aperiodic.
  - As long as the support of $q()$ includes the support of π, the chain will be irreducible and ergodic
  - Detailed balance is also satisfied.
- Note however that the sequence of samples obtained will not be independent
  - Independent samples can be obtained by (say) retaining every Mth sample and discarding the rest.