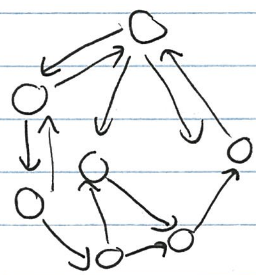
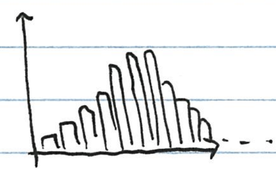
Markov Chain Monte Carlo

MCMC methods - a class of algorithms for sampling from a probability distribution

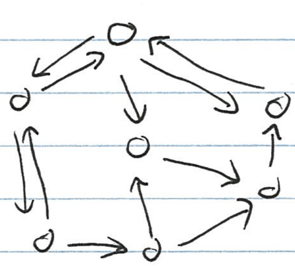
- by constructing a Markov Chain that has the desired distribution as it's equilibrium distribution

 ⇒ 

- MCMC methods are primarily used for calculating numerical approximation of multi-dimensional integrals

General Explanation - MCMC creates a sample from a continuous random variable with probability density proportional to a known function

- thes samples can be used to evaluate an integral over that variable & expected value or variance



- these chains are stochastic processes of "walkers"

Metropolis-Hastings algorithm: - generates Markov Chains using a proposal density for new steps & method for rejecting proposed moves

- obtains a sequence of random samples from which direct sampling is difficult

- used to sample from multi-dimensional distribution

Intuition - samples will be drawn from probability distribution $P(x)$, provided we know a function $f(x)$ proportional to the density $P(x)$

- $f(x)$ is key since calculating normalizing constant is difficult

- at each iteration, the algorithm picks ~~current~~ next value based on current value (Markov Chain)

- next value is either accepted or rejected, the probability of acceptance is determined by comparing the values of the function $f(x)$ of the current & candidate sample

Metropolis Algorithm (Symmetric Proposal Dist)

$f(x) \propto P(x)$ - target distribution

1. choose x_t & density $g(x|y)$ that suggests candidate for next x_{t+1} given x_t, $g(x|y) = g(y|x)$ - symmetric
   $g(x|y)$ - proposal density

2. for each iteration t
   - generate candidate x' by picking $g(x'|x^t)$
   - calculate acceptance ratio
     $\alpha = f(x')/f(x_t) \approx P(x')/P(x)$
   - Accept or Reject
     - generate uniform random $u \in [0,1]$
     - if $u \leq \alpha$ then accept the candidate by setting $x_{t+1} = x'$
     - if $u > \alpha$ then reject $x_{t+1} = x_t$

Gibbs Sampling

Gibbs Sampling - a MCMC algorithm for obtaining a sequence of observations which are from specified multivariate probability distribution, when direct sampling is difficult

- commonly used as statistical learning especially Bayesian inference

- generate Markov chain of samples

FIVE STAR

Implentation

Given a multivariate distribution, it is simpler to sample from a conditional dist

$\theta_1, \theta_2 \sim p(\theta_1, \theta_2)$

but... you can sample

1. $p(\theta_1 | \theta_2)$
2. $p(\theta_2 | \theta_1)$

Gibbs sampling

1. initial value $(\theta_1^{(0)}, \theta_2^{(0)})$

sample $\theta_1 \sim p(\theta_1 | \theta_2)$
sample $\theta_2 \sim p(\theta_1 | \theta_1)$

- Sequence θ is not indep but a Markov Chain

Gibbs sampler algorithm: - attempts to solve problem of high dimensional sampling by breaking down the problem into several lower dimensional problems