Markov Chains and MCMC

CompSci 590.04
Instructor: AshwinMachanavajjhala

Recap: Monte Carlo Method

- If U is a universe of items, and G is a subset satisfying some property, we want to estimate |G|
  - Either intractable or inefficient to count exactly

For i = 1 to N

- Choose u ε U, uniformly at random
- Check whether u ε G ?
- Let $X_i = 1$ if u ε G, $X_i = 0$ otherwise

Return $\hat{C} = |U| \cdot \frac{\sum_i X_i}{N}$

Variance: $|U| \frac{\mu(1-\mu)}{\sqrt{N}}, where\ \mu = \frac{|G|}{|U|}$

Duke UNIVERSITY

Recap: Monte Carlo Method

When is this method an FPRAS?

- $|U|$ is known and easy to uniformly sample from U.
- Easy to check whether sample is in G
- $|U|/|G|$ is small … (polynomial in the size of the input)

Theorem:

$$\forall\, 0 < \varepsilon < 1.5, 0 < \delta < 1, if\ N > \frac{|U|}{|G|} \cdot \frac{3}{\varepsilon^2} \cdot \ln\frac{2}{\delta}$$

$$then, P\left[(1-\varepsilon)|G| \leq \hat{C} \leq (1+\varepsilon)|G|\right] \geq 1-\delta$$

Recap: Importance Sampling

- In certain case |G| << |U|, hence the number of samples is not small.
- Suppose q(x) is the density of interest, sample from a different approximate density p(x)

$$\int f(x)q(x)dx = \int f(x)\left(\frac{q(x)}{p(x)}\right)p(x)dx$$

$$= E_{p(x)}\left[f(x)\frac{q(x)}{p(x)}\right]$$

$$Hence, \int f(x)q(x)dx \approx \frac{1}{N}\sum_{i=0}^{N} f(X_i)\frac{q(X_i)}{p(X_i)},$$

$$where\ X_i\ are\ sampled\ from\ p(x)$$

Recap: Metropolis-Hastings Algorithm

- Start with any initial value x_0, such that $p(x_0) > 0$
- Using current value x_{t-1}, sample a new point according some **proposal distribution** $q(x_t \mid x_{t-1})$
- Compute $\alpha(x_t|x_{t-1}) = \min\left(1, \frac{p(x_t)}{p(x_{t-1})} \frac{q(x_{t-1}|x_t)}{q(x_t|x_{t-1})}\right)$
- With probability α accept the move to x_t,
  otherwise reject x_t

Duke UNIVERSITY

Recap: Why does Metropolis-Hastings work?

- Metropolis-Hastings describes a Markov chain with transition probabilities:

$$P(x,y) = q(y\,|x)\ \min\left(1, \frac{p(y)}{p(x)}\frac{q(x|y)}{q(y|x)}\right)$$

- Satisfied the detailed balance condition with p(x) as the stationary distribution:

$$p(x)P(x,y) = p(y)P(y,x)$$

Today's Class

- Variants on MCMC
- Burn-in and Convergence

Metropolis Algorithm

- The proposal distribution is symmetric

$$q(x|y) = q(y|x)$$

- Transition probability simplifies to:

$$P(x, y) = q(x|y)\ \min\left(1, \frac{p(y)}{p(x)}\right)$$

Gibbs Sampling

- Suppose we want to sample a high dimensional point from a probability distribution p(x1, x2, …, xd)

Algorithm:

- Initialize starting value X^0 = x1, x2, …, xd
- Pick some ordering of the variables (say 1..d)
- i = 1
- Do until convergence:
  - Sample x from p(xi | x1, x2, .., xi-1, xi+1, .., xd)
  - Set xi = x
  - i = i + 1 mod d

Gibbs Sampling is a special case of MCMC

- Sampling from conditional is precisely the transition probability
- Accept move with probability 1

$$
\begin{aligned}
&P((x, \boldsymbol{x}_{-i}), (y, \boldsymbol{x}_{-i})) \\
&= p(y \mid \boldsymbol{x}_{-i})\, min\left(1, \frac{p(y, \boldsymbol{x}_{-i})}{p(x, \boldsymbol{x}_{-i})} \frac{p(x \mid \boldsymbol{x}_{-i})}{p(y \mid \boldsymbol{x}_{-i})}\right) \\
&= p(y \mid \boldsymbol{x}_{-i})
\end{aligned}
$$

Today's Class

- Variants on MCMC
- Burn-in and Convergence

Burn-in & Convergence

- MCMC eventually converges to the stationary distribution
- Period till it reaches converges is burn-in
  - Those samples are discarded.

- Estimating convergence
  - Run multiple chains in parallel and check whether their distributions are similar.