Lecture 7
Stochastic Search

Armando Solar-Lezama and Xiaokang Qiu
MCMC and Metropolis Search

The Markov Chain Monte Carlo Revolution
Persi Diaconis
Let $\chi$ be a finite set

A Markov chain is defined by a matrix $K(x, y): \chi \times \chi \to \mathbb{R}$

- $K(x, y) \geq 0$
- $\sum_y K(x, y) = 1$

Probability of a series $X_0, X_1, X_2 ...$

- $P(X_1 = y|X_0 = x) = K(x, y)$
- $P(X_1 = y, X_2 = z|X_0 = x) = K(x, y)K(y, z)$
- $P(X_2 = z|X_0 = x) = \sum_y K(x, y)K(y, z)$
  - This is matrix multiplication!
Stationary distribution

What is the probability $\pi(x)$ of being in a node $x$ at some arbitrary step?

- $\pi(x) > 0$ and $\sum_x \pi(x) = 1$
- $\pi(y) = \sum_x \pi(x) K(x, y)$
  - i.e. $\pi = \pi K$
Fundamental theorem of (finite) Markov chains

If there is an $n_0$ s.t. $n > n_0 \Rightarrow K^n(x, y) \geq 0$

- i.e. the matrix is connected.

$$\lim_{n \to \infty} K^n(x, y) = \pi(y)$$

- The n’th step of a run starting at $x$ has probability close to $\pi(y)$ of being at $y$ if $n$ is large.
Motivating example

What is the secret function $f$?

Suppose we know probabilities of two letter sequences

\[ \pi(f) = \frac{1}{z} \prod_i M(f(s_i), f(s_{i+1})) \]

If we can sample repeatedly from $\pi$ we can find the most likely function $f$.
Metropolis algorithm

Start with a markov matrix \( J(x, y) \) with \( J(x, y) > 0 \iff J(y, x) > 0 \)

\[
K(x, y) = \begin{cases}
J(x, y) & \text{if } x \neq y, \ A(x, y) \geq 1 \\
J(x, y)A(x, y) & \text{if } x \neq y, \ A(x, y) < 1 \\
J(x, y) + \sum_{z: A(x, z) < 1} J(x, z)(1 - A(x, z)) & \text{if } x = y
\end{cases}
\]

\( A(x, y) \) is the acceptance ratio \( \frac{\pi(y)J(y,x)}{\pi(x)J(x,y)} \)

Note \( \pi(x)K(x, y) = \pi(y)K(y, x) \)

- Then \( \sum_x \pi(x)K(x, y) = \sum_x \pi(y)K(y, x) = \pi(y) \sum_x K(y, x) = \pi(y) \)
Random switch of two symbols

- Note that $J(f, f^*) = J(f^*, f)$ so $A(f, f^*) = \pi(f^*)/\pi(f)$
Many recent synthesis applications

Influential work by Schkufza, Sharma Aiken.
  • Focus on program optimization

Wide variety of applications in other areas
  • Probabilistic programming
  • Cognitive Science
  • etc.
Key issues

Need good estimates of $\pi$

Where does it work?