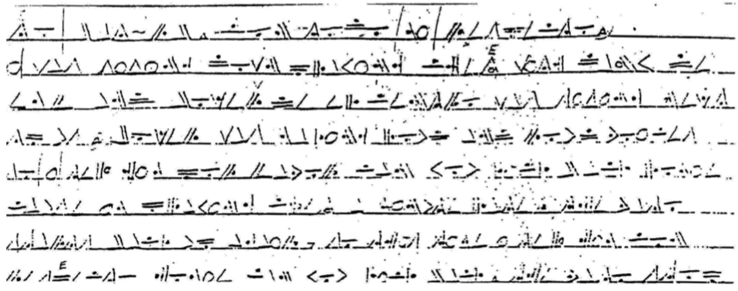
The Markov Chain Monte Carlo Revolution

Persi Diaconis

Cryptographic example



$$f : \{\text{code space}\} \to \{\text{human-readable characters}\}$$

Plausibility

Matrix of transitions in human-readable text (i.e. (english) text in War and Peace) $M(x, y)$.

$$Pl(f) = \prod_i M(f(s_i), f(s_{i+1}))$$

where s_i is the ith character in the cipher text (i.e. the code).

The Algorithm

- Start with some preliminary guess f
- Get some proposed decryptor f_*, probably using f
- accept f_* as the new f with probability $Pl(f_*)/Pl(f)$, otherwise throw it away

Using the output of the chain and our intuition, somehow get what we think is the correct f.

Intro to Markov Chains

Let $\mathcal{X}$ be finite.

- A Markov Chain is defined by a matrix $K(x, y)$ for $x, y \in \mathcal{X}$
- $K(x, y) \geq 0$ and $\sum_y K(x, y) = 1 \Rightarrow K(x, .)$ is a pmf.
- Define $K(x, y) = P(X_{i+1} = y | X_i = x)$

$$
\begin{aligned}
P(X_{i+2} = y | X_i = x) &= \sum_z P(X_{i+1} = z | X_i = x) P(X_{i+2} = y | X_{i+1} = z) \\
&= \sum_z K(x, z) K(z, y) \\
&= (K^2)_{xy}
\end{aligned}
\tag{1}
$$

- (x, y)th entry of K^n is $P(X_{i+n} = y | X_i = x)$.

Stationary and Equilibrium Distributions

If there exists a $pi(.) : \mathcal{X} \to [0,1]$ s.t. $\sum_x \pi(x) = 1$ and $\sum_x \pi(x)K(x,y) = \pi(y)$ then $\pi(.)$ is the stationary distribution of K.

Th$\underline{\text{m}}$ If there is an n_0 s.t. $K^n(x,y) \geq 0$ for all $n > n_0$, then K has a unique stationary dist$\underline{\text{n}}$ π and, as $n \to \infty$, $K^n(x,y) \to \pi(y)$ for all $x, y \in \mathcal{X}$.

Th$\underline{\text{m}}$ If π is the unique stationary distribution of K for which the Markov Chain is denoted $X_1, X_2, ...$ then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(X_i) = E_\pi(f(X))$$

so long as $E_\pi(|f|) < \infty$.

Let's Metropolize

- 1. Propose with $J(x, y)$
  2. Accept the proposed state with probability

$$A(x, y) = \min\left(1, \frac{\pi(y)J(y, x)}{\pi(x)J(x, y)}\right)$$

- Form of A implies *detailed balance*: $\pi(x)K(x, y) = \pi(y)K(y, x)$ which itself means

$$\sum_x \pi(x)K(x, y) = \sum_x \pi(y)K(y, x) = \pi(y)\sum_x K(y, x) = \pi(y)$$

Convergence

Total Variational Distance:

$$\|K^n(x,.) - \pi(.)\|_{TV} = \sup_{A \subseteq \mathcal{X}} |K^n(x,A) - \pi(A)| = \frac{1}{2}\sum_y |K^n(x,y) - \pi(y)|$$

What's the smallest n s.t. $\|K^n(x,.) - \pi(.)\|_{TV} < \epsilon$?

Let $L^2(\pi) = \{g : \mathcal{X} \to \mathbb{R}\}$ with inner product

$$\langle g, h \rangle = \sum_x g(x)h(x)\pi(x)$$

K acting on $L^2(\pi)$:

$$Kg(x) = \sum_y g(y)K(x,y)$$

K satisfies detailed balance $\Rightarrow \langle Kg, h \rangle = \langle g, Kh \rangle \Rightarrow K$ is self adjoint $\Rightarrow$ orthonormal basis of e.vectors ψ_i and e.values β_i for $0 \leq i \leq |\mathcal{X}| - 1$

Convergence II

So

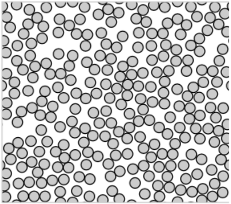
$$K(x,y) = \pi(y) \sum_{i=0}^{|\mathcal{X}|-1} \beta_i \psi_i(x)\psi_i(y)$$
$$\Rightarrow K^n(x,y) = \pi(y) \sum_{i=0}^{|\mathcal{X}|-1} \beta_i^n \psi_i(x)\psi_i(y) \tag{2}$$

and

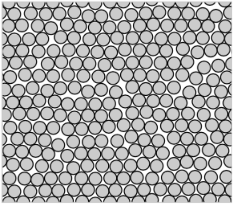
$$4\|K^n(x,.) - \pi(.)\|_{TV}^2 \leq \sum_y \frac{(K^n(x,y) - \pi(y))}{\pi(y)} = \pi(y) \sum_{i=0}^{|\mathcal{X}|-1} \beta_i^{2n} \psi_i(x)^2$$

By Cauchy-Schwartz.

Hard Discs in a Box



$\eta = 0.48$



$\eta = 0.72$

What is the topology of $\mathcal{X}(n, \epsilon)$?

Embed in $\mathbb{R}^{2n}$ and we get a natural, uniform dist$^{\underline{\text{n}}}$.

How to sample from this dist$^{\underline{\text{n}}}$? Say samples $X_1, ..., X_k$. Then

$$\int_{\mathcal{X}(n,\epsilon)} f(x)\mathrm{d}x \approx \frac{1}{K}\sum_k f(X_k)$$

Let's Metropolize: II, Electric Boogaloo

Now we want particle configurations to be distributed according to a potential $U(x)$: Let $\Omega \subseteq \mathbb{R}^d$ be a bounded, connected, open set. $p(x) = z^{-1} \exp(-U(x))$ where $z = \int_\Omega \exp(-U(x)) \mathrm{d}x$ is a probability density on Ω.

Algo$^{\underline{\text{m}}}$: for an $x \in \Omega$, fix a small, positive h

1. Choose $y \in B_h(x)$ from a normalized Lebesgue measure on the ball
2. if $p(y) \geq p(x)$:
    - Move to y
3. else:
    - Move to y with probability $p(y)/p(x)$ otherwise stay put.

Looking outside the domain

- Chemistry + Physics
  - MCMC for systems with glassy dynamics.
  - BORG for cosmology
- Biology
  - MCMC for Phylogenetics
- Theoretical Computer Science
  - $O(\exp(n))$ to $O(n^k)$ given a sampler.
  - Wigderson: We can eliminate the randomness XOR $P = NP$
- Graph Theory
  - MCMC for the Planted Clique Problem (Angelini et al. 2021)
- Literally any Bayesian inference with a complex enough posterior
  - $\pi(\theta|x) \propto L(x;\theta)\pi(\theta)$
  - Intractable likelihood? No problem! Just do more MCMC

What's Hot Right Now

- Adaptive MCMC (been hot for a long time actually)
- Non-reversible MCMC
  - Piecewise Deterministic Markov Processes
- Event Chain Monte Carlo
- Unbiased Monte Carlo via Couplings
- Kernel Stein Discrepancy
  - Thinning
  - in Adaptive MCMC