Math 456: Mathematical Modeling

Thursday, April 19th, 2018

Mixing times for Markov chains

Thursday, April 19th, 2018

Warmup

Consider our old friend, the differential equation

$$\dot{x} = Ax$$

Suppose that A is a $N \times N$ symmetric matrix.

Warmup

Denote the eigenvalues of A as

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$$

and the corresponding eigenvectors by $\phi_1, \ldots, \phi_N$ (assumed to form an orthonormal basis).

The solution to the differential equation is

$$x(t) = \sum_{k=1}^{N} e^{t\lambda_k} c_k \phi_k$$

where the coefficients c_k are determined by the initial condition

$$x(0) = \sum_{k=1}^{N} c_k \phi_k$$

Warmup

Since the vectors $\phi_1, \ldots, \phi_n$ form an orthonormal basis, we have

$$c_k = (x(0), \phi_k), \text{ for each } k.$$

Therefore, we may write

$$x(t) = \sum_{k=1}^{N} e^{t\lambda_k}(x(0), \phi_k)\phi_k.$$

Warmup

Consider the following special situation:
All of the eigenvalues are < 0, except for λ_1, which is zero.

In this scenario, we have

$$\lim_{t \to \infty} x(t) = c_1 \phi_1$$

Moreover, this convergence happens with **an exponential rate**

$$|x(t) - c_1 \phi_1| \leq e^{-\lambda_2 t} |x(0)|$$

Warmup

This means for instance, that if $|x(0)| = 1$, and we want to for $x(t)$ to be within ε of its limit $c_1\phi_1$, then it suffices to take

$$t > \tau(\varepsilon)$$

Here,

$$e^{-\lambda_2\tau(\varepsilon)} = \varepsilon$$

That is,

$$\tau(\varepsilon) = \frac{1}{\lambda_2}\ln(\varepsilon^{-1})$$

Today

1. The problem at hand: speed of convergence
2. Metrics among distributions?
3. Rates of convergence
4. What's next?

Quantifying the speed of convergence

This semester we have learned that for an irreducible, aperiodic chain we have:

The distributions $\pi_{x,n}(y) := \mathrm{p}^n(x,y)$
all converge to π **as** $n \to \infty$

That is, for any x and any y

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

where π is the unique stationary distribution of the chain.

Quantifying the speed of convergence

The following is a practical question whose importance cannot be overstated (specially, if one is using Monte Carlo methods).

Think of the vectors π and $\pi_{x,n}$ (for some fixed x), then:

$$\text{what is an upper estimate on the size of } \|\pi - \pi_{x,n}\|?$$

This is often stated as follows: how fast does the chain mix? In general, this is a very difficult question, but lots of progress has taken place in the last 50 years or so.

Quantifying the speed of convergence

Total Variation

We have the Total Variation metric

$$\mathrm{d}_{\mathrm{TV}}(\pi_1, \pi_2) = \max_{A \subset S} |\pi_1(A) - \pi_2(A)|$$

Also written as

$$\mathrm{d}_{\mathrm{TV}}(\pi_1, \pi_2) = \tfrac{1}{2} \sum_{y \in S} |\pi_1(y) - \pi_2(y)|$$

Quantifying the speed of convergence

Total Variation

If A is some set of states, then

$$\pi_2(A) - \delta \leq \pi_1(A) \leq \pi_2(A) + \delta$$

as long as $\mathrm{d_{TV}}(\pi_1, \pi_2) \leq \delta$.

Quantifying the speed of convergence

L^2 metric

The L^2 metric is given by

$$\mathrm{d}_{L^2}(\pi_1, \pi_2) = \left(\sum_x |\pi_1(x) - \pi_2(x)|^2 \right)^{\frac{1}{2}}$$

It is the same as the usual distance for vectors.

Quantifying the speed of convergence

Kullback-Leibler

The L^p metric is given by

$$\mathrm{d}_{KL}(\pi_1, \pi_2) = \sum_x \pi_1(x) \log\left(\frac{\pi_2(x)}{\pi_1(x)}\right)$$

Quantifying the speed of convergence

L^p metric

The Kullback-Leibler divergence is given by

$$\mathrm{d}_{L^p}(\pi_1, \pi_2) = \left(\sum_x |\pi_1(x) - \pi_2(x)|^p \right)^{\frac{1}{p}}$$

Here, $1 \leq p \leq \infty$

Quantifying the speed of convergence

Exponential rate of convergence

A rate of convergence theorem would look as follows:

For every $x \in S$, we have

$$d(\mathrm{p}^n(x, \cdot), \pi) \le a_1 e^{-a_2 \frac{n}{\gamma^2}}$$

Moreover,

$$\max_{x \in S} d(\mathrm{p}^n(x, \cdot), \pi) \ge a_3 e^{-a_4 \frac{n}{\gamma^2}}$$

Reversible Markov Chains

The following shows how the problem of convergence can be related to understanding eigenvalues of the transition matrix, **provided the chan is reversible**, that is

$$\pi(x)\mathrm{p}(x,y) = \pi(y)\mathrm{p}(y,x).$$

A simple case of this is when $\mathrm{p}(x,y) = \mathrm{p}(y,x)$ for all x and y.

Assuming reversibility, we introduce an inner product

$$\langle f, g\rangle_\pi := \sum_{x\in S} f(x)g(x)\pi(x)$$

defined over the vector space of functions on the state space.

Reversible Markov Chains

The reason we like this inner product is that the reversibility of the chain means that the transformation L is symmetric with respect to this inner product

$$\langle Lf, g \rangle_\pi = \langle f, Lg \rangle_\pi$$

In other words: if we were to write the matrix for L in a basis which is orthogonal with respect to π, then $L_{ij} = L_{ji}$ in this basis.

Reversible Markov Chains

$$|Lf(x)| = |\sum_{y\in S} f(y)\mathrm{p}(y,x)|$$
$$\leq \sum_{y\in S} |f(y)|\mathrm{p}(y,x)$$

Suppose that $Lf = \lambda f$ for some $f \neq 0$ and some λ

$$|\lambda|^2 |f(x)|^2 \leq \sum_{y\in S} |f(y)|^2 \mathrm{p}(y,x) \quad \text{for all } x$$

Reversible Markov Chains

Adding these up over S, we have

$$|\lambda|^2 \|f\|_\pi^2 \leq |\lambda| \|f\|_\pi^2.$$

Since $f \neq 0$, it follows that $|\lambda| \leq 1$.

Reversible Markov Chains

For any eigenvalue λ of L, we have shown

$$|\lambda| \leq 1.$$

Since L is symmetric, this means that there is a basis $\phi_1, \ldots, \phi_n$ of eigenvectors with respective eigenvalues

$$1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_N \geq -1.$$

Important: $\phi_1 = \pi$, the unique stationary distribution.

Reversible Markov Chains

The Chapman-Kolmogorov equation has the form

$$\pi_{x,n} = L^n f_x$$

Where $f_x = (0, \ldots, 1, \ldots, 0)$ with the 1 occuring at the location corresponding to the state x.

Reversible Markov Chains

Now, as we saw L has eigenvectors $\phi_1, \ldots, \phi_N$ with corresponding eigenvalues $\lambda_1, \ldots, \lambda_N$.

If $f = \sum_{k=1}^{N} c_k \phi_k$, then

$$Lf = \sum_{k=1}^{N} c_k \lambda_k \phi_k$$

Actually, for any n

$$L^n f = \sum_{k=1}^{N} c_k \lambda_k^n \phi_k$$

Reversible Markov Chains

Let us rewrite this as

$$L^n f = c_1 \pi + \sum_{k=2}^{N} c_k \lambda_k^n \phi_k$$

where, we recall that

$$\begin{aligned} c_k &= \langle f, \phi_k \rangle_\pi \\ &= \sum_y f(y) \phi_k(y) \pi(y). \end{aligned}$$

In particular, for $f = f_x$

$$c_k = \phi_k(x) \pi(x).$$

Reversible Markov Chains

If $|\lambda_k| < 1$ for $k = 2, \ldots, N$, then we have exponential convergence

$$|L^n f - \pi(x)| \leq \lambda^n \sum_{k=2}^{N} |c_k \phi_k|$$

where $\lambda = \max_{2 \leq k \leq N} |\lambda_k|$.

This requires, of course, bounding all the eigenvalues of the transition matrix, and understanding the size of its eigenvectors.

Where to go from here?

A few directions, according to your interests

1. The theory of Martingales.
2. Continuum Stochastic Processes.
3. Scientific computing (Monte Carlo methods, parallel computing, molecular simulations)
4. Analysis of large random networks.
5. Molecular biology (e.g. mechanistic models for DNA transcription)
6. Probabilistic algorithms for optimization.
7. Statistical inference and machine learning.
8. Spectral Graph theory.