Markov Chains on Finite Groups

Maarit Hietalahti

Postgraduate Seminar in Theoretical Computer Science 17.11.2003

Based on

Sections 15 and 16 of *E. Behrends. Introduction to Markov Chains, with Special Emphasis on Rapid Mixing. Vieweg & Sohn, Braunschweig Wiesbaden, 2000.*

and

P. Diaconis. Group Representations in Probability and Statistics. Institute of Mathematical Statistics, Hayward CA, 1988.

Contents

Refresher on Algebra

group $(G, \circ)$: G set, $\circ$ associative multiplication between elements of G: if $g, h \in G$ then $g \circ h \in G$. Identity: $g \circ id = id \circ g = g$. Inverse $g^{-1}g = id$.

subgroup $H \in G$: $id \in H$ and $h_1 \circ h_2 \in H$ when $h_1, h_2 \in H$. (H is closed with respect to $\circ$)

group generator $g \in G$ is said to generate the group G, if for all elements of $h \in G$ there is a k s.t. $h = g^k$.

conjugacy class H subgroup, left (right) conjugacy classes are sets of the form $H \circ g$ $(g \circ H)$ with $g \in G$.

group homomorphism is a map between two groups G, H such that 1) $f(g_1 g_2) = f(g_1)f(g_2)$ and 2) $f(id_G) = id_H$.

Markov chains on finite commutative groups

$(G, \circ)$ is a finite group, $g, h, \ldots \in G$ are the states of a Markov chain.
$\mathbb{P}_0$ is a probability measure on G.
Transition probabilities: $p_{g,h\circ g} := \mathbb{P}_0(\{h\})$

lemma 15.1

- $p_{g,h\circ g}$ are the entries of a (doubly) stochastic matrix. Thus, the uniform distribution of this matrix is the equilibrium distribution.
- H subgroup generated by $supp := \{h|\mathbb{P}_0(h) > 0\}$. The irreducible subsets of the chain are precisely the sets of the form $H \circ g$ with $g \in G$, that is, the left conjugacy classes. In particular, the chain is irreducible iff $supp\mathbb{P}_0$ generates G.
- The chain is aperiodic and irreducible iff there is a k s. t. every element of G can be written as the product of k elements, each lying in $supp\mathbb{P}_0$.

Outline: the train of thought

Problem: How fast does the chain converge to its equilibrium?

–> What is the distribution after k steps of a walk which starts at 0?
Answer: $\mathbb{P}_0^{(k*)}$.

Notion Matrix doubly stochastic: the uniform distribution is the equilibrium distribution!

–> How fast does the $\mathbb{P}_0^{(k*)}$ tend to the uniform distribution?

–> How close is a distribution $\mathbb{P}_0$ to the uniform distribution?

Notion Variation distance can be calculated with the help of the Fourier transformation

–> How small are the $\hat{\mathbb{P}_0}(\chi)$ for the nontrivial characters χ?

k-step transitions

Probability $\mathbb{P}_0$ on G for the one-step transitions.

- Start: g_0 arbitrary.
- $g_0 + h_0$ with probability $\mathbb{P}_0(\{h_0\})$ for h_0.
- $(g_0 + h_0) + h_1$ with probability $\mathbb{P}_0(\{h_1\})$ for h_1.
- and so on.

Note that h_0 and h_1 are independent. 2-step transitions:
$g_0 \to (g_0 + h_0) + h_1 = g_0 + h$ for which the probability is
$\Sigma_{h_0+h_1=h}\mathbb{P}_0(\{h_0\})\mathbb{P}_0(\{h_1\}) = \Sigma_{h_0}\mathbb{P}_0(\{h_0\})\mathbb{P}_2(\{h - h_0\}))$.

Convolutions of probability measures

Definition 15.9 Let $\mathbb{P}_1, \mathbb{P}_2$ be probability measures on G.

(i) We define the convolution $\mathbb{P}_1 * \mathbb{P}_2$ of $\mathbb{P}_1, \mathbb{P}_2$ by
$(\mathbb{P}_1 * \mathbb{P}_2)(\{h\}) := \Sigma_{h_0} \mathbb{P}_1(\{h_0\}) \mathbb{P}_2(\{h - h_0\})$

(ii) In the special case $\mathbb{P}_1 = \mathbb{P}_2 = \mathbb{P}_0$ we put $\mathbb{P}_0^{(k*)} := \mathbb{P}_0 * \mathbb{P}_0$. This is extended to a definition for arbitrary integer exponents $\mathbb{P}_0^{((k+1)*)} := \mathbb{P}_0^{(k*)} * \mathbb{P}_0$.

Characters

Relating abstract groups to complex numbers:

Definition 15.2 Denote by $(\Gamma, \cdot)$ the multiplicative group of all complex numbers of modulus one. Then a character on G is a group homomorphism χ from G to Γ: $\chi(g+h) = \chi(g)\chi(h)$ for all $g, h \in G$.

Properties of characters:

- $(\overline{\chi}(g) = \overline{\chi(g)})$ is a character. (Also, $\overline{\chi}$ is the inverse $1/\chi$ of χ.)
- $\chi_1\chi_2$ is a character when χ_1, χ_2 are.
- The trivial character: $\chi_{triv} : g \rightarrow 1$.
- $\hat{G}$, the collection of all characters, forms a commutative group with resp. to pointwise multiplication.
- If G has N elements, the range of any character on G is contained in the set of the N'th roots of unity ($exp(2\pi ij/N)$, $j = 0, \ldots, N-1, i = \sqrt{-1}$)

Lemma 15.3 and corollary

Let $(G,+)$ be a commutative group with N elements. The N-dimensional vector space of all mappings from G to $\mathbb{C}$ will be denoted by X_G, and this space will be provided with the scalar product $< f_1, f_2 >_G := \Sigma_g f_1(g)\overline{f_2(g)}/N$.

(i) Let χ be a character which is not the trivial character χ_{triv}. Then $\Sigma_g \chi(g) = 0$.

(ii) In the Hilbert space $(X_g, < \cdot, \cdot >_G)$ the family of characters forms an orthonormal system.

(iii) Any collection of characters is linearly independent

(iv) $\hat{G}$ has at most N elements.

(v) In fact there exists N different characters so that $\hat{G}$ is an orthonormal basis of X_G. Also $(G,+)$ is isomorphic with $(\hat{G},\cdot)$.

Corollary 15.4

(i) Let f be any element of X_G. Then f can be written as a linear combination of the $\chi \in \hat{G}$ as follows: $f = \Sigma_\chi < f, \chi >_G \chi$.

(ii) For different $g, h \in G$ there is a character χ s.t. $\chi(g) \neq \chi(h)$.

Fourier transform

Fourier transform of measure $\mathbb{P}_0$:

$$\hat{\mathbb{P}_0} : \hat{G} \rightarrow \mathbb{C}, \chi \mapsto \Sigma_g \chi(g)\mathbb{P}_0(\{g\})$$

Fourier transform of convolutions:
For probability measures $\mathbb{P}_1, \mathbb{P}_2$ on $(G, +)$ the Fourier transform of $\mathbb{P}_2 * \mathbb{P}_1$ is just the (pointwise) product of the functions $\hat{\mathbb{P}_1}$ and $\hat{\mathbb{P}_2}$. In particular it follows that, for any probability $\mathbb{P}_0$, the Fourier transform of $\mathbb{P}_0^{(k*)}$ is the k'th power of the Fourier transform of $\mathbb{P}_0$.

Calculating the variation distance

Lemma 15.8 Let $\mathbb{P}_0, \mathbb{P}_1, \mathbb{P}_2$ be probability measures on the finite commutative group G. By U we denote the uniform distribution.

(i) $\mathbb{P}_0 = U$ iff $\hat{\mathbb{P}}_0(\chi)$ is one for the trivial character and zero for the other χ.

(ii) The variation distance $\|\mathbb{P}_1 - \mathbb{P}_2\|$ can be estimated by $(\Sigma_\chi |\hat{\mathbb{P}}_1(\chi) - \hat{\mathbb{P}}_2(\chi)|^2)^{1/2}/2$; in particular $\|\mathbb{P}_1 - U\|$ is less than or equal to $(\Sigma_{\chi \neq \chi_{triv}} |\hat{\mathbb{P}}_1(\chi)|^2)^{1/2}/2$, where the summation runs over all nontrivial characters χ.

(iii) Conversely, the distance of $\hat{\mathbb{P}}_1$ and $\hat{\mathbb{P}}_2$ with respect to the maximum norm is bounded by $2\|\mathbb{P}_1 - \mathbb{P}_2\|$.

Rapid mixing: Conclusion

Combining previous results gives us

$$\|\mathbb{P}_0^{(k*)} - U\|^2 \leq \frac{1}{4}\Sigma_{\chi \neq \chi_{triv}} |\hat{\mathbb{P}_0}(\chi)|^{2k}$$

Remark: Generalization to arbitrary finite groups

Relating the abstract group to something more concrete is done by using *representations*. Characters will no longer do, as they are homomorphisms with commutative ranges, which cannot distinguish between different elements of a non-commutative groups.

The use of representations leads to more demanding technicalities. In other respects, the construction follows the same principles.