RANDOM WALK AND FOURIER ANALYSIS ON GROUPS

HPH@MATH.UCHICAGO.EDU

CONTENTS

ABSTRACT. How many times do we need to shuffle a deck of card to make it sufficiently random? In this talk, we will formalize this notion, from random walk to fourier analysis on groups. We will also draw similarities to a family of problems (that we do not address), that, roughly speaking, is a (more complicated) sampling on groups.

Whenever necessary, we assume that the group is finite (so that we do not need measure-theoretic languages).

0. A FEW SCENARIOS

(1) Given an encrypted text; want to recover it.
(2) Given a scrambled text; want to recover it.
(3) Given a deck of cards; want to randomize it.

Scenario	(1),(2)	(3)
Aim	Find $f \in \begin{cases} \text{Bij}(\{\text{Codeword}\}, \{\text{Usual Alphabet}\}) \\ \text{Bij}(\{\text{Usual Alphabet}\}, \{\text{Usual Alphabet}\}) \end{cases}$ in scenario $\begin{cases} (1) \\ (2) \end{cases}$ that is "most plausible", i.e. f that maximizes $\text{Pl}(f) := \prod_i M(f(s_i), f(s_{i+1}))$, where $M(\$, \#)$ as "the probability that the next letter is $\#$ given the current letter being \$".	Get a random deck of card
Algorithm	"The Metropolis Algorithm" (a Monte-Carlo-Markov-Chain method): (1) Pick a random f. (2) Perform the following repeatedly (for "sufficiently many times"): (a) $f_* :=$ random transposition $\cdot f$. • If $\text{Pl}(f_*) > \text{Pl}(f)$, goto f_* • Otherwise, flip a $\text{Pl}(f_*)/\text{Pl}(f)$; – If comes up H, goto f_*. – Otherwise (comes up T), goto f.	Repeated shuffling
Target Probability Distribution	$p(f) \propto \text{Pl}(f)$	uniform

Questions:

Q1: Why would the proposed algorithms cause our configuration converge to the target probability distribution?
A: Our algorithm above is a S_n-valued Markov chain, which target probability appears as a left eigenvector of the 1-eigenvector of the transition probability.

Q2: What does it even mean by "converge to"? How to measure the difference between 2 probability distributions?
A: There is a distance function $\|\cdot\|_{\text{TV}}$ between 2 probability distributions; p_1 and p_2 "sufficiently close to" usually means $\|p_1 - p_2\|_{\text{TV}} \leq 1/4$, but really, I'm not sure why this is a good cutoff.

Q3: Why consider (1)&(2) and (3) differently?
• The distinction of treatment is not artificial: in the latter, the transition probability only depends on the (group) difference between 2 states, while this is not the case in the former. This matters.

In the rest of the talk, we will give an upper bound to $\|P(X_t) - p\|_{\text{TV}}$ to the latter case in Q3; more importantly, we introduce Fourier analysis of group, and draw useful connection to group representation. via Fourier analysis of group, and our understanding of certain group representation.

1. Random Walk on Group

A random walk on a group G is a sequence of G-valued random variables $\{X_i\}_{i\in\mathbb{Z}_{\geq 0}}$, s.t. the transition probability $P(X_{i+1} = y|X_i = x)$ is

- independent of i
- memoryless, i.e. $P(X_{i+1}|X_i) = P(X_{i+1}|X_i, \ldots, X_1)$

In this case, we denote $P_{x,y} := P(X_{i+1} = y|X_i = x)$. Also, we denote $p_0(x) := P(X_0 = x)$ as the starting distribution.

Definition 1.1. A probability distribution on G is a function $p : G \to [0, 1]$ s.t. $\sum_{g\in G} p(g) = 1$. The set of probability distributions on G is denoted as $\text{Prob}(G)$.

Remark 1.2. 2 examples:

(1) The uniform distribution on G is defined as $u(g) = \frac{1}{|G|}$.

(2) The delta distribution on G wrt $x \in G$ is defined as $\delta_x(g) = \begin{cases} 1 & x = g \\ 0 & \text{otherwise} \end{cases}$.

Remark 1.3. (1) Suppose a random walk starts with $P_{x,y} = p(yx^{-1})$ for some $p \in \text{Prob}(G)$ and $p_0 = \delta_{\text{Id}}$, then for any $g \in G$,

- $P(X_2 = g) = \sum_{h\in G} p(gh^{-1})p(h)$.
- More generally, for any $k \in \mathbb{Z}^+$,

$$P(X_k = g) = \sum_{\substack{(g_k,\ldots,g_1)\in G^k \\ \prod_{i=k}^1 g_i = g}} \prod_{i=k}^1 p(g_i)$$

Above are "iterated convolutions" $\underbrace{p * \ldots * p}_{k \text{ times}}$ of p, for $k = 2$ and general $\mathbb{Z}^+$ respectively.

2. A Distance between Probability Distributions on G

Definition 2.1. The total variation distance on $\text{Prob}(G)$ is given by

$$\|p_1 - p_2\|_{\text{TV}} := \max_{A\subset G} |p_1(A) - p_2(A)| = \frac{1}{2}\sum_{g\in G} |p_1(g) - p_2(g)| = \frac{1}{2}\|p_1 - p_2\|_1$$

Remark 2.2. It is indeed a distance function.

3. Fourier Analysis on Group

In what follows, the set of irreducible representations of G is denoted as $\hat{G}$.

Definition 3.1. For a $f \in \text{Fun}(G, \mathbb{C})$, the fourier transform of f wrt the irreducible representation $\rho : G \to \text{Aut}_{\mathbb{C}}(V_\rho)$ is

$$\hat{f}(\rho) = \sum_{g\in G} f(g)\rho(g) \in \text{End}_{\mathbb{C}}(\mathbb{C}^{d_\rho})$$

Remark 3.2. (1) (Inverse Fourier Transform) For $f \in \text{Fun}(G, \mathbb{C})$, $g \in G$,

$$\frac{1}{|G|}\sum_{\rho\in\hat{G}} d_\rho \,\text{Tr}(\rho(g^{-1})\hat{f}(\rho)).$$

(2) For $G = \mathbb{R}$, and the irrep $\hat{t} : \mathbb{R} \to \text{GL}(1, \mathbb{C})$ s.t. $r \mapsto e^{-2\pi ixr}\cdot$, $\hat{f}(\hat{t}) = (\int_{x\in\mathbb{R}} f(x)e^{-2\pi ixt})\cdot \in \text{End}_{\mathbb{C}}(\mathbb{C})$. Note the bijection $\mathbb{R} \leftrightarrow \hat{\mathbb{R}}$, $t \leftrightarrow \hat{t}$.

(3) Similar story holds for any (locally compact) abelian group. In this case, all the irreps are 1-dimensional; moreover, $\text{Hom}(G, S^1) \leftrightarrow \hat{G}$ (Re: Pontryagin dual).

Remark 3.3. There is a familiar generalization. Suppose $\sigma : G \to \text{Aut}_{\mathbb{C}}(V)$ is a representation, then the σ-fourier transform of f wrt the irreducible representation $\rho : G \to \text{Aut}_{\mathbb{C}}(V_\rho)$ is

$$\hat{f}^\sigma(\rho) = \sum_{g\in G} (f \circ \sigma(g))\rho(g) \in \text{Fun}(V, \text{End}_{\mathbb{C}}(V_\rho))$$

(1) For $L_G : G \to \text{GL}(\mathbb{C}^G)$ left regular representation, above is recovered, i.e. $\hat{f}^{L_G}(\rho)(\text{Id}_G) = \hat{f}(\rho)$.

(2) For $\rho_{\text{sgn}} : \mathbb{Z}/2\mathbb{Z} \to \text{GL}(1, \mathbb{R})$ signed rep, $f \in \text{Fun}(\mathbb{R}, \mathbb{C})$, $\hat{f}^{\rho_{\text{sgn}}}(\rho)(x) = \begin{cases} f(x) + f(-x) & \rho = \rho_{\text{trv}} \\ f(x) - f(-x) & \rho = \rho_{\text{sgn}} \end{cases}$

Remark 3.4. (1) $p \in \text{Prob}(G)$

$$\hat{p}(\rho_{\text{trv}}) = \sum_{g\in G} p(g) = 1.$$

(2) $u \in \mathrm{Prob}(G)$ uniform distribution, for $\rho \in \hat{G}$,

$$\hat{u}(\rho) = \begin{cases} \mathrm{Id}_{V_\rho} & \rho = \rho_{\mathrm{trv}} \\ 0 & \text{otherwise} \end{cases}.$$

Definition 3.5. For $f_1, f_2 \in \mathrm{Fun}(G, \mathbb{C})$, the (left) convolution between f_1 and f_2 is given by

$$(f_1 * f_2)(g) = \sum_{h \in G} f_1(gh^{-1}) f_2(h).$$

Remark 3.6. For any probability distributions p_1, p_2 on G,

(1) $p_1 * p_2$ is also a probability distribution.
(2) In particular, iterated convolution p^{*k} ($p \in \mathrm{Prob}(G)$, $k \in \mathbb{Z}^+$) is a probability distribution.
   - On a random walk where $P_{x,y}$ only depends on yx^{-1}, i.e. $P_{x,y} = p(yx^{-1})$ for some probability distribution p on G, then $P(X_k = x) = p^{*k}(x)$.

Remark 3.7. The following holds:

(1) ("Convolution theorem") $\widehat{f_1 * f_2} = \hat{f}_1 \cdot \hat{f}_2$
(2) ("Plancheral theorem") $\|f\|_2^2 = \frac{1}{|G|} \sum_{\rho \in \hat{G}} d_\rho \|\hat{f}(\rho)\|_{\mathrm{HS}}^2 (= \frac{1}{|G|} \sum_{\rho \in \hat{G}} d_\rho \operatorname{Tr}(\hat{f}(\rho)\hat{f}(\rho)^*))$

The latter is phrased this way to highlight an isometry between G and $\hat{G}$,

Lemma 3.8. *("Upper Bound Lemma") For any $k \in \mathbb{Z}^+$, $p \in$ Prob(G),*

$$\|p^{*k} - u\|_{TV}^2 \leq \frac{1}{4} \sum_{\rho \neq \rho_{trv}} \|\hat{p}(\rho)\|_{HS}^{2n}$$

Proof.

$$\|p^{*k} - u\|_{\mathrm{TV}}^2 = \frac{1}{4}\|p^{*k} - u\|_1^2 \leq \frac{1}{4}|G|\|p^{*k} - u\|_2^2 = \frac{1}{4}\sum_{\rho \in \hat{G}} \|\hat{p}^n(\rho) - \hat{u}(\rho)\|_{\mathrm{HS}}^2 = \frac{1}{4} \sum_{\substack{\rho \in \hat{G} \\ \rho \neq \rho_{\mathrm{trv}}}} \|\hat{p}(\rho)\|_{HS}^{2k},$$

□

where the 1st "=" holds since $\|\cdot\|_{TV} = \frac{1}{2}\|\cdot\|_1$, the 1st "≤" holds by Cauchy-Schwarz, the 2nd "=" holds by "convolution theorem", "Plancheral", and linearity of "$f \mapsto \hat{f}$" and the 3rd "=" holds by Rem 3.4.

4. Applications

Thanks to our extensive knowledge of the representation theory of symmetric groups, we have:

Theorem 4.1 (Diaconis-Shahshahani, 1980). *On $G = S_{52}$, for "random transpotion" $p(g) = \begin{cases} \frac{1}{52} & g = Id \\ \frac{2}{52^2} & g = \text{ transposition}, \\ 0 & \text{otherwise} \end{cases}$, $\forall c > 0$, $k \geq 103 + 26c \Rightarrow d(\mu^{*k}, \lambda) \leq 6e^{-c}$.*

Remark 4.2. In particular, $k \geq 270 \Rightarrow d(\mu^{*k}, \lambda) \leq \frac{1}{100}$.

Theorem 4.3 (Bayer-Diaconis, 1992). *On $G = S_n$, $p \in$ Prob(S_{52}) given by the following distribution:*

> ***1st description.*** **Begin by choosing an integer c from $0, 1, \ldots, n$ according to the binomial distribution $P\{C = c\} = \frac{1}{2^n}\binom{n}{c}$. Then, c cards are cut off and held in the left hand, and $n - c$ cards are held in the right hand. The cards are dropped from a given hand with probability proportional to packet size. Thus, the chance that a card is first dropped from the left hand packet is c/n. If this happens, the chance that the next card is dropped from the left packet is $(c-1)/(n-1)$.**

$\|p^{*k} - u\|_{TV} \leq 1 - \prod_{i=1}^{n-1}(1 - \frac{i}{2^k})$.

Remark 4.4.
(1) This shuffle is riffle shuffle, and p has a name: Gilbert-Shannon-Reed distribution.
(2) $k = 2\log_2(n/c) \Rightarrow \|p^{*k} - u\|_{\mathrm{TV}} \leq c^2/2$
(3) More precise calculations on RHS of the theorem yield

k	10	11	12	13	14
Upper bound	.73	.48	.28	.15	.08

(4) More refined analysis was performed, to show that "LHS when $k \geq 7$" $\leq 1/4$.

4.1. **Back to "Scenarios (1) and (2)".** Since the transition probability P_{xy} does not just depend on yx^{-1}, we cannot describe the probability distribution at the kth timestamp by iterated convolutions, thereby cannot be cleanly described by the language above. However, for some special target distributions, an analysis can be done thanks to symmetric function theory. This is treated in references[?] in the paper "Monte Carlo Markov Chain Revolution" by Diaconis.

5. References

(1) Diaconis, P. Group Representations in Probability and Statistics.
(2) Diaconis, P. The Markov Chain Monte Carlo Revolution.
(3) Sahlsten, T. Analysis, Random Walks and Groups.
(4) Wikipedia.