CIS 700/004: Paper Writeup

Riffled Independence for Ranked Data.
Jonathan Huang, Carlos Guestrin.
In *Advances in Neural Information Processing Systems* (NIPS), 2009.

Writeup by Zachary Schutzman

February 2, 2017

1 Introduction

This paper introduces the notion of riffle independence as a method of considering distributions over S_n when the set of objects can naturally be split into a p-subset and a q-subset ($n = p + q$) by considering separate distributions S_p and S_q along with a shuffling distribution that interleaves the objects from the two subsets. The authors approach this problem with techniques from group-theoretic Fourier analysis. Using some experiments on real data sets, they demonstrate that their algorithms may be useful for analyzing certain distributions over the symmetric group.

The numbering of references corresponds with the citations in the paper.

2 Problem Setup

It is well-established that there is a correspondence between distributions on the symmetric group and the matrices corresponding to the Fourier transform at the irreducible representations of the group [2,6,7,8]. In fact, these matrices systematically encode information about the distribution. Under a particular partial ordering of the representations, the first two Fourier matrices encode the first order marginals of the distribution, the next three encode the second order marginals, and so on, where an k^{th} order marginal refers to the probability of observing an k-tuple at some position in the permutation. In a practical setting, because higher marginals require storage of a number of parameters quadratic in k, it may be impractical to store them for large n. The process and consequences of discarding higher-order marginals is called 'bandlimiting', and is not studied in this paper.

For large problems, we look to definitions of independence. If we can factor our distribution h into less complex components, we may be able to reduce the storage complexity of h. The authors define **probabilistic independence** as a distribution h which factors as $h = f \cdot g$, where f is a distribution over S_p and g over S_q. Storing this distribution only requires $O(p! + q!)$ parameters instead of $O(n!)$, but this full independence assumption may be too strong, as it asserts that there is no notion of preference across the two subsets.

The authors introduce the idea of **riffle independence**, where a distribution h factors as $h = m * f \cdot g$, where $*$ denotes the convolution operation. m is defined as a distribution over the (p, q)**-interleavings**, a strict subset of S_n containing only those permutations that preserve the order within the p- and q-subsets. Note that any

(p, q)-interleaving is fully specified by the positions of the p elements, and therefore there are $\binom{n}{p} = \binom{n}{q}$ such interleavings. The authors define the uniform riffle distribution, m^{unif}, as the uniform distribution over all such interleavings, which in effect expresses preferences within the subsets but indifference across them.

As a generalization, the authors then define a **biased** riffle shuffle, with bias parameter α. In this family of distributions, an interleaving is drawn from m^{α} by iteratively choosing to take the first item in the p-subset with probability proportional to the fraction of remaining items in that subset and α. Note that if $\alpha = .5$, we recover exactly m^{unif}. Increasing α corresponds to increased preference for the p items over the q items. $\alpha = 1$ indicates we prefer every item in p to any in q, and $\alpha = 0$ indicates the opposite.

Observe that in terms of complexity, riffle independence lives somewhere between full independence and conditional independence. If we think about the factorization of h as being conditionally independent over the selected subset, we have a number of parameters on the order of $\binom{n}{p} \cdot (p! + q!)$, as we have, for each p-subset, two riffle factors conditional on that subset. On the other hand, if h is fully riffle independent, we only have $O(\binom{n}{p} + p! + q!)$ parameters characterizing the distribution. This can be significantly fewer than in the case of conditional independence but still more than the $O(p! + q!)$ parameters required for full independence.

3 Main Results

The authors present two algorithms [6,7,8], $RiffleJoin$ and $RiffleSplit$ to analyze riffle indpendence. The first, given the Fourier matrices for two riffle factors f, g and a shuffling distribution m, constructs a full distribution h over all of S_n. The second, given a distribution h over S_n performs a dual convolution by the transpose of the m^{unif} matrices, and essentially performs the inverse of $RiffleJoin$.

The authors provide several theoretical properties and guarantees for their algorithms. The first, which is rather intuitive, is that if h factors into perfectly riffle independent f and g with respect to some shuffling distribution m, then $RiffleSplit(h)$ returns *exactly* f and g. This is a relatively direct consequence of f and g being independent implies that the Fourier matrices of h are block-diagonal with block sizes corresponding to composed representations of S_p and S_q.

Similarly, they claim that we can construct the k^{th} order marginals of h from the Fourier matrices that yield enough information to construct the k^{th} order marginals of f and g. In the perfectly riffle independent case, this also follows from the block-diagonal structure of the Fourier matrices of h. It is less obvious in the case where f and g are not perfectly riffle independent.

In the case where the distribution is not perfectly riffle independent, the authors show that $RiffleSplit(h)$ returns the factors $\hat{f}, \hat{g}$ that minimize the KL-divergence from the true distributions with respect to some shuffling distribution m. This implies that the algorithms can reveal useful information about h even when f and g are not entirely riffle independent.

The authors then use these two algorithms on two data sets to demonstrate riffle independence in real data. In the first, they examine the APA data set [3] and determine that in ranking the five candidates, voters picked between 'factions', and then chose within those factions. On the Sushi data set [10], the authors used their algorithms to show that assumptions of riffle independence can significantly reduce the necessary sample complexity for learning a distribution and that biased riffle distributions are useful as a learning bias tool for small samples.

4 Discussion

This paper is not the first to examine Fourier-theoretic methods of analysis of distributions on S_n, but it is (likely) the first to consider the notion of riffle independence. As such, this topic, its uses, and its implications are largely unexplored. The authors also come to this problem from the perspective of robotics and object tracking, so there may be a lot of future work to be done in relating these concepts to ranked and choice data.

In particular, there may be room to explore methods to determine which subsets form the riffle factors. For the APA data set, the authors performed an exhaustive search over the input to determine the two factions of the candidates. This may work for a case with five objects and 5000 observations, but as either of those items grows, exhaustive search becomes computationally intractable.

One weakness of this method is that assuming the data contains two riffle independent factors is a very strong assumption. The algorithms they present operate specifically in the case of two factors. The extension of the algorithm to three or more riffle factors may be simple, but it is not immediately obvious that their proof extends to this case. Specifically, one would need to prove that if a distribution h factors into three riffle factors f_p, f_q, f_r with a (p, q, r)-interleaving then it can be first decomposed as (f_{pq}, f_r) with a $(p + q, r)$-interleaving and then further decompose f_{pq} into f_p, f_q with a (p, q)-interleaving such that we get $h = m_{p+q,r} * (m_{p,q} * (f_p \cdot f_q) \cdot f_r) = m_{p,q,r} * (f_p \cdot f_q \cdot f_r)$. If this equality holds, then the authors' proof extends to an arbitrary number of riffle factors.

Citations and References

[2] P. Diaconis. *Group Representations in Probability and Statistics*. IMS Lecture Notes, 1988.

[3] Persi Diaconis. *A generalization of spectral analysis with application to ranked data*. The Annals of Statistics, 17(3):949979, 1989.

[6] J. Huang, C. Guestrin, and L. Guibas. *Efficient inference for distributions on permutations*. In NIPS, 2007.

[7] J. Huang, C. Guestrin, and L. Guibas. *Fourier theoretic probabilistic inference over permutations*. JMLR, 10, 2009.

[8] J. Huang, C. Guestrin, X. Jiang, and L. Guibas. *Exploiting probabilistic independence for permutations*. In AISTATS, 2009.

[10] Toshihiro Kamishima. *Nantonac collaborative filtering: recommendation based on order responses*. In KDD, pages 583588, 2003.

For a thorough mathematical presentation of representation theory: G. James and M. Liebeck, *Representations and Characters of Groups*. Cambridge, U.K.: Cambridge Univ. Press, 1993.