Random Walks in the Permutation Group

Moo K. Chung, ¹ Linhui Xie, ² Yixian Wang, ¹ Shih-Gu Huang, ¹ Jingwen Yan, ² Li Shen ³

1) University of Wisconsin-Madison, USA

2) Indiana University-Purdue University, USA

3) University of Pennsylvania, USA

email: mkchung@wisc.edu

ABSTRACT

The permutation test is an often used test procedure for determining statistical significance. However, generating every possible permutation for whole images with million voxels is extremely time consuming even for modest sample sizes [1]. The number of permutations increases *exponentially* as the sample size increases. In this study, we show how to accelerate the permutation test by exploiting the underlying *topological structure* of the permutation group.

PERMUTATION TEST

Consider ordered sets $\mathbf{x} = (x_1, x_2, \dots, x_m)$ and $\mathbf{y} = (y_1, y_2, \dots, y_n)$. The distance between \mathbf{x} and \mathbf{y} is measured by statistic $f(\mathbf{x}, \mathbf{y})$. Under the null hypothesis of the equivalence of \mathbf{x} and \mathbf{y} , the elements are permutable. Consider combined set $\mathbf{z} = (x_1, \dots, x_m, y_1, \dots, y_n)$ and its all possible permutations \mathbb{S}_{m+n} , which is a symmetric group of order m+n [3]. Split the permutation $\tau(\mathbf{z})$ of \mathbf{z} into two groups: $\tau(\mathbf{x}) = (\tau(x_1), \dots, \tau(x_m)), \tau(\mathbf{y}) = (\tau(y_1), \dots, \tau(y_n))$. Then the *exact p-value*, the probability of the observed event in \mathbb{S}_{m+n} , is given by the fraction

$$\frac{1}{(m+n)!} \sum_{\tau \in \mathbb{S}_{m+n}} \mathcal{I}\Big(f(\tau(\mathbf{x}), \tau(\mathbf{y})) > f(\mathbf{x}, \mathbf{y})\Big),$$

where $, \tau \in S_{m+n}$ and \mathcal{I} is an indicator function taking value 1 if the argument is true and 0 otherwise. The fraction is not possible to compute within reasonable time when m and n are large.

RANDOM TRANSPOSITIONS

Apply permutation π_{ij} between **x** and **y** such that only one element in each group is changed: $\pi_{ij}(\mathbf{x}) = (x_1, \dots, x_{i-1}, y_j, x_{i+1}, \dots, x_m), \pi_{ij}(\mathbf{y}) = (y_1, \dots, y_{j-1}, x_i, y_{j+1}, \dots, y_n).$ Such a permutation is called the *transposition* or *walk* [3].

Theorem 1. Any permutation in the permutation group \mathbb{S}_{m+n} of order m + n can be reachable by a sequence of transpositions.

Over transposition π_{ij} , the statistic changes from $f(\mathbf{x}, \mathbf{y})$ to $f(\pi_{ij}(\mathbf{x}), \pi_{ij}(\mathbf{y}))$, which can be computed in *constant* run time by updating the value of $f(\mathbf{x}, \mathbf{y})$ in an online fashion [2]. For

instance, computing the t-statistic with m and n samples requires 3(m+n) + 20 operations per permutation. In comparison, it only takes 35 operations to computer t-statistic per transposition.

Theorem 2. If f is an algebraic function such as addition, subtraction, multiplication, division, integer exponents, there exists a function g such that

$$f(\pi_{ij}(\mathbf{x}), \pi_{ij}(\mathbf{y})) = g(f(\mathbf{x}, \mathbf{y}), x_i, y_j),$$

where g can be computed in constant run time.

HERITABILITY OF LARGE-SCALE BRAIN BRAIN NETWORK

Diffusion weighted imaging (DWI) data of 356 twin subjects were used in the study. There are 117 monozygotic (MZ) and and 61 same-sex dizygotic (DZ) twin pairs (ages 29.2 ± 3.4). Biologically accurate 1 million white matter fiber tracts per subject were obtained. The subject-level 116×116 connectivity matrices were constructed by counting the number of tracts connecting between parcellations. The heritability index (HI) *h* was estimated using 1 million transpositions (Figure 1). The statistical significance of the HI-map was determined based on 30 million transpositions for 52 minutes in a computer (corrected *p*-value = 0.0144).



Figure 1. a) Simulation study showing the faster convergence of transposition method. b) The relative error based on the average of 1000 simulations. c-d) MZ- and DZ- twin correlation networks thresholded at 0.6. The correlation networks are obtained after averaging one million sequential random transpositions. (c) HI-map thresholded at 1 showing significant genetic heritability of the brain network.

Acknowledgements. NIH grants R01 EB022856 and EB022574. We would like to thank Michael Newton of University of Wisconsin-Madison for valuable discussion.

REFERENCES

- 1. M.K. Chung, V. Vilalta-Gil, H. Lee, P.J. Rathouz, B.B. Lahey, and D.H. Zald. Exact topological inference for paired brain networks via persistent homology. *Information Processing in Medical Imaging (IPMI), Lecture Notes in Computer Science (LNCS)*, 10265:299–310, 2017.
- 2. M.K. Chung, Y. Wang, S.-G. Huang, and I. Lyu. Rapid acceleration of the permutation test via slow random walks in the permutation group. *arXiv preprint arXiv:1812.06696*, 2018.
- 3. T.W. Hungerford. Algebra. Springer-Verlag, New York, 1980.