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

\[
\frac{1}{(m+n)!} \sum_{\tau \in S_{m+n}} I(f(\tau(x), \tau(y)) > f(x, y)),
\]

where \( \tau \in S_{m+n} \) and \( I \) 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 \( \pi_{ij} \) between \( x \) and \( y \) such that only one element in each group is changed: \( \pi_{ij}(x) = (x_1, \cdots, x_{i-1}, y_j, x_{i+1}, \cdots, x_m) \), \( \pi_{ij}(y) = (y_1, \cdots, y_{j-1}, x_i, y_{j+1}, \cdots, y_n) \). Such a permutation is called the transposition or walk [3].

Theorem 1. Any permutation in the permutation group \( S_{m+n} \) of order \( m+n \) can be reachable by a sequence of transpositions.

Over transposition \( \pi_{ij} \), the statistic changes from \( f(x, y) \) to \( f(\pi_{ij}(x), \pi_{ij}(y)) \), which can be computed in constant run time by updating the value of \( f(x, y) \) 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

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

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 61 same-sex dizygotic (DZ) twin pairs (ages $29.2 \pm 3.4$). Biologically accurate 1 million white matter fiber tracts per subject were obtained. The subject-level $116 \times 116$ 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).

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

**REFERENCES**

