Comparisons of Distance Function Based Permutation Testing in Diffusion Tensor-MRI with Multiple Sclerosis Induced Microstructural Variations

Lingchih Lin[1], Jianhui Zhong[1,2], and Walter G. O'Dell[3]

[1]*Department of Physics and Astronomy, University of Rochester, Rochester, NY, United States,* [2]*Department of Imaging Sciences, University of Rochester, Rochester, NY, United States,* [3]*Department of Radiation Oncology, University of Florida, Gainesville, FL, United States*

Introduction: To understand the relationship between the variations of distribution functions within/across region of interests (ROI) and multiple sclerosis induced structural changes, the geometry properties of Euclidean, squared Euclidean[1], and log-Euclidean distance functions [2] are investigated in this study to quantify the DTI detected variations, within and between group differences, in the permutation tests. A novel method to enhance the efficacy in computing tensor pairs as well as sensitivity in diagnosing multiple sclerosis from Diffusion Tensor-MRI is performed by extrapolating the tail distribution of the test statistics (the distance functions) according to a generalized Pareto model [3]. This method provides a means to increase the accuracy in diagnosis to cover all the relabeling data sets with a reduced number of permutations. To assess and enhance the statistical power of inference test, without assuming the exact density functions of diffusivities and diffusion anisotropy indices a priori, the median, mean, and Riemannian based distance functions provide alternatives from the pre-assumed anisotropy levels for the resampling based analysis. This analysis is applied to a small sample of human brain data-sets, which are the averages across time points, enhancing signal-to-noise ratio (SNR) without removing fine-scale structures.

Theory:. The fractions of p-value distributions are calculated from r dimensional distance functions, $\delta_{L,M} = [\sum_{h=1}^{r}(x_{hL} - x_{hM})^2]^{v/2}$(where x is the three dimensional diffusivity or can otherwise be selected from one of the DTI-derived parameters), for (1) v=2, the squared Euclidean distances associated with parametric ANOVA and two-sample t tests and, for (2) v=1, the median based Euclidean distances satisfying the triangle inequality, from objects L and M. Another family of (3) log-Euclidean distances, with analogy to the geodesic anisotropy (GA) [4] but between two diffusion tensors to be evaluated, can be derived from the assumption that $\delta_{L,M} = [\sum_{h=1}^{r}(\log(x_{hL}/x_{hM}))^2]^{1/2}$. The permutation combinations are selected from allocated permutation pairs with distances, $\delta_{intra} = \sum_{i=1}^{g} \varphi_i (C_2^{ni})_v^{-1} \delta_{L,L'}$, (φ_i is the i-th group weight evaluated as ($ni/\sum_{i=1}^{g} ni$))), within and, $\delta_{inter} = [\sum_{h=1}^{r}(< x_{hL} > - < x_{hM} >)^2]^{\bar{z}}$(or $[\sum_{h=1}^{r}(\log(< x_{hL} > / < x_{hM} >))^2]^{1/2}$), between two anisotropy levels of (n1,n2) subjects. The inference p-values can then be predicted as $p_{intra} = (\#(\delta_{intra} < \delta_{intra\_0})/(((\sum ni)!/\prod ni!))$ and $p_{inter} = (\#(\delta_{inter} > \delta_{inter\_0})/(((\sum ni)!/\prod ni!))$. Computation of all possible permutation pairs from (a small sample size) datasets is demonstrated in this study. Instead of selecting from a limiting number of pairs from a large number of possible relabelings, (ex: $> 1.5 \times 10^5$ for n1=n2=10) without considering all of the combinations, we approximate the tail distribution of p-values from the extreme value theory [5] with a generalized Pareto distribution (GPD), $CDF(x) = 1 - (1 - (kx/\alpha))^{\bar{k}}(k \neq 0), = 1 - \exp(-(x/\alpha)), (k = 0)$, according to two parameters, α, the scale parameter, and k, the shape parameter.

Methods: Inter- and Intra- distance functions are simulated in changes with two pre-defined and resampled diffusion tensors by considering only the effect of Johnson thermal noise from high to low anisotropy levels for SNR ranging from 0 to 40 in Matlab (Mathworks, Natick, MA). Then, the DT-MRI detected voxelwise structural variations are compared and quantified in a small sample dataset, including both five healthy human brains (averaging over three time points) and patients of clinically definite secondary progressive multiple sclerosis (SPMS) (averaging over six time points) with relabeled permutation pairs. The data are acquired from a 1.5T GE scanner with 21 unique diffusion gradient directions/two average baseline images (b=1000/0 s mm^{-2}) ,TR/TE=10800/80ms, FOV=24cm, voxel size= 1x1x3 mm$_3$, matrix size = 128x128. The percentages of deteriorations induced by SPMS in each of the region of interests (ROI) are calculated based on Euclidean, Squared Euclidean, and log-Euclidean distance functions to quantify the ratios of significant altered inter- and intra- p-values. The tail distributions of median, mean, and log-Euclidean based distance functions are estimated from maximizing the likelihood function of GPD, comparing the distributions with parameters estimated from methods of moments (MOM) and probability weighted moments (PWM).

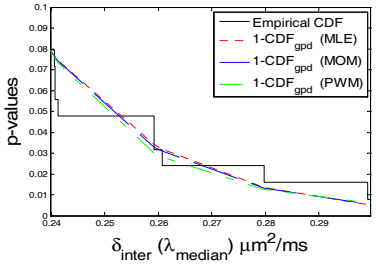

Figure 1: Solid line indicates the cumulative empirical δ to detect the total number of p-values exceeding the test statistic. Dashed lines are the tail distributions approximated by generalized Pareto distribution.

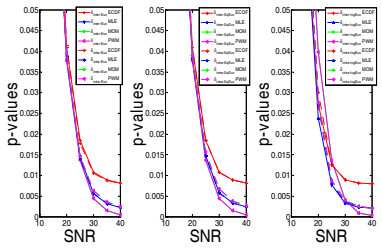

Figure2: p-values of inter-(solid line) and intra-(dashed line) distances from Euclidean, squared Euclidean, and Log-Euclidean distance functions are simulated from two structures of diffusion tensors (FA1 = 0.7 and FA2 = 0.8) and approximated by the methods of MLE, MOM, and PWM.

Results and Discussion: Figure 1 shows the inter-group distances calculated from the median diffusivity and its corresponding p-values from a voxel. The algorithm first computed the number of resamplings that exceeds the difference between the original distance across the normal controls and SPMS patients, the test statistic, with a rough p-value. If the p-value falls into the range of tail distributions as shown in Fig 1, the density is fitted by the generalized Pareto distribution, which covers the distributions from exponential, uniform, and Pareto density, with methods of MLE, MOM, and PWM according to the range of estimated parameters. Figure 2 shows the simulations from the symmetric tensors with large interval variations in high anisotropy level. Parameters estimated from PWM as well as MOM reveal a better approximation of p-value compared to estimations from methods of maximum likelihood between differences in large anisotropy interval. Figure 3 shows the percentage changes of different distance functions between healthy human brains and patients with SPMS. Log-Euclidean based distance functions exhibit more than 45% deterioration in PCR compared to the ratio derived from mean and median based distance functions. A more detailed comparison with both constrained and non-constrained tensor will be the subjects of future studies.

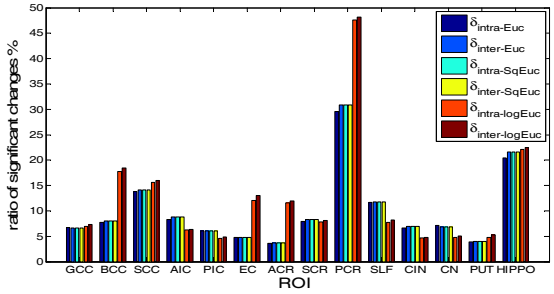

Figure 3: Ratios of Significant different voxels from each of ROI (GCC: Genu Corpus Callosum, BCC: Body Corpus Callosum, SCC: Splenium Corpus Callosum, AIC: Anterior Internal Capsule, PIC: Posterior Internal Capsule, EC: External Capsule, ACR: Anterior Corona Radiata, SCR: Superior Corona Radiata, PCR: Posterior Corona Radiata, SLF:Superior Longitudinal Fasciculus, CIN: Cingulum, CN: Caudate Nucleus, PUT:Putaman, HIPPO:Hippocampus) are evaluated and compared between inter-and intra distance functions from normal controls and patients with SPMS.

References: [1] Mielke, P.W, Berry, K.J.(2001), *Permutation Methods: A Distance Function Approach*, New York: Springer-Verlag [2] Arsigny V. *et al.* Magn. Reson. Med.56:411 (2006) [3] Knijnenburg, T.A., Bioinformatics, 25:161 (2009) [4] Batchelor, P.G., *et al.* Magn. Reson. Med. 53:221 (2005) [5] Gumbel, E.J.(1958), *Statistics of Extremes.* New York: Columbia University Press.