Modelling complex neuroimaging data: Practical approaches to longitudinal and dependent data

Wednesday, Jun 20: 2:45 PM - 4:00 PM
1194
Symposium
Wednesday - Symposia PM

There has been a steady growth in the use of designs that induce complex dependencies in group-level modelling of neuroimaging data. Longitudinal studies, rare early in brain imaging's history, are becoming more prevalent, especially studies with 3 or more time points. Data with related individuals, once only found in the occasional twin study, are now prevalent thanks to the Human Connectome Project (HCP), which uses a twins+siblings design. And stratified sampling, a staple of survey methodology, is integral to the Adolescent Brain Cognitive Development (ABCD) Study, where subjects are sampled within schools, requires modelling of random school effect. All of these designs induce dependence between scans and/or subjects in a way that violates the usual independence assumption. While the standard software packages can account for such dependence in rudimentary ways (e.g. for only 2 scans per subject; or by assuming globally homogeneous repeated measures correlation), they struggle to model arbitrary dependence.

In this workshop we will highlight a number of tools facilitating complex statistical modeling of neuroimaging data. Neuropointilist allows changing neural processes to be modeled as part of complex systems interacting with other change processes with voxel-level structural equation modeling. The Nonparametric Neuroimaging Genetic Analysis framework provides a framework for fast, voxel-level modeling while assessing heritability directly or controlling for its effect on other factors of interest. Mixed Effects for Large Datasets is a robust approach for voxel-level nonparametric hypothesis testing while controlling for sources at multiple levels of the analysis (i.e., subject and item level random effects). Finally, we will discuss bayesian hierarchical modeling as an alternative to moving null hypothesis significance testing that avoids sharp and arbitrary thresholding and promotes complete and transparent reporting of results.

Objective

Understand the importance of accurately modeling complex temporal, genetic, and sampling dependencies in neuroimaging datasets.
Learn about new tools for voxel level modeling of complex dependencies in neuroimaging datasets.
Appreciate the current debate over the utility of null hypothesis statistical testing, how it applies to neuroimaging, and the availability of alternative approaches.

Target Audience

Neuroimagers collecting or analyzing rich datasets who want to learn about the newest tools for modeling the rich relationships in their data.

Co Organizer

Thomas Nichols, University of Oxford

Organizer

Dylan Nielson, PhD, NIMH

Presentations

Neuropointillist: Bringing the interpretive power of structural equation models to longitudinal neuroimaging data (index.cfm?do=ev.viewEv&ev=1616)

The human brain is constantly changing in response to the environment, development, aging and neurodegeneration. Although fMRI has been crucial in helping us understand brain function, modeling trajectories of change over time and examining the relationship of individual differences in growth to other variables has been challenging within existing statistical GLM frameworks. There have been advances in addressing limitations of the GLM approach. For example, mixed effects models as implemented in AFNI greatly improve longitudinal modeling ability, with the ability to handle missing data, compare models, and use the underlying power of mixed effects modeling packages available the R statistical language. However, structural equation models (SEM), popular in the social sciences, give us the ability to examine change in neural processes as outcomes, predictors, correlates of other change processes, or moderators or mediators. This flexibility is currently lacking in neuroimaging software. In this session we introduce important classes of questions about longitudinal change and the kinds of models that can be used to answer them. Finally, we describe Neuropointillist, a simple framework for voxel-wise analyses using any model that can be described in R, and its potential for fostering modeling innovation.

Presenter

Tara Madhyastha, PhD, University of Washington

Modeling genetically-induced dependence in studies of families and “unrelated” individuals. (index.cfm?do=ev.viewEv&ev=1617)

The Human Connectome Project (HCP) and Adolescent Brain Cognitive Development (ABCD) studies are just two examples of large scale studies that have recruited twins or otherwise related subjects. If this dependence is ignored, the positive correlation induced by heritability will inflate false positives and exaggerate significance of true effects. While there are a number of genetics software packages to analyze data from related individuals, they cannot read neuroimaging data and do not scale to voxel-wise analyses, while the standard neuroimaging packages can’t accommodate family structure. We will present our framework for modelling relatedness in neuroimaging data via a fast, non-iterative approach we call Nonparametric Neuroimaging Genetic Analysis (NINGA). Using a 1-step optimisation we obtain approximate estimates of heritability that we’ve validated to be highly accurate. We can make inferences on the heritability itself or, make inferences on any regressor while adjusting for heritability. Specifically we’ve optimised NINGA for genome wide association (GWA) analysis, which requires millions of tests for each voxel. Our method produces parametric p-values, but is sufficiently fast to allow permutation to find FWE-corrected inference or spatial inferences like cluster size or TFCE. We will illustrate NINGA in application to the HCP, providing a survey of heritability across the wide array of measures produced by HCP, and previewing HCP GWA results.

Presenter

Habib Ganjgahi, Statistics department, University of Oxford

Voxel level control of item as fixed-effect fallacy with MELD (index.cfm?do=ev.viewEv&ev=1618)

Researchers often make two contradictory assumptions when analyzing their functional neuroimaging experiments. On the one hand, they assume that stimuli from different categories give rise to different neural responses, while on the other hand, they assume that all the stimuli within a category produce identical responses. This contradiction is the fixed-effect fallacy, well described in the psycholinguistics literature, and recently highlighted as an issue in neuroimaging by Westfall et al (2017). They have shown that failing to correct for item level variance inflates effect sizes by between 16 and 60 percent depending on the numbers of subjects and distinct items presented to those subjects. Some packages exist that allow control of both subject and item level variance at the ROI level, but they

are impractical to apply at the voxel level. We have developed a technique called Mixed Effects for Large Datasets (MELD) to determine feature level significance while controlling for item and subject variance. We achieve this via bootstrap-based feature selection with threshold free cluster enhancement and running mixed effects models in component space after applying a singular value decomposition to those sparse features. We map the parameter estimates from these models back out to voxel space and use permutation to assess voxel-level significance. This combination of feature selection and performing the analysis in component-space makes MELD much faster than element-wise mixed effects analysis while still controlling for variance from multiple levels.

Presenter

Dylan Nielson, PhD, NIMH

Simultaneously Handling Multiple Comparisons and Gaining Inference Efficiency in Neuroimaging (index.cfm?do=ev.viewEv&ev=1619)

In neuroimaging, the multiplicity issue may sneak into data analysis through multiple channels, affecting expected false positive rates (FPRs; type I errors) in diverse ways. One widely recognized aspect of multiplicity occurs when the investigator fits one model for each voxel in the brain, but other situations trigger a multiplicity issue as well. Here, we raise the question of whether the current practice of handling the multiple testing problem through controlling the overall FPR in neuroimaging under the null hypothesis significance testing (NHST) paradigm excessively penalizes the statistical power. More fundamentally, we join the ongoing debate in the statistics community as to whether the adoption of dichotomous decisions through sharp thresholding under NHST is appropriate where the null hypothesis itself is not pragmatically relevant when the effect of interest takes a continuum instead of discrete values, and furthermore is not expected to be null everywhere in the brain. When the noise inundates the signal, two different types of error are more relevant than the concept of FPR: incorrect sign (type S) and incorrect magnitude (type M). In light of these considerations, we introduce a novel strategy using Bayesian hierarchical modeling (BHM) to achieve a different goal: turning the focus of conventional NHST on FPR into quality control by calibrating type S errors while maintaining a reasonable level of inference efficiency. The performance and validity of this approach is demonstrated through an application at the region of interest (ROI) level, with all the regions on an equal footing: small regions are not disadvantaged simply because of their physical size. In addition, compared to the massively univariate approach, BHM may simultaneously achieve increased spatial specificity and detection efficiency. The benefits of BHM are illustrated in model performance and quality checking using an experimental dataset. In addition, BHM offers an alternative, confirmatory, or complementary approach to the conventional whole brain analysis under NHST, and promotes results reporting in totality and transparency. The methodology also avoids the sharp and arbitrary thresholding in the p-value funnel to which the multidimensional data are reduced. The modeling approach with its auxiliary tools will be available as part of the AFNI suite for general use.

Presenter

Gang Chen, National Institutes of Health