The Annals of Applied Statistics
2011, Vol. 5, No. 2B, 1127–1131
DOI: 10.1214/11-AOAS485

SPECIAL SECTION ON STATISTICS IN NEUROSCIENCE

BY KAREN KAFADAR

Indiana University

This article provides a brief introduction to seven papers that are included in this special section on Statistics in Neuroscience:

(1) Xiaoyan Shi, Joseph G. Ibrahim, Jeffrey Lieberman, Martin Styner, Yimei Li and Hongtu Zhu: Two-state empirical likelihood for longitudinal neuroimaging data
(2) Vincent Q. Vu, Pradeep Ravikumar, Thomas Naselaris, Kendrick N. Kay, Jack L. Gallant and Bin Yu: Encoding and decoding V1 fMRI responses to natural images with sparse nonparametric models
(3) Sourabh Bhattacharya and Ranjan Maitra: A nonstationary nonparametric Bayesian approach to dynamically modeling effective connectivity in functional magnetic resonance imaging experiments
(4) Christopher J. Long, Patrick L. Purdon, Simona Temereanca, Neil U. Desai, Matti S. Hämäläinen and Emery Neal Brown: State-space solutions to the dynamic magnetoencephalography inverse problem using high performance computing
(5) Yuriy Mishchenko, Joshua T. Vogelstein and Liam Paninski: A Bayesian approach for inferring neuronal connectivity from calcium fluorescent imaging data
(6) Robert E. Kass, Ryan C. Kelly and Wei-Liem Loh: Assessment of synchrony in multiple neural spike trains using loglinear point process models
(7) Sofia Olhede and Brandon Whitcher: Nonparametric tests of structure for high angular resolution diffusion imaging in Q-space

1. Introduction. In a lecture at Indiana University in March 2008, Peter Hall offered several valuable insights about the field of statistics, three of which are noted below:

1. Advances in statistics have come from the need to analyze different data types ("Statistics is 'reactive;' it is very responsive to new problems that arise in chemistry, biology, physics, ...").
2. Data sets continue to increase in size.
3. Computational algorithms are essential components of the analysis: "Advances in powerful computing equipment has had a dramatic impact on statistical methods and theory. It has changed forever the way data are analyzed."

Received May 2011.
Key words and phrases. Functional magnetic resonance imaging (fMRI), brain imaging, exploratory analysis, nonparametric fitting, model selection, signal detection.

The seven articles in this special section on *Statistics and Neuroscience*, together with two earlier *AOAS* articles, vividly illustrate all three principles.

Function of the human nervous system has fascinated researchers for decades, due to its complex network of interactions among critical parts of its components in the central nervous system (brain, spinal cord, retina) and periphery (nerves). The amount of data that can be collected on these individual components is truly massive, now that instruments for measuring signals (responses to stimuli) have been developed with increasing resolution (spatially and temporally) and sensitivity (weaker signals in the presence of high noise levels). The range of statistical methods that are needed to understand neural and brain development, functionality, and interactions is extremely broad. This special section includes seven articles that present useful statistical methodology designed to address various aspects of data that arise in neuroscience, specifically with brain imaging data collected via functional magnetic resonance imaging (fMRI) or other imaging techniques, and the analysis of neural spike train data. The articles demonstrate the wide variety of statistical problems, the diversity of methods that can be applied, and, most importantly, the valuable insights that are obtained through the application of sound statistical methods.

Functional magnetic resonance imaging was developed in the early 1990s for brain imaging [e.g., Ogawa et al. (1992)] and immediately presented statisticians with a huge new area of problems to be considered: the analysis of massive data sets. The data, changes in blood flow in response to neural activity [blood oxygen level dependent (BOLD) signals], can be measured and recorded with spatial resolution on the order of 2–4 millimeters, taken every 2–4 seconds. Noise reduction, image registration, outliers, image detection, spatial and time trends, and multiplicity are only some of the problems that can arise with these data. Among the first statisticians to attack these problems were Keith Worsley and Karl Friston [Worsley and Friston (1995); Worsley et al. (1996); Friston et al. (1995)] and William Eddy and his colleagues [Eddy et al. (1995); Eddy, Fitzgerald and Noll (1996)], who had sufficient computational resources at the time to handle the massive amounts of data. Since then, computational power has significantly advanced, enabling statisticians to investigate other aspects of these types of data. In addition, other imaging methods have been developed with increased sensitivity and resolution. The first three articles in this section develop methods for analyzing fMRI data: Shi et al. (1), Vu et al. (2), and Bhattacharya and Maitra (3). Three articles develop methods for analyzing data using more sensitive imaging techniques: Long et al. (4) model electromagetic source imaging data (magnetoencephalography imaging, or MEG); Mishchenko et al. (5) develop neural connectivity models from data using calcium fluorescent imaging; and Olhede and Whitcher (7) analyze brain images from measurements obtained via a type of magnetic resonance imaging known as high angular resolution diffusion imaging (HARDI). Neural spike trains collected from multielectrode recordings motivate the methods in Kass et al. (6).

Shi et al. (1) develop an *adjusted exponentially tilted empirical likelihood* method to detect differences in the morphological changes, measured via fMRI, in specific regions of the brain between two groups of patients on different treatment protocols. Beyond the development of an appropriate model that accounts for longitudinal measurements with time-varying covariates is the challenge of developing a computational algorithm to handle the data on 238 patients. The results indicate regions of important differences which provide insights into the different mechanisms of the two treatment protocols. Vu et al. (2) use exploratory data analysis and model selection procedures to improve a previously proposed model for brain activity in encoding and decoding sensory stimuli in the form of local constant energy features. Their analysis reveals nonlinearities which, when incorporated into the model, yields a 25% improvement in encoding prediction and hence greater accuracy in image identification. Bhattacharya and Maitra (3) also analyze fMRI signals to model dynamic, nonstationary neural connectivity via a first-order vector autoregressive model which, when applied to fMRI data on patients performing specific tasks, provides insights into those brain mechanisms involved in distinguishing shapes.

Data from more sensitive and higher resolution imaging techniques require more computationally intensive approaches. Long et al. (4) develop high-dimensional (in the number of parameters) state-space models to identifying magnitudes and locations of neural sources that give rise to MEG signals recorded on the surface of the head. Due to the greatly increased resolution of the data and the number of parameters to be estimated, the Kalman filter solution can be implemented only on high-performance supercomputers. The authors' Kalman filter approach can be viewed as a specific implementation of a more general approach using random field theory proposed by Taylor and Worsley (2007) and applied to MEG (and electroencephology, or EEG) data by Kilner and Friston (2010) that appeared in *The Annals of Applied Statistics* last year.

The next two articles in this special section use different sources of data to model neuronal connectivity. One source of data is calcium-sensitive fluorescent imaging, which offers much finer spatial and temporal resolution than is possible with fMRI. Mishchenko et al. (5) use such imaging data to model neural circuitry with a collection of coupled Hidden Markov models (HMMs), where each Markov chain represents the behavior of a single neuron and the coupling between the HMMs reflects the network connectivity matrix. As is the case with the other articles in this section, the vast amounts of data and the complexity of the coupled models require clever computational approaches (in this case, a blockwise Gibbs algorithm) to estimate model parameters with biologically meaningful relevance. Kass et al. (6) consider models for data from external electrodes on the brain. In the past, neural spike trains from external electrodes have been analyzed traditionally as point processes [Brillinger (1988, 1992)]. Such models usually assume stationarity and distinct events (no two events occur at the same time). Here, Kass, Kelly, and Loh enhanced these models for neural spike trains by introducing a class of continuous-time-varying loglinear models which incorporates time-varying intensities, autocovariation, and synchrony. For an approach to estimating the number of neurons involved in a multi-neuronal spike train, see Li and Loh (2011) that appeared in the most recent issue of *AOAS*.

Olhede and Whitcher (7) approach the analysis of brain images through the local estimation of the two-dimensional probability density function (pdf) of HARDI measurements (i.e., measurements of the local molecular diffusion of water molecules, obtained via high angular resolution diffusion imaging). Rather than assuming a Gaussian pdf, Olhede and Whitcher use the increased sampling rate of HARDI to estimate a nonparametric pdf using local measurements of the covariance matrix, enabling greater accuracy (less bias) at relatively little cost in terms of precision (increased variance). However, because the data come from a diffusion process, the measurements are inherently spectral in nature. The authors provide the statistical framework for estimating pdfs in the spectral domain, incorporating known properties of the diffusion process, and then use properties of Fourier transforms to invert the estimated pdf into the brain image domain. Nonparametric tests for non-uniformity, asymmetry, and ellipsoidality in the pdf lead to increased understanding of diffusion in the brain.

As Peter Hall indicated with respect to data in other fields, here the analysis of neuroscience data led to the development of new statistical methodology. Besides the common theme of neuroscience as the motivation for the methodology, all nine articles (the present seven in this issue and the two articles that appeared earlier) share two additional features: (1) the analysis of very large data sets, which thereby require (2) the development of computational algorithms to facilitate estimation of complex models needed to incorporate the nonstandard features of the data (e.g., nonlinearity, nonstationarity, etc.). Many more problems posed by these sorts of data are in need of solutions, for example, relaxing assumptions on models, designing experimental strategies to make best use of the data, developing methods to reduce noise (increase signal-to-noise ratio), etc. Useful, practical solutions can be obtained only through collaboration between scientists and statisticians. We hope that these articles will stimulate statisticians and neuroscientists to collaborate on these problems to further research in both domains.

REFERENCES

BRILLINGER, D. R. (1988). Some statistical methods for random process data from seismology and neurophysiology. *Ann. Statist.* **16** 1–54. MR0924855

BRILLINGER, D. R. (1992). Nerve cell spike train data analysis: A progression of technique. *J. Amer. Statist. Assoc.* **87** 260–271.

EDDY, W. F., FITZGERALD, M. and NOLL, D. C. (1996). Improved image registration by using Fourier interpolation. *Magnetic Resonance in Medicine* **36** 923–931.

EDDY, W. F., FITZGERALD, M., GENOVESE, C. and MOCKUS, A. (1995). The challenge of functional magnetic resonance imagining. In *Massive Data Sets: Proceedings of a Workshop* 39–45. National Academies Press, Washington, DC.

FRISTON, K. J., HOLMES, A. P., WORSLEY, K. J., POLINE, J. B., FRITH, C. and FRACKOWIAK, R. S. J. (1995). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping* **4** 189–210.

KILNER, J. M. and FRISTON, K. J. (2010). Topological inference for EEG and MEG. *Ann. Appl. Statist.* **4** 1272–1290.

LI, M. and LOH, W.-L. (2011). Estimating the number of neurons in multi-neuronal spike trains. *Ann. Appl. Statist.* **5** 176–200.

OGAWA, S., TANK, D. W., MENON, D. W., ELLERMANN, J. M., KIM, S., MERKLE, H. and UGURBIL, K. (1992). Intrinisic signal changes accompanyig sensory stimulation: Functional brain mapping using MRI. *Proc. Natl. Acad. Sci.* **89** 5951–5955.

TAYLOR, J. E. and WORSLEY, K. J. (2007). Detecting sparse signals in random fields, with an application to brain mapping. *J. Amer. Statist. Assoc.* **102** 913–928. MR2354405

WORSLEY, K. J. and FRISTON, K. J. (1995). Analysis of fMRI time-series revisited—again. *Neuroimage* **2** 173–181.

WORSLEY, K. J., MARRETT, S., NEELIN, P., VANDAL, A. C. and FRISTON, K. J. (1996). A unified statistical approach for determining significant voxels in images of cerebral activation. *Human Brain Mapping* **4** 189–210.

DEPARTMENT OF STATISTICS
INDIANA UNIVERSITY
BLOOMINGTON, INDIANA 47408-3825
USA
E-MAIL: kkafadar@indiana.edu