Chapter 5
Statistical Analysis of Functional Magnetic Resonance Imaging Data

Nicole A. Lazar
University of Georgia, USA

ABSTRACT
The analysis of functional magnetic resonance imaging (fMRI) data poses many statistical challenges. The data are massive, noisy, and have a complicated spatial and temporal correlation structure. This chapter introduces the basics of fMRI data collection and surveys common approaches for data analysis.

INTRODUCTION

Human beings have long been interested in how we ourselves work and function. Of particular fascination has been the working of the human brain. Through the ages we have had opportunity to learn about how our brain functions, mostly haphazardly as the result of illness or accident: a Roman gladiator who suffered amnesia after a blow to the skull; an elderly person who had a stroke, lost the use of one side of the body, but gradually recovered much of the original functionality; savants who were unable to speak but were musical or mathematical prodigies. From all of these isolated incidents, together with postmortem dissections of the brains of healthy individuals, scientists were able to build models of how the brain functions, how information is processed, and what specific regions in the brain are responsible for different types of tasks.

It is only relatively recently, however, that technological advances have allowed us to study the function of the human brain – healthy or diseased – in something closer to real time. Neuroimaging techniques such as positron emission tomography (PET), electroencephalography (EEG), magnetoencephalography (MEG), and functional magnetic resonance imaging (fMRI) have brought a wealth of new knowledge and new understanding to cognitive neuroscientists. Statisticians have been an important part of this exciting endeavor.

DOI: 10.4018/978-1-5225-2498-4.ch005
Consider fMRI as an example. With this imaging modality, the subject is put into a magnetic resonance machine – a large powerful magnet – and is asked to perform some task, for example to tap his or her fingers in a particular pattern, or to do a simple math problem. While the subject is carrying out the task, a complex array of machinery takes images of the brain in action; actually, the neuronal activity itself is not measured, but rather an indirect measure called the blood-oxygenation level dependent, or BOLD, signal, which is related to the oxygen requirements of the brain when it processes information. In general terms, parts of the brain that are responding to a stimulus or performing a cognitive task require more oxygen than those that are not. Hence, the blood that is delivered to different areas of the brain will differ in the ratio of oxygenated to deoxygenated hemoglobin. Oxygenated and deoxygenated hemoglobin, in turn, have different magnetic properties. Functional MRI takes advantage of that difference through the measured BOLD signal.

This is obviously a complicated process, and the data that result in the end are also complicated, in ways that make them interesting and challenging for statistical analysis. Some of the crucial features of fMRI data are:

1. They are very noisy. Some noise is intrinsic – it comes from the scanner itself. This type of noise can usually be estimated rather easily and can often be corrected for. Other noise comes from the subject. For instance, breathing and heartbeat introduce systematic noise, as they move the brain within the skull and affect blood flow to the brain. fMRI has very good spatial resolution; measurements are taken on the millimeter scale. As a result, however, even small amounts of movement on the part of the subject contaminate the data, shifting the measured signal from one location to another. Subject-related noise, particularly motion, is harder to handle, but there are procedures in place for this as well.

2. They are massive in scale. For a single subject, the magnetic resonance machine may scan the whole brain 100–200 times over the course of an experiment. Each such three-dimensional volume is made up of hundreds of thousands of volume elements, or voxels, little cubes of length 3 mm per side, typically. That is, for each subject in a study, the data may consist of hundreds of thousands of time series, and each time series may have 100-200 measurements. Of course most studies also include multiple subjects, sometimes divided into groups (for example, patients and controls). That is a lot of data; computational challenges are one important consequence.

3. They are correlated both temporally and spatially. Temporal correlation arises because the entire sequence of the brain response – the so-called hemodynamic response function – to a stimulus is slower than the time it takes to collect an image; hence, the measured signal at a given time point is affected by the signal that was measured in previous time points, and in turn affects signals measured at later time points. As for spatial correlation, all of the voxels are in the same brain; what takes place at one location is affected by - and affects - other locations. The spatial dependence is particularly interesting from a statistical perspective. Whereas many standard spatial statistics models assume that the closer in space two locations are, the more strongly correlated they are, with the strength of the dependence decreasing quickly with distance, the brain does not work that way at all. The brain has two hemispheres, and while one hemisphere tends to be dominant for specific types of processing, there will also be similar functions on the other side - language and visual processing for instance take place in both hemispheres. The result is that locations that are physically far from each other - on different sides of the brain - can still be very highly correlated if they take on similar functional roles.