Abstract

The Big Bang model is a widely held theory of the evolution of the universe and postulates that around 14 billion years ago, the universe was hot, dense and expanding. As it expanded, it started cooling down. In the first three minutes, the temperature was as high as one billion degrees Kelvin and the energy density was so high that atoms could not form, filling the space with stew of photons, baryons, electrons, neutrinos and other matter. As the temperature cooled below one billion degrees Kelvin, it resulted in formation of light-element nuclei, leaving photons as the dominant form of energy in the Universe. And as the temperature reached around 3000K, electrons and protons combined to form atoms, which decoupled the photon-baryon fluid and photons started freely propagating through space. This left-over radiation with its current mean temperature of $\sim 2.7$ Kelvin in the microwave range, is called the cosmic microwave background (CMB). CMB data therefore contains valuable information about fluctuations in matter and energy density of the early universe.

The cosmological model of the Universe that underlies the Big Bang picture is the FRW equation. The adjustable parameters in this equation (and in thermal/chemical history of the Universe) are commonly called the cosmological parameters. Accurate estimation of the cosmological parameters is essential for describing/understanding the nature of the Universe. Specifically, such estimation of the cosmological parameters can be done from observations related to the CMB, LSS, Type-1a supernovae, gravitational lensing (GL), etc. Indeed, such data and a variety of statistical methodologies for inferring cosmological parameters from data have already become essential tools for a cosmologist. Naturally, different types of observations are sensitive to different subsets of the full cosmological parameter set, and a comprehensive picture is likely to emerge only by combining inferences from different sets of observations.

Technological progress of last few decades has made cosmology a data-rich scientific enterprise – cosmology today is driven more and more by observation than ever before. Such data could, in principle, help make accurate estimates of the
cosmological parameters, thereby allowing for a discrimination between competing theories of the Universe by putting stronger bounds on the values of the cosmological parameters. A number of standard statistical techniques are already being used for inferring cosmological parameters from such data.

Cosmic microwave background (CMB) is not same in all directions and is anisotropic. This CMB temperature fluctuation carries abundant information about the early universe. The angular spectrum of CMB temperature fluctuation is a measurable quantity and the shape and locations and heights of the peaks are directly related to parameters in the underlying cosmological models. Therefore angular power spectrum has been extensively used as an acid test to validate competing cosmological models and is also a rich source of information about the cosmological parameters themselves.

The problem of cosmological parameter estimation calls for a careful application of sophisticated computational and statistical methodologies primarily because of the complex nature of noise in the data. For example, different statistical methodologies are often based on different sets of assumptions; consequently, they are likely to yield differing inferences starting from the same data. Also, a methodology may be particularly well-suited for one kind of data. As such, it makes sense to apply a variety of methodologies to the same data to assess the variability of the conclusions drawn from them. Such exercise is expected to help infer quantities of interest in a conservative fashion (i.e., minimize the risk of a wrong inference).

Cosmologists for estimating the CMB power spectrum from data conventionally use model based parametric statistical methods. These parametric regression methods require a pre-specified functional form of the unknown regression function $f$. On the other hand, nonparametric function estimation methods does not assume any specific functional form for $f$, therefore is model independent and is based on fewer and less restrictive assumptions about $f$. Nonparametric methods assumes mild regularity conditions such as smoothness assumptions, membership to a function space etc and can be used as sanity-enforcing mechanism on parametric analysis.
The focus of this thesis is on adaptation and development of sophisticated statistical methodologies for the problem of cosmological parameter estimation, assessing their comparative merit and on checking the validity of mainstream parametric results.

We develop and use an extension of the nonparametric regression method called REACT (Risk Estimation and Adaptation after Coordinate Transformation). We show how risk minimization under monotonicity constraints can be treated as a quadratic minimization. We also suggest a simpler alternative way to calculate the weight matrix in heteroskedastic case.

We apply this method on WMAP 1-, 3-, 5- and 7-year data and make model-independent estimation of the CMB temperature angular power spectrum together with a confidence set around this fit. We obtain the smooth angular power spectrum and uncertainties on locations and heights of peaks and dips by using our other extensions. The extensions include a prescription for obtaining nonparametric fits that are closer to cosmological expectations on smoothness, and a method for sampling cosmologically meaningful power spectrum variations from the confidence set of a nonparametric fit. We demonstrate how this model-independent and data-driven approach can be used to test different cosmological models.

We extend our work to make inferences about cosmological parameters using the nonparametric confidence set for the WMAP 7-year data. We devise a simple and effective way to map the confidence set around the best nonparametric fit into the cosmological parameter space. We obtain the confidence intervals for cosmological parameters, which reflect the degeneracies, and correlations that are known to be associated with CMB data.

We forecast nonparametric estimation of TT, EE and TE angular power spectrum for Planck from simulated Planck data. We present the accuracy of estimation by comparing the nonparametric estimated spectra with true base-line spectra in simulated data. We show how these results can reveal some information about the universe like initial conditions, reionization, acoustic scale, etc. The satisfactory precision in estimated power spectra promises a reasonable model-independent
inference from real Planck data, which will lead to tighter confidence intervals on cosmological parameters and breaking some degeneracies.