Astronomy Program Opening Workshop
August 22-26, 2016

SPEAKER TITLES/ABSTRACTS

Tamas Budavari
Johns Hopkins University

"Multi-Epoch Source Detection"

Observational astronomy in the time-domain era faces several new challenges. One of them is the efficient use of multiple observations. The work presented here is focusing on faint sources at the detection threshold, and seeks to find an incremental strategy for separating real objects from artifacts in ongoing surveys, where one does not have all the observations readily available. We study the detection probability of sources in single exposures and stacks in comparison with what is achievable by matching noisy catalogs to find their results to agree well in the interesting regime. Given a set of measurements, we then discriminate real sources from noise peaks using Bayesian hypothesis testing. Including the astrometric measurements boosts the evidence for the faintest sources, which makes the discrimination easier.

James Cordes
Cornell University

"Pulsar Timing Arrays and Detection of Nanohertz Gravitational Waves"

I will discuss use of an ensemble of radio pulsars as a gravitational wave detector. GW-induced timing signal levels are roughly 100 ns or less over time scales of 5 to ten years and they result from a stochastic background produced by supermassive black-hole binaries. Individual binaries may be detectable as continuous wave signals or by the gravitational wave memory effect produced during mergers. I will discuss the challenge of GW signal detection amid several source of noise that include the pulsar itself and the interstellar medium.

Ian Czekala
Harvard University

"Constructing Flexible Likelihood Functions for Spectroscopic Inference in the Systematics-Dominated Regime"

Spectroscopic observations can provide an immense amount of information about an astrophysical source. Depending on the nature of the source, however, fully extracting this information can be a difficult and meticulous process. For the specific application of inferring fundamental stellar parameters from spectroscopic observations of stars, we are now in an exciting era where synthetic stellar models can deliver high resolution spectra over a wide range of input fundamental stellar parameters (e.g., effective temperature, composition) with full wavelength coverage across the optical and infrared windows. Although these models typically show decent bulk agreement with observed spectra, there is often significant disagreement at the level of individual spectral lines, which results from intrinsic model imperfections (e.g., in the atomic molecular databases or opacity prescriptions). We present a modular, extensible likelihood framework for spectroscopic inference that aims to mitigate the influence of these spectral line outliers on the inferred fundamental stellar parameters. This framework specifically addresses the common problem of mismatches in model spectral line strengths (with respect to data) by using Gaussian process kernels to identify and self-consistently downweight pathological spectral line "outliers." I will discuss some potential applications of this approach to cooler stars (M spectral type) and young stars (variable spectrum).

Beyond systematic model imperfections, a central challenge for full-spectrum inference is the expense of calculating synthetic spectra at new input parameters. Typically, synthetic spectra are published in large libraries spanning a range of parameters (albeit sparsely sampled with respect to the changes in the emergent spectrum). Because simple interpolation schemes lead to artifacts, we developed a Bayesian emulator to provide smooth interpolations of the library while also marginalizing over uncertainties resulting from the interpolation process. Lastly, I will highlight some related challenges of spectroscopic inference, such as searching for blended light from spatially unresolved binary stars using templates, as well as other topics, which will hopefully lead to interesting discussion during the workshop.

Rebekah Dawson
Pennsylvania State University

"Introduction to Exoplanet Data"

I will give an overview of exoplanet data and the statistical challenges posed by these data. I will emphasize time series data (particularly transit and radial velocity) and exoplanet population catalogues and touch briefly on other types of data and detection techniques. I will also discuss simulation data and statistical challenges in comparing simulations to observations.

Ben Farr
University of Chicago

"Characterization of GW Transients"

Characterization efforts within LIGO are divided by source class. Robust models for the emission from compact binary mergers allow for detailed estimates of the physical parameters of the source. For transient sources which lack strong models, such as supernovae, alternative techniques are necessary to detect and constrain signal properties. I will provide a broad overview of the techniques currently used to characterize known and unknown classes of transients, which signal and noise components our models account for, and which they do not.

Will Farr
University of Birmingham

"Multidimensional Clustering, Model Comparison, Measurement Uncertainty, Selection Effects etc., in GW"

Based on the inferred rate of binary black hole coalescences following the observation of the GW150914 event in the first month of Advanced LIGO operation we expect to observe tens of coalescing BBH systems in the next few years as upgrades to the Advanced LIGO detectors improve their sensitivity. With this population of observations we would like to learn about the population of coalescing BBH sources, look for sub-populations, and compare various theoretical formation models against our observations. To do this we will have to deal with selection effects and measurement uncertainty within each source. I will discuss existing methods for performing these sorts of analyses and try to spark a discussion about possible improvements to these methods.

David Jones
SAMSI

"Disentangling Overlapping Astronomical Sources using Spatial, Spectral, and Temporal Information"

In Jones, Kashyap, van Dyk (2015) we present a powerful new algorithm that combines both spatial information (event locations and the point spread function) and spectral information (photon energies) to separate photons from overlapping sources. We use Bayesian statistical methods to simultaneously infer the number of overlapping sources, to probabilistically separate the photons among the sources, and to fit the parameters describing the individual sources. The advantages and utility of combining spatial and spectral information are demonstrated through a simulation study and a data analysis. Since many sources vary in intensity over time, we now aim to additionally model temporal data, that is, the time tags of the detected photons. We discuss the first step towards this goal which is to model spatial and temporal data together and investigate the extent to which the temporal data helps in separating sources. The final aim is to model the spatial, spectral, and temporal data together, taking into account the dependence between the spectral and temporal observations.

Elisabeth Krause
Stanford University

"Statistical Challenges in Cosmology"

Broadly speaking, the precision measurement of cosmological parameters involve the choice and estimation of suitable summary statistics from the pixel-level data, accurate modeling of these summary statistics including observational and astrophysical systematics, and parameter inference. In this talk I will highlight representative statistical challenges in these key analysis steps, using the cosmology analysis plans for the Large Synoptic Survey Telescope (LSST) as an example. In particular, I will discuss the challenges of jointly analyzing different cosmological observables, such as galaxy clustering, weak lensing, and supernovae.

Soumendra Lahiri
North Carolina State University

"Beyond Stationary Gaussian Processes: An overview of model based non-Gaussian time series analysis"

This talk will cover some of the standard time series models that form the basis of the Box & Jenkins approach to time series analysis, and their extensions in recent years. Specifically, I will introduce the Autoregressive (AR) class of non-Gaussian stationary models based on linear recursive relations and their differenced versions that can be used to model nonstationary time series data. I will also go over fractionally integrated processes that are useful for modelling time series with long range dependence, and ARCH/GARCH models that are suitable for modelling increased conditional variability. I will also discuss some simple model diagnostics based on the shape of the autocorrelation function (ACF) and some important properties of these models.

"Nonparametric methods for Irregularly Spaced Non-Gaussian Spatial Data Analysis"

I consider a class of non-Gaussian spatial processes on an Euclidean space that are discretely sampled under a stochastic design giving a finite set of irregularly spaced spatial data values. Distributional properties of the commonly used tests and estimators exhibit very nonstandard behavior due to complex interactions among (a) the irregular spacings of the data locations, (b) relative growth rates of the sampling region and the sample size, and (c) the inherent dependence structure of the underlying spatial process. I will describe some recent advances on nonparametric inference techniques that are useful for analyzing such complex dependent data. Specifically, I will describe some Spatial Bootstrap and Spatial Empirical Likelihood methodology that can be used for inference in this context, without assuming that the spatial process is Gaussian.

Tom Loredo **Eric Feigelson**
Cornell University Pennsylvania State University

"Science and Statistics in Time Domain Astronomy and Cosmic Demographics"

We provide overviews of motivating science and important statistical issues in two areas of astronomy that underlie many planned working areas for the SAMSI ASTRO program: time domain astronomy (TDA), and cosmic demographics. TDA studies the variable sky — changes in the brightness, position, and other properties of astronomical sources with time. We focus on brightness time series — light curves — and survey the wide variety of variability astronomers study, including stochastic, periodic, and transient variability, over time scales from milliseconds to centuries. Important statistical challenges include handling irregularly sampled data, accounting for heteroscedastic measurement error, and modeling complex variability, potentially requiring non-Gaussian and non-stationary models. Cosmic demographics is concerned with classifying and characterizing astronomical objects from a population perspective. Heteroscedastic measurement error is a major challenge here, too, as is accounting for selection effects. Current and forthcoming synoptic time domain surveys are generating large datasets comprised of populations of light curves. The resulting collision of TDA and cosmic demographics raises new data analysis challenges, novel in both complexity and scale, that we briefly highlight.

Ashish Mahabal
Caltech

"Transients: Gotta catch 'em all"

There is a huge variety in surveys and the parameter space that they cover. The same is true for follow-up resources. We will describe several surveys, follow-up resources and the parameter space for transients. The main aim is to raise questions that will lead to good discussions and work during the year-long working groups.

Kaisey Mandel
Harvard–Smithsonian Center for Astrophysics

"Hierarchical Bayesian Models for Type Ia Supernova Cosmology"

The accelerating expansion of the Universe was discovered by astronomers analyzing optical observations of the brightness time series (light curves) of faraway exploding stars (Type Ia supernovae) to determine cosmological distances. Current and future optical supernova surveys aim to determine the physical nature of the mysterious dark energy driving the acceleration. Supernova distances can be effectively determined by modeling the empirical relations between their peak luminosities, colors, and the shape of their light curves (time series) over multiple wavelengths. However, their accuracy is systematically limited by the dimming of their intrinsic luminosities by dust in the host galaxies of the supernovae, which makes them appear farther away. Hierarchical Bayesian models are very useful constructs for accounting for the latent physical mechanisms underlying the observed data while fitting for the properties of the population as well as the constituent individuals. I have constructed hierarchical Bayesian models for Type Ia supernova data incorporating multiple random effects and uncertainties, including host galaxy dust, measurement error, and intrinsic supernova variations across time and wavelength. I will describe applications of these models to improve the precision and accuracy of supernova distance estimates by (1) understanding the joint effects of intrinsic variations and dust, (2) incorporating infrared data, which is more transparent to dust, and (3) leveraging spectroscopic information to better constrain the intrinsic supernova colors.

Paul Robertson
Pennsylvania State University

"Separating Stellar Activity from Doppler Shifts"

In the advent of ultra-precise spectrometers designed exclusively to make Doppler (or radial velocity) measurements of nearby stars, instrumental measurement noise is no longer the chief impediment to the discovery of Earth-mass exoplanets in the Solar neighborhood. Instead, it is the astrophysical noise arising from the atmospheric motions of the stars themselves. If we are to make confident detections of Earthlike exoplanets with reasonable allotments of telescope time, we must make progress towards the efficient separation of radial velocity signals originating from stellar activity and true Doppler motion. I will describe the basic challenges of extracting exoplanet signals from astrophysical noise, and highlight some of the astrophysical phenomena that give rise to Doppler noise. I will also discuss specific examples of activity-induced noise in radial velocity studies, and some of the techniques used to minimize its impact.

Jeff Scargle
NASA

"Time Series Explorer: Bayesian Blocks with Generalized Profiles and in Higher Dimensions"

The Time Series Explorer (TSE) is a project aimed at provided new and advanced time series analysis algorithms in two forms: a tool kit and an automated pipeline applying selected tools in machine learning settings. I will present a sketch of TSE with emphasis on time-domain modeling in general and recent improvements of the Bayesian Block (BB) algorithm in particular. This includes generalizing the shape of the elementary blocks from the current constant-rate model to general shapes, such as two-sided exponentials. Related topics will include extension of BB to higher dimensions and a novel way to detect and characterize short time-scale bursts in time-tagged event data. The Fermi Gamma Ray Space Telescope light curve for the Crab Nebula will be used as an example for all of the algorithms discussed. This work is in collaboration with Tom Loredo.

David Stenning
SAMSI

"Using Bayesian Computing to Solve a Complex Problem in Astrophysics"

Computer models are becoming increasingly prevalent in a variety of scientific settings; these models pose challenges because the resulting likelihood function cannot be directly evaluated. For example, astrophysicists develop computer models that predict the photometric magnitudes of a star as a function of input parameters such as age and chemical composition. A goal is to use such models to derive the physical properties of globular clusters—gravitationally bound collections of up to millions of stars. Recent observations from the Hubble Space Telescope provide evidence that globular clusters host multiple stellar populations, with stars belonging to the same population sharing certain physical properties. We embed physics-based computer models into a statistical likelihood function that assumes a hierarchical structuring of the parameters in which physical properties are common to (i) the cluster as a whole, or to (ii) individual populations within a cluster, or are unique to (iii) individual stars. A Bayesian approach is adopted for model fitting, and we devise an adaptive MCMC scheme that greatly improves convergence relative to its precursor, non-adaptive MCMC algorithm. Our method constitutes a major advance over standard practice, which involves fitting single computer models by comparing their predictions with one or more two-dimensional projections of the data.

Hyungsuk Tak
SAMSI

"Bayesian Estimates of Time Delays between Gravitationally Lensed Stochastic Light Curves and Time Delay Challenge II"

The gravitational field of a galaxy can act as a lens and deflect the light emitted by a more distant object such as a quasar. If the galaxy is a strong gravitational lens, it can produce multiple images of the same quasar in the sky. Since the light in each gravitationally lensed image traverses a different path length from the quasar to the Earth, fluctuations in the source brightness are observed in the several images at different times. The time delay between these fluctuations can be used to constrain cosmological parameters, e.g., Ho, and can be inferred from the time series of brightness data or light curves of each image. To estimate the time delay, we construct a model based on a state-space representation for irregularly observed time series generated by a latent continuous-time damped random-walk process. We account for microlensing, an additional source of independent long-term extrinsic variability, via a polynomial regression. Our Bayesian strategy adopts a Gibbs sampler. We introduce a profile likelihood of the time delay as an approximation of its marginal posterior distribution. The Bayesian and profile likelihood approaches complement each other, producing almost identical results. We demonstrate our estimation strategy and describe the second Time Delay Challenge using simulated data of a doubly-lensed quasar, and observed data from quasars J1029+2623 and Q0957+561.

John Veitch
University of Birmingham

"Gravitational Wave Astronomy with Ground-based Detectors"

The detection by LIGO of gravitational waves from coalescing binary black holes was a milestone discovery in physics and astronomy, representing a decades-long effort to build and operate the ground-based gravitational wave observatories. These instruments are among the most sensitive measuring devices ever constructed, but as well as detecting gravitational waves their data is polluted with a multitude of environmental and fundamental noises. Analyzing this noisy data is one of the main challenges in detecting the very weak gravitational wave signals, and the proper interpretation of the data requires a good understanding of the noise in order to make astrophysical inferences. I will review the data analysis strategies used by the LIGO Scientific Collaboration to make the first detections, and highlight areas where future detections can benefit from new developments in data analysis and statistical modelling.

Angie Wolfgang
Pennsylvania State University

"Hierarchical Modeling of Exoplanet Populations"

The last few years have marked a substantial increase in the application of hierarchical models to populations of extrasolar planets. I will overview the science driving this hierarchical modeling, survey the numerical algorithms which are commonly used to evaluate these models, and outline the common elements of the hierarchical models themselves that are currently in the exoplanet literature. Building on this recent history, I will then identify future directions for hierarchical modeling of exoplanet populations, with particular attention to the challenges that will be encountered during this more advanced modeling. These challenges provide an excellent starting point for discussion and could guide the formation of ASTRO SAMSI working groups.

Robert Wolpert
Duke University

"Inference for non-Stationary, non-Gaussian, Irregularly-Sampled Processes"

Most statistical time series courses emphasize inference for regularly-sampled continuously-distributed data, often with an implicit assumption of Gaussianity (through use of least-squares methods) and an explicit assumption of stationarity. At the frontiers of science, data bearing on the shakiest pillars of our scientific understanding are usually sparse, irregularly sampled, sometimes featuring extremes, and often presented as binned counts. The Central Limit Theorem does not apply, and the methods taught in our statistics courses are not applicable. This talk presents a few methods and examples for addressing and making inference for these data. The key ideas are: 1. Even if data are sampled at discrete times, it   can be useful to model the underlying phenomena in continuous time to overcome the problem of irregularly-sampled data; 2. Poisson, negative Binomial, alpha Stable,and other Infinitely-Divisible Distributions offer a rich alternative to the Gaussian for modeling and inference, with some exciting new features; 3. Stationary processes can be used as building blocks or starting points for modeling non-stationary phenomena.