International Collaboration in Nuclear Theory

BAYESIAN WORK GROUP REPORT

Application of Bayesian Statistics to Nuclear Physics

Extracting Bulk Properties of Neutron-Rich Matter with Transport Models in Bayesian Perspective

FRIB-MSU, East Lansing, Michigan, USA

March 22 - April 12, 2017

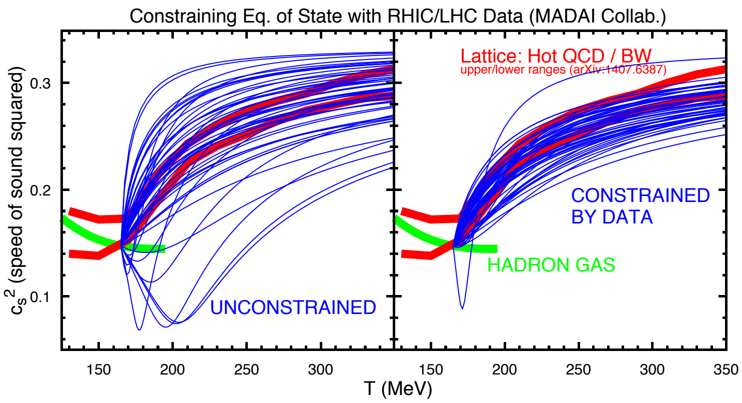
Like many fields of science, nuclear physics faces the daunting challenge of addressing critical questions by comparing large heterogenous data sets to complex and numerically intensive calculations. One oft-stated and ambitious goal of nuclear physics is to determine bulk properties of nuclear matter. As a function of temperature, baryon density and isospin composition, one wishes to know bulk properties such as the pressure, viscosities, conductivities, and neutrino absorption rates, along with surface properties. This requires the comparison of observations from heavy-ion collisions, astronomical observations, nuclear structure trends, and nuclear scattering data to nuclear structure calculations, heavy-ion simulations, nuclear reaction theory, and models of stellar evolution.

Each subfield mentioned above requires the consideration of a heterogenous set of observables. For example, super-nova observations include cooling curves, spectroscopy, rotation curves, and hopefully, neutrino measurements. For each field the disparate data sets are matched by an equivalently diverse set of theoretical models. Each model uses numerous unknown parameters, typically a few dozen, to cover the range of possible outcomes. Many of these parameters encapsulate the equation of state and other bulk properties. Thus, the success of these scientific endeavors depends critically on describing the likelihood of a highly multi-dimensional parameter space by comparing the model output to a large number of observed quantities. This is a classical problem in statistical inference, with the twist that the expense of the theoretical models makes it difficult to explore the enormous parameter space.

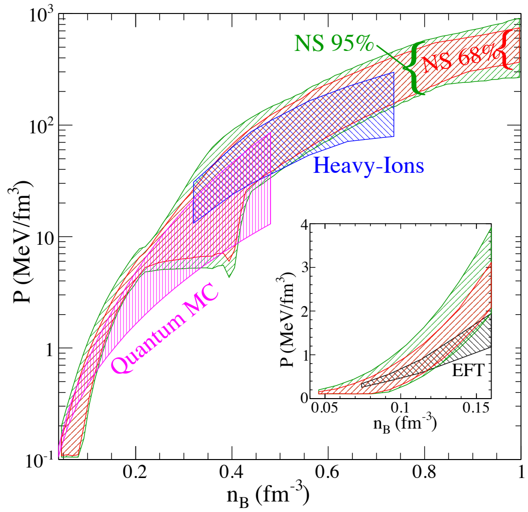
The ICNT program, along with the TRANSPORT 2017 Workshop, featured several speakers working on statistical inference:

Earl Lawrence	statistics	Los Alamos National Laboratory
Scott Pratt	nuclear theory	Michigan State University
Yingxun Zhang	nuclear theory	China Institute of Atomic Energy
Michael Grosskopf	statistics	Simon Fraser University
Pawel Danielewicz	nuclear theory	Michigan State University
Genie Jhang	nuclear experiment	Michigan State University
Pierre Morfouace	nuclear experiment	Michigan State University
Jorge Piekarewicz	nuclear theory	Florida State University
Andrew Steiner	theoretical nuclear astrophysics	University of Tennessee
Francesca Gulminelli	nuclear theory	LPC Caen

Applications of Bayesian statistical inference were presented for determinations of the equation of state for high-temperature, zero baryon density matter (Pratt), and zero-temperature neutron-rich matter (Zhang, Danielewicz, Piekarewicz, Steiner, Gulminelli) below and above normal nuclear density. Figures 1-3 display results from this workshop, which were all made possible by sophisticated multi-dimensional Bayesian statistical analysis. For the RHIC/LHC analysis uncertainties in the initial state were critical, and for neutron matter, surface effects were of special concern. The extrapolative power of both microscopic and parametric (Skyrme) models was discussed.



1. Eq. of state for high-$\boldsymbol{T}$ zero-density of matter from Bayesian analysis (S. Pratt)



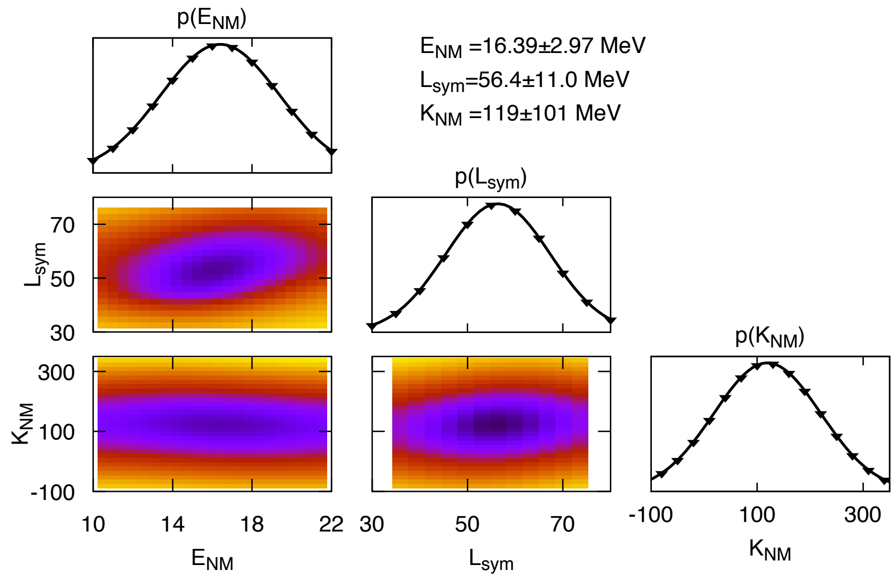
2. Eq. of state for neutron matter from neutron star observations (A. Steiner)

There were remarkable similarities between the problems encountered across all the disciplines covered here. There were also common questions concerning the statistical methods that repeatedly emerged from the other participants of the workshop. Some common challenges faced by those performing the statistical procedures were:

1. How does one state the uncertainties for a data set with many observables? Should one use error matrices? nuisance parameters?...
2. How should shortcomings of the theoretical models translate into uncertainties?
3. Most approaches applied model emulators to reduce the numerical cost of running the full model during the Markov-chain Monte Carlo explorations of parameter space. How can one assess the validity and accuracy of the emulator?
4. How should one incorporate prior knowledge of the parameter set, e.g. from previous analyses, into the procedure?
5. How can one handle the comparison of distinct theoretical models, with different parameter sets?

Beginning practitioners, their collaborators who wish to follow their work, and other participants repeatedly asked similar questions over the three-week program.

1. How do Markov-chain Monte Carlo procedures express the likelihood? Are they reliable?
2. What is Bayes theorem? – and why do we call these analyses "*Bayesian*"?
3. What is an emulator? How is it constructed and tuned?
4. What are Gaussian processes? Why does one use the assumption of Gaussian processes to construct a model emulator?



3. Parameters describing isospin dependence of eq. of state (F. Gulminelli)

Action Plan:

The similarity of the issues encountered by the work presented at this workshop, plus the similarity of these problems to other analyses encountered in nuclear physics, underscore the need for a community resource to educate nuclear physicists to Bayesian applications for statistical inference in nuclear physics. An on-line forum that connects members in the community that have faced, or are currently facing, similar problems, could be invaluable. Although answers to these questions can typically be extracted from the published scientific literature, or from text-books, establishing a community resource with these questions addressed in one easy-to-access location, and in language nuclear physicists understand, would significantly lower the barrier for new researchers to engage in state-of-the-art statistical treatments. The on-line resource could include two main components:

1. A primer for statistical analysis, with sections describing Markov-chain Monte Carlo, Bayes theorem, principal component analysis, model emulation, Gaussian processes, strategies for uncertainty assignment, etc
2. An interactive community blog related to these analyses. This would include discussion boards and a frequently-asked-questions section.

The workshop strongly endorsed this plan and believes it would be provident for FRIB and the NSCL to support the establishment and maintenance of such a resource. Such a resource would catalyze the ongoing transformation in the analysis, evaluation and interpretation of complex twenty-first-century scientific data sets using state-of-the-art computational modeling.