White Paper: Bayesian Methods in Nuclear Physics

Edited by R.J. Furnstahl, D. Higdon, N. Schunck, and A.W. Steiner

October 17, 2016

Abstract

The INT program on "Bayesian Methods for Nuclear Physics"

Contributing Authors:

Karim Bennaceur
Institut de Physique Nucleaire de Lyon

Derek Bingham
Statistics and Actuarial Science
Simon Fraser University

Andreas Ekström
Department of Fundamental Physics
Chalmers University of Technology

Dick Furnstahl
Department of Physics
Ohio State University

Harald Griesshammer
Department of Physics
George Washington University

Michael Grosskopf
Statistics and Actuarial Science
Simon Fraser University

Dave Higdon
Biocomplexity Institute
Virginia Tech

Natalie Klco
Department of Physics
University of Washington

Witold Nazarewicz
Department of Physics
Michigan State University

Daniel Phillips
Department of Physics & Astronomy
Ohio University

Nicholas Schunck
Nuclear and Chemical Science Division
Lawrence Livermore National Laboratory

Andrew W. Steiner
Department of Physics and Astronomy
University of Tennessee / ORNL

Kyle Wendt
Institut für Kernphysik
Technische Universität Darmstadt

Sarah Wesolowski
Department of Physics
Ohio State University

.

Contents

1 Overview **5**
- 1.1 Goals . . . 6
- 1.2 Questions . . . 6

2 Background **6**
- 2.1 Basics of Bayesian statistics . . . 6
- 2.2 Conceptual or philosophical issues . . . 7

3 Themes **7**
- 3.1 Likelihood function and Bayes theorem . . . 7
- 3.2 Gaussian Process (GP) as a tool . . . 8
- 3.3 Discrepant functions / random effects . . . 9
- 3.4 χ^2 and dofs . . . 9
- 3.5 Pitfalls . . . 9
- 3.6 Model selection / comparison (or metrics) . . . 9
- 3.7 Selection of priors . . . 9
- 3.8 Use of MCMC . . . 10
- 3.9 Specific issues . . . 11
- 3.10 Understanding of model function . . . 11

A Guide to the literature (written and software) **13**
- A.1 General Bayesian statistics . . . 13
- A.2 Gaussian Process (GP) models . . . 13
- A.3 Hypothesis Testing and Model Comparison . . . 14

B Summary of notation and terminology **14**
- B.1 Lexicon . . . 14

C Using git and github **14**

References **15**

Building this document

[*This subsection is only part of the working document, not the final version.*]

- Editors are the organizers of the program (Furnstahl, Higdon, Schunck, Steiner). In this capacity they are to provide a coherent narrative and make the final decisions on what content is included.
- Contributors are self-selected participants in the program. They might be designated the Bayes-INT Working Group.

Guidelines for building the document:

- Do not make firm decisions on the order/organization of the content until later, to make it flexible.
- Possible LaTeX formats:
  - We will build the document with a convenient structure (e.g., with a table of contents) and later decide on a publication plan.
  - The Journal of Physics G LaTeX format is a possibility because this journal is a likely target for a published version.
  - To get started, we will borrow a format from an existing white paper.
- Organization of files
  - Use Git, hosted on Github. Each contributor can make their own branch, and resolution of conflicts from merging will be made by the Editors.
  - Standardized list of macros in macros.tex. Use a macro whenever possible to facilitate developing a standardized notation, which can be easily modified.
  - Use BibTeX for references (newer version?). Include the titles in the style to more easily identify sources.
  - For early passes through new materials, use bullets (or enumerated lists) to enter content. This makes it modular and easy to adjust the ordering and structure.
- Figures
  - Plan to use original figures for schematic illustration as much as possible
  - But point specifically to figures in the literature
- Associated computer codes. We can include as part of the supplementary material Mathematica and Ipython notebooks.

1 Overview

With the maturation of calculational methods such as lattice QCD for hadronic physics, ab initio, shell model and density functional theory approaches for nuclear structure and reactions (with applications to astrophysics and fundamental symmetries), and viscous hydrodynamic modeling of relativistic heavy-ion collisions, nuclear theory is entering an era of precision calculations. In all these various subfields of nuclear physics, however, theories always depend on several free parameters that are not defined within the theory and must be adjusted by comparing model predictions to experimental data. In addition, experimental data may suffer from weak statistics and poor signal-to-noise ratio, or may have been extracted through a model-dependent procedure. Such issues lead to increased demand for sophisticated uncertainty quantification, to effectively interface with, inform, and analyze experiments. Although the methods used to quantify errors are often based on frequentist statistical analysis, Bayesian methods have recently become increasingly popular.

Bayesian statistics is a well-developed field, although it has not been part of the traditional education of nuclear theorists. In schematic form, Bayesian statistics treats the parameters or the model/theory as genuine random variables. It then uses Bayes theorem of probabilities to provide a recipe to compute their probability distribution (the "posterior") in terms of prior information (e.g., about the data) and a likelihood function. For applications to fitting ("parameter estimation"), the posterior lets us infer, given the data we have measured, the most probable values of the parameters and predict values of observables with confidence intervals. Other applications involve deciding between alternative explanations or parameterizations ("model selection"). In practice, there are pitfalls in the implementation of this formalism and it is often a computationally hard problem.

Interest in Bayesian statistics has increased significantly in the past 10 years. The wide availability of large-scale computing resources has made the computation of the integrals needed for Bayesian inference easier. Modern experimental and observational facilities generate large amounts of data, often best analyzed and characterized through Bayesian methods. Bayesian methods are often preferred for under-constrained fits and inverse convolutions. In nuclear science, Bayesian methods have found their way into such areas as nuclear data, lattice QCD, dense matter, effective field theory, nuclear reactions, and parton distribution functions. These sub-fields have generally turned to Bayesian inference methods independently and in some cases without access to expert advice and guidance from professional statisticians.

The INT program on "Bayesian Methods for Nuclear Physics" brought together statisticians and nuclear practitioners, principally theorists, to explore how Bayesian inference could enable progress on the frontiers of nuclear physics and open up new directions for the field. The program also served as ISNET-4, the fourth meeting in a series helping researchers bridge the gap between experiment and theory (ISNET stands for Information and Statistics in Nuclear Experiment and Theory). The goal of this whitepaper is to summarize the topics that have been discussed during the program and serve as a reference for applications of Bayesian statistics in nuclear theory.

1.1 Goals

The Bayesian approach to statistics is not necessarily a part of the curriculum of physics studies. The first goal of the workshop was to provide the opportunity for nuclear physicists unfamiliar with Bayesian methods with an in-depth introduction to Bayesian methods by experts in the field. In this document, we pay special attention to presenting Bayesian methods in a language understandable by most nuclear physicists, and with examples taken from actual applications – mostly in nuclear theory.

Contrariwise, nuclear physics is a source of non-trivial problems for experts in Bayesian statistics. Large datasets, models that are by definition incorrect, non-linearities of the models, computationally-intensive models present opportunities to test statistical methods in non-traditional settings. A second goal of the program was to facilitate cross communication, fertilization, and collaboration on Bayesian applications among the nuclear sub-fields.

One of the reasons Bayesian methods have gained popularity is the increased availability of large-scale computing capabilities. In practice, constructing the Bayesian posterior distribution often comes down to Markov Chain Monte Carlo simulations of the model. When the model is too time- or resource-consuming to run, it is sometimes possible to build an emulator and treat the parameters of the emulator on the same footing as the parameters of the original model, i.e., obtain a full estimate of uncertainties of both model and emulator parameters. Nuclear theory presents several case studies for such approaches, and a third goal of the program was to learn from the experts about advanced computational tools and methods.

1.2 Questions

In the following subsections are (partial) lists of questions that were considered during the INT program.

- What are Bayesian and Frequentist approaches and how different are they?
- What are the advantages/pitfalls of Bayesian approaches in nuclear physics?
- What is the best literature (for physicists) on Bayesian approaches?
- Does nuclear physics present opportunities for "Bayesists" to advance their field?

2 Background

[*It is not yet clear how to organize the subsections, which currently overlap significantly.*]

2.1 Basics of Bayesian statistics

- Not an exhaustive recounting, but the minimum needed to make sense of the review.
- Pointers to the literature (which is listed later).

2.2 Conceptual or philosophical issues

- In comparison to frequentist approaches, the emphasis is on sampling of posteriors rather than optimization (e.g., finding the maximum of the likelihood).
- One of the essential, philosophical, difference between Bayesian and frequentist approach is the emphasis put on prior knowledge. Consider the textbook example of coin tossing: *If* the toss is perfectly random and the coin is perfect, *then* the probability P of next head or tail is 0.5 *irrespective of the history* of all previous tosses. For example, there could already have been one million tosses that all gave head: in a strict frequentist approach, the probability of observing head in the next toss would remain 0.5 exactly. I suppose we could introduce bias in our statistical model. By contrast, Bayesian statistics provides a very natural framework to relax this strict interpretation. In our example, putting a (non-uniform) prior distribution to reflect the observation of these one millions heads implies that we accept that the coin may not be perfect and/or that the tosses were not truly random. As a consequence, the probability of observing a next head can be $\neq 0.5$.
- In a scientific context, the notion of prior knowledge makes the Bayesian approach very pertinent, since there is almost always something we know about our model/experiment.
- Younger generation statisticians do not have the baggage of the frequentist-Bayesian wars; they freely use both approaches as tools to do statistical analysis.
- Is exchangability a purely formal aspect or an important consideration for practical purposes?

3 Themes

3.1 Likelihood function and Bayes theorem

- Recall Bayes theorem introduced (probably) in sec. 2.1
- Posing the problem (and the notations): data = model + random error

$$y(\boldsymbol{x}) = \eta(\boldsymbol{x}) + \epsilon \tag{1}$$

  with η a model that depends on parameters $\boldsymbol{x}$, and ϵ the deviation between the output $\eta(\boldsymbol{x})$ of the model (a prediction) and the experimental data $y(\boldsymbol{x})$
- Define the likelihood function in both simple (1D) and advanced (ND) case
- Everything else follows: GP + MCMC to actually determine x, variations on the same theme for model defects ($\epsilon \to \epsilon + \delta(x)$ and augmented χ^2, choice of prior for x, etc.
- Example: nuclear forces $\hat{V}_{NN}$, $\hat{V}_{NNN}$, etc., are characterized by a set of parameters $\boldsymbol{x} = (x_1, \ldots, x_N)$. The predictive power of any quantum many-body method, i.e., of a given model η, to reproduce nuclear properties y such as masses, radii, separation

energies, transition rates, etc., rely on our ability to obtain a reliable parametrization $\boldsymbol{x}$ which minimizes the deviation ϵ.

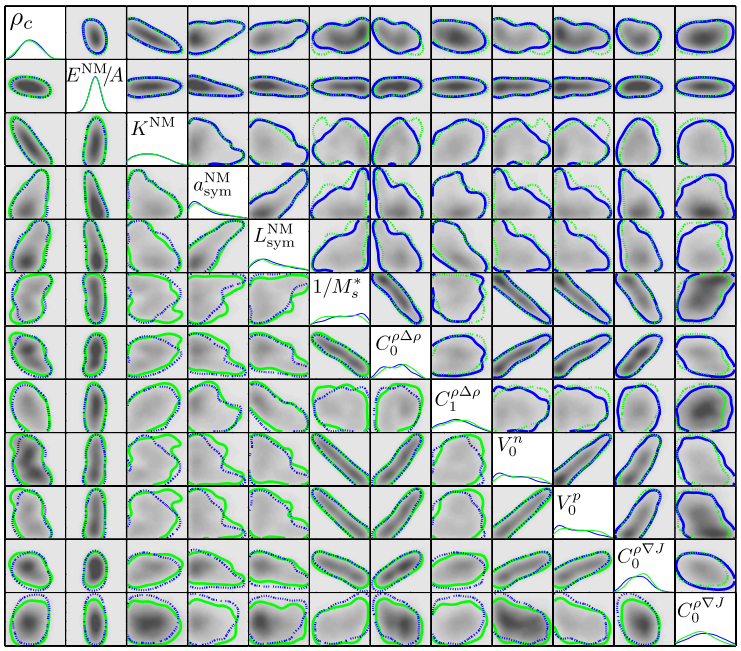

Figure 1: (Color online) Univariate and bivariate marginal estimates of the posterior distribution for the 12-dimensional DFT parameter vector of the UNEDF1 parameterization of the Skyrme energy functional. The blue lines enclose an estimated 95% region for the posterior distribution found when only the original UNEDF1 data are accounted for; the green-outlined regions represent the same region for the posterior distribution found when additional data on recent mass measurements are also included. For the range of parameter variation, see [1].

3.2 Gaussian Process (GP) as a tool

- GP simulates correlations between parameters
- emulators
- discrepant functions / random effects
- linear regression $\rightarrow$ GP
- limitations: bad at emulating waveforms (Lackey)

3.3 Discrepant functions / random effects

- Model discrepancy: mock up what we know about the intrinsic limitations of our model into the posterior
- data = model + truly random error + systematic bias

$$y(x) = \eta(x) + \epsilon + \delta(x)$$

- Example of nuclear masses: deviations near closed shells (particle-vibration coupling, pairing phase transition), at $N = Z$ (Wigner energy), in transitional nuclei (shape coexistence) are known but sometimes very hard (=computationally) to model
- Using model discrepancies allow "professional" fitting of a model at some approximation while resulting parametrization is valid at higher resolution/order. Ex.: fit at SR-EDF level, but parameters hopefully valid at MR-EDF
- Avoid overfitting

3.4 χ^2 and dofs

- χ^2/dof is not meaningful for Bayesian statistics
- dofs when there are priors
- how should one count dofs?
- augmented χ^2: distance from the prior (Bartek's talk?)
- When does χ^2/dof make sense?
- AWS: Bayes' factors and Occam's razor

3.5 Pitfalls

- using same data to estimate priors and determine uncertainty

3.6 Model selection / comparison (or metrics)

- laundry list of approaches (see Vera's talk)
- new approach for nuclear physics: mixture models (supermodels)

3.7 Selection of priors

- uniform vs. where it makes a difference
- knowledge of underlying model
- running against the boundary
- AWS: Jeffrey's priors and invariance with respect to parameter transformations

3.8 Use of MCMC

Monte Carlo integration methods are based on the idea that, if one has a way to generate a random deviate from a probability distribution defined by the integrand, then one can replace the integral by a sum (the same formalism applies trivially to multi-dimensional integrals)

$$\int_a^b f(x)dx \approx \frac{(b-a)}{N}\sum_i p_i \tag{2}$$

where p_i is a list of N random numbers selected from the probability distribution $f(x)$ with $x \in [a, b]$. Monte Carlo is typically an efficient method of integration over direct integration for problems with large dimensionality. In some cases, selecting a random deviate from $f(x)$ is not straightforward. This problem can be handled by importance sampling: one decomposes $f(x) = g(x)h(x)$ where $g(x)$ is slowly varying with x and $h(x)$ is a function which one can more easily sample random deviates from.

Markov chain Monte Carlo (MCMC) is a type of importance sampling where one generates random deviates from $h(x)$ by generating a "Markov chain", a set of numbers whose distribution asymptotically approaches $h(x)$. A common approach to generating a Markov chain is the Metropolis algorithm, but several other methods exist. The key to the Metropolis algorithm is ergodicity, and this property guarantees that values in the chain indeed converge to the desired distributions[1].

Because points in the Metropolis algorithm are often chosen based on a random step (typically with a fixed maximum step size) from the previous point, adjacent points in the chain are not statistically independent. These autocorrelations are the scourge of MCMC methods. There are two principal methods [2] for overcoming these autocorrelations: (i) "thinning" the chain and selecting only those values which are farther apart than the autocorrelation length, and (ii) block averaging over blocks which have a size larger than the autocorrelation length. Autocorrelations can often be decreased by increasing the maximum stepsize, but this comes at a cost of more Metropolis rejections. The optimal step size is that which minimizes the computational time between statistically independent points in the chain [3].

For a one-dimensional function, computing the autocorrelation length can be done directly using the

$$C(k) = \frac{1}{(N-k)}\sum_i^{N-k}(x_i - \bar{x})(x_{i+k} - \bar{x}) \tag{3}$$

the autocorrelation length, $\hat{k}$, is smallest value of k for which $C(k)$ is zero (within the noise limits set by the size of the data set). The quantity $C(k)$ is often normalized by dividing by $C(0)$, which is just the variance. The Kubo formula leads to an estimate of the autocorrelation length

$$\hat{k} = 1 + 2\sum_{k=1}^{M}\frac{C(k)}{C(0)} \tag{4}$$

[1] When coding the Metropolis algorithm, we have found it relatively easy to make subtle errors which violate ergodicity, so the reader is warned to test their code.

where $M < N$ ensures that the sum only includes the values of $C(k)$ not dominated by noise. There is no clear best procedure for choosing M. One method is to choose a relatively small value, then increase M and keep increasing until $M \geq 5\hat{k}$. For multidimensional functions, it is important to note that the autocorrelation lengths of a function's parameters can be wildly different. In this case, the autocorrelation length of the chain is the largest autocorrelation length among all of the parameters.

MCMC is particularly useful in Bayesian inference because one is often required to compute many integrals based on the same kernel, e.g.

$$I_1 = \int g_1 h d\vec{x} \quad ; \quad I_2 = \int g_2 h d\vec{x} \quad ; \quad \ldots \tag{5}$$

where the g_i are slowly varying functions. In this case, MCMC methods generate a Markov chain for h, and then evaluate the functions g_i for each (statistically independent) element in the chain. Even if the function g_1 is difficult to compute, if the functions $g_{i,i\neq 1}$ are easy to compute given the result of g_1, there is a significant time savings to computing all of the integrals at one time.

[AWS: I'll talk about Metropolis-Hastings here.] One of the most significant recent improvements in the Metropolis is the use of affine-invariant MCMC [**?**]. This method uses several individual MCMC walkers and steps are generated by modifing the distance between a walker and the rest of the ensemble of walkers. Over the course of the simulation, the ensemble of walkers self-organize along the correlated directions in parameter space near the object function's maximum value.

- must be able to evaluate quickly enough
- fast model vs. good emulator vs. slow function (dictates what you do)
- burn-in as usual; is checking autocorrelation and skipping critical?
- nested sampling?
- AWS: MCMC pitfalls: Skipping points when the likelihood function is zero (bad idea)

3.9 Specific issues

- Instabilities in parameter space
  - undefined χ^2
  - diagnostics (e.g., response function)
  - alternatives?
- Bayesian inference for functions rather than for parameters
- AWS: Normalization of conditional probabilities?

3.10 Understanding of model function

- ???

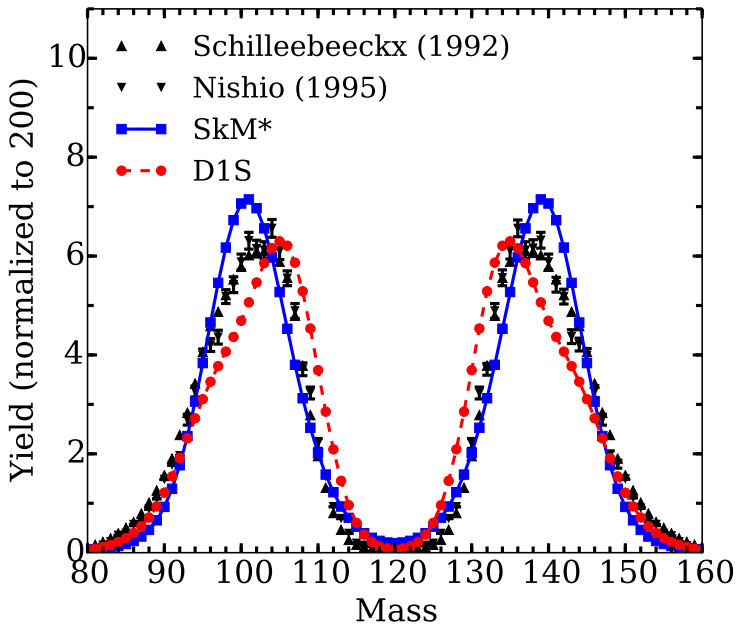

Figure 2: (Color online) Pre-neutron mass yields for ^{239}(n,f). The SkM* and D1S calculations are compared with two experimental datasets [**?**, **?**]. The data from Nishio are plotted with their statistical uncertainties. Reproduced from [4].

A Guide to the literature (written and software)

- The vastness of the literature on statistical methods poses a problem for the nuclear physicist: there is not an adequate sampling algorithm. This is compounded by unfamiliar notation and terminology used by the statistics community, as well as conceptual differences in the formulation of problems.
- Here we propose a sampling of books and articles.

A.1 General Bayesian statistics

The following are often recommended as introductory guides for physicists, because the examples and language are drawn from physics:

- *Data Analysis: A Bayesian Tutorial* by Devinderjit Sivia and John Skilling [5].
- *Bayesian Logical Data Analysis for the Physical Sciences* by Phil Gregory [6].
- *Bayes in the sky: Bayesian inference and model selection in cosmology* by Roberto Trotta [7].

Recommended more advanced texts include:

- *Bayesian Data Analysis* by Gelman et al. [8].

A.2 Gaussian Process (GP) models

Software:

- The "Sample Size Calculations for Computer Experiments" app provides a sandbox for playing with Gaussian processes. Read the "About" tab first; it includes definitions of parameters used in the GP "correlation family" (which define the covariance). The "Sample Path Plots" tab has an app that lets you draw samples of functions that depend on user selected parameters (actually "hyperparameters") that specify the details of the correlation family.
- Dan Foreman-Mackey's Gaussian Process Python Library and Tutorial steps you through using GPs with emcee.

References:

- *Efficient Global Optimization of Expensive Black-Box Functions*, by Donald Jones, Matthias Schonlau, and William Welch [9].
- *A Bayesian approach for parameter estimation and prediction using a computationally intensive model* by Dave Higdon et al. [10].
- *Gaussian Processes for Machine Learning*, by Carl Edward Rasmussen and Christopher K. I. Williams. The full text is available (free!) as a pdf.
- *Fast Direct Methods for Gaussian Processes*, by Sivaram Ambikasaran, Daniel Foreman-Mackey, Leslie Greengard, David W. Hogg, and Michael O'Neil [11].

A.3 Hypothesis Testing and Model Comparison

B Summary of notation and terminology

B.1 Lexicon

hyperparameter This is ...

C Using git and github

[*This section is just for building the white paper. It will not be included in the production version.*]

References

[1] J. D. McDonnell, N. Schunck, D. Higdon, J. Sarich, S. M. Wild, and W. Nazarewicz, *Uncertainty Quantification for Nuclear Density Functional Theory and Information Content of New Measurements*, Phys. Rev. Lett. **114** (2015) no. 12, 122501. 8

[2] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*. Oxford Science Publications, 1989. 10

[3] A. D. Sokal, *Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms*. 1980. 10

[4] D. Regnier, M. Verrière, N. Dubray, and N. Schunck, *FELIX-1.0: A finite element solver for the time dependent generator coordinate method with the Gaussian overlap approximation*, Comput. Phys. Comm. **200** (2016) 350. 12

[5] D. Sivia and J. Skilling, *Data Analysis: A Bayesian Tutorial*. Oxford University Press, 2006. 13

[6] P. Gregory, *Bayesian Logical Data Analysis for the Physical Sciences*. Cambridge University Press, 2005. 13

[7] R. Trotta, *Bayes in the sky: Bayesian inference and model selection in cosmology*, Contemp. Phys. **49** (2008) 71–104, arXiv:0803.4089 [astro-ph]. 13

[8] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*. Chapman and Hall/CRC, 2003. 13

[9] D. R. Jones, M. Schonlau, and W. J. Welch, *Efficient Global Optimization of Expensive Black-Box Functions*, J. of Global Optimization **13** (1998) no. 4, 455–492. http://dx.doi.org/10.1023/A:1008306431147. 13

[10] D. Higdon, J. D. McDonnell, N. Schunck, J. Sarich, and S. M. Wild, *A Bayesian Approach for Parameter Estimation and Prediction using a Computationally Intensive Model*, J. Phys. G **42** (2015) no. 3, 034009, arXiv:1407.3017 [physics.data-an]. 13

[11] S. Ambikasaran, D. Foreman-Mackey, L. Greengard, D. W. Hogg, and M. O'Neil, *Fast Direct Methods for Gaussian Processes*, ArXiv e-prints (2014) , arXiv:1403.6015 [math.NA]. 13