Bayesian Machine Learning

Master of Science thesis project

Nov 30, 2019

Bayesian Machine Learning, Level Densities and Probability

The level density $\rho(E)$ as function of energy E plays a central in many physics applications, ranging from the modeling of nuclear astrophysics reactions central to the synthesis of the elements to the classification and understanding of phases and phase transition in for example condensed matter physics.

In statistical physics it defines the thermodynamical potential in the microcanonical ensemble and thereby the entropy. For a finite isolated many-body system (for example an atomic nucleus), the correct thermodynamical ensemble is the microcanonical one. In this ensemble, the nuclear level density, the density of eigenstates of a nucleus at a given excitation energy, is the important quantity that may be used to describe thermodynamic properties of nuclei, such as the nuclear entropy, specific heat, and temperature. Bethe first described the level density using a non-interacting fermi gas model for the nucleons. Modifications to this picture, such as the back-shifted fermi gas which includes pair and shell effects not present in Bethe's original formulation, are in wide use. The level density ρ defines the partition function for the microcanonical ensemble and the entropy through the well-known relation $S(E) = k_B ln(\rho(E))$. Here k_B is Boltzmann's constant and E is the energy. In the microcanonical ensemble, we could then extract expectation values for thermodynamical quantities like temperature T, or the heat capacity C. The temperature in the microcanonical ensemble is defined as

$$\langle T \rangle = \left(\frac{dS(E)}{dE} \right)^{-1} . \tag{1}$$

It is a function of the excitation energy, which is the relevant variable of interest in the microcanonical ensemble. However, since the extracted level density is given only at discrete energies, the calculation of expectation values like T, involving derivatives of the partition function, is not reliable unless a strong smoothing over energies is performed. Another possibility is to perform a transformation to the canonical ensemble. The partition function for the canonical ensemble is

related to that of the microcanonical ensemble through a Laplace transform

$$Z(\beta) = \int_0^\infty dE \rho(E) \exp\left(-\beta E\right). \tag{2}$$

Here we have defined $\beta = 1/k_B T$. Since we will deal with discrete energies, the Laplace transform of Eq. (2) takes the form

$$Z(\beta) = \sum_E \Delta E \rho(E) \exp\left(-\beta E\right), \tag{3}$$

where ΔE is the energy bin used. With Z we can evaluate the entropy in the canonical ensemble using the definition of the free energy

$$F(T) = -k_B T \ln Z(T) = \langle E(T) \rangle - TS(T). \tag{4}$$

Note that the temperature T is now the variable of interest and the energy E is given by the expectation value $\langle E \rangle$ as a function of T. Similarly, the entropy S is also a function of T. For finite systems, fluctuations in various expectation values can be large. In nuclear and solid state physics, thermal properties have mainly been studied in the canonical and grand-canonical ensemble. In order to obtain the level density, the inverse transformation

$$\rho(E) = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} d\beta Z(\beta) \exp\left(\beta E\right), \tag{5}$$

is normally used. Compared with Eq. (2), this transformation is rather difficult to perform since the integrand $\exp\left(\beta E + \ln Z(\beta)\right)$ is a rapidly varying function of the integration parameter. In order to obtain the density of states, approximations like the saddle-point method, viz., an expansion of the exponent in the integrand to second order around the equilibrium point and subsequent integration, have been used widely.

For the ideal Fermi gas, this gives the following density of states

$$\rho_{\text{ideal}}(E) = \frac{\exp\left(2\sqrt{aE}\right)}{E\sqrt{48}}, \tag{6}$$

where a in nuclear physics is a factor typically of the order $a = A/8$ with dimension MeV^{-1}, A being the mass number of a given nucleus.

To obtain an experimental level density is a rather hard task. Ideally, we would like an experiment to provide us with the level density as a function of excitation energy and thereby the *full* partition function for the microcanonical ensemble. It is only rather recently that experimentalists have been able to develop methods for extracting level densities at low spin from measured γ-spectra. These measurements were performed at the Oslo Cyclotron Laboratory.

The Oslo cyclotron group has developed a method to extract nuclear level densities at low spin from measured γ-ray spectra. The main advantage of utilizing γ-rays as a probe for level density is that the nuclear system is likely

thermalized prior to the γ-emission. In addition, the method allows for the simultaneous extraction of level density and γ-strength function over a wide energy region.

With the level density we can in turn define a probability distribution function (PDF) in say for example the canonical ensemble. Alternatively, if we have the PDF we can find the level density. Having a PDF allows us also to quantify in a rigorous way statistical confidence intervals, statistical errors and other statistical quantities with far reaching consequences for our understanding of a specific physics problem. In experiments we do however normally not have the above quantities. This means that we need to translate experimental results via some theoretical modeling into suitable quantities that can be used to define either a PDF or the density of states.

A typical situation which occurs in for example nuclear reaction experiments performed at the cyclotron of the University of Oslo, is that one can extract the number of counts as function of the excitation energy E_x of a given nucleus and the resulting gamma energy E_γ from Compton scattering. This quantity, labelled $N(E_x, E_\gamma)$ can in turn be used to define either a PDF or the density of state.

In this project we will use Bayesian statistics and Bayesian machine learning to extract first the PDF based on the above experimental data in order to define a posterior distribution $P(E_x|E_\gamma)$, that is the likelihood of being in a state with energy E_x given a certain γ-energy. This quantity will in turn be used to identify a density of states. A short note on Bayes' rule and some other elements of statistics are included at the end here.

Thesis Projects

The aim of this thesis project is to employ Bayesian machine learning to define a PDF, either from experiment or from theoretical simulations. Eventually, based on the PDF, one can attempt to define the level density $\rho(E)$. The first step is to use an already available model for extracting the level density from exact diagonalization. This model, a so-called simplified pairing model is described in detail in the references below. The data from these theoretical calculations will then be used to define a posterior distribution based on a Bayesian machine learning approach.

Specific tasks and milestones. The project can easily be split into several parts and form the basis for collaborations among several students. The milestones are as follows

1. Spring 2020: Use the simple pairing model to generate training data on the density of states from numerical diagonalization (existing code) and develop a Bayesian Neural Network code and algorithm to extract a PDF. This PDF expresses the likelihood for finding the system at a given energy.

3

2. Fall 2020: Based on the experience with the theoretical model, the next step is to use experimental data from the Oslo cyclotron (see discussions above) in order to extract $P(E_x|E_\gamma)$ using Bayesian Machine Learning.

3. Spring 2021: Analysis of results and determination of level density. Finalize thesis.

The thesis is expected to be handed in May/June 2021.

References.

1. **Pairing in nuclear systems: from neutron stars to finite nuclei**, DJ Dean, M Hjorth-Jensen, Reviews of Modern Physics 75, 607 (2003).

2. Morten Hjorth-Jensen, M.P. Lombardo and U. van Kolck, Volume **936**, (2017), see chapter 8

Appendix: Brief note on Bayesian Statistics.   The aim is to assess hypotheses by calculating their probabilities $p(H_i|\ldots)$ conditional on known and/or presumed information using the rules of probability theory. Bayes' theorem is based on the standard Probability Theory Axioms:

1. Product (AND) rule : $p(A, B|I) = p(A|I)p(B|A, I) = p(B|I)p(A|B, I)$. Should read $p(A, B|I)$ as the probability for propositions A AND B being true given that I is true.

2. Sum (OR) rule: $p(A + B|I) = p(A|I) + p(B|I) - p(A, B|I)$. $p(A + B|I)$ is the probability that proposition A OR B is true given that I is true.

3. Normalization: $p(A|I) + p(\bar{A}|I) = 1$. $\bar{A}$ denotes the proposition that A is false.

Bayes' theorem follows directly from the product rule

$$p(A|B, I) = \frac{p(B|A, I)p(A|I)}{p(B|I)}.$$

The importance of this property to data analysis becomes apparent if we replace A and B by hypothesis(H) and data(D):

$$p(H|D, I) = \frac{p(D|H, I)p(H|I)}{p(D|I)}. \tag{7}$$

The power of Bayes' theorem lies in the fact that it relates the quantity of interest, the probability that the hypothesis is true given the data, to the term we have a better chance of being able to assign, the probability that we would have observed the measured data if the hypothesis was true.

The various terms in Bayes' theorem have formal names.

- The quantity on the far right, $p(H|I)$, is called the *prior* probability; it represents our state of knowledge (or ignorance) about the truth of the hypothesis before we have analysed the current data.

- This is modified by the experimental measurements through $p(D|H, I)$, the *likelihood* function,

- The denominator $p(D|I)$ is called the *evidence.* It does not depend on the hypothesis and can be regarded as a normalization constant.

- Together, these yield the *posterior* probability, $p(H|D, I)$, representing our state of knowledge about the truth of the hypothesis in the light of the data.