# Machine Learning: 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 as

$$S(E) = -k_B \ln\left(\rho(E)\right),$$

and the partition function $Z(\beta)$ (with $\beta = 1/k_BT$) as

$$Z(\beta) = \int dE \exp\left(-\beta E\right)\rho(E),$$

and the expectation values of various moments of the energy as

$$\mathbb{E}^n(\beta) = \frac{\int dE E^n \exp\left(-\beta E\right)\rho(E)}{Z(\beta)}.$$

We can rewrite this equation as

$$\mathbb{E}^n(\beta) = \int dE E^n P(E|\beta),$$

where $P(E|\beta)$ is the likelihood of being in a state with energy $E$ with temperature $\beta$. The probability is defined as

$$P(E|\beta) = \frac{\exp\left(-\beta E\right)\rho(E)}{Z(\beta)}.$$

With the density of states 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 density of states. 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 gmamma energy $E_\gamma$ 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 $P(E_{x\gamma})$, that is the likelihood of the state of energy $E_x$ given a certain $\gamma$-energy. This quantity will in turn be used to identify a density of states.

In order to get familiar

## Thesis Projects

The aim of this thesis project is employ Bayesian machine learning to define a PDF, either from experiment or from theoretical simulations. Eventually, based on the PDF, can attempt to define a a level density $\rho(E)$, or the other way around. The first step is to use an already available model for extracting the level density from exact diagonalization. These data will then be used to define a posterior distribution based on a Bayesian machine learning approach.

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

1. Spring 2020:
2. Fall 2020:
3. Spring 2021:

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

**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($D$):

$$p(H|D, I) = \frac{p(D|H, I)p(H|I)}{p(D|I)}. \tag{1}$$

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.