

Bayesian techniques and applications to QCD

Alexander Rothkopf*

Faculty of Science and Technology, University of Stavanger, NO-4036 Stavanger, Norway

E-mail: alexander.rothkopf@uis.no

Realizing the full potential of interconnecting the large amounts of data created in physics experiments, phenomenological models and theory simulations requires robust tools for statistical inference. Here I review a particularly promising branch, Bayesian statistics, which over the past decade has found manifold use in high-energy physics. After a brief introduction to Bayesian statistics I will present two concrete examples, where Bayesian thinking has led to progress in understanding strongly interacting matter: unfolding problems in the form of lattice QCD spectral functions (in spirit similar to detector corrections), as well as the efficient estimation of quark-gluon-plasma parameters from a systematic comparison of experimental heavy-ion collision data and phenomenological models.

*XIII Quark Confinement and the Hadron Spectrum - Confinement2018
31 July - 6 August 2018
Maynooth University, Ireland*

*Speaker.

1. Motivation

With the advent of modern high energy collider facilities, be it RHIC at BNL or the LHC at CERN, high energy physics is able to tap into vast amounts of empirical data. On the other hand high-performance computing has made possible the computation and prediction of many phenomenologically relevant quantities from first principles (e.g. hadron masses [1] and thermodynamics [2, 3]). Concurrently phenomenology has progressed in constructing more and more refined models, take for example the treatment of the dynamical evolution of a heavy-ion collision (HIC), which allows us to reproduce many features observed in experiment (for an introduction see e.g. [4]). Making these three treasure troves mutually accessible using efficient methods of statistical inference, in part supported by rapidly developing technologies such as machine learning, is a central goal for theoretical physics in the 21st century. As one of the first high-energy theory conferences, ConfXIII hosts a full session¹ dedicated to this topic.

Bayesian statistics and inference based upon it is one particularly useful piece in the arsenal available to the avid physicist in approaching the challenge of distilling novel physics insight from empirical and simulation data. For those interested in joining this field [5, 6] can serve as a first starting point.

The intention of this talk is to contribute towards demystifying the concept of Bayesian statistics and to provide two concrete examples, where it has helped improve our understanding of strongly interacting matter and QCD.

2. Bayesian inference

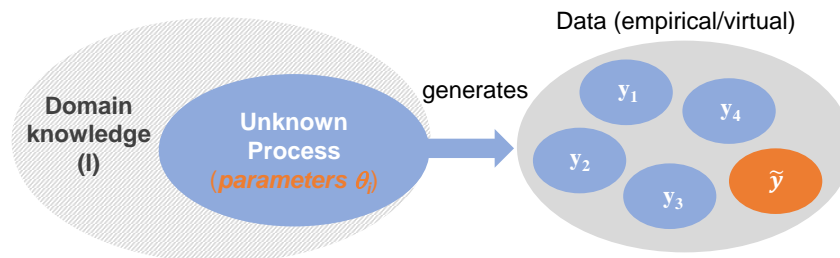


Figure 1: Generic setup for statistical inference. Bayesian inference explicitly acknowledges that a model of the unknown process is embedded in a domain to which prior knowledge I exists.

Let me start with a very general scenario (see Fig.1), where we are observing some unknown process generating data y_k . This may be an experimental setup or a simulation based on Monte-Carlo sampling. The central goal of *statistical inference* now is to draw conclusions about unobserved quantities, based on that empirical data. What we mean by unobserved quantities are on the one hand the unknown parameters of the generating process θ_i and on the other hand future, i.e. as of yet unobserved data \tilde{y}_l .

Bayesian inference is a particular form of statistical inference, which uses a generalized concept of probability. Probability in the Bayesian sense is understood as a measure of uncertainty and does not need to be directly related to a large number of repeatable trials. The flexibility of the

¹ConfXIII - Statistical Methods for Physics Analysis in the XXI Century session

Bayesian approach may be glimpsed when considering the question: What is the chance of the first launch of the new Ariane 6 rocket to succeed. Bayesian statistics sets out to answer in the form of a probability statement. At first it may seem an improper task, as we cannot base our answer on a large number of trials of first launches (not to mention the economic difficulty).

However whenever we describe a process in nature, we are forced to model it using the language of physics and mathematics. Such a model in turn exists within the context of the process of interest, in this case aerospace engineering. And of course there exists prior *domain knowledge* I that informs our model. E.g. engineers have tested individual components of the Ariane rocket, such as valves, for their reliability using e.g. frequentist approaches. As we shall see, Bayesian statistics asks us to explicitly state what the uncertainties of our model parameters are and helps us propagate these uncertainties into the final answer in the language of a generalized probability.

The starting point for Bayesian inference is the joint probability distribution of the involved quantities $p(\theta, y, I)$. Constructing it can be a formidable task, as it requires knowledge about the data generating process and the physics context it is embedded in. Using the rules for conditional probabilities we may reexpress $p(\theta, y, I)$ to give way to *Bayes theorem*

$$\underbrace{p(\theta|y, I)}_{\text{posterior}} = \underbrace{p(y|\theta, I)}_{\text{likelihood}} \underbrace{p(\theta|I)}_{\text{prior}} / \underbrace{p(y|I)}_{\text{evidence}}, \quad (2.1)$$

whose parts are straightforwardly matched to our problem at hand. The likelihood encodes how probable it is that a datapoint y has been generated from a process with fixed set of parameters θ . The prior probability on the other hand tells us how probable it is that a certain set of θ parameters arises from our prior domain knowledge. The evidence (its name is simply a historic convention) enters as a normalization factor independent of the parameters θ .

To the frequentist scholar the prior probability might be a stumbling block, so let us have a closer look. In the literature $p(\theta|I)$ is often parameterized using two quantities, the *default model* m and a weight α . The former refers to the extremum of the prior $\partial_{\theta} p(\theta|I)|_{\theta=m} = 0$ while α denotes an overall weight factor $p(\theta|I) = \exp(\alpha P(\theta|I))$, i.e. they take the role of a *hyperparameter*. Bayesian statistics now forces us to make explicit the parameters of the prior in the joint probability distribution $p(\theta, y, m, \alpha)$. In modern approaches to Bayesian statistics one acknowledges that model parameters themselves carry uncertainties, so that we must endow both m and α with distributions of their own, so called hyperpriors $p(m)$ and $p(\alpha)$. This is the simplest form of a so called *hierarchical model* in which successive layers inform us about the interdependencies and uncertainties of the parameters entering our model.

Once the prior and posterior have been determined we are able to close in on the central task of Bayesian inference, to compute (or more likely numerically estimate) the posterior distribution, which provides access to

$$p_{\text{MP}}(\theta_j|y, I) = \prod_{i \neq j} \int d\theta_i p(\theta|y, I), \quad p_{\text{PP}}(\tilde{y}|y, I) = \int d\theta p(\tilde{y}|\theta, I) p(\theta|y, I), \quad (2.2)$$

the *marginal posterior* $p_{\text{MP}}(\theta_j|y, I)$ and the *posterior predictive* distribution $p_{\text{PP}}(\tilde{y}|y, I)$. The former constitutes the final say of Bayesian inference on how probable a certain parameter of our unknown process is given the observed data and prior knowledge. The latter is valuable for assessing the validity of the inference, as its predictions for future measurements can be compared against actual collected data.

At this point I would like to stress that in its modern incarnation Bayesian inference does not suffer from what is often called subjectivity in the literature. We have seen that prior knowledge is nothing more than domain knowledge with the appropriate uncertainties encoded via hyperpriors.

What makes the Bayesian approach very efficient is the fact that once new measurements y' become available they are easily incorporated in an updated analysis using the old posterior as the new prior $p(\theta|y', y, I) \propto p(y'|\theta, I)p(\theta|y, I)$. Obviously knowledge of the full posterior probability $p(\theta|y, I)$ is needed in this case.

The past decades have seen major progress in the application of Bayesian statistics, often due to improved computational capabilities. Among them is the consistent adaption of priors arising from genuine domain knowledge and not from computational convenience (i.e. a liberation from conjugate priors). The use of multi-layered hierarchical models nowadays allows us to self-consistently incorporate the full inter dependencies and uncertainties of model parameters. Last but not least, the availability of powerful numerical libraries based on Monte-Carlo techniques has made it possible to straight forwardly sample the full posterior distribution instead of settling for just determining its extremum (maximum a posteriori, MAP) as was often the case in the older literature.

Indeed a full Bayesian analysis today is only one download away with e.g. the [MC-STAN](#) library² [7]. This software, maintained and supported both by scientific organizations, such as the Max-Planck Society, as well as industry in the form of Google, provides a high level language to formulate hierarchical models and to evaluate their posterior via an efficient Hamiltonian (hybrid) Monte Carlo sampler (No-UTurn).

After having briefly gone through the main ingredients to Bayesian inference let us continue by considering two concrete examples where Bayesian thinking has proven insightful in the context of QCD.

3. Applications to QCD

3.1 Lattice QCD spectral function unfolding

The topic discussed in this subsection falls into the category of inverse problems that come in the form of an unfolding task. Extracting from the readings of a detector the actual trajectory of a charged particle (see related contributions [8, 9]), as well as the extraction of spectral functions in lattice QCD simulations (see related contributions [10, 11, 12, 13]) are but two examples.

The starting point is to consider how our data is obtained, i.e. how to formulate the likelihood. In both cases the observed data \vec{y} (track in detector, simulated correlation function) is related to the actual quantity of interest \vec{x} (actual track of particle, spectral function) via a linear transformation $\vec{y} = \hat{K}\vec{x} + \vec{\eta}$. The kernel K represents the imperfections of the detector or is a manifestation of Euclidean quantum field theory. It may be obtained from e.g. detector simulations or in the case of lattice QCD correlators from QCD directly. In both experiment and Monte-Carlo simulations the data carry a finite precision, which may be captured by an additive noise term $\vec{\eta}$. In a lattice QCD simulation it often turns out to be Gaussian distributed, which leads to a likelihood that itself is of

²Thanks go out to N. Wink from Heidelberg for bringing this library to my attention

Gaussian form

$$p(y|x) \propto \exp\left[-(y - Kx)^2/2\sigma^2\right]. \quad (3.1)$$

Here σ denotes the uncertainty in the input data y . Of course in a real-world detector setup the type of noise may differ significantly.

If we were to attempt a simple χ^2 fit of \vec{x} from the data \vec{y} , it would correspond in the language of Bayes to setting the prior probability to unity and determining the maximum of the corresponding posterior. This is what is known as a *maximum likelihood* estimate. The difficulty of inverse problems lies in the fact that they are often ill-posed, so that the maximum of the likelihood is not unique and further regularization of the problem is required.

Take the task of reconstructing spectral functions from lattice QCD simulations as concrete example [14]. Numerical simulations of strongly interacting matter on the lattice act very similar to a very distorting detector setup. On the left in Fig.2 I have sketched a hadronic spectrum in vacuum in gray, where we see two well defined bound states and a continuum structure at higher frequencies. In a lattice QCD simulation, which is carried out in an artificial imaginary time direction, this spectrum is translated into a Euclidean correlation function, plotted on the right in blue. The

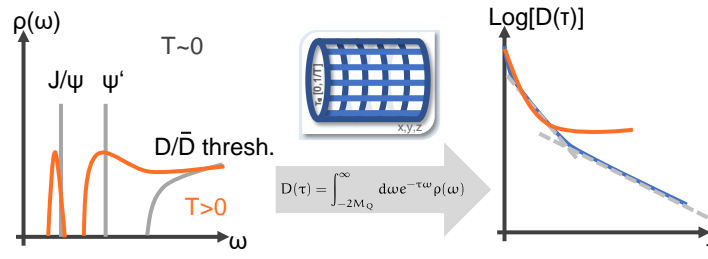


Figure 2: Translating a physical hadron spectral function into a Euclidean correlator via lattice QCD.

kernel in this case (the example pertains to NRQCD [15] fermions) amounts to a Laplace transformation and thus is exponentially damped with frequency ω and Euclidean time τ . In vacuum and in the presence of well separated peaks the bound states manifest themselves as well pronounced exponential falloffs, indicated by the gray dashed lines on the right. This may allow an exponential fitting procedure to reveal many of the vital properties encoded in such a correlator.

Once we move to finite temperature, which is where we wish to learn about QCD for heavy-ion collisions, the spectrum will change significantly. In-medium bound states will show broad peaks, some will already have melted or persist as a mere threshold enhancement of a much more dominant continuum (solid orange line). This translates into a correlator where no more clear exponential behavior is visible and which instead shows overall curvature. The most limiting aspect of the Euclidean formulation of thermal field theory is that temperature and the length of the accessible Euclidean time axis are inversely proportional, indicated by the orange correlator only reaching to a much smaller maximum τ value.

The task at hand is to extract from the simulated correlators the spectral function, containing the relevant physical information about hadrons at finite temperature. This inverse Laplace transform is *the* classic ill-posed inverse problem, in particular, since the intricate structures present in

the spectrum require us to discretize frequencies with many more bins N_ω than we have simulation data points N_τ .

$$D(\tau_i) = D_i = \sum_{l=1}^{N_\omega} \Delta\omega_l \exp[-\omega_l \tau_i] \rho_l. \quad N_\omega \gg N_\tau, \quad D \sim N(\hat{K}\vec{\rho}, C) \quad (3.2)$$

Thanks to the law of large numbers the correlator data D_i are distributed according to a (correlated) Gaussian distribution.

In order to provide a regularization in the Bayesian sense, we need to specify a prior probability which will yield a well behaved, i.e. unimodal posterior. There are several proposals on the market on how to construct such a prior. One proposal arises from what otherwise is known as TV regularization [16], which in the language of Bayes can be recast as $p_{\text{TV}}(\rho|m, \alpha) = \text{Laplace}[m, 1/\alpha]$. Standard Tikhonov regularization [17] on the other hand is equivalent to $p_{\text{TK}}(\rho|m, \alpha) = N[0, 1/\alpha]$ and may be straight forwardly modified to actually take into account the values of the default model $p_{\text{TK}'}(\rho|m, \alpha) = N[m, 1/\alpha]$. The statement that standard Tikhonov regularization does not introduce a default model is incorrect, as it simply sets the values of the default model to zero, leading to the phenomenon of parameter shrinkage.

The most popular choice to date is the Maximum Entropy Method (MEM) [18] with $p_{\text{MEM}}(\rho|m, \alpha) \propto e^{S_{\text{SJ}}}$, which employs the Shannon-Jaynes entropy $S_{\text{SJ}}(m, \alpha) = \alpha \int (\rho - m - \rho \log[\rho/m]) d\omega$. This prior is based on four axioms motivated from considerations of image reconstruction. Contrary to the previous two proposals it actually enforces positivity, which is a genuine piece of prior information in case of hadronic spectra in QCD. It has been argued that the principle of maximum entropy which motivates S_{SJ} leads this prior to favor smooth spectral functions, an assertion which has however been proven incorrect [19, 20].

A recent addition to this collection of priors is the BR prior [21] $p_{\text{BR}}(\rho|m, \alpha) = \text{Gamma}[\alpha + 1, \alpha/m]$ (abbreviated from Bayesian Reconstruction), which is also based on four axioms. In their motivation however the one-dimensional reconstruction task takes center stage. Again positivity is a vital ingredient. In addition the requirement of scale invariance plays a central role, which takes care of the fact that a spectral function (contrary to statements in the literature) is not in general a probability distribution. (In particular hadronic spectral functions diverge with positive powers of the frequency, as shown in perturbation theory.) From all the proposals so far the BR prior imprints prior information in the most gentle fashion, while still leading to a unimodal posterior. A generalization of the BR prior to non-positive spectral functions [22], encountered e.g. in the study of parton spectra [23, 24] has also been put forward.

Once the prior is specified the standard procedure in the literature up to now is to compute the maximum of the posterior (MAP) and declare this function the Bayesian answer to the most probable spectrum, given simulation data and prior information. As long as the posterior is not heavy tailed this is a valid strategy. (For the technical details on how the hyperparameter α is treated in the different methods see [18, 21]). It turns out that the MAP solution may suffer from ringing artifacts as long as local regulators, such as those in the MEM and BR method [19] are used. The exploration of non-local regulators (see e.g. [19, 20]), which in the language of Bayesian statistics correspond to correlated distributions among the different parameters ρ_l is work in progress.

Nowadays we are however not limited to the simple MAP estimate and can easily explore the full posterior distribution using e.g. the MC-STAN framework. This also liberates us from some of

the rather ad-hoc approximations deployed when handling the hyperparameter α e.g. in the MEM. Let me give a concrete example here using actual lattice NRQCD correlation functions for the Bottomonium 3S_1 state Upsilon (data from [19]) to carry out a Bayesian reconstruction with the BR prior.

We start by computing the N_τ means D_i (see top panel of Fig.3) of the N_{data} correlator samples D_i^k (see 2nd panel in Fig.3 for D_1^k) and their correlation matrix $C_{ij} = \frac{1}{N_{\text{data}}(N_{\text{data}}-1)} \sum_n (D_i^k - D_i)(D_j^k - D_j)$ (note that this is the correlation matrix of the mean with respect to the correct value D_i^{true}). The relation between this data and the discretized spectral function is given by eq.(3.2). For computational efficiency we can apply a coordinate transformation on the kernel matrix \hat{K} and the data vector \vec{D} to a basis in which the correlation matrix C is diagonal. Then we may consider each \tilde{D}_i as truly independently Gaussian distributed with an uncertainty given by the corresponding eigenvalue of C . This sets the likelihood of the problem.

Now we turn to the BR prior, to which we here for simplicity supply a constant default model $m = 1$. In practice one can incorporate into m e.g. the large ω behavior of the spectrum known from lattice perturbation theory. The prior includes a hyperparameter α , whose values we do not know apriori. This uncertainty we make explicit by using a very broad half Gaussian distribution as hyperprior $p(\alpha) = \text{HalfNormal}[0, 10]$ (3rd panel of Fig.3). For $\alpha = 10$ the prior should already fully dominate the posterior. This is all that is necessary to set up the MC-STAN model which reads

```
stan_model_BR = """
data {
  int sNt; // # of tau input data
  int sNw; // # of rho bins
  matrix[sNt, sNw] Kernel;
  vector[sNt] D; // simulation data
  real<lower=0> ConstDM; // default model
  vector[sNt] Uncertainty; // corr. error
}
parameters {
  vector<lower=0>[sNw] rho;
  real<lower=0> alpha; // Hyperparameter
}
model {
  alpha ~ normal(0, 10); //hyperprior
```

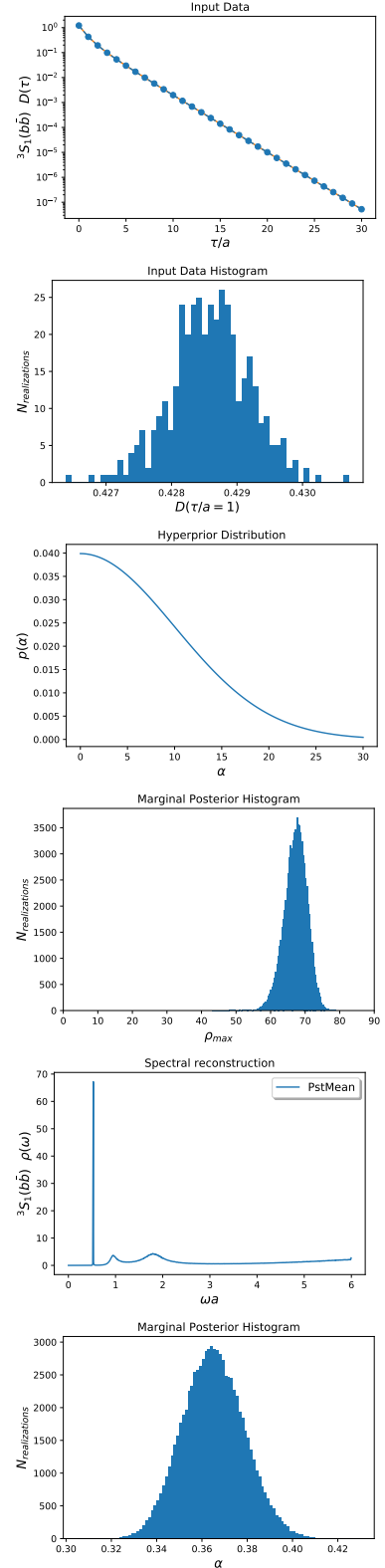


Figure 3: Relevant parts of the Bayesian analysis: input data mean, histogram of D_1 , $p(\alpha)$, $p(\rho_{\text{max}}|D, I)$, ρ_{Bayes} , $p(\alpha|D)$.

```

rho ~ gamma(alpha+1, alpha/ConstDM);
D ~ normal(Kernel * rho, Uncertainty);
}
"""

```

These lines are all that is needed to specify a full Bayesian analysis. Using $N_\omega = 1000$ frequency bins with $N_\tau = 32$ and running with 40 MC chains of 2500 iterations each (500 warmup steps) we obtain a well defined histogram for each of the marginal posteriors $p(\rho_l|D, I)$ (4th panel in Fig.3 for the maximum ρ_l). Via their mean a highly accurate image of the encoded spectral function emerges (panel 2nd from bottom in Fig.3). Note the dominant peak, which corresponds to the exponential falloff in the correlator.

Since we here treated α as one of the parameters of the joint probability distribution we may actually learn self-consistently, how strongly the regulator affects the posterior by looking at the marginal posterior $p(\alpha|D)$ shown in the bottom panel of Fig.3.

Even if the implementation of a Bayesian analysis nowadays is straight forward several conceptual challenges remain. One is related to the question of the information content within the lattice QCD correlation functions. Especially at $T > 0$ the finite Euclidean time extend limits the accuracy of the reconstruction and going to the continuum limit does not remedy this issue (for a recent discussion see [19, 25]). The question to answer is how to set up simulations to improve the relevant information content. Possible strategies involve going to anisotropic lattices or more exploratory routes, such as attempting to generalize the multi-level algorithm to full QCD [26] or to explore modified time contours in the complex plane [25].

The other challenge is how to incorporate analytic properties of the spectra and correlators into the regulator. Current priors use mostly concepts unspecific to QCD, e.g. smoothness. On the other hand analytic properties may be treated by introducing e.g. restricted sets of basis functions among which to select the MAP [23]. This however is not Bayesian in nature, where prior information must be provided in terms of a functional or correspondingly a probability distribution. One possible strategy is to deploy machine learning to construct an appropriate regulator from prior knowledge, which is work in progress.

3.2 Quark-gluon-plasma parameter estimation

The second example of Bayesian inference in QCD connects the progress made in the experimental investigation of heavy-ion-collisions with that of phenomenologically modelling the dynamical evolution of the bulk matter created therein. Its physics goal is to generate new insight into the properties of the quark-gluon-plasma.

Over the past decade a *standard model* of HICs has emerged whose success rests on chaining together several effective descriptions of the physics relevant at different stages of the collision. It starts from fluctuating initial conditions, modern incarnations of which are based on classical statistical simulations of Yang-Mills fields. These smoothly connect via kinetic theory to a description of the bulk matter in terms of viscous relativistic hydrodynamics that in turn terminates at hadronization, where hadronization models connect to transport models for the hadronic phase. At the same time highly precise measurements of the yields and correlations among light hadrons (e.g. harmonic flow) have been obtained in experiment.

It is natural to ask how statistical inference can help us identify the most probable values for the parameters entering the phenomenological models of HICs by employing Bayesian reasoning. We here follow the strategy laid out in the first [27] in a series of studies [28] approaching this question in a genuinely Bayesian fashion (see related contribution [29]). The problem at hand is well suited for Bayesian statistics as we have to deal with many sources of uncertainties, stemming in part from finite precision in theory input, such as e.g. in the equation of state of QCD obtained from lattice QCD. On the other hand one is faced with e.g. parametrization choices for quantities, such as the specific viscosities, since no theory consensus has been reached on their temperature dependence so far. The one uncertainty that is not part of such an analysis is the one related to the range of validity of the deployed models themselves, i.e. it cannot be quantified in how far e.g. viscous hydrodynamics is an appropriate description of the intermediate time evolution of the QGP.

The authors of [27] choose nine parameters among all model parameters which are left free and fix all others using best estimates from the literature. While strictly speaking this does not amount to a fully Bayesian approach it is a very important first step into the direction of an urgently needed comprehensive Bayesian parameter estimation. The free parameters come in two groups, one related to the initial conditions containing four quantities: p entropy deposition, k shape fluctuations, w Gaussian nucleon width and a normalization. The other five parameters are genuine QGP properties: T_{switch} hadronization temperature, η/S_{HRG} specific viscosity in the hadronic phase, η/S_{min} minimal specific viscosity at T_{switch} , the slope of the specific viscosity modeled as linear increasing function of T and ζ/S the magnitude of the specific bulk viscosity.

Given these 9 input parameters θ ; the chain of models is able to produce a plethora of simulated data y_{sim} , such as yields dN_{π}/dy or v_2 for different centrality bins which may then be compared to experimental measurements. Bayes theorem in this case then takes the form

$$p(\theta|y_{\text{exp}}) \propto p(y_{\text{exp}}|\theta)p(\theta), \quad (3.3)$$

where we have suppressed the dependence of the likelihood and posterior on the choice of models used.

The main focus in the following lies on how to efficiently describe the likelihood $p(y_{\text{exp}}|\theta)$ in this problem for which [27] deploys a technique that is well established in the realm of machine learning: Gaussian processes. An introduction to the general topic can be found in [30].

The challenge lies in the fact that to exhaustively sample the posterior we need to evaluate the likelihood for a very large range of parameter values. At the same time the computational cost involved in running the chain of models prevents us from doing so in any reasonable time frame. In order to overcome this limitation, Gaussian processes provide a strategy of how to efficiently estimate the true likelihood based on a small number of simulated training data sets. The starting point is to regard the simulated data y_{sim} as distributed according to a correlated Gaussian, where the non-trivial information is contained in a usually dense correlation matrix Σ . The corresponding likelihood then also takes the form of a correlated Gaussian

$$p(y_{\text{exp}}|\theta) \approx \exp\left[-(y_{\text{sim}}(\theta) - y_{\text{exp}})\Sigma_{\text{GP}}^{-1}(y_{\text{sim}}(\theta) - y_{\text{exp}})\right]. \quad (3.4)$$

As preparatory step one carries out full model computations on a set of $\mathcal{O}(100)$ model parameter sets θ_{train} distributed for efficiency on a *Latin hypercube*. These will serve as training data y_{train} for the Gaussian process.

Our goal is to set up an approximate Gaussian distribution for $y_{\text{sim}}(\theta)$ which is fully specified by its mean $m(\theta, \theta_{\text{train}})$ and its correlations $K(\theta, \theta_{\text{train}})$ interpolating between the values θ_{train} . The starting point is to assume that the training data belongs to the same Gaussian distribution from which we will draw our estimates $y_{\text{sim}}(\theta)$ for some other parameter set

$$\begin{pmatrix} y_{\text{train}} \\ y_{\text{sim}} \end{pmatrix} = N \begin{pmatrix} m(\theta) & K(\theta_{\text{train}}, \theta_{\text{train}}) & K(\theta_{\text{train}}, \theta_{\text{sim}}) \\ m(\theta, \theta') & K(\theta_{\text{sim}}, \theta_{\text{train}}) & K(\theta_{\text{sim}}, \theta_{\text{sim}}) \end{pmatrix}. \quad (3.5)$$

All the important information resides in the model correlation functions K , which we here take to also have Gaussian form. This is a particular choice and independent from the Gaussian character of the joint probability distribution

$$K(\theta_k, \theta'_l) = \sigma_{\text{GP}}^2 \exp\left[-\frac{(\theta_k - \theta'_l)^2}{2l^2}\right] + \sigma_n^2 \delta_{kl}. \quad (3.6)$$

We have three hyperparameters in this model, σ_{GP} specifying the strength of correlation among the training data, the correlation lengths l_k characterizing the distance between parameters and σ_n related to the uncertainty of the y_{train} 's. It is the values for these, which we need to fit using the training data.

Using $p(z|x) = p(x,z)/\int dz p(x,z)$ we then arrive at the conditional probability for our estimated y_{sim} as $p(y_{\text{sim}}(\theta)|y_{\text{train}}) = N(\mu_{\text{GP}}(\theta), \Sigma_{\text{GP}}(\theta))$ with expressions for

$$\begin{aligned} \Sigma_{\text{GP}} &= K(\theta, \theta) - K(\theta, \theta_{\text{train}}) - K(\theta_{\text{train}}, \theta_{\text{train}})^{-1} K(\theta_{\text{train}}, \theta_{\text{train}}), \\ \mu_{\text{GP}} &= m(\theta_{\text{train}}) + K(\theta, \theta_{\text{train}}) K(\theta_{\text{train}}, \theta_{\text{train}})^{-1} (y_{\text{train}} - m(\theta_{\text{train}})). \end{aligned}$$

I.e. we now are in possession of a means to quickly draw an approximate y_{sim} to be used in eq.(3.4) together with the appropriate correlation function Σ_{GP} .

Before we deploy the Gaussian process let us make sure that the approximation made so far is not too simplistic and manages to capture the relevant physics encoded in y_{train} . To this end one can evaluate the full model computations at a new set of parameters and compare those to the predictions. As shown in the top panel of Fig.4 the correlation between prediction and actual values is excellent, the values scatter tightly along the diagonal.

The last missing piece to the Bayesian analysis is to specify the prior distribution. Here the authors resort to uniform distributions for the parameter values setting the upper and lower limits large enough so that the relatively large uncertainty e.g. in the specific viscosities (gray shaded region in Fig.4 bottom panel) is well represented. Combining these together with the likelihood we constructed before allows us to systematically interrogate the experimental data and infer the most probable value of the

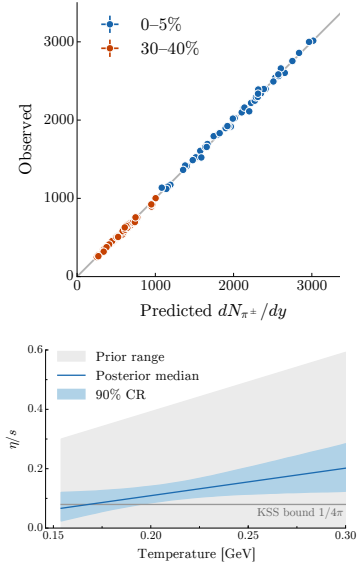


Figure 4: (top) Validation of the Gaussian process model using experimental data (bottom) specific shear viscosity estimate from the posterior (figures taken from [27])

parameters (blue line in bottom panel of Fig.4) accompanied by a significantly reduced uncertainty band.

4. Conclusion

In this contribution my goal has been to argue that Bayesian inference provides a flexible approach to extract insight from empirical data based on the consistent application of Bayes theorem. In the introductory part I showed that in its modern incarnation Bayesian statistics does not suffer from subjectivity, as prior information is understood as domain knowledge informing our modelling with the appropriate uncertainties properly encoded in hyperparameters of hierarchical models. With the availability of efficient and open source Monte-Carlo samplers no more hurdles exists to embark on an exploration of the full posterior distribution of such realistic hierarchical models.

I presented two concrete examples, where Bayesian inference helps us in furthering our understanding of strongly interacting matter. The first is the extraction of lattice QCD spectral functions, which is an ill-posed inverse unfolding problem similar to the unfolding of detector tracks. Here the main challenges reside in the design of improved prior distributions that incorporate analytic properties of the spectra known from QCD, where machine learning may play an important role in the near future. At the same time there exists the need to improve the information content of simulated correlators. The use of anisotropic lattices or exploring more unconventional strategies, such as extensions of the multilevel algorithm to full QCD or deformed temporal paths in the complex plane are currently being considered. The second example concerned the estimation of model parameters for the QGP from a systematic comparison with experimental measurements from HICs. The central focus there lay on the efficient sampling of the likelihood, which was achieved by use of a Gaussian process model.

This talk is only able to scratch at the surface of the many fascinating ways how (Bayesian) statistical analysis can help us push forward the boundaries of our knowledge of strongly interacting matter. The interested reader is cordially invited to peruse the many other contributions to this conference in the realm of statistical methods and machine learning located in the corresponding dedicated session.

References

- [1] S. Borsanyi et al., *Ab initio calculation of the neutron-proton mass difference*, *Science* **347** (2015) 1452 [1406.4088].
- [2] A. Bazavov et al., *The QCD Equation of State to $\mathcal{O}(\mu_B^6)$ from Lattice QCD*, *Phys. Rev.* **D95** (2017) 054504 [1701.04325].
- [3] S. Borsanyi et al., *Calculation of the axion mass based on high-temperature lattice quantum chromodynamics*, *Nature* **539** (2016) 69 [1606.07494].
- [4] W. Florkowski, *Basic phenomenology for relativistic heavy-ion collisions*, *Acta Phys. Polon.* **B45** (2014) 2329 [1410.7904].

- [5] A. Gelman, J. B. Carlin, H. S. Stern and D. B. Rubin, *Bayesian Data Analysis*. Chapman and Hall/CRC, 2nd ed. ed., 2004.
- [6] R. McElreath, *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*, Chapman & Hall/CRC Texts in Statistical Science. CRC Press, 2016.
- [7] B. Carpenter, A. Gelman, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt et al., *Stan: A probabilistic programming language*, *Journal of Statistical Software, Articles* **76** (2017) 1.
- [8] M. Krüger, *Bayesian unfolding of charged-particle p_T spectra with ALICE at the LHC*, in *Proceedings, 13th Conference on Quark Confinement and the Hadron Spectrum (Confinement XIII)*, PoS, 2018.
- [9] M. Kuusela, *Unfolding: Point Estimation, Uncertainty Quantification and Future Directions*, in *Proceedings, 13th Conference on Quark Confinement and the Hadron Spectrum (Confinement XIII)*, PoS, 2018.
- [10] O. Kaczmarek, *Spectral and transport properties from Lattice QCD*, in *Proceedings, 13th Conference on Quark Confinement and the Hadron Spectrum (Confinement XIII)*, PoS, 2018.
- [11] N. Astrakhantsev, *Temperature dependence of $SU(3)$ -gluodynamics bulk and shear viscosities within lattice simulation*, in *Proceedings, 13th Conference on Quark Confinement and the Hadron Spectrum (Confinement XIII)*, PoS, 2018.
- [12] R. Quinn, *Spectral properties of light and charm mesons from $N_f = 2 + 1$ anisotropic lattice QCD*, in *Proceedings, 13th Conference on Quark Confinement and the Hadron Spectrum (Confinement XIII)*, PoS, 2018.
- [13] A. Rothkopf, *In-medium heavy quarkonium from lattice NRQCD*, in *Proceedings, 13th Conference on Quark Confinement and the Hadron Spectrum (Confinement XIII)*, PoS, 2018.
- [14] M. Jarrell and J. E. Gubernatis, *Bayesian inference and the analytic continuation of imaginary-time quantum Monte Carlo data*, *Phys. Rept.* **269** (1996) 133.
- [15] G. P. Lepage, L. Magnea, C. Nakhleh, U. Magnea and K. Hornbostel, *Improved nonrelativistic QCD for heavy quark physics*, *Phys. Rev.* **D46** (1992) 4052 [hep-lat/9205007].
- [16] L. I. Rudin, S. Osher and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, *Physica D: Nonlinear Phenomena* **60** (1992) 259 .
- [17] D. J. C. MacKay, *Information Theory, Inference & Learning Algorithms*. Cambridge University Press, New York, NY, USA, 2002.
- [18] M. Asakawa, T. Hatsuda and Y. Nakahara, *Maximum entropy analysis of the spectral functions in lattice QCD*, *Prog. Part. Nucl. Phys.* **46** (2001) 459 [hep-lat/0011040].
- [19] S. Kim, P. Petreczky and A. Rothkopf, *Quarkonium in-medium properties from realistic lattice NRQCD*, 1808.08781.

- [20] C. S. Fischer, J. M. Pawłowski, A. Rothkopf and C. A. Welzbacher, *Bayesian analysis of quark spectral properties from the Dyson-Schwinger equation*, *Phys. Rev.* **D98** (2018) 014009 [1705.03207].
- [21] Y. Burnier and A. Rothkopf, *Bayesian Approach to Spectral Function Reconstruction for Euclidean Quantum Field Theories*, *Phys. Rev. Lett.* **111** (2013) 182003 [1307.6106].
- [22] A. Rothkopf, *Bayesian inference of nonpositive spectral functions in quantum field theory*, *Phys. Rev.* **D95** (2017) 056016 [1611.00482].
- [23] A. K. Cyrol, J. M. Pawłowski, A. Rothkopf and N. Wink, *Reconstructing the gluon*, 1804.00945.
- [24] E.-M. Ilgenfritz, J. M. Pawłowski, A. Rothkopf and A. Trunin, *Finite temperature gluon spectral functions from $N_f = 2 + 1 + 1$ lattice QCD*, *Eur. Phys. J.* **C78** (2018) 127 [1701.08610].
- [25] J. Pawłowski and A. Rothkopf, *Thermal dynamics on the lattice with exponentially improved accuracy*, *Phys. Lett.* **B778** (2018) 221 [1610.09531].
- [26] M. Ce, L. Giusti and S. Schaefer, *A local factorization of the fermion determinant in lattice QCD*, *Phys. Rev.* **D95** (2017) 034503 [1609.02419].
- [27] J. E. Bernhard, J. S. Moreland, S. A. Bass, J. Liu and U. Heinz, *Applying Bayesian parameter estimation to relativistic heavy-ion collisions: simultaneous characterization of the initial state and quark-gluon plasma medium*, *Phys. Rev.* **C94** (2016) 024907 [1605.03954].
- [28] J. Auvinen, J. E. Bernhard, S. A. Bass and I. Karpenko, *Investigating the collision energy dependence of η/s in the beam energy scan at the BNL Relativistic Heavy Ion Collider using Bayesian statistics*, *Phys. Rev.* **C97** (2018) 044905 [1706.03666].
- [29] V. Kovalenko, *Determination of the quark-gluon string parameters from the data on pp, pA and AA collisions at wide energy range using Bayesian Gaussian Process Optimization*, in *Proceedings, 13th Conference on Quark Confinement and the Hadron Spectrum (Confinement XIII)*, PoS, 2018.
- [30] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*, Adaptive computation and machine learning series. University Press Group Limited, 2006.