Bayesian Monte Carlo Evaluation Framework for Imperfect Nuclear Data[1]

Jesse M. Brown*, Goran Arbanas*, Andrew Holcomb*, Dorothea Wiarda*

*Nuclear Data & Criticality Safety Group, Oak Ridge National Laboratory, P.O. Box 2008, Oak Ridge, TN 37831

INTRODUCTION

Bayesian evaluation of resolved resonance region (RRR) nuclear data has historically been carried out using the generalized least squares (GLS) formalism, as implemented in, e.g., SAMMY [1]. We have recently developed a prototype of Bayesian Monte Carlo (BMC) evaluation framework [2], implemented using a Markov Chain Monte Carlo (MCMC) method with a Metropolis-Hastings (MH) [3, 4, 5] acceptance criterion. This was done in order to remove the approximations underlying the conventional GLS evaluations, namely, the linear approximation, and the approximation that all probability density functions (PDFs) are of the normal kind. Recent works by others have used similar stochastic approaches to quantify cross section uncertainties from ENDF evaluated covariances, and/or, from integral benchmark data [6, 7, 8, 9, 10], but those have not been conceived as an evaluation framework like the one presented here.

In this iteration of Bayesian MC evaluation framework, we deploy a generalization of the conventional Bayesian posterior PDF, along the trajectory outlined by the authors in [11], and provide formal expressions by which evaluators could account for inherent imperfections in nuclear data, besides providing a novel form of the posterior PDF that is amenable to computation by stochastic methods, while elegantly retaining harmony with the Bayes theorem[1]. This generalization yields a formal Bayesian posterior PDF which, once computed, is to be used for computations of all posterior expectation values, including, but not limited to, of posterior covariance matrices. The generalized method is intended to obviate the need for manual increases to the posterior covariance matrices, which are currently resorted to in order to make otherwise unreasonably small evaluated uncertainties acceptable. Furthermore, this evaluation framework provides a natural way of preserving the original data and covariance information, by clearly demarcating it from information or expert judgement introduced by the evaluator. However, this comes at the cost of introducing Lagrange multipliers that serve as a formal vehicle for accommodating, and preserving, any expert judgement(s) made by an evaluator during the evaluation.

In this abstract the new evaluation framework is illustrated by a generalized cost function, and its application to some imperfect cross section data. Furthermore, the connection to the conventional form of the Bayes theorem has been derived formally, and verified numerically on a line-fitting problem.

GENERALIZED BAYESIAN POSTERIOR PDF

One compact way to introduce generalized Bayesian posterior PDF, and the cost function associated with it, is based on the concept of *generalized* data, z, that is a union of the model parameters P, and of data, D,

$$z \equiv \begin{pmatrix} P \\ D \end{pmatrix}, \tag{1}$$

whose prior expectation values, $\langle z \rangle$, and the prior covariance matrix, $\mathbf{C}$, are usually provided to define a (normal) prior PDF, namely,

$$p(z\,|\,\langle z\rangle, \mathbf{C}) \propto e^{-\frac{1}{2}(z-\langle z\rangle)^{\intercal}\mathbf{C}^{-1}(z-\langle z\rangle)}. \tag{2}$$

Furthermore, it is useful to define a vector of deviations, δ, between the model, $T(P)$, and measured data D, namely

$$\delta \equiv T(P) - D, \tag{3}$$

and its covariance matrix, $\mathbf{\Delta}$,

$$\mathbf{\Delta} \equiv \langle(\delta - \langle\delta\rangle)(\delta - \langle\delta\rangle)^{\intercal}\rangle. \tag{4}$$

Using these definitions, we observe that the conventional Bayes' theorem is obtained by constraining the posterior expectation values and the covariance matrix of δ to be equal zero. Consequently, a generalization of the Bayes' theorem originally introduced in [11] is achieved by imposing non-zero constraints on the same posterior expectation values and covariance matrix, namely $\langle\delta\rangle'$ and $\mathbf{\Delta}'$. The freedom to set these constraining values is given to the evaluator. In our notation, the posterior expectation values, computed using the posterior PDF, $p'(z|...)$, are indicated by a prime, as in $\langle\cdot\rangle'$, while prior expectation values, computed using the prior PDF, $p(z|...)$, omit the prime, as in $\langle\cdot\rangle$. The posterior PDF is constructed by using Lagrange multipliers, labeled by a (constant) vector λ and a (constant) matrix $\mathbf{\Lambda}$, whose values are set so that the posterior expectation values $\langle\delta\rangle'$ and Δ' are equal to the values set by the evaluator:

$$p'(z|\langle\delta\rangle', \mathbf{\Delta}', \langle z\rangle, \mathbf{C}) \propto \mathcal{L}(\langle\delta\rangle', \mathbf{\Delta}'|z, \langle z\rangle, \mathbf{C}) \times p(z|\langle z\rangle, \mathbf{C}) \tag{5}$$

[1]Notice: This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE).

[1]Attempts to to compute posterior PDF of the conventional form of Bayes theorem by stochastic methods become computationally intractable for sufficiently large data sets because the posterior MC weights become increasingly dominated by a single instance of the entire ensemble, namely the one with the least value of the cost function, i.e., χ^2. (This failing is analogous to the "curse of dimensionality" in the case of large parameter sets, and the method presented here may be able to address it, too.) This behavior is conventionally avoided by modifying the expression for the cost function (e.g. dividing it by the number of data points) in order to make the posterior PDF easier to explore by stochastic methods, but sacrificing the harmony with the Bayes theorem by doing so. Mathematical harmony with the Bayes theorem is completely preserved in the framework presented here.

where the likelihood function turns out to be

$$\mathcal{L}(\langle\delta\rangle', \mathbf{\Delta}'|z, \langle z\rangle, \mathbf{C}) = e^{-\frac{1}{2}(\delta-\lambda)^{\intercal}\mathbf{\Lambda}^{-1}(\delta-\lambda)}, \tag{6}$$

and where values of λ and $\mathbf{\Lambda}$ are such that $\langle\delta\rangle'$ and $\mathbf{\Delta}'$ computed using this posterior PDF are exactly equal to the values chosen by an evaluator. This is the posterior PDF that should be used for calculations of all posterior expectation values, including covariances, and these ought not be adjusted by hand, as is often done when evaluation is performed using a conventional form of Bayes's theorem. This is posterior PDF of generalized data, i.e., of model parameters, P, and, of data, D, in contrast to the conventional Bayesian posterior PDF that is a PDF of parameters, P, only, as elaborated further below.

This generalized form of the Bayes' theorem, unlike the conventional one, is amenable to Monte Carlo methods, even as the number of data points becomes infinitely large. This feature can be illustrated by setting $\lambda = \langle\delta\rangle$, i.e. the prior expectation value of δ, and $\mathbf{\Lambda} = \mathbf{\Delta}$, i.e. the prior covariance matrix of δ. To simplify the illustration further, lets assume that the prior covariance matrix, $\mathbf{C}$, is an identity matrix. Then the application of the central limit theorem to computation of our new cost function $X^2(z)$ implies that the stochastic ensemble generated by randomly drawing values of z from the prior PDF, will converge to a normal PDF (of unit variance) with increasing number of data points. Consequently, the posterior expectation values can straightforwardly be computed as weighted averages of the said stochastic ensemble.

The expressions for calculating λ and $\mathbf{\Lambda}$ from given $\langle\delta\rangle'$, $\mathbf{\Delta}'$, $\langle z\rangle$, and $\mathbf{C}$, will be presented in a future publication. (To illustrate some of the features of this posterior PDF, we will set λ and $\mathbf{\Lambda}$ to some instructive set of values, and show the results for each such set.) Collecting the exponents of the likelihood and the prior function, renders the posterior PDF as

$$p'(z|\langle\delta\rangle', \mathbf{\Delta}', \langle z\rangle, \mathbf{C}) \propto e^{-\frac{1}{2}X^2(z)}, \tag{7}$$

where $X^2(z)$ is a generalized cost function:

$$X^2(z) \equiv (\delta-\lambda)^{\intercal}\mathbf{\Lambda}^{-1}(\delta-\lambda) + (z-\langle z\rangle)^{\intercal}\mathbf{C}^{-1}(z-\langle z\rangle), \tag{8}$$

which becomes practical for linear models and normal PDFs, since it could be minimized analytically as a function of z.

The conventional Bayesian posterior PDF is obtained by setting $\lambda = 0$ and $\mathbf{\Lambda} = 0$ (corresponding to constraints $\langle\delta\rangle' = 0$ and $\mathbf{\Delta}' = 0$), for which the likelihood function becomes a Dirac delta function

$$\mathcal{L}(\langle\delta\rangle' = 0, \mathbf{\Delta}' = 0\,|\,z, \langle z\rangle, \mathbf{C}) = \delta_{\mathrm{Dirac}}(T(P) - D). \tag{9}$$

Thanks to this particular form of the likelihood function, the data, D, can be integrated out of the posterior PDF, and thus yield a (conventional) posterior PDF of model parameters, P, alone:

$$p'(P\,|\,\langle z\rangle, \mathbf{C}) \propto e^{-\frac{1}{2}\chi^2(P)} \tag{10}$$

where $\chi^2(P)$ is the conventional cost function:

$$\chi^2(P) = (\hat{z}(P) - \langle z\rangle)^T\,\mathbf{C}^{-1}\,(\hat{z}(P) - \langle z\rangle)\,, \tag{11}$$

where

$$\hat{z}(P) \equiv \begin{pmatrix} P \\ T(P) \end{pmatrix}. \tag{12}$$

This shows that the conventional form of the Bayes' theorem is indeed a special case of the generalized one. In the following section we demonstrate the equivalence numerically, by verifying that the generalized BMC solution is equivalent to the analytical solution of a line-fitting problem, for a sufficiently small $\mathbf{\Lambda}$ matrix.

LINK TO CONVENTIONAL χ^2

In order to numerically demonstrate the equivalence between the new cost function and the conventional χ^2 approach, we needed to calculate a posterior distribution in which the value of $\Lambda \to 0$ and $p(f|\langle z\rangle, \mathbf{C}) \to \delta_{Dirac}$. Since we can analytically solve Bayes' equation for linear models, we compare GLS to the MH solutions of a simple linear model. If we had simply set $\mathbf{\Lambda}$ to a value << 1, e.g. `1e-20`, we would have needed an enormous amount of time for the MH algorithm to converge (and of course if $\mathbf{\Lambda} = 0$ we would encounter divide-by-zero errors). Instead we iteratively solved (iterating outside of the MC loop) Bayes' equation by MHMC, where upon each successive iteration we set:

$$\begin{aligned} \langle z\rangle_{n+1} &\leftarrow \langle z\rangle'_n, \\ \mathbf{C}_{n+1} &\leftarrow \mathbf{C}'_n, \\ \mathbf{\Lambda}_{n+1} &\leftarrow \mathbf{\Delta}'_n, \end{aligned} \tag{13}$$

where n is the index of iterations in the MC sequence (not the internal MC index). On each Bayesian solve for the linear model we reduced the values of δ and therefore $\mathbf{\Delta}$. By setting $\mathbf{\Lambda}_{n+1} \leftarrow \mathbf{\Delta}'_n$ we incrementally reduced $\mathbf{\Lambda}_{n+1}$ and more closely estimated $\mathbf{\Lambda} \to 0$. Each of the P and D PDFs are shown in Fig. 1 for their prior distributions and posteriors as determined by conventional GLS χ^2, MHMC, and our iterative MHMC approximation of the conventional GLS χ^2.

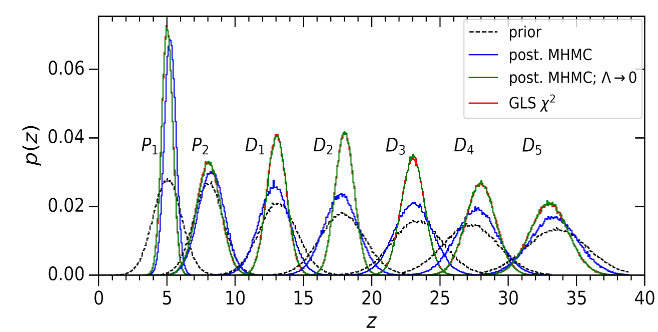

Fig. 1. The posterior distribution of parameters in z for MHMC (blue) are wider than the posteriors of the conventional χ^2 method (red). We also showed that we could approximate the conventional method with an iterative solution of MHMC (green), where we run a MC algorithm with iteratively smaller values of $\mathbf{\Lambda}$. These are all compared to the prior distributions of z (black, dashed).

The similarity between the conventional GLS method and the iterative MHMC posterior PDFs (which are difficult to differentiate as they lie almost entirely on each other) provides a measure of confidence in the new method. The increased variance of the normal MHMC posterior PDF as compared to that of the GLS χ^2 demonstrates the capability of the new method, where the evaluator can provide input on their confidence in a given set of data by providing an appropriate $\mathbf{\Lambda}$. In Fig. 1 we selected a $\mathbf{\Lambda}$ such that the algorithm "knows" that that the data will have an imperfect match to the model. This could be applied, e.g., to two discrepant datasets.

SIMPLE ^{233}U EVALUATION

In order to demonstrate the functionality of the BMC method with the new cost function applied to a real dataset and non-linear nuclear interaction model we evaluated a neutron transmission measurement of ^{233}U. In order to model this experimental data, we use the R-matrix code SAMMY. SAMMY is a well-established program for modeling energy-differential cross section data and evaluating it using a conventional GLS Bayesian analysis. Recently, SAMMY has undergone modernization efforts that allowed us to create a beta (in-development) version of the code that can incorporate arbitrary fitting methods such as MHMC. We can then consistently compare the GLS method to MHMC since we are using precisely the same compiled software to model the cross section (and therefore transmission) in each case.

We selected the ^{233}U neutron transmission dataset by Guber [12] for our example. The SAMMY model calculates cross section via the R-matrix formalism for the RRR and then Doppler and resolution broadens it to match the transmission experiment. Only a single resonance at $\approx$1.75 eV is evaluated (and therefore allowed to vary) in this example. The prior uncertainties on the energy E_λ, radiative width Γ_γ, neutron width Γ_n, and two fission widths $\Gamma_{f,1}$ and $\Gamma_{f,2}$ are from the ENDF/B-VIII.0 library [13]. We then fit using the same model and same priors for both the GLS and BMC methods to compare. The MHMC evaluation used very simple posterior constraints: $\lambda = 0$ and diag($\mathbf{\Lambda}$) = 0.25 (all off-diagonal equal to zero). In Fig. 2 we show the various transmission models for this experiment.

The bold blue line represents the mean model given the posterior distribution of P, while in light blue we've plotted a random set of models tested by the MHMC method to demonstrate the full distribution of the transmission models. Note that the light blue curves are showing samples beyond a single standard deviation from the mean. The model $T(P)$ for which P provides the minimum value of δ (in chartreuse) among the MHMC iterations is shown alongside the GLS solution demonstrating that they are quite similar. This is somewhat intuitive as the GLS algorithm gives significant weight to δ. The comparison of the uncertainty between the two models is, however, quite different. This is shown in Fig. 3, where we see that the standard deviation from the mean for GLS and MHMC are proportional to each other, and the MHMC is approximately an order of magnitude larger. This is expected due to the simplistic choice of a constant λ and $\mathbf{\Lambda}$.

CONCLUSIONS

The Monte Carlo method for solving Bayes' equation has the opportunity to provide posterior distributions (normal or non-normal) of RRR parameters from which any order of moment could be extracted, not only the mean values and covariances, as are currently stored in common nuclear data formats. In addition to using the BMC method in the SAMMY framework to evaluate RRR parameters, the development of this new cost function will, once fully incorporated, give evaluators greater freedom to address imperfections in nuclear data cross section models.

REFERENCES

1. N. M. LARSON, "Updated User's Guide for Sammy: Multilevel R-Matrix Fits to Neutron Data Using Bayes' Equations," Tech. Rep. ORNL/TM-9179/R8, Oak Ridge National Laboratory, Oak Ridge, Tennessee, United States (2008).
2. G. ARBANAS, J. M. BROWN, D. WIARDA, and A. HOLCOMB, "Bayesian Monte Carlo Evaluation Framework for Cross Sections Nuclear Data and Integral Benchmark Experiments," in "Transactions of the American Nuclear Society, Vol. 123, 2020 ANS Virtual Winter Meeting, November 16-19, 2020," Washington, D.C. (2020).
3. J. C. SPALL, "Estimation via Markov Chain Monte Carlo," in "Proceedings of the 2002 American Control Conference (IEEE Cat. No.CH37301)," (2002), vol. 4, pp. 2559–2564 vol.4.
4. S. CHIB and E. GREENBERG, "Understanding the metropolis-hastings algorithm," *The American Statistician*, **49**, *4*, 327–335 (1995).
5. S. D. HILL and J. C. SPALL, "Stationarity and Convergence of the Metropolis-Hastings Algorithm: Insights into Theoretical Aspects," *IEEE Control Systems Magazine*, **39**, *1*, 56–67 (2019).
6. L. FIORITO, G. ŽEROVNIK, A. STANKOVSKIY, G. VAN DEN EYNDE, and P.-E. LABEAU, "Nuclear Data Uncertainty Propagation to Integral Responses Using SANDY," *Annals of Nuclear Energy*, **101**, 359–366 (2017).
7. P. HELGESSON, H. SJÖSTRAND, A. KONING, J. RYDÉN, D. ROCHMAN, E. ALHASSAN, and S. POMP, "Combining Total Monte Carlo and Unified Monte Carlo: Bayesian Nuclear Data Uncertainty Quantification from Auto-generated Experimental Covariances," *Progress in Nuclear Energy*, **96**, 76–96 (2017).
8. D. ROCHMAN, E. BAUGE, A. VASILIEV, H. FERROUKHI, S. PELLONI, A. KONING, and J. C. SUBLET, "Monte Carlo Nuclear Data Adjustment via Integral Information," *The European Physical Journal Plus*, **133**, *12*, 537 (2018).
9. D. ROCHMAN and A. KONING, "Random Adjustment of the H in H2O Neutron Thermal Scattering Data," *Nuclear Science and Engineering*, **172**, *3*, 287–299 (2012).
10. R. CAPOTE and D. SMITH, "Unified Monte Carlo and Mixed Probability Functions," *Journal of the Korean Phys-*

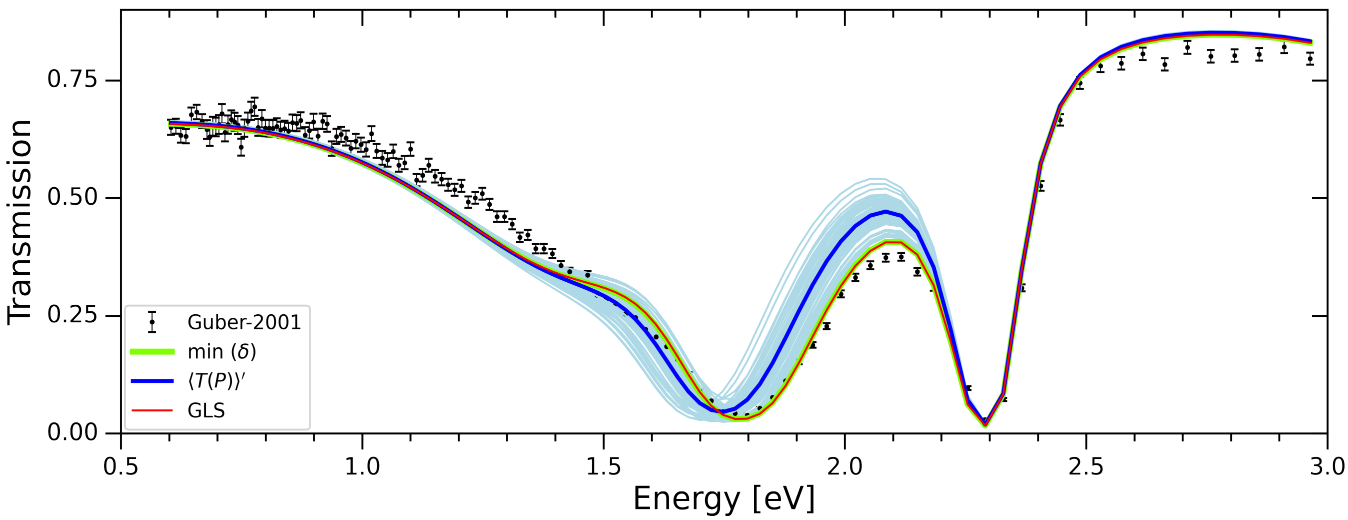

Fig. 2. Shown here are various neutron transmission models created by the SAMMY code. The MHMC algorithm produces a model $T(P)$ at every iteration, a sample of these models is shown in light blue to provide the reader some intuition on what the variance of the model might be, based on the spread of $T(P)$. The posterior expectation value of the model $\langle T(P)\rangle'$ is shown in a bold blue. The conventional χ^2 solution to Bayes' equation (normal SAMMY operation) is shown in red, and matches very well to the MC $T(P)$ corresponding to a minimum δ value shown in chartreuse.

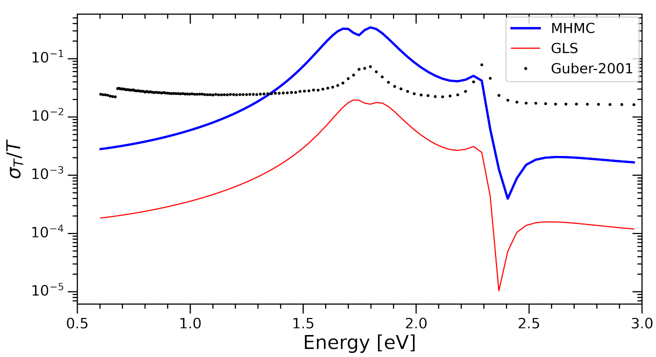

Fig. 3. The relative posterior model uncertainty (σ_T/T) for the MHMC method (in blue) is approximately an order of magnitude greater than the GLS method (in red). This came from the evaluator input quantifying the degree of belief in how "perfect" the data D would match the model $T(P)$. Note that the uncertainty around the $\langle T(P)\rangle'$ is proportional to the GLS uncertainty. This is because $\mathbf{\Lambda}$ is a constant value along the diagonal and zeros elsewhere.

ical Society, **59**, *2*, 1284–1287 (2011).

11. G. ARBANAS ET AL., "Bayesian Optimization of Generalized Data," *EPJ Nuclear Sciences and Technology*, **4**, 30 (2018).
12. K. H. GUBER ET AL., "High-resolution transmission measurements of ^{233}U using a cooled sample at the temperature T= 11 K," *Nuclear science and engineering*, **139**, *2*, 111–117 (2001).
13. D. BROWN ET AL., "ENDF/B-VIII.0: The 8th major release of the nuclear reaction data library with CIELO-project cross sections, new standards and thermal scattering data," *Nuc. Dat. Sheets*, **148**, 1 – 142 (Feb. 2018. Accessed: Jul. 2, 2019).