Project 1

Effective field theory modeling (Bayesian parameter estimation)

Number of pages (excluding title page, references & appendix): 7

Hampus Hansen
CID: `hhansen`

Erik Karlsson Öhman
CID: `erikohm`

1 Problem description

Quantum chromodynamics is a theory concering strong force interaction between quarks, the fundamental particle of hadrons. Calculating the physical properties of these systems are challenging due to the nature of the QCD theory. For example, the large valued coupling constant at low energies renders pertubative methods inefficient [1] and the many-body nature of the systems complicates lattice QCD calculations immensly [2].

An efficient solution to this problem is to use an effective field theory (EFT), thus changing the effective degrees of freedom from quarks and gluons to hadrons. This change of perspective keeps the symmetries while decreasing the degrees of freedom of the system drastically, which allows computation without the need for increadible computational power. In order to model the nuclei with the effective field theory we need to assign values to the parameters called Low Energy Constants (LECs) since these dictate the strong force interaction at nuclear energies [1].

The purpose of this project is to infer the values of two such LECs, c_{D} and c_{E}, given some experimental data of four observables, [3]

$$\boldsymbol{y}_{\mathrm{exp}} = (E\,(^4He), r\,(^4He), E\,(^3H), f\,T_{1/2}), \qquad \text{where} \quad \begin{cases} E\,(^4He) = -28.296\,\mathrm{MeV} \\ r\,(^4He) = 1.4552\,\mathrm{fm} \\ E\,(^3H) = -8.482\,\mathrm{MeV} \\ f\,T_{1/2} = 1129.6\,\mathrm{s}, \end{cases}$$

and to some extent reproduce the results of the article *Rigorous constraints on three-nucleon forces in chiral effective field theory from fast and accurate calculations of few-body observables* by Wesolowski et al. [3] This is done via Bayesian parameter estimation, and the methods to achieve this will be descibed in section 2.

2 Methods description

Since the parameters we need to infer are physical constants and not random variables, it is unwise to use a frequentist method. We should instead use a method of infering the probability of these parameters given the data such that the probability is maximized. This motivates the use of the Bayesian parameter estimation, the process of infering the model parameters $\boldsymbol{\theta} = (c_{\mathrm{D}}, c_{\mathrm{E}})$ given the data $\boldsymbol{y}_{\mathrm{exp}}$. The posterior is given by

$$p(\boldsymbol{\theta}|D, I) = \frac{p(D|\boldsymbol{\theta}, I)p(\boldsymbol{\theta}|I)}{p(D|I)}, \tag{1}$$

where $p(D|\boldsymbol{\theta}, I)$ is the data likelihood meaning the likelihood that the data is produced given the parameters and background information, $p(\boldsymbol{\theta}|I)$ is the model prior and $p(D|I)$ is a normalization factor. Each factor of the numerator is a probabilitiy density function (pdf) and the idea is to analyze the posterior pdf as a function of $\boldsymbol{\theta}$, thus allowing us to analyze how the choice of $\boldsymbol{\theta}$ effects the models accuracy conditioned to some observables. [4]

The model prior is the probability of the parameter $\boldsymbol{\theta}$ before any data is gathered. We know that the model parameters are of natural size and could be either positive of negative. This motivates the choice of each parameter prior to be represented by a Gaussian distribution centered around zero and with $\sigma_{\mathrm{prior}} = 5$. A prior assumption we make is that the parameters are independent of each other, so the joint prior can be written as a product of the two separate Gaussian distributions [4]

$$p(\boldsymbol{\theta}|I) = \mathcal{N}(c_{\mathrm{D}}; \mu = 0, \sigma_{\mathrm{prior}} = 5)\mathcal{N}(c_{\mathrm{E}}; \mu = 0, \sigma_{\mathrm{prior}} = 5). \tag{2}$$

The data likelihood is computed at the experimental values $\boldsymbol{y}_{\mathrm{exp}}$ and is normally distributed, centered around the point predicted by the model $\boldsymbol{y}_{\mathrm{th}}$, with the adopted uncertainty from [3] used as the standard deviation

$$\boldsymbol{\sigma}_{\text{Adopted Uncertainty}} = (0.005\,\mathrm{MeV},\ 0.0062\,\mathrm{fm},\ 0.0015\,\mathrm{MeV},\ 3.0\,\mathrm{s}), \tag{3}$$

which includes both the experimental errors and the errors from the model prediction. The model prediction is calculated with the function `vfewnucleonEmulator()` from the `quantumsolver` library, which takes the parameters c_{D}, c_{E} as input and computes observables for a $3-4$ nucleon system, including the relevant observables [5]. This results in the data likelihood

$$p(D|\boldsymbol{\theta}, I) = \mathcal{N}(\boldsymbol{y}_{\mathrm{exp}}; \boldsymbol{y}_{\mathrm{th}}, \boldsymbol{\sigma}_i), \tag{4}$$

where $\boldsymbol{\sigma}_i$ specifies the width of the distribution for the observables. The truncation performed in the EFT expansion yields a truncation error that can be quantified by the model discrepancy term

$$\boldsymbol{\sigma}^2_{\text{EFT}} = \frac{(\boldsymbol{y}_{\text{exp}} \bar{c} Q^{k+1})^2}{1 - Q^2}, \tag{5}$$

where the parameter are set to $\bar{c} = 1$, $k = 3$ and $Q = 0.33$. This model discrepancy term can be added to the aforementioned $\boldsymbol{\sigma}_i$ in order to account for this truncation error. Results for both cases ($\boldsymbol{\sigma}_i = \boldsymbol{\sigma}_{\text{Adopted Uncertainty}}$ and $\boldsymbol{\sigma}^2_i = \boldsymbol{\sigma}^2_{\text{Adopted Uncertainty}} + \boldsymbol{\sigma}^2_{\text{EFT}}$) will be presented in section 3.1.

The posterior is then calculated as seen in equation 1. This pdf will be a function of the two model parameters. Evaluating the posterior in a two dimensional grid in the model parameter space yields the posterior pdf as a function of $(c_{\text{D}}, c_{\text{E}})$, the analysis of which is invaluable when estimating the parameters, an essential component of the EFT modelling of nuclear properties.

Assuming that the evaluation of the model in each point takes no more than $\sim 1\,\text{ms}$ a reasonable upper limit on number of grid points would be around 500x500 points, given that computations should take no more than a few minutes. In reality computations were slightly less expensive than anticipated and a grid of 1000x1000 points turned out to be computationally feasible.

This grid search method is useful since it evaluates the posterior in the grid with some step size. This means that a grid search will find all modes of the posterior, even if there are multiple modes, assuming the stepsize is small enough. However, two major problems can arise from this. The number of grid points increases exponentially with the number of dimensions, resulting in very inefficient computation. Another problem is if the posterior distribution is narrow it can appear as multiple local modes. The Markov Chain Monte Carlo method (MCMC) is therefor prefered over the direct grid search method.

MCMC builds on multiple "walkers", each taking a step randomly taken from a proposal step function $S(\boldsymbol{\theta}, \boldsymbol{\Phi})$. We start by picking a random position and evaluating the posterior at this point, $p(\boldsymbol{\theta}_0)$. At step j, a new position, $\boldsymbol{\Phi}$ is proposed, specified by a random "angle" and the hyperparameter stepsize. The posterior is evaluated at this new proposed point, $p(\boldsymbol{\Phi})$. The step is then accepted or rejected depending on the MH-ratio, r, defined by

$$r = \frac{p(\theta_1)}{p(\theta_0)} \frac{S(\boldsymbol{\theta}_1, \boldsymbol{\theta}_0)}{S(\boldsymbol{\theta}_0, \boldsymbol{\theta_1})}. \tag{6}$$

The last factor cancels if S is a symmetric step proposal function. Acceptance of step $j+1$, given the new position $p(\boldsymbol{\Phi})$, is determined by the algorithm

(i) if $r \geq 1:$ ACCEPT and set $\boldsymbol{\theta}_{j+1} = \boldsymbol{\Phi}$

(ii) if $r < 1:$ draw random sample, $s \sim U[0,1]$
- if $s \leq r:$ ACCEPT and set $\boldsymbol{\theta}_{j+1} = \boldsymbol{\Phi}$
- if $s > r:$ REJECT and set $\boldsymbol{\theta}_{j+1} = \boldsymbol{\theta}_j$

If this process is repeated enough times the distribution of the samples will converge to the posterior pdf. The first number of steps are generally not representative of the distribution, and thus the hyperparameter "burn-in" is introduced. This burn-in specifies the number of initial steps that should be excluded from the recording. This ensures that the walkers "finds" the main distribution before the samplings are recorded, and thereby are representative of the distribution itself. It converges quicker if multiple walkers are used and the algorithm is run for each walker after each new step. This method decreases the computational power needed for higher dimensionell computations drastically [4]. This method was implemented with the function `EnsembleSampler()` from the `emcee` library.

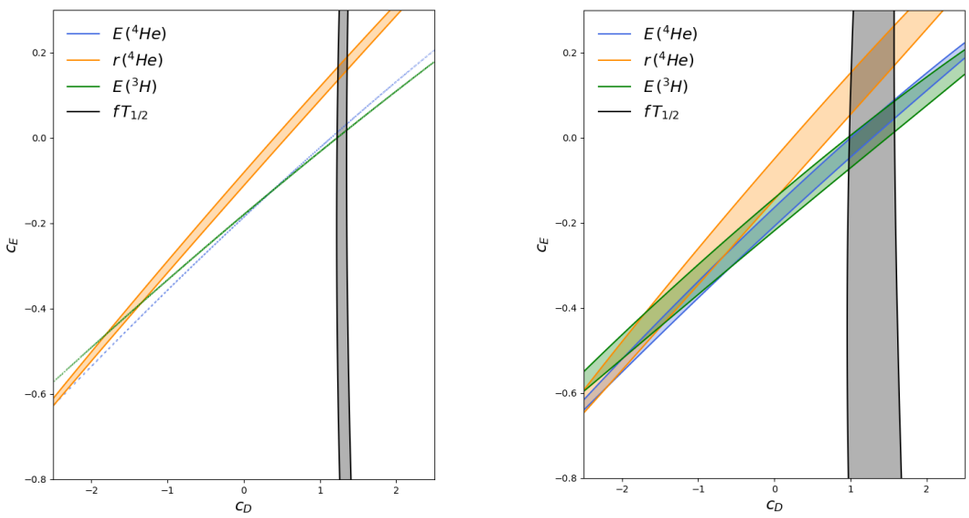
A posterior predictive method can be used to make predictions in regions where there is no available experimental data. This is done by introducing a posterior predictive distribution, conditioned on the posterior infered from the existing experimental data. But this method is also useful as a form of model checking. It is clearly of interest to quantify the quality of the model, i.e. check how accurate its predictive capabilities are. If the posterior predictive distribution is conditioned on the same observables as the ones that were used for conditioning the posterior distribution, we can check how well the model reflects the available experimental data. If the predictive posterior distribution predicts the experimental data of the observables, the model can be regarded as consistent.

3 Results and discussion

The results and the following discussion are presented below under subsections corresponing to each task. The values of the hyperparameters for each relevant method are specified with the corresponding result.

3.1 Task 1

The results from task 1 is presented in figure 1. Figure 1b is generated with the model discrepancy term included in the variation of the data likelihood around the model's theoretical value. This explains the wider distribution seen in this figure, compared to figure 1a. The grid search for the separate observables generates a maximum joint posterior along a ridge, resulting in an approximatly linear dependence of the two model parameters. The distributions of $E\,(^4He)$, $r\,(^4He)$ and $E\,(^3H)$ are positively correlated, while the distribution for $f\,T_{1/2}$ is independent of c_{E} in this particular parameter range.



(a) Excluding model discrepancy term $\boldsymbol{\sigma}_{\mathrm{EFT}}$

(b) Including model discrepancy term $\boldsymbol{\sigma}_{\mathrm{EFT}}$

Figure 1: The iso-probability curves of the joint posterior distibutions as a function of c_{D} and c_{E}, obtained with grid search method, where the colored regions correspond to the 1-σ region, i.e. 68 % of the probability mass is located within these lines. Each line correspond to one observable. Both figures are generated with 1000x1000 grid points.

The best choice of the model parameters with regards to all the observables would be the intersect of their distributions. But even when considering the truncation error from the EFT-approximation, the four separate distributions do not have a single intersect. In order to set the model parameters to the values that maximizes the posterior conditioned on all four observables, we would need to pick a point somewhere between the intersect of $E\,(^4He)$ and $f\,T_{1/2}$ and the intersect of $r\,(^4He)$, $E\,(^3H)$ and $f\,T_{1/2}$. This will be discussed further in section 3.2.

The result compared to those presented by Wesolowski et al. are rather similar [3]. This is expected since the parameters are infered based on the same experimental data. There are differences however, especially for the $r\,(^4He)$ observable. These differences are due to differences in the modelling, aswell as some additional approximations in this project.

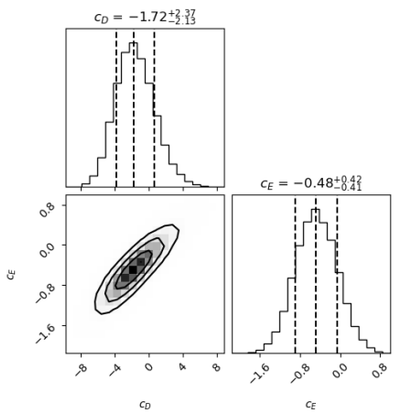
The problem of a narrow distribution in the grid searching method can be seen in figure 1a. The distribution of $E\,(^4He)$, and to some extent $E\,(^3H)$, appears as discrete local modes. This does not truly represent the positive distribution. In order to solve this, one could plot more than one iso-probability curve, for example the $2-\sigma$ would be wide enough to close the gaps between the modes. Another solution would be to generate the figure with more grid points. However, this solution is not prefered since more grid points would increase the computational time greatly, as mentioned before. One should instead opt for using the MCMC method in this case.

3.2 Task 2

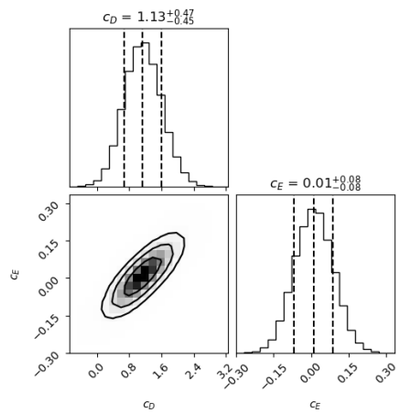
The following two posterior distributions were generated by MCMC sampling. One distribution is conditioned on two obervables: $E\,(^4He)$ and $r\,(^4He)$, whilst the other is conditioned

on all observables: $E\,(^4He)$, $r\,(^4He)$, $E\,(^3H)$ and $f\,T_{1/2}$, figure 2a and 2b respectively. Key assumptions during the generation of the posterior distributions were that all errors were independent and that these errors were given by $\sigma^2 = \sigma^2_{\text{Adopted Uncertainty}} + \sigma^2_{\text{EFT}}$.

50 walkers and 3000 steps were used for the MCMC sampling. Out of those 3000 steps, 100 steps were used as a burn-in. The walkers starting position were randomized numbers between zero and one, which in this context turned out to be an appropriate choice. As such the walkers reached the distribution quickly, and not many samples had to be removed, see appendix A for detailed convergence analysis.



(a) Sampled posterior conditioned on two observables, $E(^4He)$ and $r(^4He)$, case 1.



(b) Posterior conditioned on all observables, case 2.

Figure 2: Joint posterior distributions of c_{D} and c_{E}. Dotted lines mark the median and the 68% degree of belief intervals. The contours in the figures correspond to 0.5-, 1-, 1.5- and 2-sigma intervals.

In the joint posterior conditioned on $E(^4He)$ and $r(^4He)$ we can infer the parameter medians to $c_{\text{D}} = -1.72$ and $c_{\text{E}} = -0.48$. Their corresponding 68% degree of belief intervals are $-3.85 \leq c_{\text{D}} \leq 0.65$ and $-0.89 \leq c_{\text{E}} \leq -0.06$. It is also clear from the corner plot that the parameters c_{E} and c_{D} are positively correlated. In the second posterior, when we used all available observables we found the parameter medians to be $c_{\text{D}} = 1.13$ and $c_{\text{E}} = 0.01$. The 68% degree of belief intervals are in this case $0.68 \leq c_{\text{D}} \leq 1.60$ and $-0.07 \leq c_{\text{E}} \leq 0.09$. We note that c_{D} and c_{E} are positively correlated in this case as well.

The parameter values differ quite a lot from case 1 to case 2. By comparing the result of task 2 to task 1 we can be assured that this is an expected result. The point of intersection for the posterior distributions of $E(^4He)$ and $r(^4He)$ in figure 1 lay between $-2.0 \lesssim c_{\text{D}} \lesssim -1.0$ and $-0.6 \lesssim c_{\text{E}} \lesssim -0.4$ which corresponds qualitatively to the parameter values infered in task 2, case 1. Similarly the area where almost all posterior distributions intersect, $1.0 \lesssim c_{\text{D}} \lesssim 1.6$ and $-0.1 \lesssim c_{\text{E}} \lesssim 0.1$ agree quite well with the values from task 2, case 2.

3.3 Task 3

By using the previously infered parameter distribution from task 2, case 2 as input values to the `vfewnucleonEmulator()` from the `quantumsolver` library we can obtain posterior predictive distributions of the values of the observables.

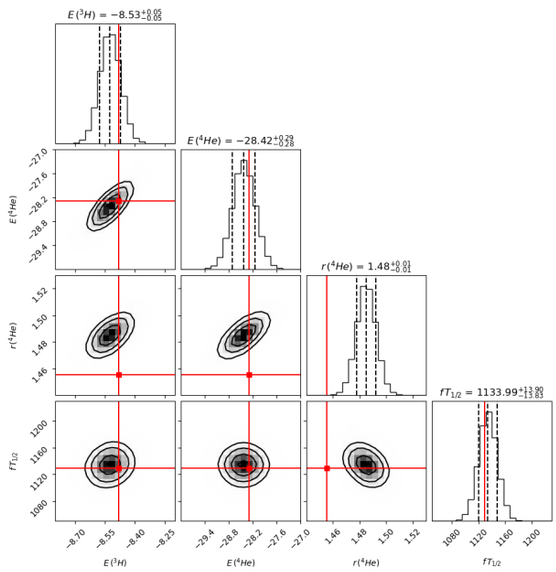


Figure 3: Posterior predictive distribution of observable values. The red lines and squares denote experimental observable values. Dotted lines mark the median and the 68% degree of belief intervals. The contours in the figures correspond to 0.5-, 1-, 1.5- and 2-sigma intervals.

Notable is that the experimental values of the observables fall in the 68% degree of belief interval for all observables except $r\,(^4He)$. This is not surprising as we saw that the grid searched posterior distribution of $r\,(^4He)$ in task 1 does not intersect $f\,T_{1/2}$ in the same region as $E(^4He)$ and $E(^3H)$ does. As such one can deduce that the joint posterior distribution conditioned on all variables will be somewhat more skewed towards the intersection of $f\,T_{1/2}$, $E(^4He)$ and $E(^3H)$, and the resulting posterior predictive distributions will be more accurate for the $f\,T_{1/2}$, $E(^4He)$ and $E(^3H)$ observables. If a posterior predictive distribution were to be calculated from the posterior conditioned on only $r\,(^4He)$ and $E(^4He)$, task 2 case 1, we would likely see that the experimental values of those observables would fall in a 68% degree of belief interval (whilst the posterior predictions for the remaining observables would be worse).

We can also note that there is little to no correlation in the joint posterior predictive distribution of $f\,T_{1/2}$ and both $E\,(^3H)$ and $E\,(^4He)$. This is evident from the absence of a unique symmmetry line for the respective subplots in figure 3. By examining the individual distributions attained in task 1 we can gather that this is because the overlaps between $f\,T_{1/2}$ and the other observables are small and almost at a right angel. As such the $f\,T_{1/2}$ posterior distributions always constrains the parameter values of c_{D} and c_{E} to the same area of overlap. In contrast the joint posterior predictive distribution of $E\,(^4He)$ and $E\,(^3H)$ is clearly positively correlated, a result of their individual posterior distributions overlapping to a great extent. In that case parameters values infered from one individual posterior distribution are consistent with infering those parameter values from the other distribution. Assuming a similar dependence on parameters for $E\,(^4He)$ and $E\,(^3H)$ the resulting posterior predictive distributions will be positively correlated.

4 Extra Task

For the following distribution the value of Q is no longer fixed to 0.33, rather it is treated as an additional parameter that is infered during the MCMC sampling. For this task we are now using 100 walkers and 1000 steps. In a similar manner to task 2 the convergence of the walkers was analyzed and 100 steps was found to be a sufficient burn-in period, see appendix A for detailed convergence analysis.

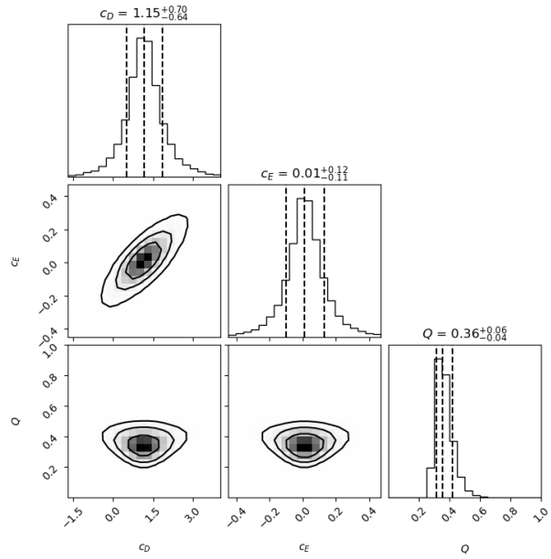


Figure 4: Joint posterior distributions of c_{D}, c_{E} and Q. Dotted lines mark the median and the 68% degree of belief intervals. The contours in the figures correspond to 0.5-, 1-, 1.5- and 2-sigma intervals.

From the sampling we learn that the median of Q is 0.36, with its 68% degree of belief being $0.32 \leq Q \leq 0.42$. We note that the median value of the model parameters change to $c_{\mathrm{D}} = 1.15$ and $c_{\mathrm{E}} = 0.01$ and that their 68% degree of belief intervals have increased to $0.51 \leq c_{\mathrm{D}} \leq 1.85$ and $-0.10 \leq c_{\mathrm{E}} \leq 0.13$. It is expected that the uncertainties of our parameters have increased as we now have introduced a new parameter Q with its own associated uncertainty. The prior distribution for Q is set to a β-distribution (with $a = 3$, $b = 5$). This is an appropriate choice as Q is a parameter expected to range between zero and one. $a < b$ also reflects that we believe Q to be somewhat closer to zero than one. From the figures we also learn that their is no strong correlation between Q and the remaining parameters, c_{D} and c_{E}. From our new joint posterior distribution we can create new predictive posterior distributions.

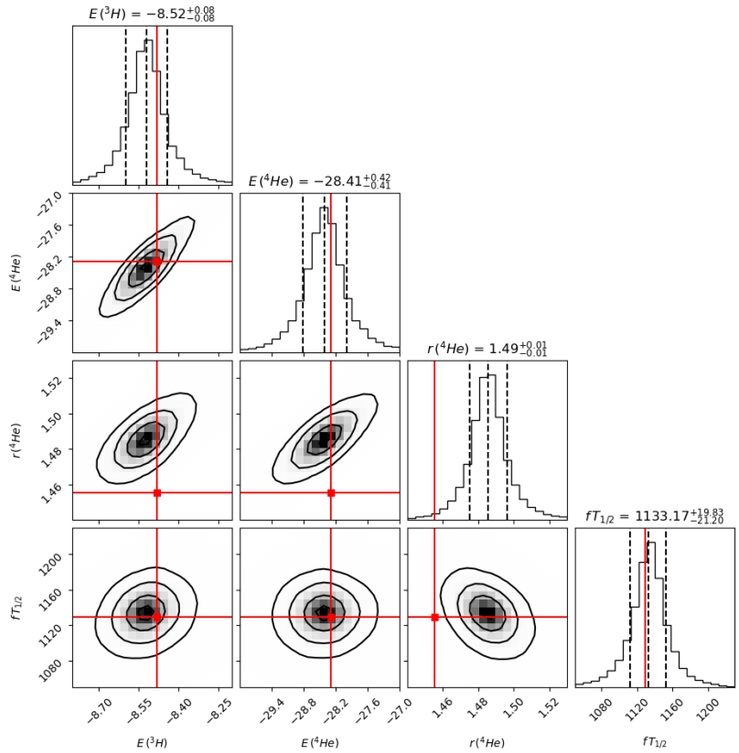


Figure 5: Posterior predictive distribution generated from the posterior distribution where Q was treated as an additional parameter. The red lines and squares denote experimental observable values. Dotted lines mark the median and the 68% degree of belief intervals. The contours in the figures correspond to 0.5-, 1-, 1.5- and 2-sigma intervals.

These distributions are quite similar to those aquired in task 3. The most notable difference is that the uncertainty has increased somewhat for the observables. This is, as touched upon earlier, due to the fact that we have introduced further uncertainty into the model by not fixing our parameter Q to 0.33.

5 Conclusions

We have shown with different methods how the model parameters values change the posterior. We have seen the benefits of using MCMC over a direct grid search method. Even though the grid search was done in only two dimensions, it was hard to justify the additional computational time from increasing the grid points to truly represent the narrow distributions in figure 1a. We've seen how the joint posteriors changes depending on which observables it is conditioned on. During the model checking we noted that the model predicts the experimental values within one standard deviation, except for the observable $r\,(^4He)$. But the model should most likley still be considered relatively good in the context of this project. The modeling in the article by Wesolowski et al. [3] is more sophisticated, and included fewer approximations, which reflects the differences in their results compared to the ones presented here. The results presented in this project have therefore, to some extent, reproduced the results presented by Wesolowski et al. [3].

6 References

References

[1] Thim O. *Renormalization of Chiral Effective Field Theory in the Nucleon-Nucleon Sector*, [Internet], 2021 [Available from: https://odr.chalmers.se/bitstreams/9602afe8-17b4-4099-8047-9c1389a42194/download], (2023-09-26)

[2] Gupta R. *Introduction to Lattice QCD*, [Internet], 1997. [Available from: https://arxiv.org/abs/hep-lat/9807028], (2023-09-26)

[3] Wesolowski S., Svensson I., Ekström A., Forssén C., Furnstahl R. J., Melendez J. A., Phillips D. R. *Rigorous constraints on three-nucleon forces in chiral effective field theory from fast and accurate calculations of few-body observables*, [Internet], 2022. [Available from: https://arxiv.org/pdf/2104.04441.pdf], (2023-09-26)

[4] Forrsén C. *Learning from data*, [Internet], 2023 [Available from: https://cforssen.gitlab.io/tif285-book/content/Intro/welcome.html], (2023-09-27)

[5] Forssén C. *Project I: Effective field theory modeling (Bayesian parameter estimation)*, [Internet, Jupyter Notebook], 2023. [Available from: https://github.com/physics-chalmers/tif285/tree/master/doc/pub/Projects/Project1], (2023-09-27)

A Convergence analysis of walkers in MCMC sampling

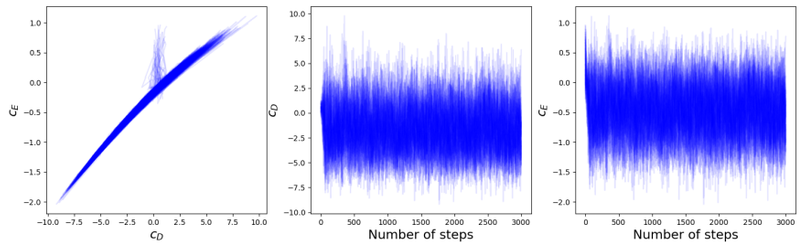


Figure 6: Convergence of walkers during MCMC sampling: Task 2, case 1.

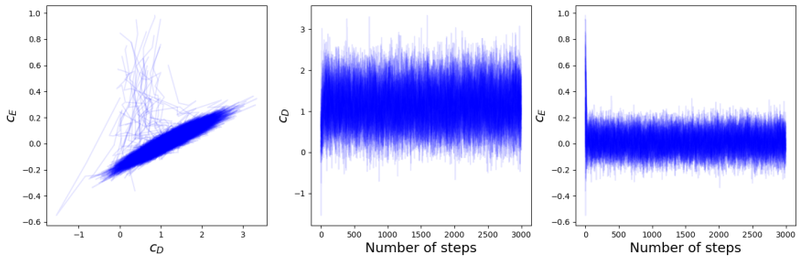


Figure 7: Convergence of walkers during MCMC sampling: Task 2, case 2.

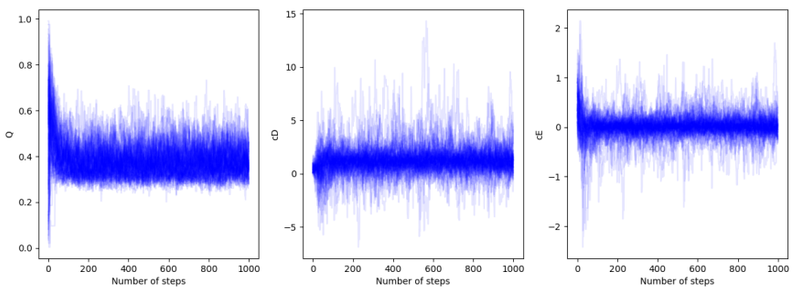


Figure 8: Convergence of walkers during MCMC sampling: Extra task.