Contents

1 Introduction and problem formulation

The goal of this project is to reproduce the result published in paper [2] by Wesolowski et al. regarding the Bayesian approach to the parameters estimation problem. The focus is on the linear toy Effective field theory (EFT) model with coefficients that are unknown but a theoretical expectation of their values might be known in advance.

Bayesian statistical analysis employs the use of the Bayes' theorem of probability to develop the conditional probability distributions of model parameters given a set of experimental data. These probability distributions of model parameters are known as posteriors, the probability of experimental data outcomes (in a distribution sense) is known as the likelihood and the presumed knowledge of the model parameters' probability is known as the prior. The function used to generate the data for this project is given by the eq.(1),

$$g(x) = \left(\frac{1}{2} + \tan\left(\frac{\pi}{2}x\right)\right)^2 \tag{1}$$

and the toy model for the effective field theory is a Taylor expansion of $g(x)$ around $x_0 = 0$. The coefficients in this Taylor series are known as Low Energy Coefficients, or LECs.

The order of the model parameters values might be known in advance and so the prior that best impresses this information on the posterior is expected to reproduce the more accurate result. Additionally, an assessment of how the posterior probability density functions develop with increase of the number of Taylor series terms, called the model complexity, as a result of the selection of prior will serve to further demonstrate the importance of a Bayesian approach to predictions and data processing.

2 Methodology

As a whole the mathematical process can be divided into the following steps: developing the likelihood as a chi-squared probability distribution on the experimental data set, formulating the two priors as functions of these parameters, adding the logarithms of these probabilities and plotting the logarithmic posterior intensity as a function of parameter tuples in a set of corner plots. Monte Carlo Markov Chain (MCMC) sampling will provide sets of the parameters for posteriors, then Bayes evidence ratios could be computed.

2.1 χ^2 measure for likelihood function

As the likelihood $p(D|a, I)$ is defined as the total conditional probability of a data set D being measured given a model parameter vector a, it is natural to suppose that it is derived from a Gaussian distribution known as the χ^2 measure. The measure assumes the role of the variance to be fulfilled by the data measurement error and the square error to correspond to the mean of the distribution. As such, a higher measurement error increases the width of the Gaussian curve while differences between model a prediction and measurement data result in shifts of the mean.

The formulation of the χ^2 measure is given by eq.(2),

$$p_\chi(D_j|a, I) = \frac{1}{\sqrt{2\pi\varepsilon_i^2}} \exp\left\{-\frac{(y_j - f_a(x_j))^2}{2\varepsilon_j^2}\right\} \tag{2}$$

where f_a corresponds to the model given parameter vector a and the points (x, y) are data points of D of the measurement set with error vector ϵ and j iterates over the entire set of data D.

The total likelihood of a given set D is the product $\Pi_j p_\chi(D_j|a, I)$, and is an essential factor in the evaluation of the posterior probability $p_\chi(a|D, I)$ in Bayesian analysis. Because the values of the total likelihood are formed as a product of exponential functions, it is very convenient to define the logarithmic total likelihood,

$$L_\chi(D|a, I) = \sum_j -\frac{1}{2}\log(2\pi\varepsilon_j^2) - \frac{(y_j - f_a(x_j))^2}{2\varepsilon_j^2} \tag{3}$$

As is the nature of the evaluation of posterior probabilities, the logarithmic prior must be added to the logarithmic likelihood. The χ^2 measure implies that predictions for a model f_a very close to the measurement set will heavily dominate the posterior unless the prior is biased against the model.

2.2 Uniform and Gaussian probability distributions for flat and natural priors

In this project, two logarithmic priors are used to evaluate the posterior probability function distributions. The first is a flat distribution prior, which carries the assumption that all model parameters **a** within an interval $U = 100$ are equally valid. The second is a natural distribution prior which carries the assumption that the LECs **a** average to 0, and have a standard deviation of 5.

The logarithmic contribution of the flat prior follows from the uniform distribution:

$$|a| \leq U,\ L_{U=100}(a) = 0, \quad |a| > U,\ L_{U=100}(a) = -\infty \tag{4}$$

The flat prior has no biases on the interval U, but sets the posterior strictly 0 outside of it. The uniform prior is a natural choice in situations where the model parameter vector a is physically equivalent to a state of a system in equilibrium, and is also used as baseline prior in the cases where no restrictive knowledge of **a** is available. Together with the χ-squared distribution this prior creates a posterior virtually identical to the least-squares error ([2], p. 5).

The logarithmic contribution of the natural prior follows from the Gaussian distribution:

$$L_{N(0,5),d}(a) = \sum_d -\frac{1}{2}\log(\pi 50) - \frac{(a_d)^2}{50} \tag{5}$$

Where d iterates over the length of vector a. The natural prior is heavily biased to suppress posteriors for a vectors with values outside the interval $a_j \in [-5,5]$. The choice of mean 0 and standard deviation 5 are motivated by supposing a symmetry for the negative and positive perspectives of the models, and the natural-size of the LECs that suggests their values should lie closer to order 10^0 than 10^1. Together with the χ-squared distribution as likelihood, this creates a more narrow probability distribution function of the posterior ([2], p. 7, fig. 2).

2.3 MCMC sampling and plotting the posterior

Generally, it is infeasible to evaluate multivariate integrals used in Bayesian inference [1], thus in order to circumvent this limitation and to attain a graphical representation of the probability distribution functions of the posteriors, MCMC is used to sample the logarithmic posteriors. This returns a substantial set of potential a vectors and posteriors $L(\mathbf{a}|D,I)$. The corner plotting algorithm is then used to perform a simple marginalization of their associated total logarithmic posterior, producing a set of projected probability distribution functions $L(i,j|D,I)$ where a_i and a_j are select LECs. The algorithm is also used to perform a histogram of the posterior distribution $L(i|D,I)$ producing the pdf for each individual LEC. The results can be found in section 3.1.

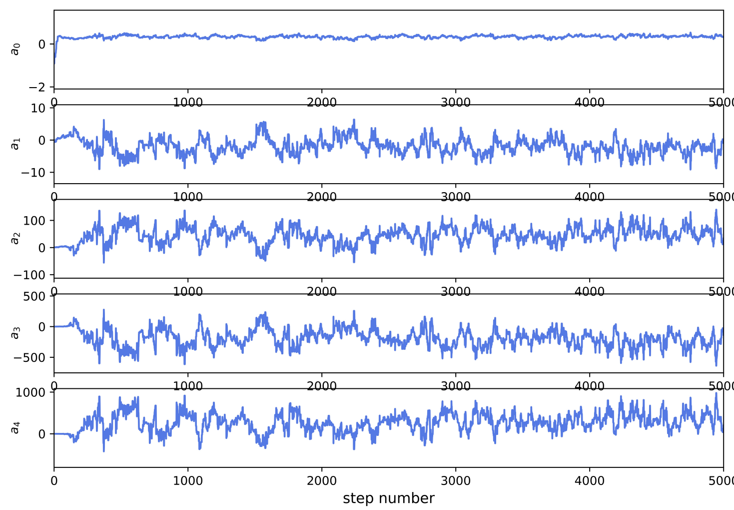

Figure 1: MCMC traces for estimated parameters $a_0, a_1, ..., a_{k+1}$, Model: $k = 4$, $k_{max} = 4$, uniform prior, dataset D1 from [2]. Only iterations up to 5000 are shown.

2.4 Evaluating the prior for fit and complexity

To determine the adequacy of these models and justify the selection of prior, it is necessary to compare the quality of the prediction the models provide with regard to the true function $g(x)$ from eq.(1). The fit of the uniform prior model and the Gaussian prior model to the function $g(x)$ is demonstrated in section 3.2. The validity for model complexity, that is the dimension d of vector $\mathbf{a}$ for models f_a, is determined differently for the flat and the natural priors.

The measure used for the flat prior is χ-square test of independence, evaluated for the following χ^2 value,

$$\chi^2 = \frac{1}{D} \sum_j \frac{(y_j - f_a(x_j))^2}{\varepsilon_j^2} \tag{6}$$

where the degrees of freedom D are counted as $(N - d)$ where N is the total number of measurement points. A higher level of independence indicates the LECs are not greatly affected by other terms in the model f_a. The natural prior is evaluated with a direct comparison of LECs found with higher degrees of model complexity, where the independence of the LEC values from the model complexity provides a qualitative indication of the natural prior's ability to reduce the impact of higher order terms in the Taylor expansion of equation 1 on the posteriors of LECs due to the prior dominating the posterior curve, instead of the likelihood ([2], p. 8). Tabulated evaluation of LECs with higher model complexity is shown in section 3.3.

Additionally, a measure of the Bayesian evidence can be found by multiplying the peak likelihood of an additional LEC with the ratio of the width of this likelihood peak to the width of the new Gaussian prior:

$$p_N(D|M_{d+1}, I) = p_{peak}(D|a_{d+1}, M_{d+1}, I)\frac{\delta_{likelihood}a_{d+1}}{\Delta_{prior}a_{d+1}} \tag{7}$$

If the prior dominates, the evidence will be greatly diminished as the likelihood peak becomes independent of complexity. If the likelihood dominates, the evidence for increased complexity is high when the likelihood peak greatly surpasses the Occam factor $\frac{\delta_{likelihood}a_{d+1}}{\Delta_{prior}a_{d+1}}$ ([2], p. 8). Bayesian evidence will not be evaluated explicitly in the report, but it serves as the hidden mechanism for why the choice of prior affects the consistency of LECs for increased model complexity.

3 Reproduced results

The results from Wesolowski et al. [2] are obtained with the pseudo-random generator seed set to 42, number of walkers is four times the number of estimated parameters ([2], Appendix A). Number of iterations per each walker was set to 25.000 with 50 cutoff length for the emcee Markov Chain Monte Carlo (MCMC) sampling implementation.

3.1 Posterior probability distribution functions

The corner plots represent the projections of the joint pdfs from the MCMC sampling of the logarithmic posteriors for LECs. The logarithmic flat and natural priors used for these posteriors are represented by uniform and Gaussian probability distributions described in section 2.2, respectively. The logarithmic likelihood was derived from the data sets according to the χ^2 measure in section 2.1.

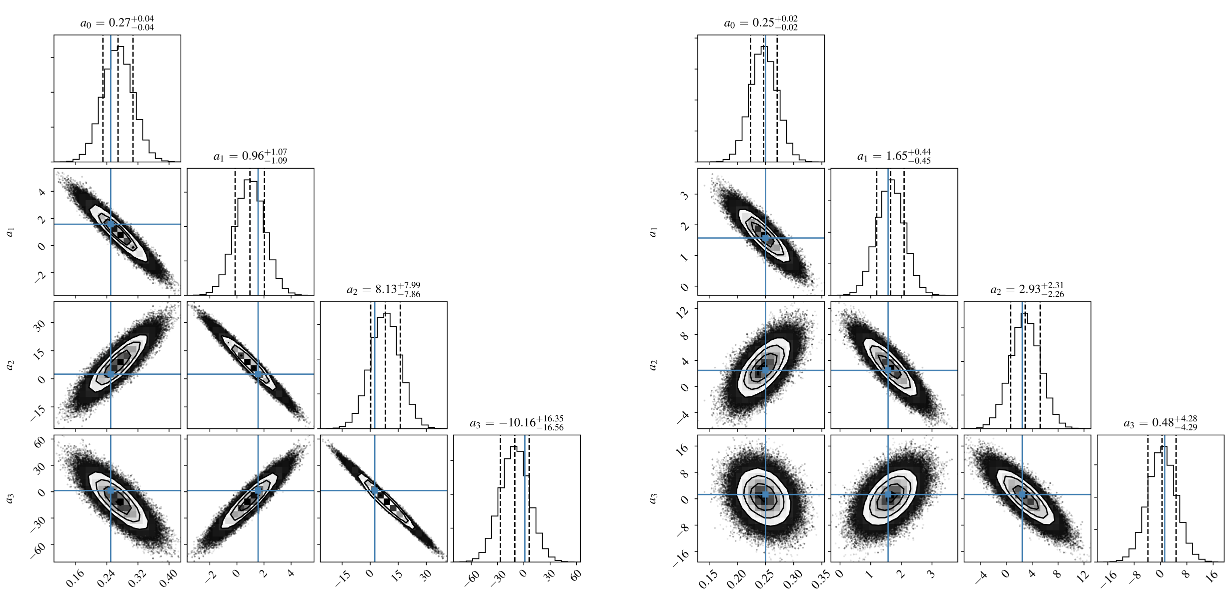

Figure 2: Projected posterior corner-plots based on data set D1_c_5 given along with [2]. This data set has 10 random points with a relative error 5%. *Left:* posteriors pdfs with flat prior. *Right:* posterior pdfs with natural prior. The joint pdfs are represented on the lower triangle plots, marginalized histograms for each LEC are on the diagonal.

The flat prior in Figure 2 gives significant LECs' error margins for higher order terms in the Taylor series f_a, meaning the posterior attempts to account for the truncation error for terms higher than the model complexity with a higher value of these LECs. The natural prior suppresses these contributions and reduces the truncation error, thereby lowering the LEC modes and errors to the natural order ([2], p. 2).

3.2 Prediction plots

In this section we present and compare the the outcomes of the models f_a derived from flat and natural priors at model complexity $k_{max} = 3,\, 4$ with the true function $g(x)$ in the eq.(1) and train data-set.

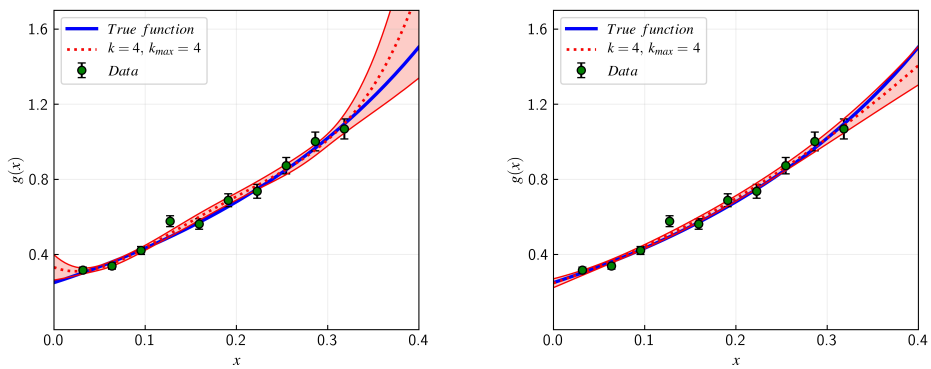

Figure 3: Comparison plot for the model $k = 3$, $k_{max} = 3$ (median and $\pm 68\%$ interval) prediction to the true function $g(x)$ for the default dataset D1. *Left:* model with the flat (uniform) prior. *Right:* model with the natural (Gaussian) prior.

Figure 3 is consistent with the figures in the Wesolowski et al. article ([2], p. 11). The diverging credibility intervals in the flat prior model for x-values outside the data set range is a direct demonstration of underfitting due to least-squares estimation. This is greatly diminished for the natural prior and the model remains a good fit for the entire $[1,\, 1/\pi]$ interval.

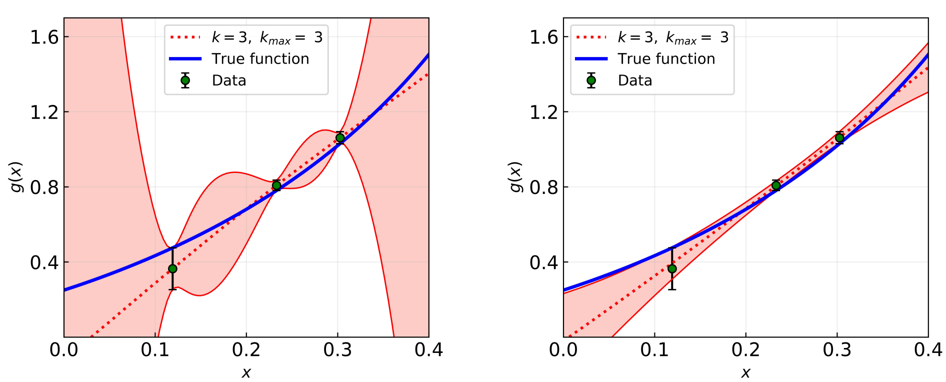

Figure 4: Comparison plot for the model $k = 3$, $k_{max} = 3$ (median and $\pm 68\%$ interval) prediction to the true function $g(x)$ for the newly generated dataset ($N = 3$, $\sigma = 0.1$). *Left:* model with the flat prior. *Right:* model with the natural (Gaussian) prior.

Figure 4 demonstrates the effect of a reduction of data points in establishing the model. The flat prior is shown to have even more under-fitting issues as the median prediction is linear for an evidently non-linear true function, as well as the divergent errors. Because of the low number of data, the model for the natural prior is dominated by the prior, which is why the credibility interval is narrower. However, the natural prior is unable to establish a higher order prediction with only three data points, so it does not reproduce the true function curve.

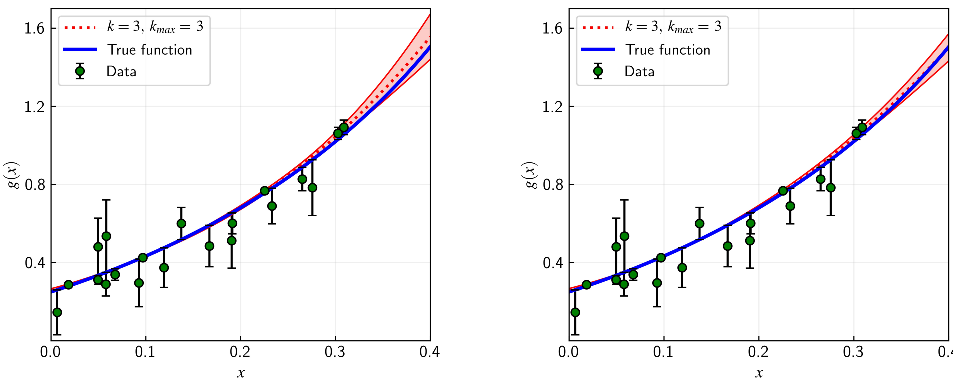

Figure 5: Comparison plot for the model $k = 3$, $k_{max} = 3$ (median and $\pm 68\%$ interval) prediction to the true function $g(x)$ for the newly generated dataset ($N = 20$, $\sigma = 0.1$). *Left:* model with the flat (uniform) prior. *Right:* model with the natural (Gaussian) prior.

Figure 5 demonstrates the effect of increased number of data points in establishing the model. The likelihood is dominating in the models behind both plots due to the higher volume of data, however in the region outside the data set range the natural prior still retains a narrower credibility interval, meaning that the potential of extrapolation for the least squared estimate capability does not increases as much as that of the Bayesian analysis when taking more measurements in the same region.

3.3 Tabulated LEC estimates and goodness of fit

The Table 1 lists estimated medians and 68% credibility intervals for the coefficients $a_0, a_1, \ldots, a_{k+1}$ given different orders of the regression model. The fit of the model with the uniform prior is evaluated via χ^2/dof measure, while predictions with the Gaussian prior are supposed to be evaluated in terms of the Bayesian evidence. Results in the Table 1 are controlled entirely by the k_{max} values.

Table 1: Estimates of the linear regression coefficients – samples from the $L(\mathbf{a}|D,k,k_{max})$ with the corresponding values of the measure of fit.

		Uniform prior					Gaussian prior			
k	k_{max}	$\frac{\chi^2}{dof}$	a_0	a_1	a_2	a_3	a_0	a_1	a_2	a_3
0	0	66.6	0.48 ± 0.01				0.48 ± 0.01			
1	1	2.2	0.20 ± 0.02	2.56 ± 0.12			0.20 ± 0.02	2.56 ± 0.13		
2	2	1.6	0.25 ± 0.03	1.57 ± 0.41	3.33 ± 1.34		0.25 ± 0.03	1.63 ± 0.40	3.14 ± 1.30	
2	3	1.9	0.27 ± 0.04	0.97 ± 1.10	8.06 ± 8.07	-9.95 ± 16.73	0.25 ± 0.03	1.65 ± 0.46	2.95 ± 2.33	0.45 ± 4.37
2	4	2.0	0.33 ± 0.07	-1.84 ± 2.69	44.06 ± 32.56	-179.05 ± 149.30	0.25 ± 0.03	1.65 ± 0.46	2.97 ± 2.38	0.33 ± 4.39
2	5	2.4	0.33 ± 0.07	-1.79 ± 2.59	45.32 ± 32.68	-202.86 ± 174.47	0.25 ± 0.03	1.65 ± 0.47	2.93 ± 2.41	0.47 ± 4.41
2	6	3.2	0.33 ± 0.07	-1.84 ± 2.63	46.33 ± 33.50	-210.32 ± 180.03	0.25 ± 0.03	1.65 ± 0.46	2.97 ± 2.39	0.37 ± 4.42
True values			0.25	1.57	2.47	1.29	0.25	1.57	2.47	1.29

The natural prior surpasses the flat prior for LECs consistency at higher orders of the regression model as can be seen from results in Table 1. The values of a reach the natural size order at the model order $k_{max} = 2$ while the credibility intervals diverge as k_{max} is increased. This makes it clear that the Bayesian analysis that informs the evaluation of LECs with naturalness conditions requires far lower model complexity to provide a satisfyingly accurate outcomes, while the ordinary least squares result in the greatly imprecise estimates with modes far higher than expected as the complexity is increased i.e. having the over-fitted model. The lower value of χ^2 test of independence further indicates the higher order terms in f_a impact the value of lower LECs, causing under-fitting.

The differences for χ^2 test of independence between the different data sets indicate that the number of data points has an impact on the co-dependency of the LECs in the case of a flat prior. The difference of LEC values for the different data sets suggests that the Bayesian evidence takes a different level of model complexity to reach saturation. As we discussed in section 2.4 this may depend on the Occam factor being directly impacted by the differences in width of likelihood peaks between the data sets.

4 Conclusion

In this report we have reproduced the published results on LEC estimation for the elementary EFT model by means of the Bayesian inference. We illustrated that within the Bayesian framework it becomes possible to naturally incorporate in our models all the knowledge about the phenomenon of interest including expectations as well as the uncertainties associated to experimental data. The Bayesian approach demonstrates a significantly lower rate of overfitting, a lesser dependence of confidence interval on the amount of data, and a greater consistency for evaluated LEC values with increasing model complexity. Therefore the conclusion is that the use of Bayesian statistical analysis opens the way towards more interpretable mathematical models by making them consistent with experiments and background theory.

References

[1] RJ Furnstahl, DR Phillips, and S Wesolowski. A recipe for eft uncertainty quantification in nuclear physics. *Journal of Physics G: Nuclear and Particle Physics*, 42(3):034028, 2015.

[2] S Wesolowski, N Klco, R J Furnstahl, D R Phillips, and A Thapaliya. Bayesian parameter estimation for effective field theories. *Journal of Physics G: Nuclear and Particle Physics*, 43(7):074001, May 2016.