EPJ Web of Conferences **245**, 06026 (2020) https://doi.org/10.1051/epjconf/202024506026
CHEP 2019

Constraining effective field theories with machine learning

Johann Brehmer[1], *Kyle* Cranmer[1], *Irina* Espejo[1], *Alexander* Held[1,*], *Felix* Kling[2], *Gilles* Louppe[3], and *Juan* Pavez[4]

[1]New York University, United States
[2]SLAC National Accelerator Laboratory, United States
[3]University of Liège, Belgium
[4]Federico Santa María Technical University, Chile

Abstract. An important part of the Large Hadron Collider (LHC) legacy will be precise limits on indirect effects of new physics, framed for instance in terms of an effective field theory. These measurements often involve many theory parameters and observables, which makes them challenging for traditional analysis methods. We discuss the underlying problem of "likelihood-free" inference and present powerful new analysis techniques that combine physics insights, statistical methods, and the power of machine learning. We have developed `MadMiner`, a new Python package that makes it straightforward to apply these techniques. In example LHC problems we show that the new approach lets us put stronger constraints on theory parameters than established methods, demonstrating its potential to improve the new physics reach of the LHC legacy measurements. While we present techniques optimized for particle physics, the likelihood-free inference formulation is much more general, and these ideas are part of a broader movement that is changing scientific inference in fields as diverse as cosmology, genetics, and epidemiology.

1 The inference challenge

Measurements of particle collisions at the Large Hadron Collider (LHC) provide a wealth of high-dimensional data x. This data is used to measure phenomena predicted by the Standard Model of particle physics (SM), and to search for indications of physics from beyond the SM. With an effective field theory (EFT) approach, possible variations from the SM can be parameterized and studied systematically [1–3]. This requires the simultaneous measurement of many parameters θ. In the EFT context, these parameters are known as Wilson coefficients, but the following discussion applies similarly for other physical properties.

The relation between parameters and the data x is given by the differential cross-section $\mathrm{d}\sigma(x|\theta)/\mathrm{d}x$. Normalized by the total cross-section, the expression becomes

$$p(x|\theta) = \frac{1}{\sigma(\theta)} \frac{\mathrm{d}\sigma(x|\theta)}{\mathrm{d}x}, \tag{1}$$

which is also called likelihood $L(\theta) = p(x|\theta)$. Likelihood ratios

$$r(x|\theta, \theta_0) = \frac{p(x|\theta)}{p(x|\theta_0)} \tag{2}$$

*e-mail: alexander.held@nyu.edu

are a central ingredient for statistical inference at the LHC [4].

Samples of collision processes at the LHC are generated via a chain of Monte Carlo simulators. They describe the the hard scattering of partons z_p, subsequent parton showering and hadronization, as well as interactions with the detector and signal propagation up to a set of measurements of reconstructed quantities x. The likelihood can be factorized into different contributions,

$$p(x|\theta) = \int \underbrace{p(x|z_\mathrm{d})p(z_\mathrm{d}|z_\mathrm{s})p(z_\mathrm{s}|z_\mathrm{p})p(z_\mathrm{p}|\theta)}_{p(x,z|\theta)} \mathrm{d}z_\mathrm{p}\mathrm{d}z_\mathrm{s}\mathrm{d}z_\mathrm{d}. \tag{3}$$

The parton-level configuration is summarized as z_p, z_s represents the state after parton shower and hadronization, and z_d the state after interactions with the detector. Performing the integral over the huge space of latent variables in equation (3) is impossible, the likelihood $p(x|\theta)$ is intractable. By running the simulation chain, it is possible to generate samples drawn from $p(x|\theta)$ for various settings of θ.

2 Established inference techniques

Several methods are commonly used to perform inference in cases with intractable likelihoods. This section highlights two examples; a discussion of additional methods can be found in reference [5]. Inference in these settings is called *likelihood-free* and *simulation-based* [6].

2.1 Histograms of summary statistics

For certain kinds of parameter measurements, most of the information about a parameter of interest can be contained in summary statistics $v(x)$. Examples for this are an invariant mass when measuring a resonance, and the output of a multivariate method that distinguish between signal and background processes when measuring the cross-section of the signal. The likelihood $p(v(x)|\theta)$ can be obtained with density estimation techniques. Histograms are often used for this purpose in high energy physics.

This approach has two severe limitations. The number of samples required to fill the histograms scales exponentially with the dimension of v, and can thus not be used for simultaneous measurements of many parameters. It furthermore requires the design of appropriate summary statistics, a difficult task when measuring many parameters that affect the measurement x in diverse ways.

2.2 Matrix element method

The matrix element method (MEM) [7] provides an approximate expression for the intractable likelihood $p(x|\theta)$ given by equation (3). At parton-level, the likelihood $p(z_\mathrm{p}|\theta)$ can be calculated. The relation between parton-level kinematics and the measured kinematics x is approximated with simple so-called transfer functions $\hat{p}(x|z_\mathrm{p})$. The full likelihood is obtained by marginalizing over the parton-level kinematics,

$$\hat{p}(x|\theta) = \int \hat{p}(x|z_\mathrm{p})p(z_\mathrm{p}|\theta)\mathrm{d}z_\mathrm{p}. \tag{4}$$

Several approximations are made in practice to limit the computational expense of calculating the above integral. The transfer functions cannot capture all shower and detector effects, and Dirac delta functions are commonly used for several integration dimensions to limit the time required for integration. The matrix element calculation is usually done at leading order, and extensions to next-to-leading order come at a considerable computational cost [8]. There is no amortization of the integration time, an expensive integration is required for every event that needs to be processed.

3 New techniques: mining gold

A recent set of publications outlines a new approach to likelihood-free inference [9–12]. Core ideas and results are briefly summarized in the following. While the likelihood $p(x|\theta)$ in equation (3) is intractable, the joint likelihood ratio

$$r(x, z|\theta, \theta_0) = \frac{p(x, z|\theta)}{p(x, z|\theta_0)} = \frac{p(z_\mathrm{p}|\theta)}{p(z_\mathrm{p}|\theta_0)} \tag{5}$$

is tractable. The required parton-level likelihoods $p(z_\mathrm{p}|\theta)$ are also used in the MEM, they are obtained as the product of phase space factors, squared matrix elements and parton distribution functions (PDFs). The joint likelihood ratio is scattered around the likelihood ratio of interest, $r(x|\theta, \theta_0)$.

3.1 The usefulness of gold

Consider the functional

$$L_r\left[\hat{r}(x)\right] = \int p(x, z|\theta_0)\left[r(x, z|\theta, \theta_0) - \hat{r}(x)\right]^2 \mathrm{d}x\mathrm{d}z. \tag{6}$$

Calculus of variations shows that it is minimized by

$$r^*(x) = \arg\min_{\hat{r}} L_r\left[\hat{r}(x)\right] = \frac{1}{p(x|\theta_0)} \int p(x, z|\theta_0) r(x, z|\theta, \theta_0)\mathrm{d}z = r(x|\theta, \theta_0). \tag{7}$$

The minimization of L_r thus yields an expression for the intractable likelihood ratio $r(x|\theta, \theta_0)$. This requires sampling from $p(x, z|\theta_0)$, achieved by running the relevant simulators, and calculation of the tractable joint likelihood ratio $r(x, z|\theta, \theta_0)$.

Similarly to the joint likelihood ratio, the joint score

$$t(x, z|\theta_0) = \nabla_\theta \log p(x, z|\theta)\Big|_{\theta_0} = \frac{\nabla_\theta p(z_\mathrm{p}|\theta)}{p(z_\mathrm{p}|\theta)}\Big|_{\theta_0} \tag{8}$$

can also be calculated. It describes how the joint likelihood varies with the parameters θ. The joint score can be used to define a functional L_t, which can be minimized to approximate the score,

$$t(x|\theta_0) = \arg\min_{\hat{t}} L_t\left[\hat{t}(x)\right]. \tag{9}$$

One possibility to obtain approximations for the likelihood ratio $r(x|\theta, \theta_0)$ and score $t(x|\theta_0)$ is implementing $\hat{r}(x)$ and $\hat{t}(x)$ as neural networks, and minimizing the loss functions L_r and L_t. The crucial ingredient to the loss functions is the "gold" mined from the simulators; it consists of the parton-level kinematics z_p, as well as the joint likelihood ratio and the joint score.

3.2 General approach

Figure 1 shows the general workflow of this new approach. Events are simulated for parameter settings θ, resulting in parton-level kinematics z_p and measured observables x. A neural network learns an approximate expression $\hat{r}(x|\theta)$ for the likelihood ratio, using the information extracted from the simulators. This approximate likelihood ratio can then be used directly for inference.

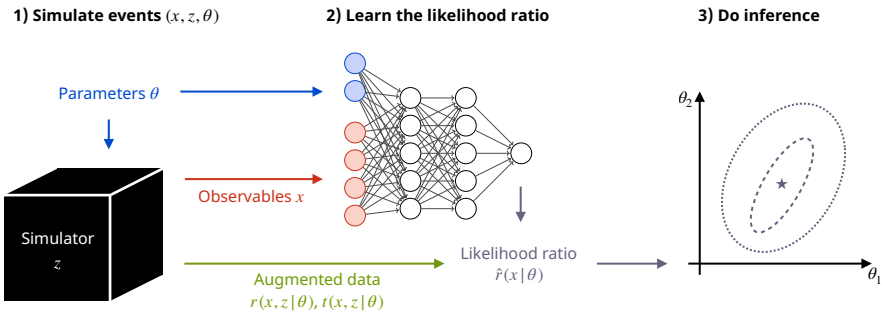

Figure 1. The likelihood ratio $r(x|\theta)$ can be approximated with a neural network by minimizing a suitable loss function which makes use of the joint likelihood ratio $r(x, z|\theta)$ and the joint score $t(x, z|\theta)$ extracted from the simulators. Inference can then be performed directly with the approximation $\hat{r}(x|\theta)$.

3.3 Available algorithms

A range of algorithms is available to provide approximate likelihood ratios. They differ in their objectives; for some algorithms the loss function includes the joint likelihood ratio, for others the joint score, or even both. CARL [13] relies on neither of these two quantities. ROLR [10] and ALICE [12] make use of the joint likelihood ratio. CASCAL [10] uses the joint score, while RASCAL [10] and ALICES [12] use both the joint likelihood ratio and the joint score. The two methods SALLY and SALLINO [10] use the joint score to provide locally optimal observables.

4 The MadMiner package

`MadMiner` is a software package that implements the ideas and techniques described in the previous section in a ready-to-use python package. It is available on GitHub [14] and documented in reference [15]. Interactive tutorials are included in the repository, and provide a good way for getting started with `MadMiner`. The package wraps around `MadGraph5_aMC@NLO` [16], `Pythia` 8 [17] and `Delphes` [18]. Its modular nature allows to extend this to additional simulators.

4.1 Exemplary workflow

An exemplary `MadMiner` workflow starts with the generation of events on parton-level, using `MadGraph5_aMC@NLO`. The dependence on parameters θ is expressed by event weights, which can be used to reweight the generated set of events to describe distributions at various benchmark settings of θ [19]. A morphing procedure [10, 20] allows to interpolate the weights between benchmark points and to extrapolate to any point in parameter space. The event simulation concludes with `Pythia` 8 and `Delphes`. Observables x are then defined and calculated for every event. They serve as the input to the neural network that will approximate the likelihood ratio. A set of unweighted datasets is produced to have representative measurements of the observables for various settings of θ. The originally produced dataset, alongside the reweighting and morphing procedure, forms the input to these new datasets. With these unweighted datasets available, the machine learning step can proceed. Various architectures

and loss functions are available to make use of the different methods summarized in section 3. The neural networks are implemented in `PyTorch` [21]. When approximating the likelihood ratio $r(x|\theta, \theta_0)$, the output of the neural network can then be used for statistical inference. The network output can also be treated as an observable for a traditional histogram-based approach.

5 Applications: constraining effective field theories

Significant sensitivity increases for multi-parameter measurements are possible when using the new techniques outlined in section 3 instead of the established methods described in section 2. This section briefly describes examples of this.

5.1 Achieving optimal performance

The vector boson fusion process, with Higgs boson decays to a final state of four charged leptons, is sensitive to the two Wilson coefficients f_W and f_{WW}. The sensitivity to these two parameters is studied in references [9, 10].

Figure 2 (left) shows a comparison of the sensitivity of multiple approaches to these coefficients. The exclusion contours obtained with an approach based on Sally significantly improve the sensitivity compared to a measurement based on a two-dimensional histogram. Additional improvements are obtained with Rascal. It produces results indistinguishable from those obtained by using the true likelihood ratio $r(x|\theta)$, which can be calculated in this example at parton-level. In order to obtain the same exclusion limits with the histogram-based approach as with Rascal, around twice the amount of integrated luminosity is needed [9].

5.2 Inclusion of systematic uncertainties

The Wilson coefficients c_{uG}, c_u, c_G affect the top quark and Higgs boson couplings in $t\bar{t}H$ production. Reference [15] studies the sensitivity of various approaches to these parameters, using the $t\bar{t}H$ process with Higgs boson decays to diphoton final states. The analysis includes systematic uncertainties in the form of PDF and scale variations. These uncertainties can be parameterized, such that the neural networks used in this study can learn the dependence of the likelihood ratio $r(x|\theta)$ on both the parameters of interest and nuisance parameters [22].

Figure 2 shows 68% confidence level limits as a function of two of the Wilson coefficients when setting $c_u = 0$. An approach based on a histogram of the diphoton transverse momentum, shown in green, improves the sensitivity over just using rate information, shown in gray. The multivariate approach with Sally results in a significant increase in sensitivity over the histogram-based measurement. When including systematic uncertainties into the measurement, the sensitivity slightly decreases. The solid lines show the confidence level limits with profiled systematic uncertainties; these limits are decreased compared to the measurement without systematic uncertainties, visualized with dashed lines.

5.3 High energy physics and beyond

A study of the performance of a `MadMiner`-based approach for the measurement of EFT operators via the WH process is provided in reference [23]. Reference [24] demonstrates an application outside the field of high energy physics. It describes the use of the same techniques to infer dark matter substructure via gravitational lensing and demonstrates how the ideas behind `MadMiner` are broadly applicable.

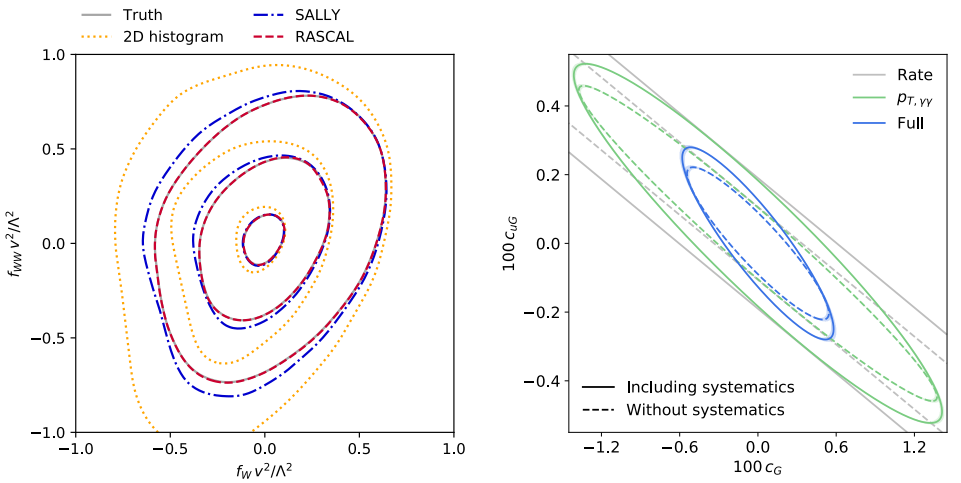

Figure 2. Left: Expected 68%, 95%, and 99.7% confidence level limits as a function of the two Wilson coefficients f_W and f_{WW}, rescaled with the Higgs vacuum expectation value v and scale of new physics Λ. The measurement is performed with events from Higgs boson production via vector boson fusion. The exclusion limits with SALLY improve the performance over a histogram-based approach. The limits with RASCAL are indistinguishable from those obtained by using the true likelihood ratio. Additional details are provided in references [9, 10]. **Right:** Expected 68% confidence level limits as a function of two Wilson coefficients, c_G and c_{uG}, from an analysis of the ttH process with Higgs boson decays to a diphoton final state. The results shown in green are based on using a histogram binned in the transverse momentum of the diphoton system, the results in blue use SALLY. The gray line corresponds to the limits when only making use of rate information. More details are provided in reference [15].

6 Conclusions

Traditional methods for likelihood-free inference do not scale well to simultaneous measurements of many parameters. A new family of methods has been developed recently to address this challenge. By combining matrix element information and machine learning techniques, likelihood ratios can be approximated efficiently and used directly for inference. Studies of this new approach for multi-parameter measurements show promising performance.

`MadMiner` is a new python package that automates the relevant workflow. It implements the full analysis chain for phenomenological studies, and is planned to be integrated into the software frameworks for use by the LHC experiments. The ideas behind `MadMiner` are applicable in many fields, and not constrained to use in high energy physics.

This work was supported by the U.S. National Science Foundation (NSF) under the awards ACI-1450310, OAC-1836650, and OAC-1841471. It was also supported through the NYU IT High Performance Computing resources, services, and staff expertise. JB, KC, and GL are grateful for the support of the Moore-Sloan data science environment at NYU. KC is also supported through the NSF grant PHY-1505463, FK is supported by the U.S. Department of Energy grant DE-AC02-76SF00515, while JP is partially supported by the Scientific and Technological Center of Valparaíso (CCTVal) under Fondecyt grant BASAL FB0821. GL is recipient of the ULiège - NRB Chair on Big Data and is thankful for the support of NRB.

References

[1] S. Weinberg, Phys. Rev. Lett. **43**, 1566 (1979)

[2] W. Buchmuller, D. Wyler, Nucl. Phys. B **268**, 621 (1986)

[3] B. Grzadkowski, M. Iskrzynski, M. Misiak, J. Rosiek, JHEP **10**, 085 (2010), 1008.4884

[4] G. Cowan, K. Cranmer, E. Gross, O. Vitells, Eur. Phys. J. C **71**, 1554 (2011), [Erratum: Eur. Phys. J. C **73**, 2501 (2013)], 1007.1727

[5] J. Brehmer, K. Cranmer, I. Espejo, F. Kling, G. Louppe, J. Pavez, *Effective LHC measurements with matrix elements and machine learning*, in *19th International Workshop on Advanced Computing and Analysis Techniques in Physics Research* (2019), 1906.01578

[6] K. Cranmer, J. Brehmer, G. Louppe, *The frontier of simulation-based inference* (2019), 1911.01429

[7] K. Kondo, J. Phys. Soc. Jap. **57**, 4126 (1988)

[8] T. Martini, P. Uwer, JHEP **09**, 083 (2015), 1506.08798

[9] J. Brehmer, K. Cranmer, G. Louppe, J. Pavez, Phys. Rev. Lett. **121**, 111801 (2018), 1805.00013

[10] J. Brehmer, K. Cranmer, G. Louppe, J. Pavez, Phys. Rev. D **98**, 052004 (2018), 1805.00020

[11] J. Brehmer, G. Louppe, J. Pavez, K. Cranmer, *Mining gold from implicit models to improve likelihood-free inference* (2018), 1805.12244

[12] M. Stoye, J. Brehmer, G. Louppe, J. Pavez, K. Cranmer, *Likelihood-free inference with an improved cross-entropy estimator* (2018), 1808.00973

[13] K. Cranmer, J. Pavez, G. Louppe, *Approximating Likelihood Ratios with Calibrated Discriminative Classifiers* (2015), 1506.02169

[14] J. Brehmer, F. Kling, I. Espejo, K. Cranmer, *MadMiner*, https://github.com/diana-hep/madminer

[15] J. Brehmer, F. Kling, I. Espejo, K. Cranmer, Comput. Softw. Big Sci. **4**, 3 (2020), 1907.10621

[16] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.S. Shao, T. Stelzer, P. Torrielli, M. Zaro, JHEP **07**, 079 (2014), 1405.0301

[17] T. Sjostrand, S. Mrenna, P.Z. Skands, Comput. Phys. Commun. **178**, 852 (2008), 0710.3820

[18] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, M. Selvaggi (DELPHES 3), JHEP **02**, 057 (2014), 1307.6346

[19] O. Mattelaer, Eur. Phys. J. C **76**, 674 (2016), 1607.00763

[20] The ATLAS Collaboration, *A morphing technique for signal modelling in a multidimensional space of coupling parameters* (2015), ATL-PHYS-PUB-2015-047

[21] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga et al., *Pytorch: An imperative style, high-performance deep learning library* (2019), 1912.01703

[22] P. Baldi, K. Cranmer, T. Faucett, P. Sadowski, D. Whiteson, Eur. Phys. J. C **76**, 235 (2016), 1601.07913

[23] J. Brehmer, S. Dawson, S. Homiller, F. Kling, T. Plehn, JHEP **11**, 034 (2019), 1908.06980

[24] J. Brehmer, S. Mishra-Sharma, J. Hermans, G. Louppe, K. Cranmer, *Mining for Dark Matter Substructure: Inferring subhalo population properties from strong lenses with machine learning* (2019), 1909.02005