The DL Advocate: Playing the devil's advocate with hidden systematic uncertainties

Aleksandr Iniukhin[1] **Andrei Golutvin**[2] **Andrea Mauri**[2] **Patrick Owen**[3] **Nicola Serra**[3]
Andrey Ustyuzhanin[1,4,5]
[1]Yandex School of Data Analysis
[2]Imperial College of London, United Kingdom
[3]Physik-Institut, Universität Zürich, CH-8057 Zürich, Switzerland
[4]Constructor University, 28759 Bremen, Germany
[5]Institute for Functional Intelligent Materials, National University of Singapore, 117544, Singapore
aleksandr.iniukhin@cern.ch

Abstract

We propose a new method based on machine learning to *play the devil's advocate* and investigate the impact of unknown systematic effects in a quantitative way. This method proceeds by reversing the measurement process and using the physics results to interpret systematic effects under the Standard Model hypothesis. We explore this idea through a combination of gradient descent and optimisation techniques, its application and potentiality is illustrated with an example that studies the branching fraction measurement of a heavy-flavour decay. We find that the size of a hypothetical hidden systematic uncertainty strongly depends on the kinematic overlap between the signal and normalisation channel.

1 Introduction

Addressing hidden systematic effects in physics, such as "unknown-unknowns," is challenging. Independent confirmations, like the Higgs boson discovery by ATLAS and CMS, enhance confidence, but future confirmations might be financially restrictive due to collaboration size and experiment complexity. The discovery beyond the Standard Model (SM) aims to resolve its open questions and demands rigorous validation, especially for unique experiments with no direct counterparts. Typically, validation is qualitative, based on control channel tests or observing multiple new physics signatures. Our paper proposes a method for quantitative assessment, employing deep learning to challenge deviations from the SM, effectively treating observed measurements as potential experimental effects within SM parameters. Concretely, one can consider the set of measurements of the experiment as a system of equations:

$$\mathcal{F}(\tilde{\eta}, \Omega_i) = M_i \Longrightarrow \begin{cases} \mathcal{F}(\tilde{\eta}, \Omega_1) = M_1 \\ \mathcal{F}(\tilde{\eta}, \Omega_2) = M_2 \\ \dots \end{cases}, \tag{1}$$

where $\mathcal{F}$ represents the measurement process, $\tilde{\eta}$ are parameters in common to all measurements, such as the detector response, and Ω_i are parameters specific to a particular measurement, such as theoretical parameters.

When a measurement deviates from its SM prediction, it is tempting to interpret the observed deviation as a sign of physics beyond the SM. However, such an important claim must be supported by an equally strong confidence in the understanding of the experimental apparatus. The idea developed in this paper reverses the classical reasoning and, instead of attributing the observed deviation to

Machine Learning and the Physical Sciences Workshop, NeurIPS 2023.

physics beyond the SM, it starts from the SM hypothesis by fixing the theory parameters to their predicted values, uses the simulation to model $\mathcal{F}$ and uses a neural network to find possible values of $\tilde{\eta}$ that reproduce the observed measurements M_i. In other words, it tries to find possible detector effects that can cause the observed deviation.

As demonstrated, these effects could represent mismodellings of detector efficiency but can also encompass broader analysis assumptions. Our method's strength lies in generating falsifiable mismodelling predictions via parameter values. Such values can embody either "high-level" data, like particle kinematics which demand specific tuning, or "low-level" metrics, such as material budget, offering broader applicability but challenging real-time tuning. Recent trends leverage machine learning for systematic uncertainties, mainly focusing on optimizing statistical inference or integrating known systematic effects. Our method, however, seeks to quantify unacknowledged or undervalued systematic effects, ensuring a rigorous control vital for authentic scientific discoveries.

We apply this method to a simple toy example of a branching fraction measurement of a particle decay and restrict our attention to a potential mis-modelling of the efficiency. Such measurements are often normalised to a decay mode with a known branching fraction and ideally the same final state as the signal, which cancels systematic uncertainties due to efficiency mismodelling to a high degree. It is therefore an ideal testing ground for our approach.

2 Methodology

Taking the example laid out in Eq. 1, we consider a set of measurements M_i, where one is the measurement of interest and the others are control channels which are used to validate the analysis and to check for systematic uncertainties. For each of these measurements there are some observed candidates N_i and an associated efficiency e_i, so that $M_i = \frac{N_i}{e_i}$. The candidates of each channel are characterised by a set of variables (features) such as the kinematics of the produced particles. Differences in these distributions, together with a detection efficiency which can depend on the same kinematic variables, can result in different total efficiencies between the signal and control channels.

We describe possible mismodelling of the efficiency with a weighting function $w(x)$, which depends on the kinematic variables x of the event. Values of $w(x) = 1$ correspond to a perfect modelling of the efficiency, while values below/above unity correspond to efficiency under/over estimated. The key idea is that, while the detector response depends entirely on the kinematics of the single event, the total signal/control channel efficiency can suffer from different biases once integrated over the individual kinematic distribution of each decay channel. We can then define the *true* total efficiency for a given channel i as

$$e_i = \mathbb{E}_{x\sim p(x|i)}\left[w(x) \times \hat{\epsilon}(x)\right] , \tag{2}$$

where $\hat{\epsilon}(x)$ is the per-event *estimated* efficiency in the experiment and the expectation value indicates the weighted average over the kinematic distribution of each decay channel $p(x|i)$. Since our goal is to study the impact of possible mismodelling of the efficiency in a given set of measurements we can safely assume $\hat{\epsilon}(x) = 1$ without loss of generality. This simplifies the expression of the per-channel efficiency to

$$e_i = \mathbb{E}_{x\sim p(x|i)}\left[w(x)\right] \equiv \frac{1}{n_i}\sum_{k=1}^{n_i} w(x_{k,i}) , \tag{3}$$

where we approximated the expectation value with a sum over a large number n_i of simulated events $\{x_k\}_i$, where i labels the different decay channels.

Control channels provide important constraints on how well the efficiency is estimated. They are typically selected with topology and kinematics similar to the signal decay mode in order to maximise the phase space overlap between channels. The result of the measurements obtained on such control channels can then be compared to known reference values, *e.g.* existing precise measurements from other experiments or clean SM predictions. If a good agreement is found, a certain level of confidence can be ascribed to the estimation of the efficiency, at least for what concerns the kinematic regions populated by the control channels. The requirement that measurements performed on the control channels must be compatible with a certain reference can be formulated as

$$M_i \in [M_i^{low};\ M_i^{high}] \tag{4}$$

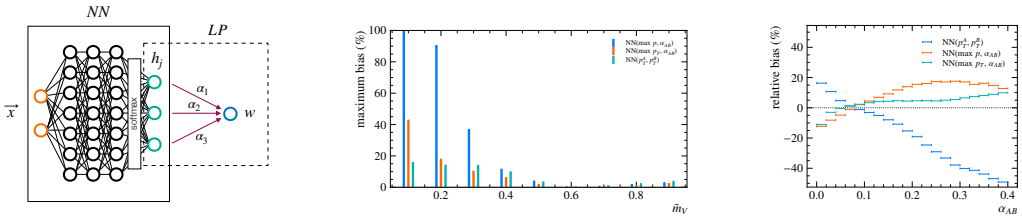

Figure 1: **Left**: Schematic view of the DL Advocate algorithm. The algorithm is formed by two parts: the output of a neural network (NN) is passed to a linear programming (LP) solver which returns the final weight for each event. **Center**: maximum allowed bias obtained on $P \to VC$ branching ratio by running the DL Advocate algorithm with the three studied sets of features at different masses. **Right**: Expected relative bias on differential branching fraction measurement for the $P \to XC$ control channel as function of the opening angle. The algorithm has been trained with a signal of $\tilde{m}_V = 0.3$, while the result for the three tested configurations is shown with different colours.

which reduces to

$$e_i \in [V_i^{low};\ V_i^{high}]\,, \tag{5}$$

once we restrict the attention to the sole role of the efficiency in the measurement. Here V_i^{low} and V_i^{high} are the values that bound the efficiency for the control measurements to pass scrutiny.

The goal of this paper is to find regions of the kinematic space $\vec{x}$ where a mismodelling of the efficiency can have a significant impact in the signal measurement while the effect on the control channels remains within the constraints of Eq. 5. Given these constraints, the intuition is that this problem can be thought as a classification task between the signal and control channels using the kinematic variables provided in the space $\vec{x}$.

2.1 DL Advocate algorithm

The algorithm is formed of two main parts, as shown schematically in Fig. 1. The first is a fully connected neural network which resembles a multi-classification algorithm. The inputs are a set of features x whereas the output h_j is a classification score for each decay hypothesis. The last layer of the NN is normalised with a softmax activation function which enforce $h_j(x) \geq 0$ and $\sum_j h_j(x) = 1$. Details on the technical implementation of the neural network are given in [1].

The second part of the algorithm consists of a linear combination of the NN output, also referred to as linear programming (LP), which defines the final per-event weight

$$w(x) = \sum_j \alpha_j h_j(x) \quad \text{with} \quad \alpha_j \geq 0\,. \tag{6}$$

Combining this with Eq. 3 and moving to a vectorial notation we can express the total per-channel efficiency as

$$\vec{e} = H\vec{\alpha}\,, \tag{7}$$

where we have introduced the H matrix which is defined as

$$H_{i,j} = \frac{1}{n_i}\sum_k h_j(x_{k,i})\,. \tag{8}$$

In such a scenario H takes the form of the identity matrix and α_j can be individually chosen to satisfy all possible combination of efficiencies, *i.e.* we can choose $\alpha_{(j|j\in\mathcal{C})} = 1$ for all control channels $\mathcal{C}$ in order to perfectly satisfy their efficiency constraints, and arbitrarily move $\alpha_{(j|j=s)}$ for the signal channel s to get any possible values for the signal efficiency e_s.

In general, however, the H matrix will have non-diagonal terms that correlate the efficiency of the different decay channels. The measurements carried out on the control channels, therefore, provide non trivial constraints on the signal efficiency.

2.2 Optimisation procedure

The goal of this algorithm part is to find the solution that maximises possible shifts in the signal efficiency while maintaining the control measurements within their allowed range, as defined in Eqs. 4 and 5. This is achieved with an iterative procedure:

i) the NN is pretrained as a simple classifier between channels;

ii) for a given θ we derive the matrix H and comput the optimal values of $\vec{\alpha}$;

iii) with $\vec{\alpha}$, the NN parameters are updated to improve the solution.

Details of the algorithm are provided in [1].

3 A particle physics example: measurement of a branching fraction

We illustrate the use of the DL Advocate method with the example of a branching fraction measurement. Branching fractions are typically measured as a ratios of a given signal decay channel with respect to a normalisation channel

$$\mathcal{B}_{\rm sig} = \frac{N_{\rm sig}}{N_{\rm norm}} \cdot \frac{e_{\rm norm}}{e_{\rm sig}} \cdot \mathcal{B}_{\rm norm} \tag{9}$$

where $\mathcal{B}$ is the branching fraction, N is the observed yield and e is the detector efficiency. It is evident from Eq. 9 that a problem in the efficiency estimation would unavoidably lead to an incorrect determination of the branching ratio.

For this example, we consider the branching fraction measurement of a hypothetical decay of the type $P \to VC$, where $V \to AB$ is a hypothetical intermediate resonance decaying into two given particles A and B, and C is a companion final-state particle. Branching ratios are typically normalised relative to decays with similar topology, which we denote here with $P \to XC$ channel, with $X \to AB$, where the branching ratios $P \to XC$ and $X \to AB$ are assumed to be well known. For instance the Particle Data Group (PDG) lists several branching fractions for various systems, some of which have uncertainties as low as 3% [2]. In addition to the normalisation itself, crosschecks involving other control channels with well known branching fractions are often performed during experimental analyses. In the following, we consider the existence of a second control channel denoted with $P \to YC$ with $Y \to AB$.

With the help of the method described, we can now quantitatively define the observed bias in each decay mode as the deviation of its integrated efficiency with respect to the hypothesis of perfect efficiency modelling, *i.e.* $(1 - e_i)$. The maximum allowed shift on the determination of the signal efficiency, and hence the signal branching ratio, obtained in the different tested configurations is illustrated in Fig. 1 (center). We can draw the following conclusions:

- as expected, the lower the normalised signal mass $\tilde{m}_V$ is, the bigger the bias is allowed to be; this is due to the large separation in the feature space between signal and control channels, which allows a mismodelling of the efficiency that only affects the signal channel;
- vice versa, it is nearly impossible for signal with mass similar to the ones of the normalisation/control channels to suffer from uncontrolled systematic effects;
- in all cases, the set of variables that allows the largest bias is given by the combination $\{\max p,\, \alpha_{AB}\}$, which is natural being the one with the largest correlation with the mass.

Figure 1 (right) shows the bias expected for the different obtained solutions as function of the opening angle of the considered control channel. The strong visible trend demonstrates that, assuming a sufficient statistical precision, the systematic uncertainty postulated by the DL Advocate algorithm.

4 Summary

In summary, we have introduced a method to place quantitative bounds for hidden systematic effects using machine learning.

The result is based on using a fast simulation, but a definitive and reliable statement can only be obtained by thoroughly examining all possible effects of the detector through a comprehensive simulation of the entire detector.

References

[1] A. Golutvin, A. Iniukhin, A. Mauri, P. Owen, N. Serra, and A. Ustyuzhanin. The DL advocate: Playing the devil's advocate with hidden systematic uncertainties. *Eur. Phys. J. C*, 83(779), 2023. doi: 10.1140/epjc/s10052-023-11925-w.

[2] R. L. Workman et al. Review of Particle Physics. *PTEP*, 2022:083C01, 2022. doi: 10.1093/ptep/ptac097.