PROCEEDINGS OF SCIENCE

Imposing exclusion limits on new physics with machine-learned likelihoods

Ernesto Arganda,[a,b] **Martín de los Rios,**[a,c] **Andres D. Perez**[b] **and Rosa María Sandá Seoane**[a,*]

[a] *Instituto de Física Teórica UAM-CSIC,*
C/ Nicolás Cabrera 13-15, Campus de Cantoblanco, 28049, Madrid, Spain

[b] *IFLP, CONICET - Dpto. de Física, Universidad Nacional de La Plata,*
C.C. 67, 1900 La Plata, Argentina

[c] *Departamento de Física Teórica, Universidad Autónoma de Madrid,*
E-28049 Cantoblanco, Madrid, Spain

E-mail: ernesto.arganda@csic.es, andres.perez@iflp.unlp.edu.ar, martin.delosrios@uam.es, r.sanda@csic.es

Machine-Learned Likelihood (MLL) is a method that, by combining modern machine-learning techniques with likelihood-based inference tests, allows estimating the experimental sensitivity of high-dimensional data sets. Here we extend the MLL method by including the exclusion hypothesis tests and study it first on a toy model of multivariate Gaussian distributions, where the true probability distribution functions are known. We then apply it to a case of interest in the search for new physics at the LHC, in which a Z' boson decays into lepton pairs, comparing the performance of MLL for estimating 95% CL exclusion limits with respect to the prospects reported by ATLAS at 14 TeV with a luminosity of 3 ab^{-1}.

41st International Conference on High Energy physics - ICHEP2022
6-13 July, 2022
Bologna, Italy

*Speaker

PoS(ICHEP2022)1226

1. Introduction

Modern machine learning (ML) has become a fundamental tool in experimental and phenomenological analyses of high-energy physics. In order to estimate the experimental sensitivity to potential new-physics signals at colliders, it was shown in Ref. [1] that the calibration of classifiers trained to distinguish signal and background samples under the relevant hypotheses ensures to properly estimate the likelihood ratio and consequently can be used to compute a statistical significance. In Ref. [2], a simplification of Ref. [1] has been proposed, the so-called Machine-Learned Likelihoods (MLL), which computes the expected experimental sensitivity by means of the use of ML classifiers, utilizing the entire discriminant output. A single ML classifier estimates the individual probability densities and subsequently one can calculate the statistical significance for a given number of signal and background events (S and B, respectively) with traditional hypothesis tests. By construction, the output of the classifier is always one-dimensional, so we reduce the hypothesis test to a single parameter of interest, the signal strength μ. On the one hand, it is simply and reliably applicable to any high-dimensional problem. On the other hand, by using all the information available from the ML classifier, it does not require defining working points like traditional cut-based analyses. The ATLAS and CMS collaborations incorporate similar methods in their experimental analyses, but consider only the classifier output as a good variable to bin and fit the binned likelihood formula (see, for instance, Ref. [3]).

The MLL method developed in Ref. [2] only includes the calculation of the discovery hypothesis test, although the expressions needed to calculate the exclusion limits were provided. In this work we extend the MLL method by adding the exclusion hypothesis test and applying it to two cases of interest: the case where the true probability distributions functions (PDFs) are known, through a toy model with multivariate Gaussian distributions; and an LHC study of Sequential Standard Model (SSM) [4] Z' bosons decaying into lepton pairs. We also compare the MLL performance for estimating 95% CL exclusion limits to the prospects reported by the ATLAS collaboration for a LHC center-of-mass energy of $\sqrt{s} = 14$ TeV with a total integrated luminosity of $\mathcal{L} = 3$ ab^{-1} [5].

2. Method

In this section, we present the corresponding formulae for the estimation of exclusion sensitivities with the Machine-Learned Likelihood method, first introduced in Ref. [2]. We also summarize the main features of the method which allows dealing with data of arbitrarily high dimension, while using the traditional inference tests to compare a null hypothesis (signal-plus-background) against an alternative one (background-only).

Following the statistical model in Ref. [6], we can define the likelihood $\mathcal{L}$ of N independent measurements with an arbitrarily high-dimensional set of observables x as

$$\mathcal{L}(\mu, s, b) = \text{Poiss}\big(N|\mu S + B\big) \prod_{i=1}^{N} p(x_i|\mu, s, b)\,, \tag{1}$$

where S (B) is the expected total signal (background) yield, Poiss stands for a Poisson probability mass function, and $p(x|\mu, s, b)$ is the probability density for a single measurement x, where μ defines the hypothesis we are testing for.

We can model the probability density containing the event-by-event information as a mixture of signal and background densities

$$p(x|\mu, s, b) = \frac{B}{\mu S + B}\, p_b(x) + \frac{\mu S}{\mu S + B}\, p_s(x)\,, \tag{2}$$

where $p_s(x) = p(x|s)$ and $p_b(x) = p(x|b)$ are, respectively, the signal and background probability densities for a single measurement x, and $\frac{\mu S}{\mu S+B}$ and $\frac{B}{\mu S+B}$ are the probabilities of an event being sampled from said probability densities.

To derive upper limits on μ, and in particular considering models with $\mu \geq 0$ and our choice for the statistical model in Eq. (1), we need to consider the test statistic for exclusion limits [8]:

$$\tilde{q}_\mu = \begin{cases} 0 & \text{if } \hat{\mu} > \mu\,, \\ -2\ \mathrm{Ln}\ \frac{\mathcal{L}(\mu,s,b)}{\mathcal{L}(\hat{\mu},s,b)} & \text{if } 0 \leq \hat{\mu} \leq \mu\,, \\ -2\ \mathrm{Ln}\ \frac{\mathcal{L}(\mu,s,b)}{\mathcal{L}(0,s,b)} & \text{if } \hat{\mu} < 0\,, \end{cases} \tag{3}$$

where $\hat{\mu}$ is the parameter that maximizes the likelihood in Eq. (1)

$$\sum_{i=1}^{N} \frac{p_s(x_i)}{\hat{\mu} S\, p_s(x_i) + B\, p_b(x_i)} = 1\,. \tag{4}$$

Since $p_{s,b}(x)$ are typically not known, the base idea of our method presented in Ref. [2] is to replace these densities for the one-dimensional signal and background manifolds obtained with a ML classifier. As it is well known, after training the classifier with a large and balanced dataset, the classification score $o(x)$ that maximizes the binary cross-entropy approaches [1, 7]

$$o(x) = \frac{p_s(x)}{p_s(x) + p_b(x)}\,, \tag{5}$$

as the classifier approaches its optimal performance. Then, the dimensionality reduction is introduced by dealing with $o(x)$ instead of x, using

$$p_s(x) \rightarrow \tilde{p}_s(o(x))\,, \qquad \text{and} \qquad p_b(x) \rightarrow \tilde{p}_b(o(x))\,, \tag{6}$$

where $\tilde{p}_{s,b}(o(x))$ are the distributions of $o(x)$ for signal and background, computed by evaluating the classifier on a set of pure signal or background events, respectively. Notice that this allows us to approximate both signal and background distributions individually, retaining the full information contained in both densities, without introducing any working point. We highlight again that these distributions are one-dimensional, and therefore can always be easily obtained by binning $o(x)$ and incorporated into in Eq. (3)

$$\tilde{q}_\mu = \begin{cases} 0 & \text{if } \hat{\mu} > \mu \\ 2(\mu - \hat{\mu})S - 2\sum_{i=1}^{N}\ \mathrm{Ln}\ \left(\frac{B\tilde{p}_b(o(x_i)) + \mu S \tilde{p}_s(o(x_i))}{B\tilde{p}_b(o(x_i)) + \hat{\mu} S \tilde{p}_s(o(x_i))}\right) & \text{if } 0 \leq \hat{\mu} \leq \mu \\ 2\mu S - 2\sum_{i=1}^{N}\ \mathrm{Ln}\ \left(1 + \frac{\mu S \tilde{p}_s(o(x_i))}{B\tilde{p}_b(o(x_i))}\right) & \text{if } \hat{\mu} < 0; \end{cases} \tag{7}$$

as well as into the condition on $\hat{\mu}$ from Eq. (4)

$$\sum_{i=1}^{N} \frac{\tilde{p}_s(o(x_i))}{\hat{\mu} S\, \tilde{p}_s(o(x_i)) + B\, \tilde{p}_b(o(x_i))} = 1\,. \tag{8}$$

The test statistic in Eq. (7) is estimated through a finite dataset of N events and thus has a probability distribution conditioned on the true unknown signal strength μ'. For a given hypothesis described by the μ' value, we can estimate numerically the $\tilde{q}_\mu$ distribution. When the true hypothesis is assumed to be the background-only one ($\mu' = 0$), the median expected exclusion significance med $[Z_\mu|0]$ is defined as

$$\text{med}\,[Z_\mu|0] = \sqrt{\text{med}\,[\tilde{q}_\mu|0]}\,, \tag{9}$$

where we estimate the $\tilde{q}_\mu$ distribution by generating a set of datasets with background-only events. Then, to set upper limits to a certain confidence level, we select the lowest μ which achieves the required median expected significance.

Concerning the method a final comment is in order, binning is not the only way to extract the PDFs $\tilde{p}_{s,b}(o(x))$. Instead, a non-parametric method such as Kernel Density Estimators (KDE) can be used without binning $o(x)$, as we are going to probe in a forthcoming update of this work [9].

3. Application examples

3.1 Known true PDFs: multivariate Gaussian distributions

To show the power of our method, we start with a toy model in abstract space (x_1, x_2). Events are generated by Gaussian distributions $\mathcal{N}_2(\boldsymbol{m}, \boldsymbol{\Sigma})$, so generative functions $p_{s,b}(x)$ are known for validation. We set no correlation between x_1 and x_2 (i.e., covariance matrices $\boldsymbol{\Sigma} = \mathbb{I}_{2\times 2}$) and $\boldsymbol{m} = +0.3(-0.3)$ for S (B).

We trained a supervised per-event classifier, `XGBoost`, with 1M events per class (balanced dataset), to distinguish S from B. The classifier output can be found in the left panel of Figure 1, while in the right panel we show the results for the MLL exclusion significance considering an example with a fixed background of $\langle B \rangle = 50$k and different signal strengths. We also include the significance calculated using the true probability density functions in Eq. (3), and the results employing a binned Poisson log-likelihood of the original 2-dimensional space (x_1, x_2), which is possible to compute in this simple scenario. Results with the MLL approach are close to the true PDF scenario and the Binned-Poisson method. It is important to highlight that the ML output is always one-dimensional regardless of the dimensionality of the data and can be easily binned or extracted with non-parametric methods.

In Figure 2 (left panel) we present the exclusion significance for higher dimensional data generated with $\mathcal{N}_{dim}(\boldsymbol{m}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} = \mathbb{I}_{dim\times dim}$, and $\boldsymbol{m} = +0.3(-0.3)$ for S (B). Once again, results with MLL method approach the ones with the true generative functions, while Binned-Poisson Likelihood becomes intractable. A common alternative to using the ML output is to choose a working point to define a signal-enriched region and to calculate the significance with the naive formula $S/\sqrt{B}$, which is exceeded by the MLL limits in all cases.

3.2 SSM Z' boson decaying into lepton pairs at the HL-LHC

Finally, we focus on a simple collider example, the search of a SSM Z' boson decaying into lepton pairs at the HL-LHC. We compare our results with ATLAS prospects for 95% CL exclusion limits at $\sqrt{s} = 14$ TeV and 3 ab^{-1} [5] in the right panel of Figure 2. Masses above 7 TeV could be excluded with our method, exceeding the ATLAS prospect for this search.

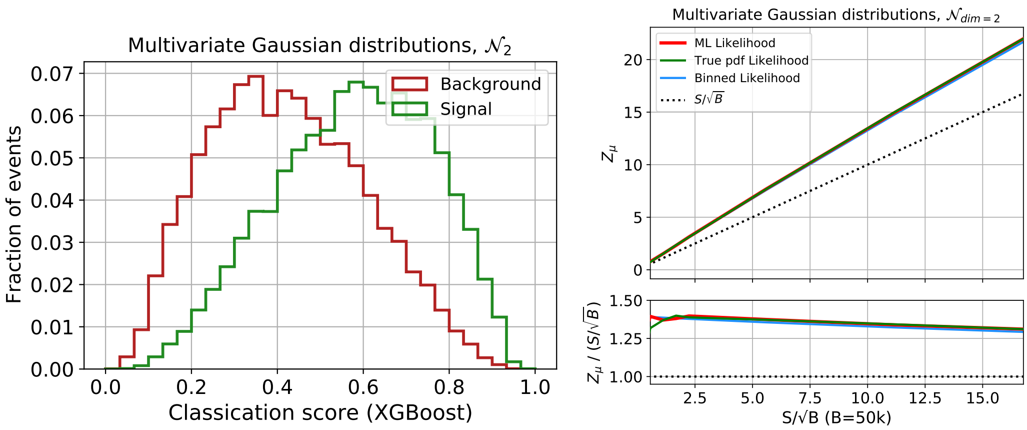

Figure 1: *Left panel*: classification score, $o(x)$, for a binary classifier using the `XGBoost` algorithm. *Right Panel:*: Exclusion significance calculated with various methods for the $dim = 2$ example for fixed background, $\langle B \rangle = 50$k, and different signal strengths $\langle S \rangle$. Red lines show the results implementing the MLL method with `XGBoost` used to estimate the probability density for single events and green lines when using the true multivariate distributions. The black dotted line represents the result of the usual counting method, $S/\sqrt{B}$, in the entire range of interest (only one bin), and the light blue curve is the result of a binned counting experiment.

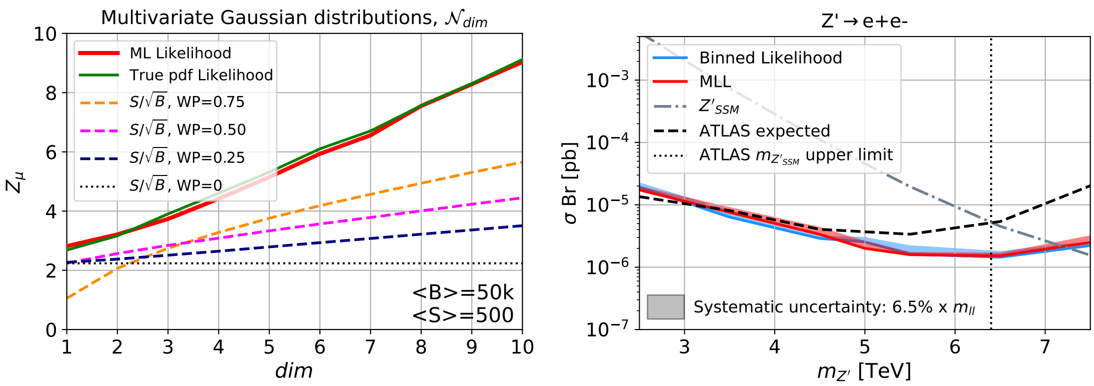

Figure 2: *Left Panel:* Exclusion significance calculated with various methods as a function of *dim*. For every case, the background and signal strengths were fixed, $\langle B \rangle = 50$k and $\langle S \rangle = 500$. Solid lines show the results implementing the method described in this work with `XGBoost` used to estimate the probability density for single events (red), and the true multivariate distributions (green). The dashed curves represent the result of the usual counting method (only one bin, $S/\sqrt{B}$), but for a subsample of the original data found with `XGBoost` assuming several working points, WP= $0.75, 0.5, 0.25$ to obtain signal enriched regions. The black dashed line also represents the result of the usual counting procedure but considering the entire dataset (equivalent to WP= 0). *Right Panel:* Exclusion significance for the Z'_{SSM} with MLL method. The red-shaded area includes the variation in the MLL significance caused by the mass variation introduced by the systematic uncertainty estimated by ATLAS for the invariant mass, as a naive estimation of the impact of systematics. Masses above 7 TeV could be excluded from our method.

4. Conclusions

Machine-Learned Likelihood method allows to obtain exclusion (and discovery) significances for additive new physics scenarios. It uses a single classifier and its full one-dimensional output, which allows the estimation of the S and B PDFs needed for statistical inference, avoiding also the use of working points. The binning of the output is always possible, irrespective of the dimensionality of the problem (unlike the Binned Likelihood method), although we emphasize that with our method it is not strictly necessary to bin in order to extract the signal and background PDFs. The results presented in this proceeding (part of a work in progress to be submitted soon [9]) improve results obtained by traditional techniques in toy models and a realistic Z' analysis, approaching (when possible) the ones computed with true generative functions.

Acknowledgments

We thank Pietro Vischia and Sergio Sanchez Cruz for fruitful feedback at the ICHEP2022.

References

[1] K. Cranmer, J. Pavez, and G. Louppe, *Approximating Likelihood Ratios with Calibrated Discriminative Classifiers*, [arXiv:1506.02169].

[2] E. Arganda, X. Marcano, V. M. Lozano, A. D. Medina, A. D. Perez, M. Szewc, and A. Szynkman, *A method for approximating optimal statistical significances with machine-learned likelihoods*, [arXiv:2205.05952].

[3] G. Aad *et al.* [ATLAS], *Measurement of the t-channel single top-quark production cross section in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Phys. Lett. B* **717** *(2012), 330-350* [arXiv:1205.3130].

[4] G. Altarelli, B. Mele, and M. Ruiz-Altaba, *Searching for New Heavy Vector Bosons in $p\bar{p}$ Colliders*, *Z. Phys. C* **45** (1989) 109 [Erratum: *Z. Phys. C* **47** (1990) 676], .

[5] **ATLAS** Collaboration, *Prospects for searches for heavy Z' and W' bosons in fermionic final states with the ATLAS experiment at the HL-LHC*, ATL-PHYS-PUB-2018-044 (2018).

[6] K. Cranmer et al., *Publishing statistical models: Getting the most out of particle physics experiments*, *SciPost Phys.* **12** (2022) 037 [arXiv:2109.04981].

[7] J. Neyman and E. S. Pearson, *On the Problem of the Most Efficient Tests of Statistical Hypotheses*, *Phil. Trans. Roy. Soc. Lond. A* **231** *(1933), no. 694-706 289–337*.

[8] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur. Phys. J. C* **71** *(2011) 1554* [Erratum: *Eur. Phys. J. C* **73** *(2013) 2501*] [arXiv:1007.1727].

[9] E. Arganda, M. de los Rios, A. D. Perez, R. M. Sandá Seoane, *Machine-learned exclusion limits without binning.*, work in progress.