Machine-Learning Collider Analysis of Radiative Neutralino Decays at the LHC

Rosa María Sandá Seoane

Departamento de Física Teórica &
Instituto de Física Teórica UAM-CSIC
Universidad Autónoma de Madrid

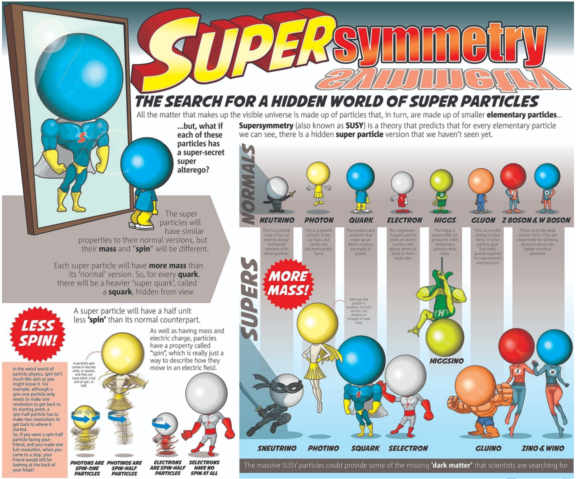
SUSY 2024
June 10, 2024

References

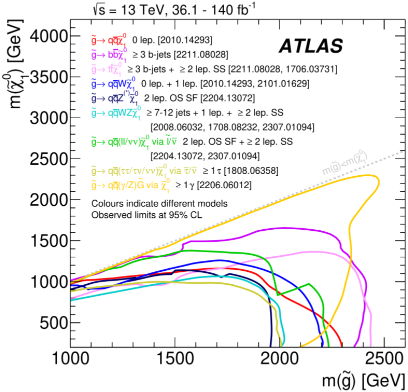
- E. Arganda, M. Carena, M. de los Ríos, A. D. Perez, D. Rocha, RMSS, C. Wagner, "Machine-Learning Collider Analysis of Radiative Neutralino Decays at the LHC," [arXiv:2406.XXXXX [hep-ph]].
- S. Baum, M. Carena, T. Ou, D. Rocha, N. R. Shah and C. E. M. Wagner, "Lighting up the LHC with Dark Matter," JHEP **11** (2023), 037 [arXiv:2303.01523 [hep-ph]].

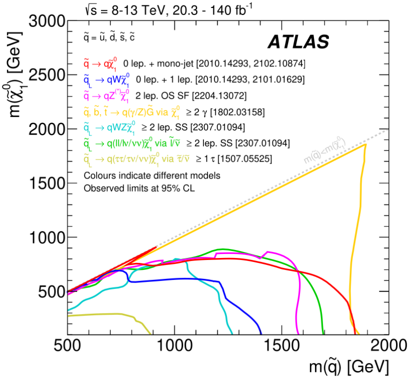
Why searching for radiative neutralino decays at the LHC?

- SUSY provides an explanation for the scale of EWSB (soft SUSY breaking scale) and for the DM (with the LSP $\tilde{\chi}_1^0$ in the presence of R-parity).

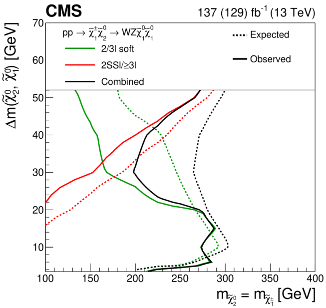


- Strong constraints at LHC for colored supersymmetric partners.



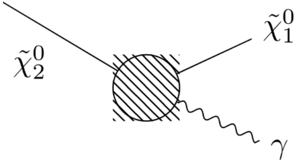


- Weakly interacting particles, instead, may be light and can be probed at the HL-LHC. In the MSSM, are also motivated to explain $(g-2)$.



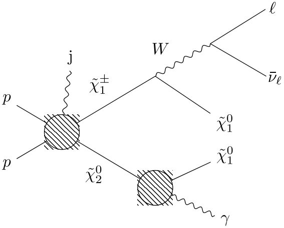
- In this scenario, the proper cosmological relic density can be achieved in the co-annihilation/compressed spectra, where the mass of the LSP in close to other weakly interacting particles, like the second lightest neutralino $\tilde{\chi}_2^0$ and the charginos $\tilde{\chi}_1^{\pm}(m_{\tilde{\chi}_1^0} \sim m_{\tilde{\chi}_2^0} \sim m_{\tilde{\chi}_1^{\pm}})$.

- If the direct detection cross-section of DM (LSP) is suppressed within the compressed region, the second lightest neutralino tends to decay into the LSP and a photon.
- Radiative decaying neutralinos at the LHC are highly suppressed by backgrounds (yet unexplored).



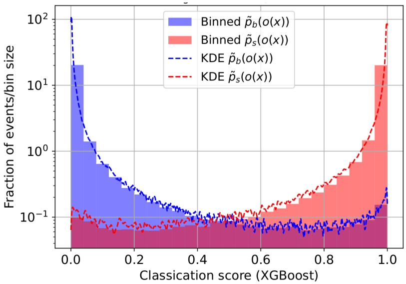
Proposal

Search for radiative decaying neutralinos $\sqrt{s} = 14$ TeV and a total integrated luminositiy of $\mathcal{L} = 100$ fb^{-1}. We require a highly energetic ISR jet in association with the electroweakino pair $\tilde{\chi}_1^{\pm}\,\tilde{\chi}_2^0$ to increase the MET signature.



Baum et al, JHEP **11** (2023), 037

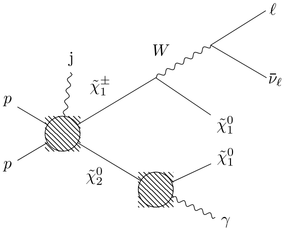
- First analysis with an optimized cut-and-count strategy.
- Main analysis adding a ML binary classifier to exploit correlations and increase the sensitivity of signal over background. Discovery significance reported for two different approaches:
  - Binned Likelihood (BL) method.
  - Machine-Learned Likelihoods (MLL) method (unbinned fit with Kernel Density Estimators -KDE).



In all cases, we compare scenario-dependent analysis (for specific set of parameters) vs. scenario-independent one (extensive to all parameter space).

Signal

$$p\,p \to \tilde{\chi}_1^{\pm}\,\tilde{\chi}_2^{0} j \to \tilde{\chi}_1^{0}\,\ell\,\nu_\ell + \tilde{\chi}_1^{0}\,\gamma + j$$



BP #	$m_{\tilde{\chi}_2^0}$ [GeV]	$(m_{\tilde{\chi}_2^0} - m_{\tilde{\chi}_1^0})$ [GeV]	$\mathrm{Br}(\tilde{\chi}_2^0 \to \tilde{\chi}_1^0 + \gamma)$	$\sigma(pp \to \tilde{\chi}_1^{\pm}\tilde{\chi}_2^0)$
1	200	34	15%	190 fb
2	200	19	37%	190 fb
3	200	10	73%	190 fb
4	250	37	15%	92 fb
5	250	22	36%	92 fb
6	250	13	67%	92 fb
7	300	39	16%	48 fb
8	300	24	36%	48 fb
9	300	15	62%	48 fb
10	350	41	17%	27 fb
11	350	26	35%	27 fb
12	350	17	58%	27 fb
13	400	43	16%	16 fb
14	400	27	32%	16 fb
15	400	18	52%	16 fb

- BPs representative of the compressed region. All BPs alleviate tension in $(g-2)$. Only BPs $\{2, 5, 8, 11, 14\}$ produce the cosmological relic density. A few BPs are naively excluded by CMS multilepton searches (the ones with the lowest $m_{\tilde{\chi}_2^0}$).
- Event selection criteria: at least one charged light lepton ($\ell = e, \mu$), at least one photon, and at least one jet. Leading jet with $p_T > 100$ GeV and $E_T^{\mathrm{miss}} > 100$ GeV.

Backgrounds

Process	Yield
$W + \text{jets}$	60058
$W\gamma$	58462
$t\bar{t}\gamma$	2877
$t\bar{t} + \text{jets}$	$11k$
$Z + \text{jets}$	$3.3k$
VV	$2.3k$
$Single - t$	$3.2k$
Total background	121397

BP #	Yield	$S/\sqrt{B}$
1	202	0.58
2	459	1.31
3	637	1.82
4	111	0.31
5	235	0.67
6	334	0.95
7	66	0.18
8	129	0.37
9	179	0.51
10	40	0.11
11	74	0.21
12	102	0.29
13	23	0.06
14	41	0.11
15	57	0.16

Only dominant backgrounds were used for the preliminary results presented today (red ones will be included in work in progress).

Kinematic variables

Simple set of variables to characterize the kinematics of the studied final state including low-level detector variables (p_T, η and ϕ of the leading objects $\{j_1, \ell_1, \gamma_1\}$, E_T^{miss}, and object multiplicities), and several high-level observables:

$$H_T^{\text{jets}} = \sum p_T^{\text{jets}},$$

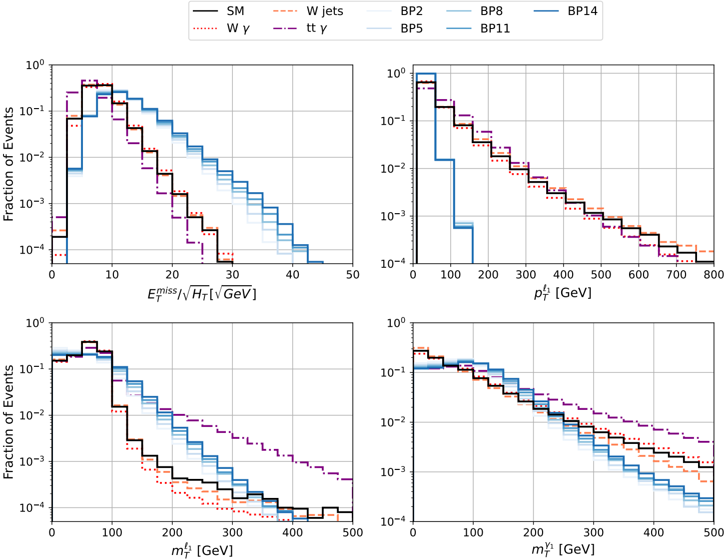
$$H_T = \sum p_T^{\text{jets}} + \sum p_T^{\tau} + \sum p_T^{e} + \sum p_T^{\mu} + \sum p_T^{\gamma},$$

$$m_T^A \equiv m_T\left(\mathbf{p}_T(A), \mathbf{E}_T^{\text{miss}}\right) = \sqrt{2p_T(A)E_T^{\text{miss}}\left(1 - \cos\Delta\phi\left(\mathbf{p}_T(A), \mathbf{E}_T^{\text{miss}}\right)\right)},$$

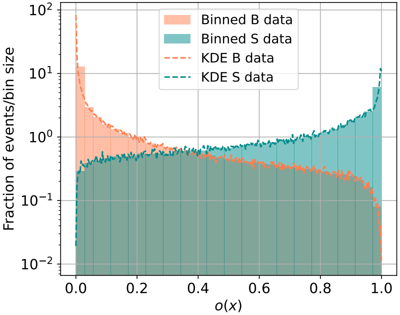
$$s_T^1 = p_T^{\ell_1} + p_T^{j_1} + p_T^{\gamma_1},$$

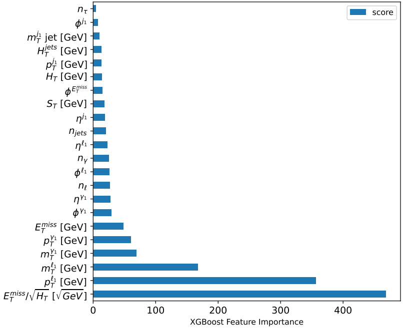
$$E_T^{\text{miss}}/\sqrt{H_T}.$$

Most relevant variables for discrimination:



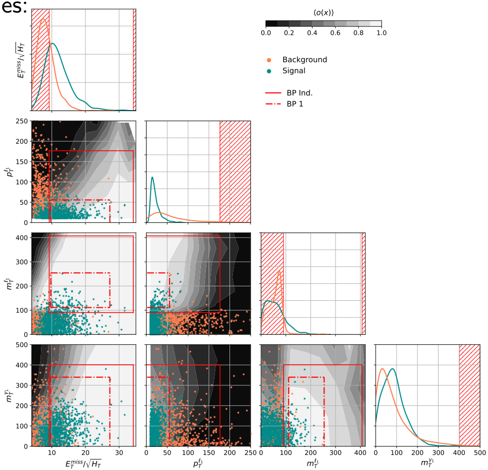
- Training of supervised XGBoost classifier, with balanced dataset, and all low and high-level variables as input features.





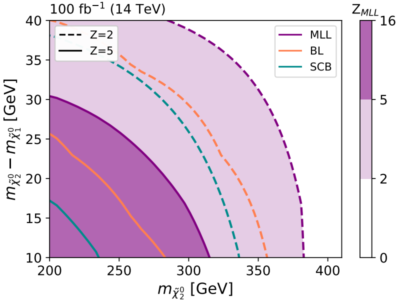
- Results for BP 1. Although there are small changes in the hierarchy of the feature importance training with a different BP or with the BP-independent dataset, the general trend is not modified.
- KDE for unbinned fit of the ML output within the MLL method, and histogram-based analysis within the BL method.

Correlation plot of the 4 most important features comparing the cut-and-count and ML strategies:



- BP-dependent cuts are more stringent and leave only $\mathcal{O}(10)$ of expected events than the BP-independent cuts.
- The ML classifier captures better the underlying physics than the signal-enriched regions described by rectangular cuts.

Projected discovery significance in the $[m_{\tilde{\chi}_2^0},\ m_{\tilde{\chi}_2^0} - m_{\tilde{\chi}_1^0}]$ plane with all methods for the BP-dependent (left) and the BP-independent (right) approaches:

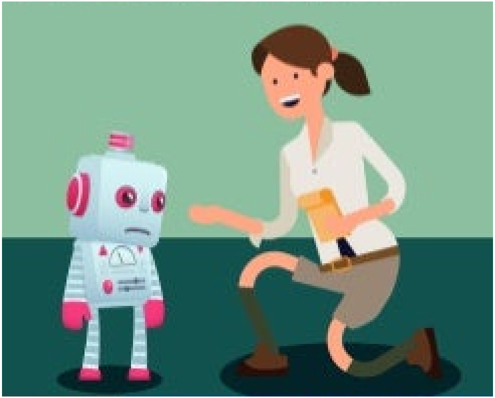


- The ML strategies (MLL and BL) are more sensitive than the cut-based one (SCB). Unbinning signal and background posteriors (MLL) provide more constraining limits than binning output (BL).
- The ML strategies provide a BP-independent sensitivity similar to the BP-dependent one, unlike the cut-and-count strategy.
- Proof-of-concept results, not systematic uncertainties included! Also still necessary to include other subdominant backgrounds.

Conclusions

- We explore an alternative channel for searching neutralinos, including photons and a hard ISR jet, never explored at the LHC. Well-motivated to explain $(g-2)$. (NEW!)
- The channel dominates in the co-annihilation region of the MSSM, where the direct DM detection cross-section is suppressed.
- The significance of these searches for the HL-LHC may be greatly improved using machine learning methods.
- Projected significances for the HL-LHC are promising, although a complete study including systematic uncertainties is still needed.
- Results may be modified when including other subdominant backgrounds (work in progress).

Thank you!



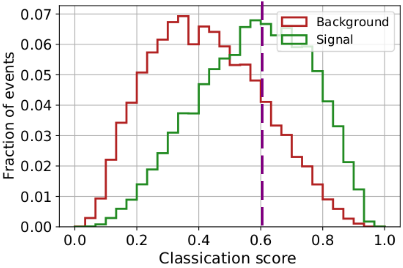
More about the model:

- M_1 (Bino mass parameter), M_2 (Wino mass parameter) and μ (Higgsino mass parameter) all takes values of a hundred GeV with $|M_1| \leqslant |M_2| \leqslant |\mu|$ (compressed spectra).
- Also $M_2 \times \mu > 0$ (prefered by $(g-2)$), and $M_1 \times \mu < 0$.
- Mass of all gluinos and squarks set to 2.5 TeV (generation universal).
- $\tan(\beta) = 50$.
- $m_h \sim 125$ GeV and $m_A = 2.5$ TeV.

Traditional vs ML search of New Physics

Distinguish SM (bckg) vs BSM (signal) in collider data:

- Design observables, define control regions... $\longrightarrow$ ML classifiers ✓
- For experimental significances, selection cuts $\longrightarrow$ Working points ✗



Traditional vs ML search of New Physics

Distinguish SM (bckg) vs BSM (signal) in collider data:

- Design observables, define control regions... ⟶ ML classifiers ✓
- For experimental significances, selection cuts ⟶ Working points ✗

Is it possible to connect the ML classifier output with the standard statistical tests without defining working points?

⟶ Machine-Learned Likelihood (MLL) Method

Traditional vs ML search of New Physics

Distinguish SM (bckg) vs BSM (signal) in collider data:

- Design observables, define control regions... $\longrightarrow$ ML classifiers ✓
- For experimental significances, selection cuts $\longrightarrow$ Working points ✗

Is it possible to connect the ML classifier output with the standard statistical tests without defining working points?

$\longrightarrow$ Machine-Learned Likelihood (MLL) Method

Can we avoid the information loss from binning the output?

$\longrightarrow$ +Kernel Density Estimators (KDE)

Method: Machine-Learned Likelihood

E. Arganda, X. Marcano, V. Martín Lozano, A. D. Medina, A. D. Perez, M. Szewc, A. Szynkman

Eur. Phys. J. C **82**, no.11, 993 (2022)

A method for approximating optimal statistical significances with machine-learned likelihoods

Ernesto Arganda,[a,b*], Xabier Marcano,[a,c†] Víctor Martín Lozano,[d,e‡] Anibal D. Medina,[b§] Andres D. Perez,[b¶] Manuel Szewc[f,g‖] and Alejandro Szynkman[b**]

[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
[d]Departament de Física Teòrica and IFIC, Universitat de València-CSIC,

E. Arganda, M. de los Rios, A. D. Perez, RMSS

PoS **ICHEP2022** (2022) 1226

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,

E. Arganda, M. de los Rios, A. D. Perez, RMSS

arXiv: 2211.04806

Machine-Learned Exclusion Limits without Binning

Ernesto Arganda[a,b*] Andres D. Perez,[b†] Martin de los Rios[a,c‡] 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

The MLL method

Statistical model for N independent measurements, with a high-dimensional set of observables x

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

where S (B) is the expected total signal (background) yield, and

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

The relevant test statistic to derive discovery significances corresponds to $\mu = 0$

$$\tilde{q}_0 = \begin{cases} 0 & \text{if } \hat{\mu} < 0 \\ -2 \text{ Ln } \frac{\mathcal{L}(0,s,b)}{\mathcal{L}(\hat{\mu},s,b)} = -2\hat{\mu}S + 2\sum_{i=1}^{N} \text{ Ln } \left(1 + \frac{\hat{\mu}S p_s(x_i)}{B p_b(x_i)}\right) & \text{if } \hat{\mu} \geqslant 0 \end{cases}$$

where $\hat{\mu}$ is the parameter that maximizes the likelihood

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

The MLL method

Statistical model for N independent measurements, with a high-dimensional set of observables x

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

where S (B) is the expected total signal (background) yield, and

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

The relevant test statistic to derive discovery significances corresponds to $\mu = 0$

$$\tilde{q}_0 = \begin{cases} 0 & \text{if } \hat{\mu} < 0 \\ -2\ \text{Ln}\ \frac{\mathcal{L}(0,s,b)}{\mathcal{L}(\hat{\mu},s,b)} = -2\hat{\mu}S + 2\sum_{i=1}^{N}\ \text{Ln}\ \left(1 + \frac{\hat{\mu}S p_s(x_i)}{B p_b(x_i)}\right) & \text{if } \hat{\mu} \geqslant 0 \end{cases}$$

where $\hat{\mu}$ is the parameter that maximizes the likelihood

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

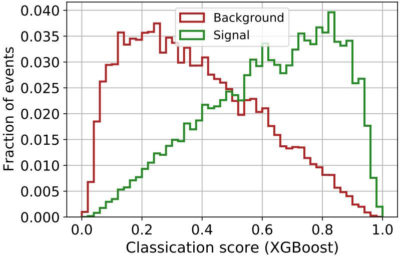
Solution: train classifier to distinguish signal from bckg with a balanced dataset. The classification score maximizes the binary cross-entropy and thus approaches

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

Dimensional reduction by dealing with $o(x)$ instead of x

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

Cranmer et al, arXiv: 1506.02169



where $\tilde{p}_{s,b}(o(x))$ are the distributions of $o(x)$ for signal and background, obtained by evaluating the classifier on a set of pure signal or background events, respectively.

The relevant test statistic for exclusion limits

$$\tilde{q}_0 = \begin{cases} 0 & \text{if } \hat{\mu} < 0 \\ -2 \text{ Ln } \frac{\mathcal{L}(0,s,b)}{\mathcal{L}(\hat{\mu},s,b)} = -2\hat{\mu}S + 2\sum_{i=1}^{N} \text{ Ln } \left(1 + \frac{\hat{\mu}S p_s(x_i)}{B p_b(x_i)}\right) & \text{if } \hat{\mu} \geqslant 0 \end{cases}$$

with $\hat{\mu}$ such us

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

The median expected discovery significance when the true hypothesis is assumed to be the signal-plus-background ($\mu' = 1$) is

$$\text{med } [Z_0|1] = \sqrt{\text{med } [\tilde{q}_0|1]}$$

Summary of MLL

- MLL method allows to obtaining exclusion (and discovery) significances for additive new physics scenarios.
- Uses a single `XGBoost` classifier and its full 1D output (no working points), which allows the estimation of the S and B pdfs needed for statistical inference. Not strictly necessary to bin the output to extract the PDFs.
- Inclusion of KDE as an extension of the MLL method to avoid the binning of the ML classifier output.
- Improves results obtained by traditional techniques in toy models and realistic analysis, approaching (when possible) the ones computed with true generative functions.
- Possible improvements: unsupervised analysis, systematic uncertainties...

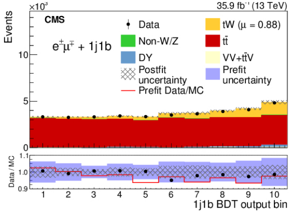
Traditional Binned-Likelihood (BL) method

$p_{s,b}(x)/\tilde{p}_{s,b}(o(x_i))$ are not known and are approximated by discrete binned distributions

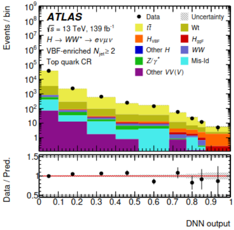
$$\mathcal{L}(\mu, s, b) = \prod_{d=1}^{D} \text{Poiss}\left(N_d|\mu S_d + B_d\right)$$

The median exclusion significance using Asimov datasets is given by

$$\text{med}[Z_\mu|0] = \sqrt{\tilde{q}_\mu|0} = \left[2\sum_{d=1}^{D}\left(B_d \ln\left(\frac{B_d}{S_d+B_d}\right) + S_d\right)\right]^{1/2} \xrightarrow[\sqrt{B}\gg 1]{S\ll B} \frac{S}{\sqrt{B}}$$



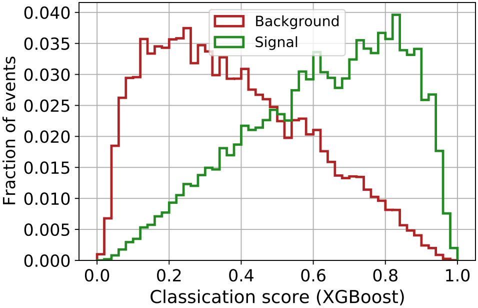
J HEP **10** (2018) 117



arXiv: 2207.00338

Density Estimation

What is the best way to extract $\tilde{p}_s(o(x))$ and $\tilde{p}_b(o(x))$?



Density Estimation

⟶Density estimation in a sense is the reverse of sampling: from given samples we want to retrieve the density function from which the samples were generated.

⟶Two types of methods for density estimation

- Parametric: model the density function as a specified functional form with a fixed number of tunable parameters.
- Non-parametric: specify a model whose complexity grows with the number of training datapoints.

Kernel Density Estimators

Kernel Density Estimators (KDE) is a non-parametric method for extracting $\tilde{p}_s(o(x_i))$ and $\tilde{p}_b(o(x_i))$

$\longrightarrow$Smoothed version of the empirical distribution $q_o(x)$ of the training data $\{x_i, i = 1, ..., N\}$

$$q_o(x) \;= \frac{1}{N}\sum_{i}^{N}\delta\left(x - x_i\right)$$

$\longrightarrow$We can smooth out the empirical distribution and turn it into a density by replacing each delta distribution with a smoothing kernel

$$\kappa_\epsilon(u) = \frac{1}{\epsilon^D}\;\kappa_1\left(\frac{u}{\epsilon}\right)$$

where $\epsilon > 0$ (bandwidth parameter) controls the width of the kernel and $\kappa_1(u)$ is a density function bounded from above (as $\epsilon \to 0$, $\kappa_\epsilon(u)$ approaches $\delta(u)$)

$$q_\epsilon(x) \;= \frac{1}{N}\sum_{i}^{N}\kappa_\epsilon\left(x - x_i\right)$$

$$\tilde{p}_{s,b}(o(x)) = \frac{1}{N}\sum_{i}^{N} \kappa_\epsilon \left[o(x) - o(x_i)\right]$$

Several options for κ_ϵ, e.g.

$$\kappa_\epsilon(u) = \begin{cases} \frac{1}{\epsilon}\frac{3}{4}\left(1-(u/\epsilon)^2\right), & \text{if } |u| \leqslant \epsilon \\ 0 & \text{otherwise} \end{cases} \qquad \textit{Epanechnikov kernel}$$