# Imposing exclusion limits on new physics with machine-learned likelihoods

Rosa María Sandá Seoane

Instituto de Física Teórica UAM-CSIC

XLI International Conference on High Energy Physics
July 9, 2022

# Method: Machine-Learned Likelihood

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

arXiv: 2205.05952

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,

and work in progress (E. Arganda, M. de los Rios, A. D. Perez, RMSS, arXiv:2207.XXXXX)

# 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... $\longrightarrow$ 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

# 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 to derive exclusion limits on $\mu$ (considering models with $\mu \geqslant 0$)

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

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

# 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 to derive exclusion limits on $\mu$ (considering models with $\mu \geqslant 0$)

$$\tilde{q}_\mu = \begin{cases} 0 & \text{if } \hat{\mu} > \mu \\ 2(\mu - \hat{\mu})S - 2\sum_{i=1}^{N} \text{Ln}\left(\frac{Bp_b(x_i)+\mu S p_s(x_i)}{Bp_b(x_i)+\hat{\mu} S p_s(x_i)}\right) & \text{if } 0 \leqslant \hat{\mu} \leqslant \mu \\ 2\mu S - 2\sum_{i=1}^{N} \text{Ln}\left(1 + \frac{\mu S p_s(x_i)}{Bp_b(x_i)}\right) & \text{if } \hat{\mu} < 0; \end{cases}$$

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 to derive exclusion limits on $\mu$ (considering models with $\mu \geqslant 0$)

$$\tilde{q}_\mu = \begin{cases} 0 & \text{if } \hat{\mu} > \mu \\ 2(\mu - \hat{\mu})S - 2\sum_{i=1}^{N} \text{Ln}\left(\frac{Bp_b(x_i)+\mu S p_s(x_i)}{Bp_b(x_i)+\hat{\mu} S p_s(x_i)}\right) & \text{if } 0 \leqslant \hat{\mu} \leqslant \mu \\ 2\mu S - 2\sum_{i=1}^{N} \text{Ln}\left(1 + \frac{\mu S p_s(x_i)}{B p_b(x_i)}\right) & \text{if } \hat{\mu} < 0; \end{cases}$$

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

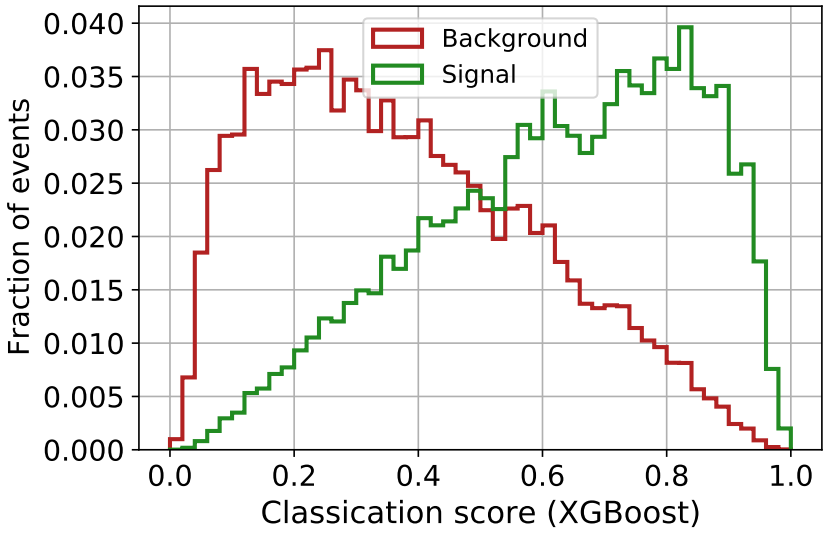
$$\sum_{i=1}^{N} \frac{p_s(x_i)}{\hat{\mu} S\, p_s(x_i) + B\, p_b(x_i)} = 1$$

Solution: train classifier to distinguish signal from bckg with a balanced dataset. The classification score maximizing 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 $x$

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




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

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

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 exclusion significance when the true hyphothesis is assumed to be the bckg-only one ($\mu' = 0$) is

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

# Traditional Binned-likelihood method

In general $p_{s,b}(x)$ are not known and are usually approximated by discrete binned distributions

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

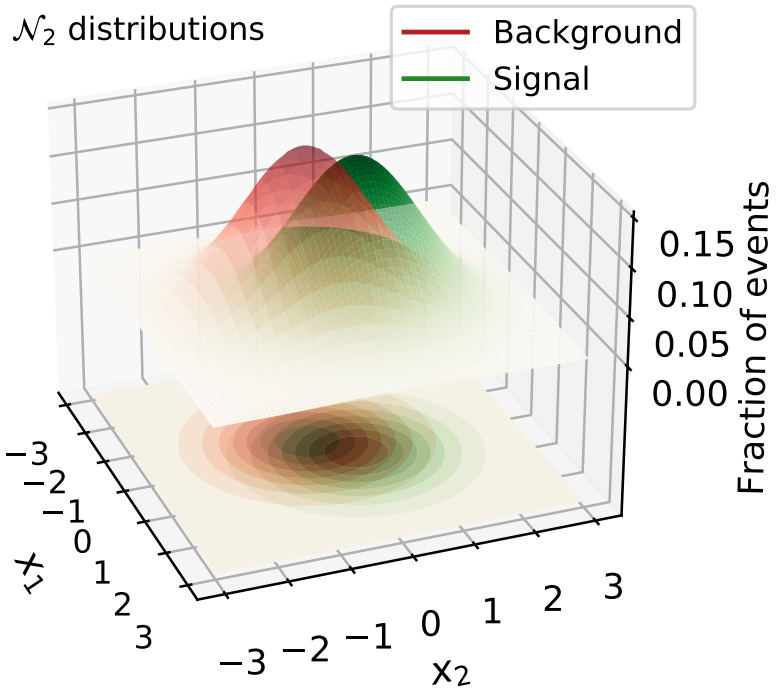
$S_d$ ($B_d$): expected number of signal (bckg) events in bin $d$.

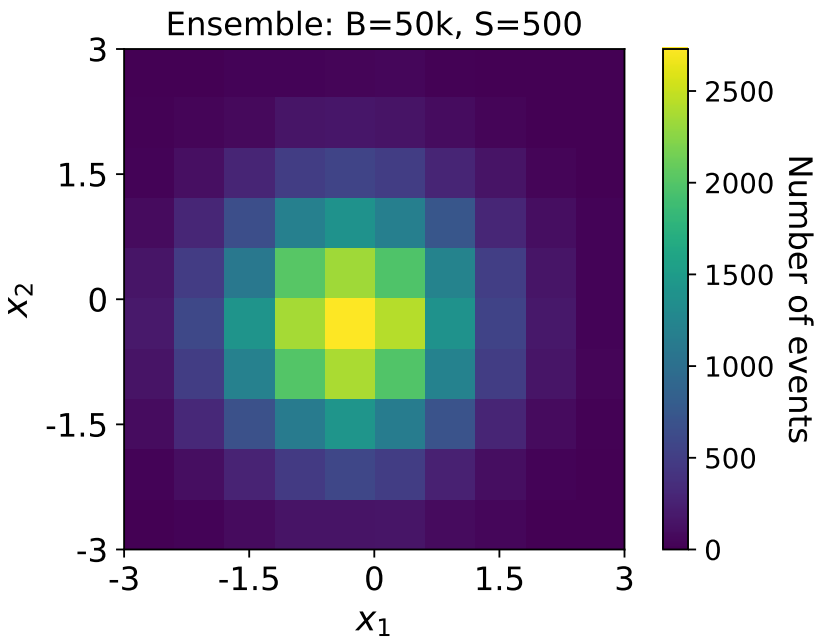
The median exclusion significance when the true hypothesis is assumed to be the bckg-only hypothesis, for the binned likelihoods (Asimov datasets) is given by

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

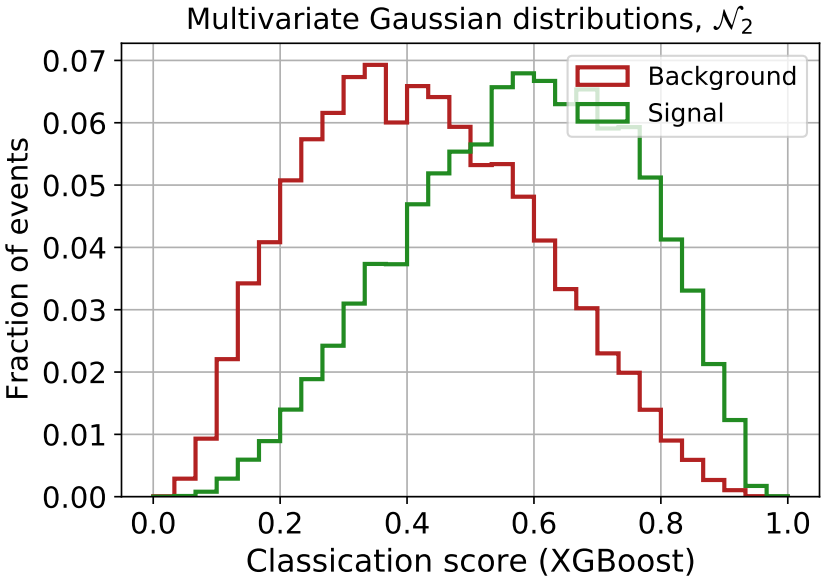
# Toy Model: Multivariate Gaussian distributions

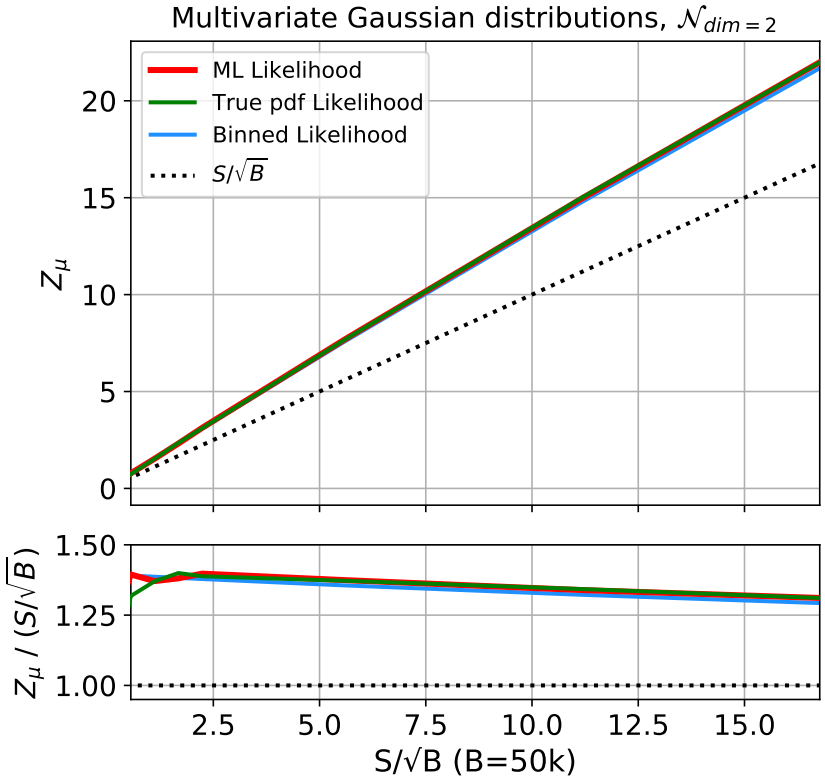
- Toy model in abstract space $(x_1, x_2)$. Events generated by $\mathcal{N}_2(\boldsymbol{m}, \boldsymbol{\Sigma})$ (known generative functions $p_{s,b}(x)$).
- Covariance matrices $\boldsymbol{\Sigma} = \mathbb{I}_{2\times 2}$ (no correlation) and $\boldsymbol{m} = +0.3(-0.3)\ \mathbb{1}_2$ for $S$ $(B)$.




- Training of supervised per-event classifier, `XGBoost` with 1M events per class (balanced dataset), to distinguish $S$ from $B$.

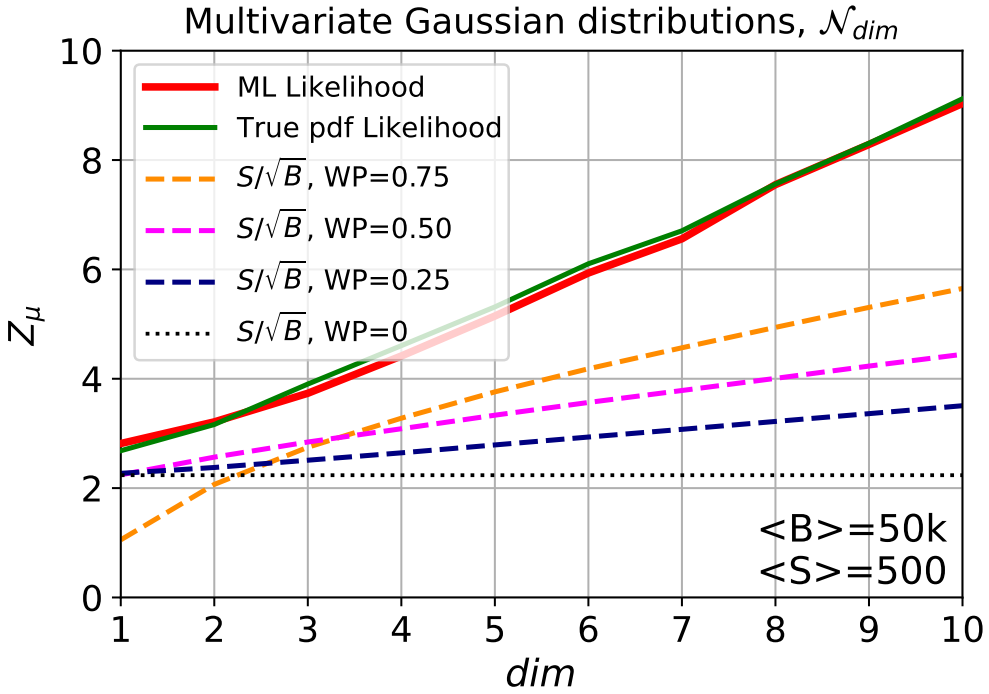



- Results with MLL approach are close to the true pdf scenario and outperforms the Binned-Poisson method.
- The ML output is always 1D regardless the dimensionality of the data and can be easyly binned.

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

For higher dimensional data with $\mathcal{N}_{dim}(\boldsymbol{m}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} = \mathbb{I}_{dim \times dim}$, $\boldsymbol{m} = +0.3(-0.3)$ $\mathbb{1}_{dim}$ for $S$ $(B)$:

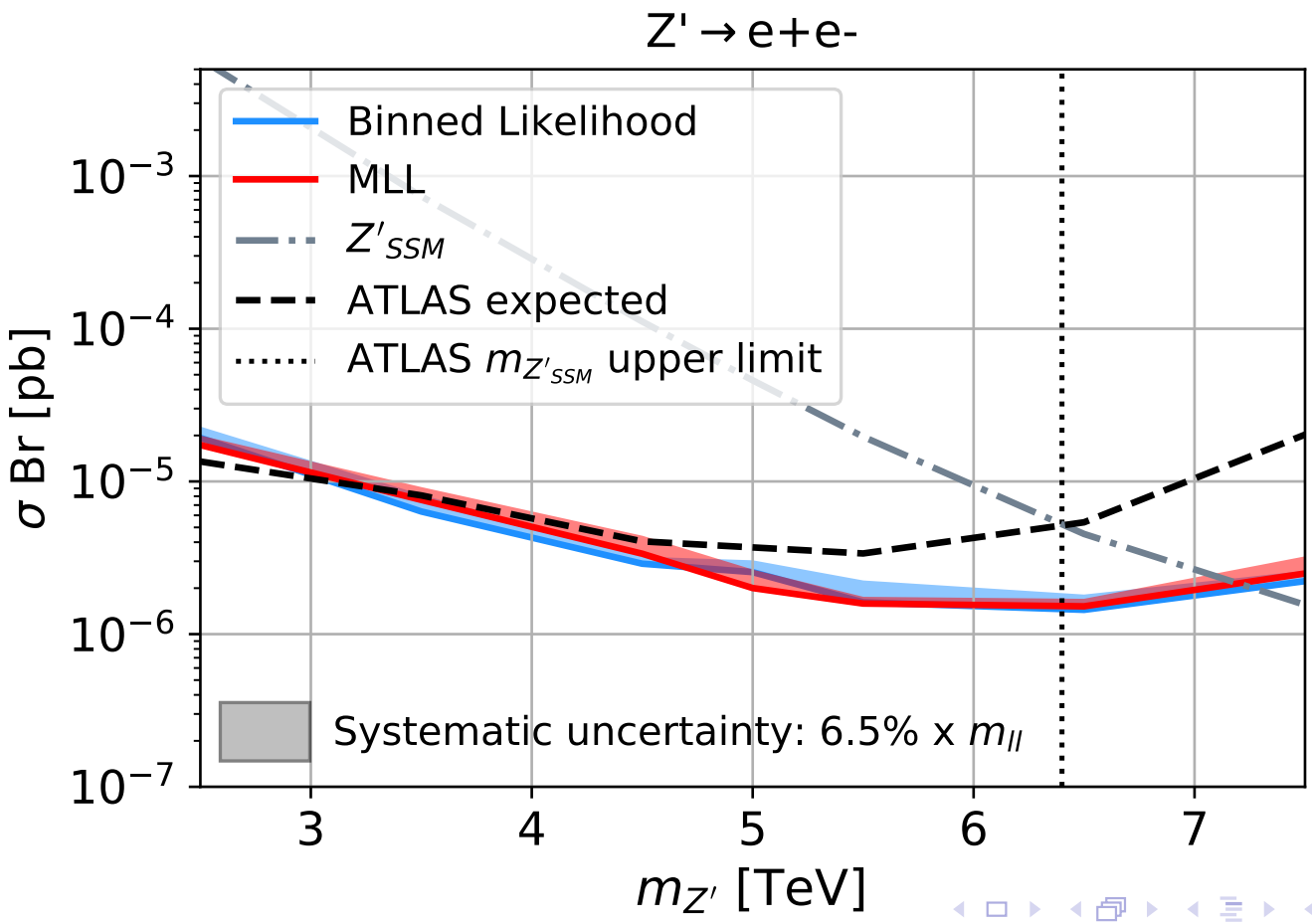



- Results with MLL method approach the ones with the true generative functions.
- Binned Poisson Likelihood intractable. Common alternative: use the ML output and define a working point to define a signal enriched region and use the naive formula for the significance.

# Search for a heavy SSM $Z'$ in dilepton final states at the HL-LHC

ATLAS prospects at $\sqrt{s} = 14$ TeV and 3000 fb$^{-1}$ for 95% CL exclusion limits on a $Z'_{SSM}$ (ATLAS-PHYS-PUB-2018-014).

S: $p\,p \to Z' \to e^+ e^-$ B: $p\,p \to e^+ e^-$



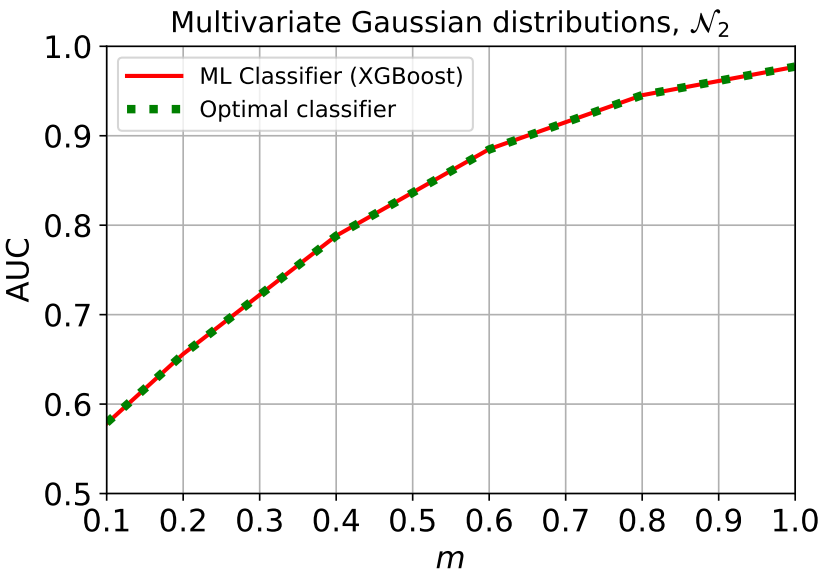
# Conclusions

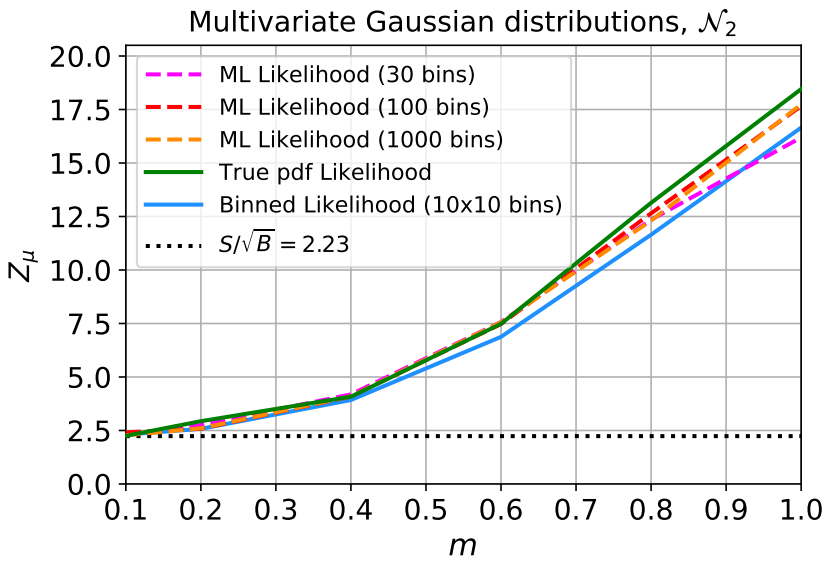
- MLL method allows to obtain exclusion (and discovery) significances for additive new physics scenarios.
- Uses a single `XGBoost` classifier and its full 1D output (no working points), which allows the estimation of the $S$ and $B$ pdfs needed for statistical inference.
- Always easy to bin the output, irrespectively of the dimensionality of the problem (unlike the Binned Likelihood method).
- 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...

Thank you!

# Extra:

Impact of the performance of the classifier and the classification score binning:




- With increasing values of $m$ the classifier performs better.
- Difference between ML Likelihood and the True Likelihood increases with the AUC ($o(x)$ more concentrated in the boundaries and the binning is not able to capture the granularity).
- Dependence with the choice of binning (compromise between approximating optimal results and computational cost).