Accelerating the search for mass bumps using the Data-Directed Paradigm

Jean-François Arguin
Georges Azuelos
Émile Baril
Fannie Bilodeau
Ali El Moussaouy
Muhammad Usman
Bruna Pascual

Shikma Bressler
Etienne Dreyer
Nilotpal Kakati
Amit Shkuri

Samuel Calvet
Julien Noce Donini
Eva Mayer

Université de Montréal

מכון ויצמן למדע
WEIZMANN INSTITUTE OF SCIENCE

LPC Particules pLasmas Univers applicationS
Laboratoire de Physique de Clermont

6th Inter-experiment Machine Learning Workshop

Jan 29 - Feb 2, 2024

Introduction

Motivation

What? → Train a neural network to **identify mass bumps in real data** without the need of simulation or analytical fit to estimate the background

- Rapidly scan many different regions of the observable-space
- Complementary to the standard analysis approach

Why? → Exploit the **discovery potential of the data**

- Impossible to cover all possible searches with the traditional analysis
- Many possible resonances in unexplored final states

Existing searches for two-body resonances[1]

	e	μ	τ	q/g	b	t	γ	Z/W	H	BSM → $SM_1 \times SM_1$				BSM → $SM_1 \times SM_2$				BSM → complex			
										q/g	γ/π^0's	b	$\cdots$	tZ/H	bH		$\cdots$	$\tau qq'$	eqq'	$\mu qq'$	$\cdots$
e	[37, 38]	[39, 40]	[39]	∅	∅	∅	[41]	[42]	∅	∅	∅	∅		∅	∅	∅		∅	[43, 44]	∅	
μ		[37, 38]	[39]	∅	∅	∅	[41]	[42]	∅	∅	∅	∅		∅	∅	∅		∅	∅	[43, 44]	
τ			[45, 46]	∅	[47]	∅	∅	∅	∅	∅	∅	∅		∅	∅	∅		[48, 49]	∅	∅	
q/g				[29, 30, 50, 51]	[52]	∅	[53, 54]	[55]	∅	∅	∅	∅		∅	∅	∅		∅	∅	∅	
b					[29, 52, 56]	[57]	[54]	[58]	[59]	∅	∅	∅		[60]	∅	∅		∅	∅	∅	
t						[61]	∅	[62]	[63]	∅	∅	∅		[64]	[60]	∅		∅	∅	∅	
γ							[65, 66]	[67–69]	[68, 70]	∅	∅	∅		∅	∅	∅		∅	∅	∅	
Z/W								[71]	[71]	∅	∅	∅		∅	∅	∅		∅	∅	∅	
H									[72, 73]	[74]	∅	∅		∅	∅	∅		∅	∅	∅	
BSM → $SM_1 \times SM_1$: q/g										∅	∅	∅		∅	∅	∅		∅	∅	∅	
γ/π^0's											[75]	∅		∅	∅	∅		∅	∅	∅	
b												[76, 77]		∅	∅	∅		∅	∅	∅	
$\vdots$																					
$\vdots$																					

[1] J. H. Kim et al., J. High Energ. Phys. **2020**, 30 (2020), arXiv:1907.06659 [hep-ph]

Data-Directed Paradigm

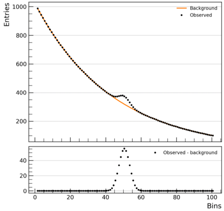
- The **Data Directed Paradigm (DDP)** is a search strategy to efficiently identify regions of interest in the data. It requires two ingredients:

Property of the SM
on which deviations can be searched for

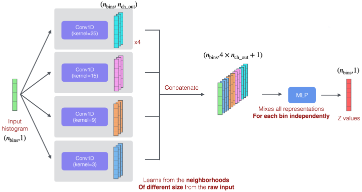
+

Tool to scan the observable-space
in search for deviations

e.g. smoothly falling invariant mass



e.g. NN getting statistical significance for bumps

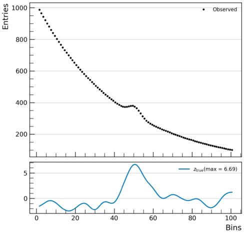


- Proof of concept performed for **symmetries**[2] and **bump searches**[3]

[2] S. Volkovich et al., Eur. Phys. J. C **82**, 265 (2022), arXiv:2107.11573 [hep-ex]
[3] M. Birman et al., Eur. Phys. J. C **82**, 508 (2022), arXiv:2203.07529 [hep-ph]

Bump search with DDP

- Bump search performed with a **network mapping invariant mass distribution to statistical significance**
  - **Input**: vector of bin entries from invariant mass histogram
  - **Target**: vector of statistical significance Z from likelihood-ratio test
  - Representative of an ideal analysis without modelling uncertainties



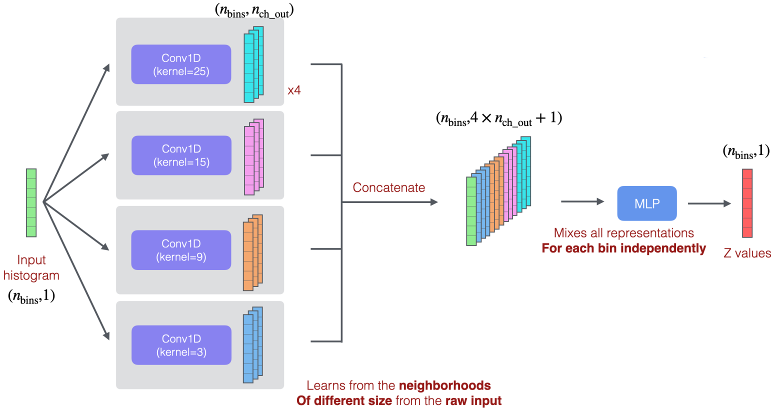
- Exploit its full potential with as **many final states** as possible
  - Multiple combinations of **objects** at the LHC
  - **Selections** on variables such as H_T, E_{Tmiss}, leading object p_T, etc

Electron	Leptonic Z
Muon	Boosted hadronic W/Z
Photon	Boosted top
Jet	High mass jet (m > 200 GeV)

Neural network and synthetic data generation

Implementation

- **Architecture:**
  - Use of 1D convolution layers followed by a dense layer
  - Intuitive and agnostic to the number of bins in the histogram

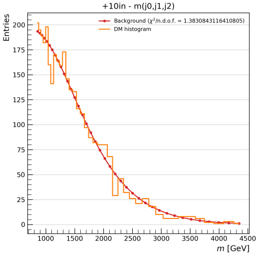


Synthetic data generation

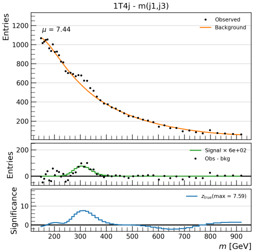
- **Data generation workflow**:
  1. Obtain background shapes
     - Analytical functions

$$be^{-ax},\quad ax+b,\quad \frac{1}{ax}+b,\quad \frac{1}{ax^2}+b,\quad \frac{1}{ax^3}+b,$$
$$\frac{1}{ax^4}+b,\quad a\left(x-x_2\right)^2+y_2,\quad -a\cdot\ln\left(x\right)+b,$$
$$\left(y_1-y_2\right)\cos\left(a\left(x-b\right)\right)+y_2,\quad \cosh\left(a\left(x-x_2\right)\right)+b.$$

     - Fits to simulation data



  2. Inject signal
     - Select background
     - Generate Gaussian signal
     - Combine both into **observed histogram**
     - Poisson fluctuate the histogram
     - Calculate **true significance** with likelihood-ratio test



- **Training data**:
  - Histograms with 30 to 100 bins
  - Broad dynamic range, from 10 to 100k entries per bin
  - Signals injected from broad significance range, from 1 to 20σ

Dark Machines sample description

- Using the **Dark Machines** dataset[4]
  - Designed to test anomaly detection techniques
  - Contains all of the **highest cross-section processes** at the LHC
  - Generation with Madgraph and Pythia, including fast detector simulation using Delphes
  - Events divided into signal regions/channels
    e.g. channel 3, which is more inclusive with cuts on $E_{Tmiss} > 100$ GeV and $H_T \geq 600$ GeV
  - Dataset equivalent to 10 fb^{-1}

SM processes			
Physics process	Process ID	σ (pb)	N_{tot} ($N_{10\,\text{fb}^{-1}}$)
$pp \to jj(+2j)$	njets	$19718_{H_T>600\text{GeV}}$	415331302 (197179140)
$pp \to l^{\pm}\nu_l(+2j)$	w_jets	$10537_{H_T>100\text{GeV}}$	135692164 (105366237)
$pp \to \gamma j(+2j)$	gam_jets	$7927_{H_T>100\text{GeV}}$	123709226 (79268824)
$pp \to l^+l^-(+2j)$	z_jets	$3753_{H_T>100\text{GeV}}$	60076409 (37529592)
$pp \to t\bar{t}(+2j)$	ttbar	541	13590811 (5412187)
$pp \to t + \text{jets}(+2j)$	single_top	130	7223883 (1297142)
$pp \to \bar{t} + \text{jets}(+2j)$	single_topbar	112	7179922 (1116396)
$pp \to W^+W^-(+2j)$	ww	82.1	17740278 (821354)
$pp \to W^{\pm}t(+2j)$	wtop	57.8	5252172 (577541)
$pp \to W^{\pm}\bar{t}(+2j)$	wtopbar	57.8	4723206 (577541)
$pp \to \gamma\gamma(+2j)$	2gam	47.1	17464818 (470656)
$pp \to W^{\pm}\gamma(+2j)$	Wgam	45.1	18633683 (450672)
$pp \to ZW^{\pm}(+2j)$	zw	31.6	13847321 (315781)
$pp \to Z\gamma(+2j)$	Zgam	29.9	15909980 (299439)
$pp \to ZZ(+2j)$	zz	9.91	7118820 (99092)
$pp \to h(+2j)$	single_higgs	1.94	2596158 (19383)
$pp \to t\bar{t}\gamma(+2j)$	ttbarGam	1.55	95217 (15471)
$pp \to t\bar{t}Z$	ttbarZ	0.59	300000 (5874)
$pp \to t\bar{t}h(+1j)$	ttbarHiggs	0.46	200476 (4568)
$pp \to \gamma t(+2j)$	atop	0.39	2776166 (3947)
$pp \to t\bar{t}W^{\pm}$	ttbarW	0.35	279365 (3495)
$pp \to \gamma\bar{t}(+2j)$	atopbar	0.27	4770857 (2707)
$pp \to Zt(+2j)$	ztop	0.26	3213475 (2554)
$pp \to Z\bar{t}(+2j)$	ztopbar	0.15	2741276 (1524)
$pp \to t\bar{t}t\bar{t}$	4top	0.0097	399999 (96)
$pp \to t\bar{t}W^+W^-$	ttbarWW	0.0085	150000 (85)

[4] T. Aarrestad et al., SciPost Phys. **12**, 043 (2022), arXiv:2105.14027 [hep-ph]

Histogram production

- Consider all possible **combinations of objects and selections**
  - With 0 to 4 objects per type
  - Additional kinematic cuts:
    $E_{Tmiss} > 200, 500$ GeV; leading object $p_T > 100, 200, 400$ GeV, ...
- Split the sub-dataset according to **jet multiplicity**
  - 0 jet, 1 jet, 2 jets, ... , ≥ 6 jets (depends on the available stat)
  - Should allow to improve S/B
  - Help reducing look-elsewhere effect (since bump should appear at the same place in neighboring jet multiplicities)
- **Build variables** from available objects
  - Mass distributions of the objects and their combinations
  - Transverse mass distributions including E_{Tmiss}
  - For jets, use only the 4 leading jets using b-tagging information
- Find the **maximum of the histogram** and start from there

- $1\mu + 3j$ + $E_{Tmiss} > 200$ GeV
  + $0e, 0\gamma, 0T, 0Z, 0Wh, 0HM$
- $1W + 1Z + 3j$ + $p_T(Z) > 100$ GeV
  + $0e, 0\gamma, 0T, 0HM$
- $\ldots$

Histogram production

- Consider all possible **combinations of objects and selections**
  - With 0 to 4 objects per type
  - Additional kinematic cuts:
    $E_{Tmiss} > 200, 500$ GeV; leading object $p_T > 100, 200, 400$ GeV, ...
- Split the sub-dataset according to **jet multiplicity**
  - 0 jet, 1 jet, 2 jets, ... , ≥ 6 jets (depends on the available stat)
  - Should allow to improve S/B
  - Help reducing look-elsewhere effect (since bump should appear at the same place in neighboring jet multiplicities)
- **Build variables** from available objects
  - Mass distributions of the objects and their combinations
  - Transverse mass distributions including E_{Tmiss}
  - For jets, use only the 4 leading jets using b-tagging information
- Find the **maximum of the histogram** and start from there

$1\mu + 3j + E_{Tmiss} > 200$ GeV
$+ 0e, 0\gamma, 0T, 0Z, 0Wh, 0HM$

$1W + 1Z + 3j + p_T(Z) > 100$ GeV
$+ 0e, 0\gamma, 0T, 0HM$

...

Histogram production

- Consider all possible **combinations of objects and selections**
  - With 0 to 4 objects per type
  - Additional kinematic cuts:
    $E_{Tmiss} > 200, 500$ GeV; leading object $p_T > 100, 200, 400$ GeV, ...
- Split the sub-dataset according to **jet multiplicity**
  - 0 jet, 1 jet, 2 jets, ... , ≥ 6 jets (depends on the available stat)
  - Should allow to improve S/B
  - Help reducing look-elsewhere effect (since bump should appear at the same place in neighboring jet multiplicities)
- **Build variables** from available objects
  - Mass distributions of the objects and their combinations
  - Transverse mass distributions including E_{Tmiss}
  - For jets, use only the 4 leading jets using b-tagging information
- Find the **maximum of the histogram** and start from there

- $1\mu + 3j$ + $E_{Tmiss} > 200$ GeV
  + $0e, 0\gamma, 0T, 0Z, 0Wh, 0HM$
  - $m(\mu j_1)$, $m(\mu j_1 E_{Tmiss})$, ...
  - $m(\mu j_1 j_2)$, $m(\mu j_1 j_2 E_{Tmiss})$, ...
  - ...
- $1W + 1Z + 3j$ + $p_T(Z) > 100$ GeV
  + $0e, 0\gamma, 0T, 0HM$
  - $m(Wh, j_1)$, $m(Wh, j_1, E_{Tmiss})$...
  - $m(j_1 j_2)$, $m(j_1 j_2, E_{Tmiss})$...
  - ...
- ...

Histogram production

- Consider all possible **combinations of objects and selections**
  - With 0 to 4 objects per type
  - Additional kinematic cuts:
    $E_{Tmiss} > 200, 500$ GeV; leading object $p_T > 100, 200, 400$ GeV, ...
- Split the sub-dataset according to **jet multiplicity**
  - 0 jet, 1 jet, 2 jets, ... , ≥ 6 jets (depends on the available stat)
  - Should allow to improve S/B
  - Help reducing look-elsewhere effect (since bump should appear at the same place in neighboring jet multiplicities)
- **Build variables** from available objects
  - Mass distributions of the objects and their combinations
  - Transverse mass distributions including E_{Tmiss}
  - For jets, use only the 4 leading jets using b-tagging information
- Find the **maximum of the histogram** and start from there

$1\mu + 3j$ + $E_{Tmiss} > 200$ GeV
+ $0e, 0\gamma, 0T, 0Z, 0Wh, 0HM$
- $m(\mu j_1)$, $m(\mu j_1 E_{Tmiss})$, ...
- $m(\mu j_1 j_2)$, $m(\mu j_1 j_2 E_{Tmiss})$, ...
- ...

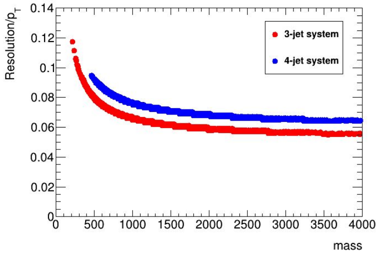
$1W + 1Z + 3j$ + $p_T(Z) > 100$ GeV
+ $0e, 0\gamma, 0T, 0HM$
- $m(Wh, j_1)$, $m(Wh, j_1, E_{Tmiss})$...
- $m(j_1 j_2)$, $m(j_1 j_2, E_{Tmiss})$...
- ...

...

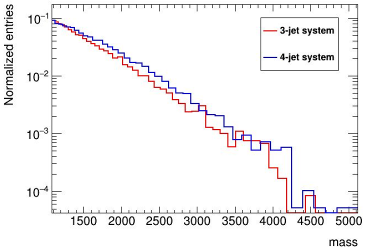
Total of **30 000 mass histograms**

Processing and calibration

- In real data, signal width determined by **detector resolution**
  - Different for different final states and/or mass
  - Narrow signal should produce a bump in few bins
- **Rebin histogram** to reflect the detector resolution we would see in actual experimental data
  - e, μ, γ: Delphes Card formulas, depends on p_T and η
  - jets: ATLAS report on jet resolution, depends on p_T
  - *HM*/*T*/*Wh* accounts for 3/3/2 jets
  - Using $p_T \approx m/2$ approximation



Resolution is higher for $m(4j)$ than $m(3j)$, also larger for small masses



Binning reflects this with larger bin width when resolution is smaller

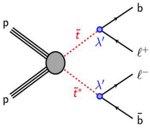
Dark Machines datasets

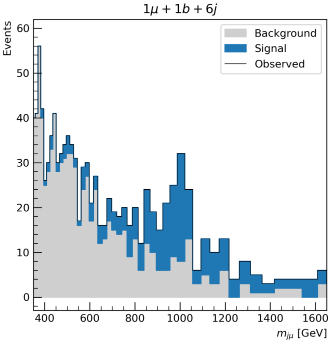
- **Background-only datasets**:
  - Obtain shapes to be included in the training sample
  - Assess false positive rate

- **BSM signal datasets**:
  - Simulated signal data added on top of the backgrounds
  - Test the network in a more realistic scenario
  - Different levels of difficulty (e.g. cross-section, mass values, etc)

Some of the new physics models we have available include:

- RPV stop $\longrightarrow b\ell$
- $W' \longrightarrow WZ \longrightarrow \ell\nu qq, qq\nu\nu$
- $LQ \longrightarrow beb\mu, bebe, t\nu t\nu$
- $Z' \longrightarrow 3l$

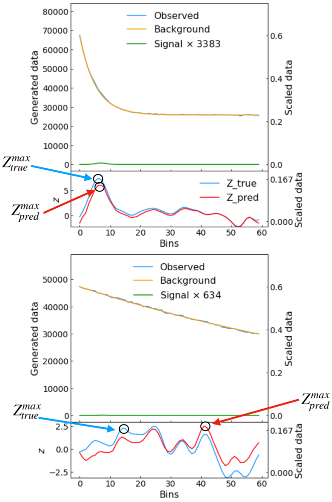




Performance and finding BSM signals

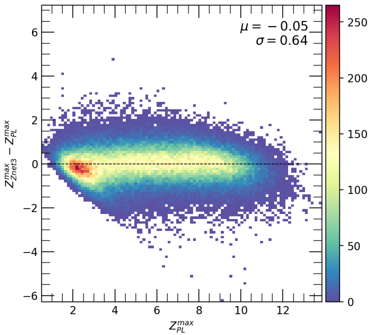
Performance over synthetic data

- Performance quantified in terms of the **difference between predicted and true maximum significance**
  - Z^{max}_{true}: maximal significance calculated via the likelihood ratio test
  - Z^{max}_{pred}: maximal predicted significance
- Majority of entries should have $Z^{max}_{pred} - Z^{max}_{true}$ close to 0 with the smallest variance possible
- **Non-zero significance** for background-only histograms
  - Due to look-elsewhere-effect
  - Could artificially bias the performance at low significance

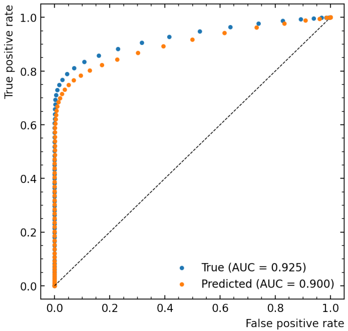


Performance on the testing sample

- Trained model **accurately predicts maximum significance** with no bias and a variance of ± 0.64

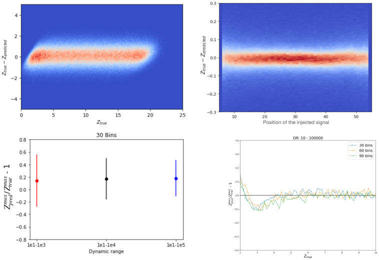


- **Excellent discriminating performance** of signal and background with an AUC of 0.900

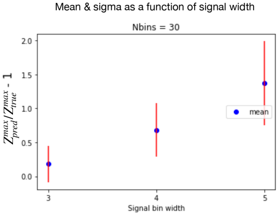


Performance stability

- Good agreement between Z_{pred}^{max} and Z_{true}^{max} is **stable**
  - For all Z_{true}^{max}
  - Over all mass range
  - For all dynamic ranges
  - For linear combinations of the functions

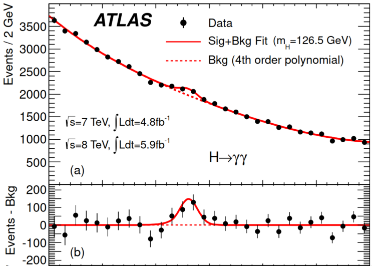


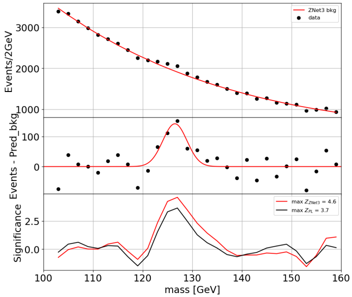
- Training with fixed signal width has **increased bias as signal width increases**
  - Same behaviour regardless of number of bins and dynamic range
  - Accurate calibration takes time (and the DDP should work fast)
  - Room for improvement, yet well defined behaviour



Finding BSM signals

- Promising results when **finding the Higgs bump**
  - Data sampled from the ATLAS plot[5] using a digitizer





- ✓ Predicted resonance at the correct mass
- ! Predicted significance of 4.6σ whereas the ATLAS significance is 3.7σ

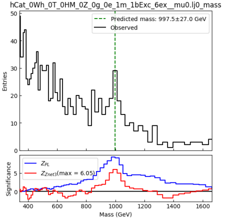
[5] ATLAS Collaboration, Physics Letters B **716**, 1–29 (2012), arXiv:1207.7214 [hep-ex]

Finding BSM signals

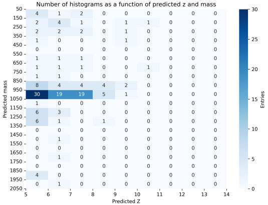
- Tested over **simulated BSM signals** added to the Dark Machines background

RPV stop $\rightarrow b\ell$

$m(j\mu)$ at 1μ+1b+6j



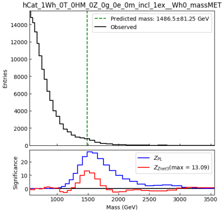
Histograms with $Z_{pred}^{max} \geq 5$



- Successfully finds an excess at the expected mass of the stop at 1 TeV

$W' \rightarrow WZ$

$m(Wh, E_{Tmiss})$ at 1Wh+1j



- Successfully finds bump for W' (with a boosted Z in the final state)

- Other signals tested and successfully found include $LQ \rightarrow beb\mu$, $bebe$ and $Z' \rightarrow 3\ell$
- **False-positive rate of 0.1%** when tested over background-only sample

Conclusion

Final remarks

- **Data directed paradigm bump search** to scan unexplored final states in search for resonances
  - Target smoothly falling invariant mass spectra across a variety of final states
  - Exploit full potential of data without the need of simulation or analytical fit
- **Network implementation and performance**
  - Using Dark Machines datasets with highest cross-section processes at the LHC
  - Produce mass histograms with binning that reflects detector resolution
  - Successfully finds Higgs bump and BSM signals on top of the Dark Machines background, such as RPV stop and W'
- **Future developements**
  - Application to real experimental data, focusing on Run 2
  - Use full MC simulation data with basic selections
  - First iteration using single-lepton trigger and same objects
  - Eventually adding more objects, such as large-R jets

Appendix

Calculation of target significance

For each bin i, we perform a **hypothesis test** and obtain a significance value z_i (see arXiv:1007.1727)

- **Step 1:** Define the shape of the signal S for which we are looking, e.g. a Gaussian centered on bin i
  - We are currently using the mass values on the x-axis

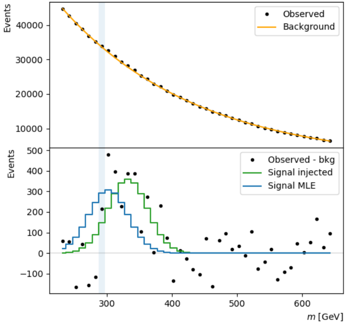
$$s_i = \int_{\mathrm{bin}\,i} f(x; x_0, \sigma)\,\mathrm{d}x \approx f(x; x_0, \sigma)\Delta x$$

$$\begin{cases} x_0 = \text{center of bin } i \\ \sigma = \text{width of bin } i \times \text{given number of bins} \end{cases}$$

- **Step 2:** Obtain the **maximum likelihood estimator** $\hat{\mu}$ for this signal S
  - Expected value is $N_{\mathrm{exp}} = b_i + \mu s_i$

$$\hat{\mu} = \arg\min - \ln[L(\mu)]$$

$$-\ln[L(\mu)] = -\sum_i N_{\mathrm{obs}} \ln(N_{\mathrm{exp}}) - N_{\mathrm{exp}} + \ln\left(\frac{1}{N_{\mathrm{obs}}!}\right)$$



- **Observed data (N_{obs})** is the synthetic data
- **Unfluctuated background (b)** is our null hypothesis
- **Injected signal** is the one added into the observed data
- **Test signal** is the signal that maximizes the likelihood for each bin

Calculation of target significance

For each bin i, we perform a **hypothesis test** and obtain a significance value z_i (see arXiv:1007.1727)

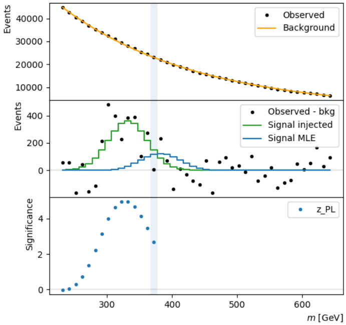
- **Step 3:** Obtain the **likelihood** for both the background-only hypothesis and the maximum-likelihood estimator (MLE) signal hypothesis
  - $-\ln[L(\mu = 0)] \rightarrow$ Background-only ($N_{\mathrm{exp}} = b_i$)
  - $-\ln[L(\mu = \hat{\mu})] \rightarrow$ Signal MLE ($N_{\mathrm{exp}} = b_i + \hat{\mu} s_i$)
- **Step 4:** Obtain the significance using the **profile likelihood ratio** test

If $\hat{\mu} \geq 0$:

$$z_i = \sqrt{q_0} \quad \text{with} \quad q_0 = -2 \ln \left[\frac{L(0)}{L(\hat{\mu})} \right]$$

If $\hat{\mu} < 0$:

$$z_i = -\sqrt{-q_0} \quad \text{with} \quad q_0 = \begin{cases} -2 \ln \left[\frac{L(\hat{\mu})}{L(0)} \right], \text{if two-sided} \\ 0, \text{otherwise} \end{cases}$$



- **Observed data (N_{obs})** is the synthetic data
- **Unfluctuated background (b)** is our null hypothesis
- **Injected signal** is the one added into the observed data
- **Test signal** is the signal that maximizes the likelihood for each bin

Background functions in the framework

$ax + b$

```
def linear(x, p0, p1):
    return p0 * x + p1
```

ae^{-bx}

```
def exponential(x, p0, p1):
    return p0 * np.exp(-p1 * x)
```

$\frac{1}{ax} + b$

```
def one_over_x(x, p0, p1):
    return 1/(p0*x) + p1
```

$a(x - x_{max})^2 + y_{min}$

```
def parabola_half(x, p0, p1, x2, y2):
    return p0*(x - x2)**2 + y2
```

$\frac{1}{ax^2} + b$

```
def one_over_x_squared(x, p0, p1):
    return 1/(p0*x**2) + p1
```

$\Delta y \cos[a(x - b)] + y_{max}$

```
def cos_quarter(x, p0, p1, y1, y2):
    return (y1 - y2)*np.cos(p0*(x - p1)) + y1
```

$\frac{1}{ax^3} + b$

```
def one_over_x_cubed(x, p0, p1):
    return 1/(p0*x**3) + p1
```

$-a \ln(x) + b$

```
def ln_negative(x, p0, p1):
    return -p0*np.log(x) + p1
```

$\frac{1}{ax^4} + b$

```
def one_over_x_to_4th(x, p0, p1):
    return 1/(p0*x**4) + p1
```

$\cosh[a(x - x_{max})] + b$

```
def cosh_half(x, p0, p1, x2):
    return np.cosh(p0*(x - x2)) + p1
```

$\frac{1}{ax^n} + b$

```
def one_over_x_to_nth(x,p0,p1,n):
    return 1/(p0*x**n) + p1
```

$\rightarrow$ where n is taken randomly from the range [0.01, 10]