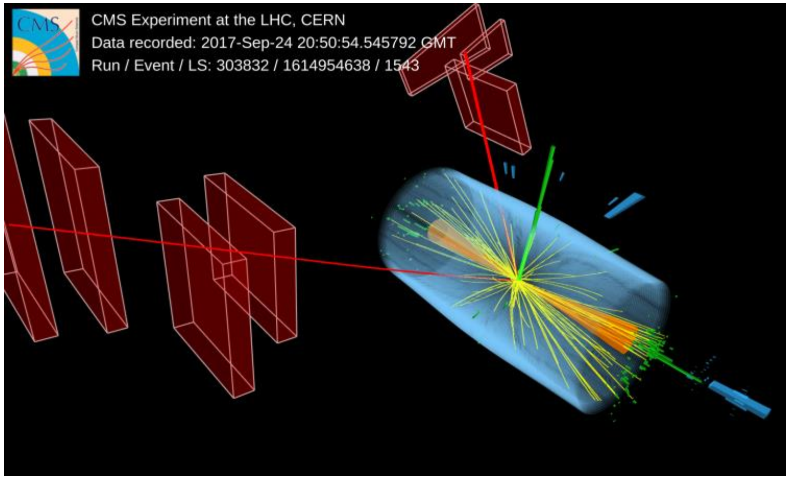
Combining Neural Network predictions with Hypothesis Testing for discovery in the LHC

Michael Soughton (University of Sussex)
With Veronica Sanz and Charanjit Kaur Khosa

ML4JETS: 06/07/2021

Observables within the LHC

- Many observables measured from LHC collisions
- We test theories by looking for deviations away from the background in certain regions.
- We like to express the significance of a discovery in terms of "sigma" of a unit Gaussian.
- No discovery made yet!



Machine Learning

- Able to take multidimensional features as input.
- Can learn complex patterns across multidimensional space
- Gives an output when applied to new data based on what it learned from training data. Could be:
  - Classification label (supervised classification)
  - Some continuous quantity (supervised regression)
  - Attempted reconstruction of input (unsupervised)

Output of ML

- In classification tasks output is

$$P(y_i = 1 \mid \boldsymbol{x}_i, \boldsymbol{\theta}) = \frac{1}{1 + \mathrm{e}^{-\mathbf{x}_i^T \boldsymbol{\theta}}} \qquad P(y_i = 0 \mid \boldsymbol{x}_i, \boldsymbol{\theta}) = 1 - P(y_i = 1 \mid \boldsymbol{x}_i, \boldsymbol{\theta}),$$

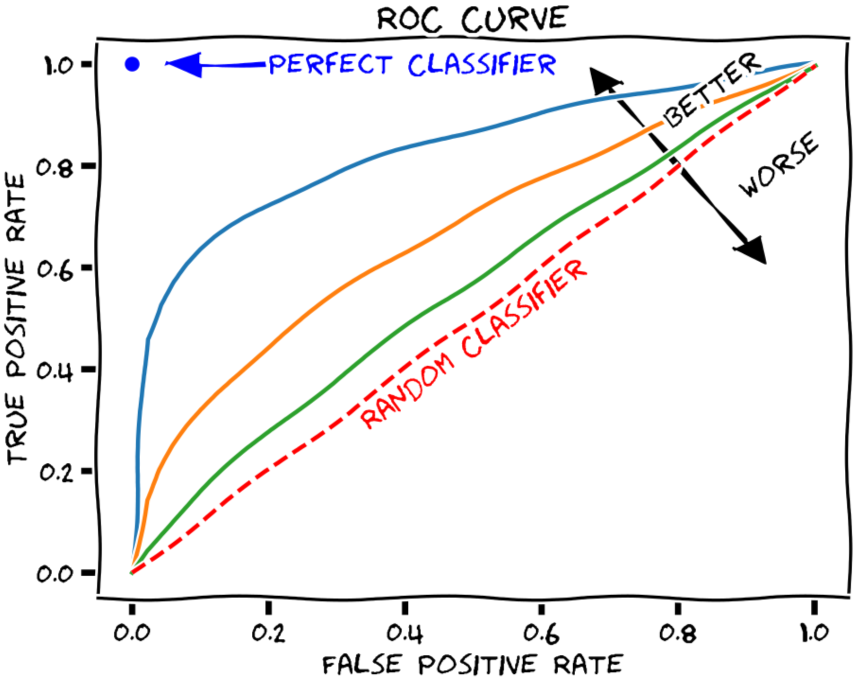
Actual Values

Predicted Values		Positive (1)	Negative (0)
	Positive (1)	TP	FP
	Negative (0)	FN	TN

$$\mathrm{Acc} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FN} + \mathrm{FP}}$$

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{TN} + \mathrm{FP}}$$



Hypothesis testing (simple hypotheses)

- Given statistical models $p(x|H_0)$ and $p(x|H_1)$ we want a test to either accept or reject H_0.
- α = probability (with data generated according to H_0) of rejecting H_0 when it is true (type I error).
- β = probability (with data generated according to H_1) of not rejecting H_0 when it is false (type II error).
- There is freedom to choose α and more considerations when doing so.

The Neyman-Pearson Hypothesis Test

- From the PDF we can construct the likelihood $L = \frac{\mu^n \exp(-\mu)}{n!} \cdot \prod_{i=1}^{n} p\left(x_i \mid H\right)$
- If the Type I error probability α is specified in a test of simple hypothesis H_0 against simple hypothesis H_1, then the Type II error probability β is minimized by ordering *x* according to the likelihood ratio

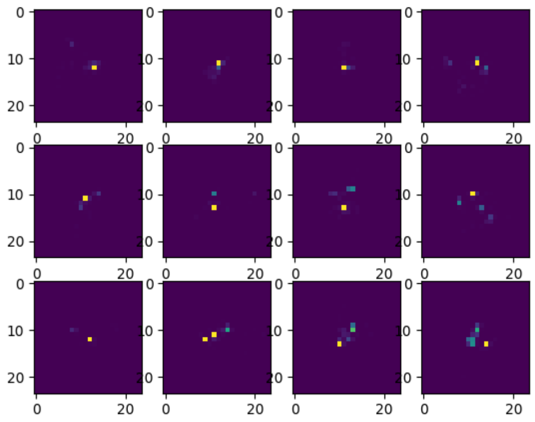
$$\lambda_{\mathrm{H}_0} = \frac{L_{\mathrm{H}_0}^{S(\mathrm{H}_0)}}{L_{\mathrm{H}_1}^{S(\mathrm{H}_0)}} \qquad \lambda_{\mathrm{H}_1} = \frac{L_{\mathrm{H}_0}^{S(\mathrm{H}_1)}}{L_{\mathrm{H}_1}^{S(\mathrm{H}_1)}} \qquad \alpha = \frac{\int_{-\infty}^{\lambda_{\mathrm{cut}}} f_1(\lambda)\mathrm{d}\lambda}{\int_{-\infty}^{\infty} f_1(\lambda)\mathrm{d}\lambda} \qquad \beta = \frac{\int_{\lambda_{\mathrm{cut}}}^{\infty} f_2(\lambda)\mathrm{d}\lambda}{\int_{-\infty}^{\infty} f_2(\lambda)\mathrm{d}\lambda}$$

- We calculate α and β as the area under the tails of the -2ln λ distributions and take α = β (symmetrical Neyman testing) or without β (Fisher testing).

Data generation

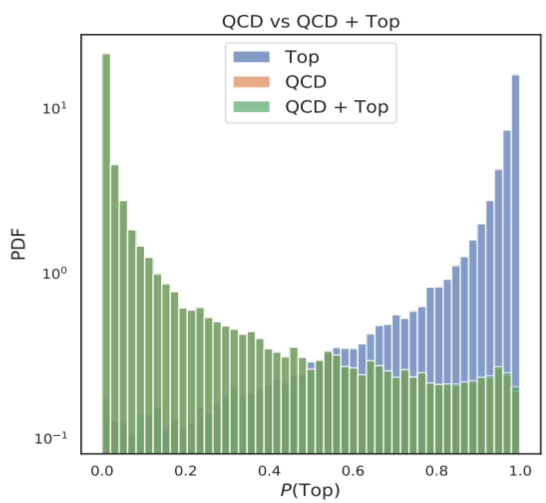
- Considered:
  - Showered jet images:
    - QCD and Top jets
    - Events generated with MadGraph5 and jet images showered with Pythia8.
  - Parton kinematics:
    - Higgs EFT [arXiv:1310.5150] :
      p p > H Z, (H > b b~), (Z > l+ l-)
    - Same channel used for SM background
    - $p_t^{b1}, p_t^{b2}, p_t^{l1}, p_t^{l2}, p_t^{h}, \eta_h, \phi_h, \Delta R_{ll}, \Delta R_{bl}, \mathrm{MET}_{vh}, p_t^{vh}, \Delta\phi^{l1b1}, \Delta\phi^{l1b2}$

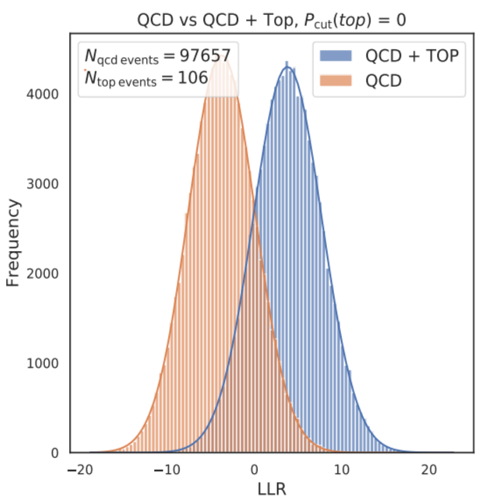
Raw SM tt~ decay leading jet



Obtaining the Log-Likelihood Ratio (LLR)

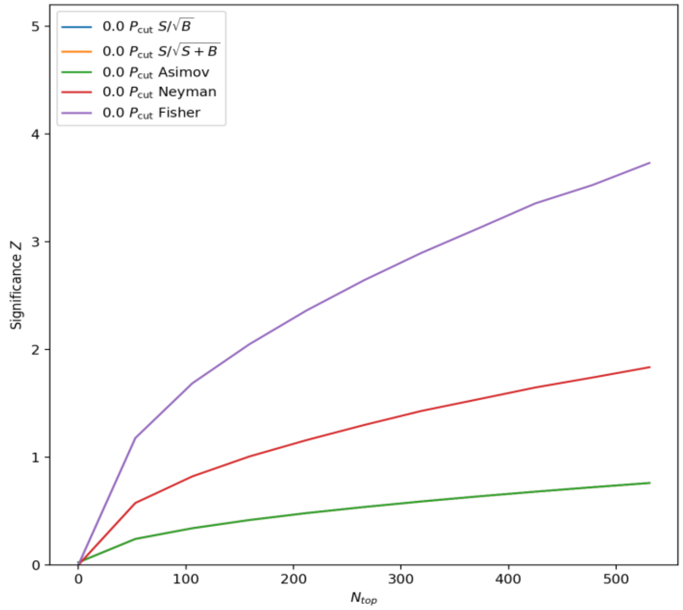
- Sample a number of toy experiments with a number of events within each to compute LLR for each experiment.
- Done both for experiments with observed data is QCD and QCD + Top.
- LLR dependant on N events ∝ Luminosity





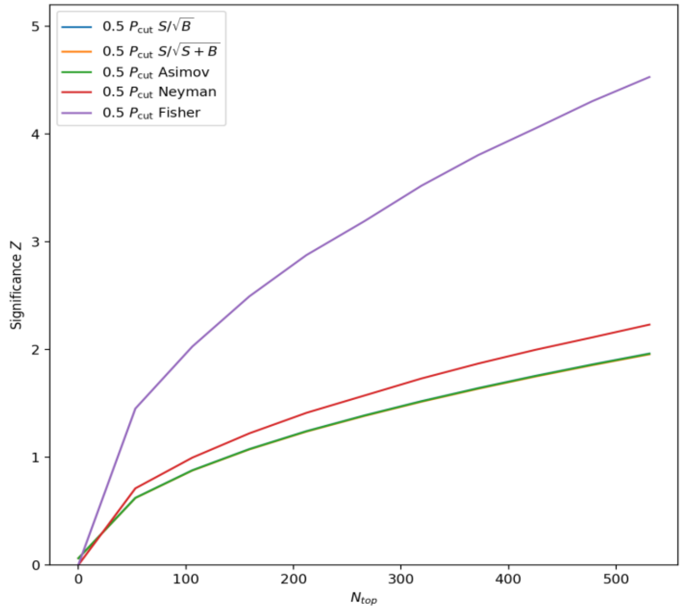
Discovery significance

- Separation significance Z calculated by solving $\alpha = \frac{1}{\sqrt{2\pi}} \int_Z^\infty \exp\left(\frac{-x^2}{2}\right).$
- Our LLR is Gaussian but Z does not need assume a Gaussian LLR distribution.



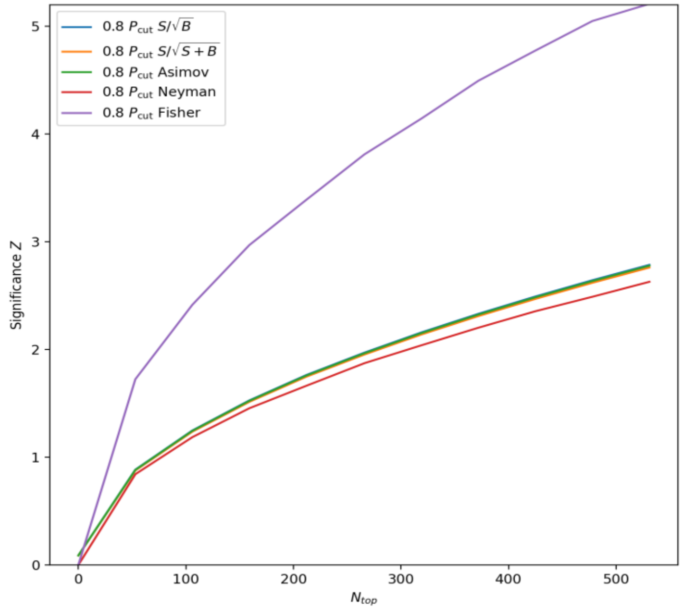
Discovery significance

- Separation significance Z calculated by solving $\alpha = \frac{1}{\sqrt{2\pi}} \int_Z^\infty \exp\left(\frac{-x^2}{2}\right).$
- Our LLR is Gaussian but Z does not need assume a Gaussian LLR distribution.



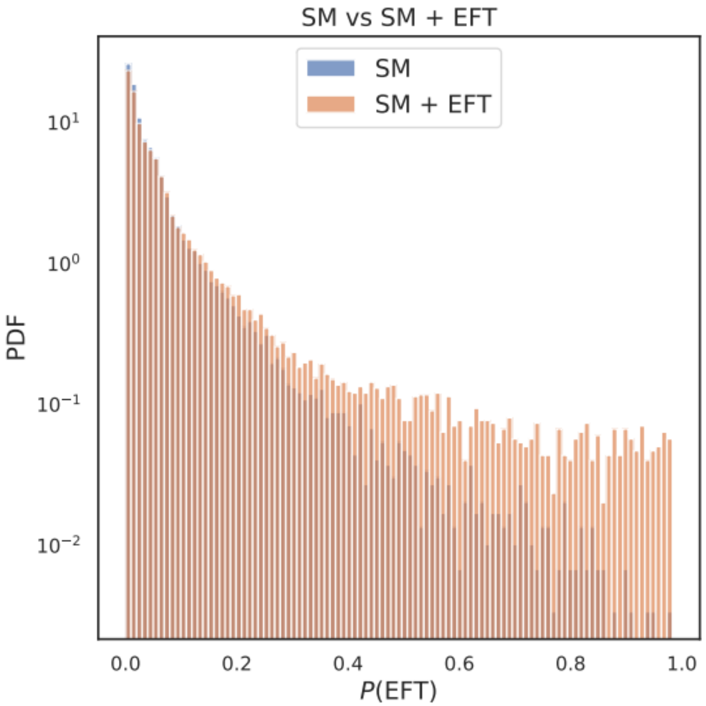
Discovery significance

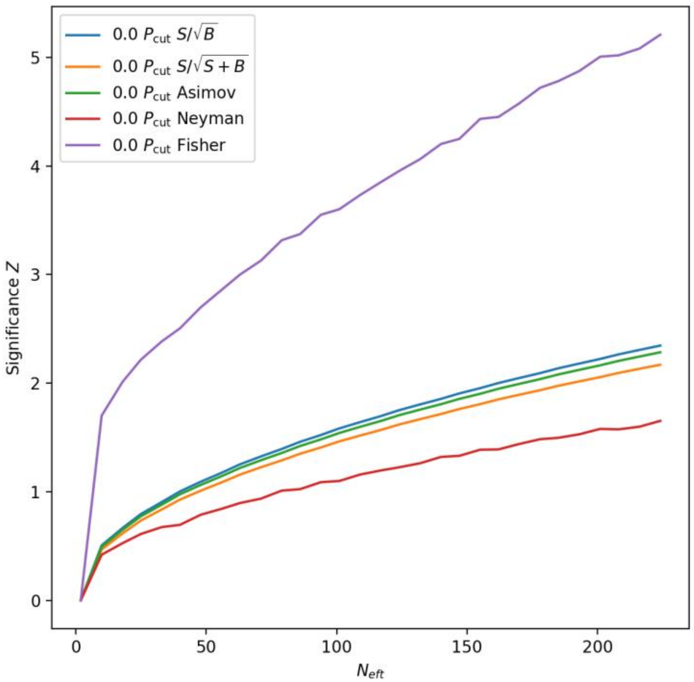
- Separation significance Z calculated by solving $\alpha = \frac{1}{\sqrt{2\pi}} \int_Z^\infty \exp\left(\frac{-x^2}{2}\right).$
- Our LLR is Gaussian but Z does not need assume a Gaussian LLR distribution.



SM vs SM + EFT supervised

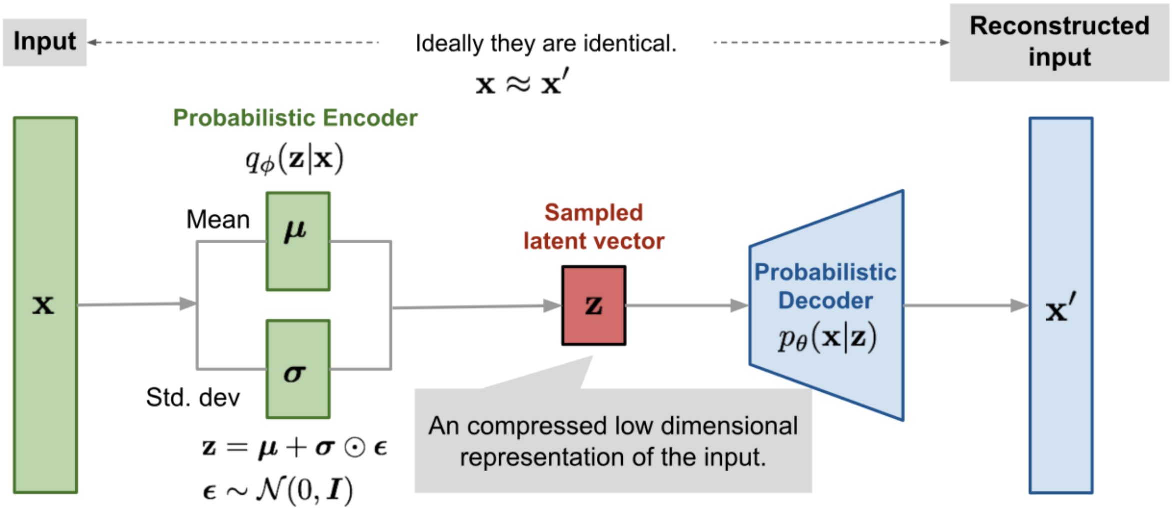
- Using 13 kinematic kinematic variables with a DNN.





SM vs SM + EFT unsupervised

- Variational Autoencoder (VAE) used to reconstruct input images, having been trained on SM data.
- Reconstruction error $R = |p_\theta(q_\phi(x)) - x|^2$ is larger for more anomalous images.



The Generalised Likelihood Ratio Test

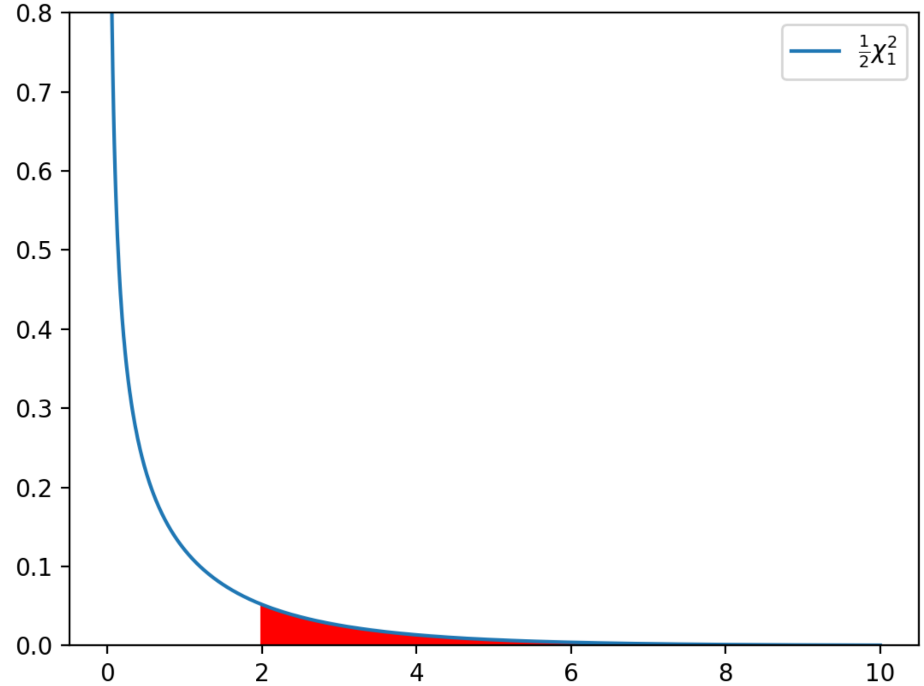
- For anomaly detection we assume the background H_0 is known but the alternative H_1 is not.
- In other words the parameter which defines the signal model, here c_{HW}, is unknown.
- Use the Generalised LRT $\lambda(\mu) = \frac{L(\mu)}{L(\hat{\mu})}, \quad \mu = l\sigma(c_{\mathrm{HW}})$; with $\mu = 0$:

$$q_0 = \begin{cases} -2\ln\lambda(0) & \text{for } \hat{\mu} \geq 0 \\ 0 & \text{for } \hat{\mu} < 0 \end{cases}$$

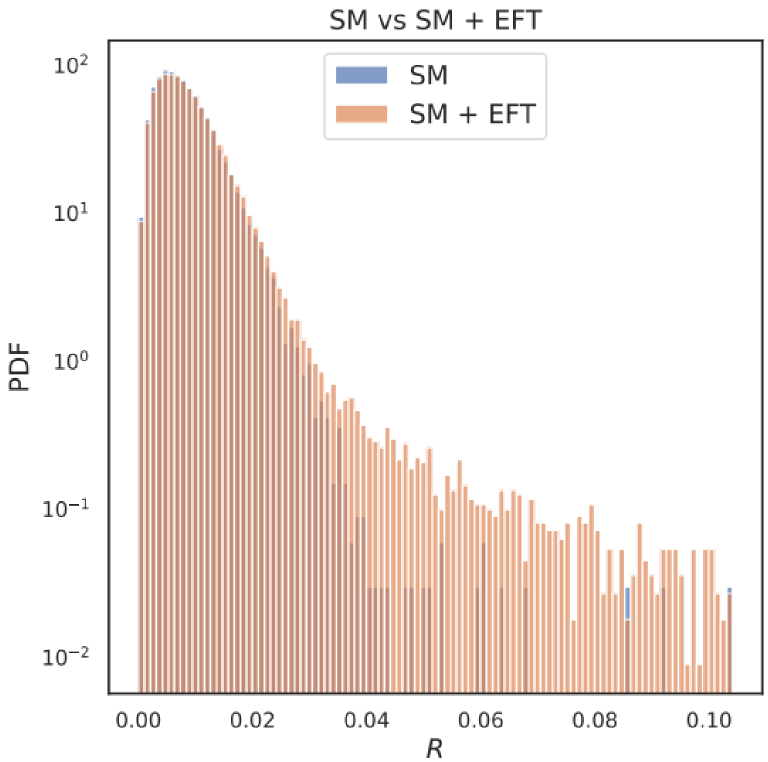
- q_0 under the null has a half-χ_1^2 distribution as shown by Wilks and Wald.

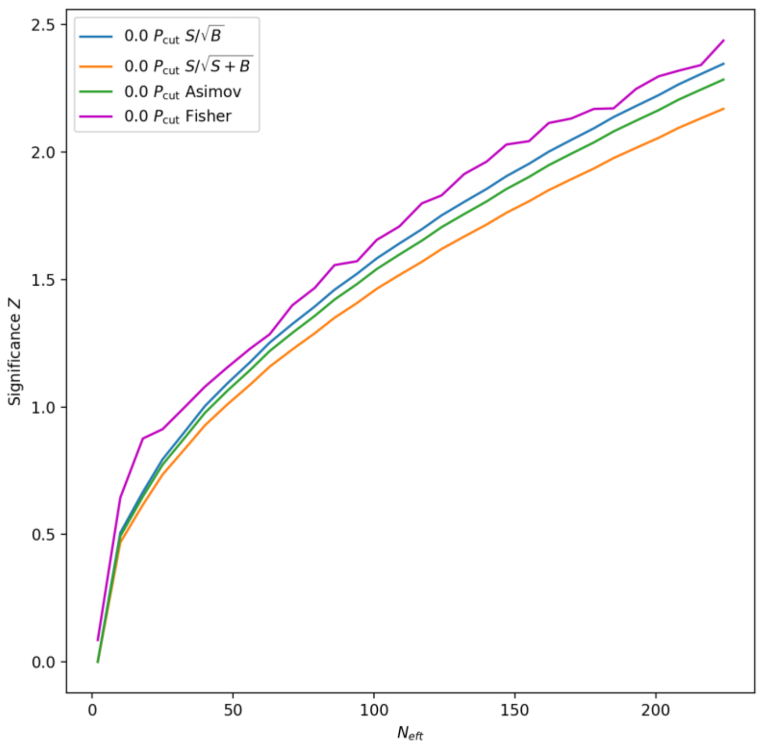
The Generalised Likelihood Ratio Test

- For discovery we obtain the p-value from comparing $q_{0,\mathrm{obs}}$ to the half-chi-square distribution.



SM vs SM + EFT unsupervised





Thank you!