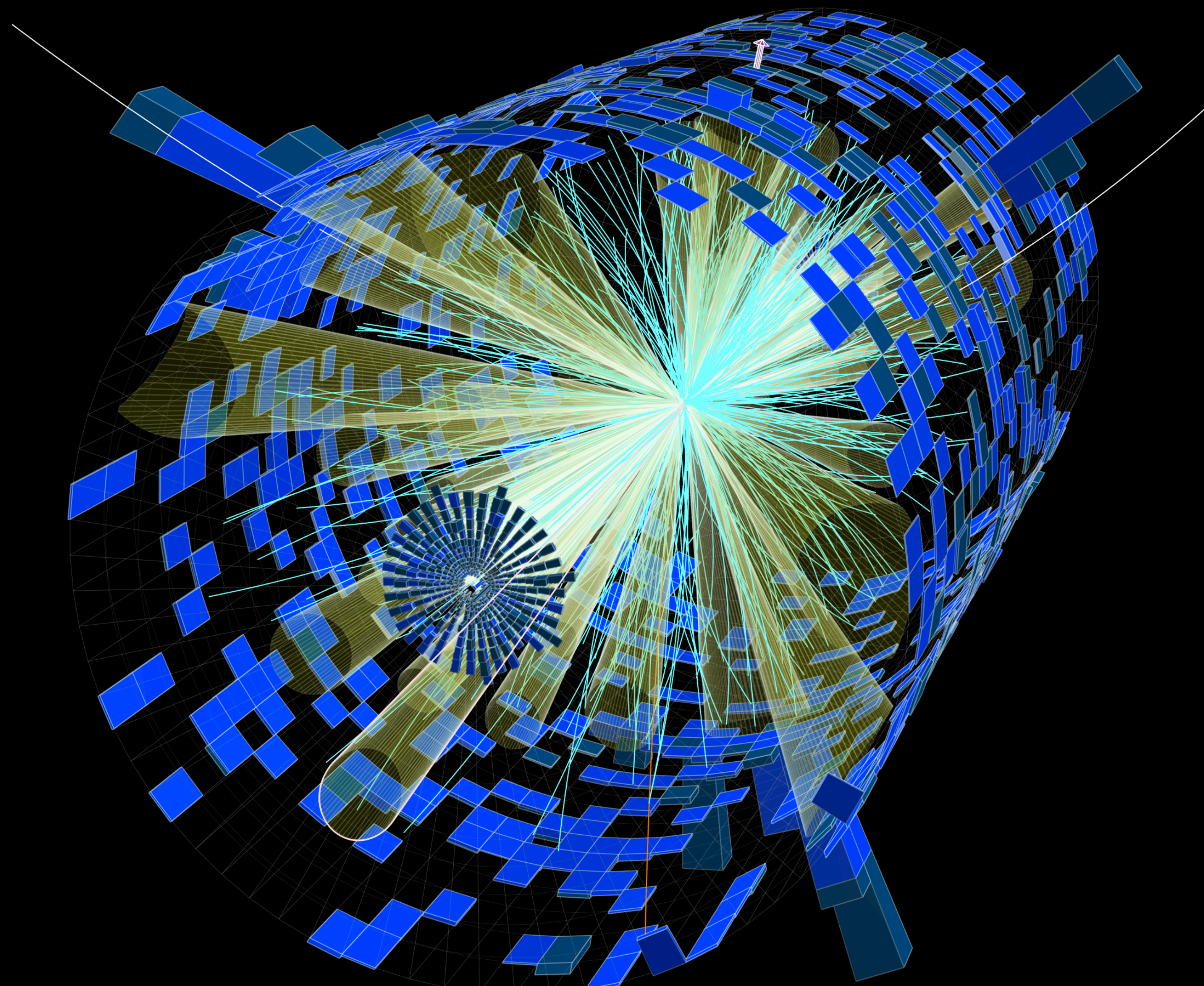
NYU Center for Data Science

Center for Cosmology and Particle Physics

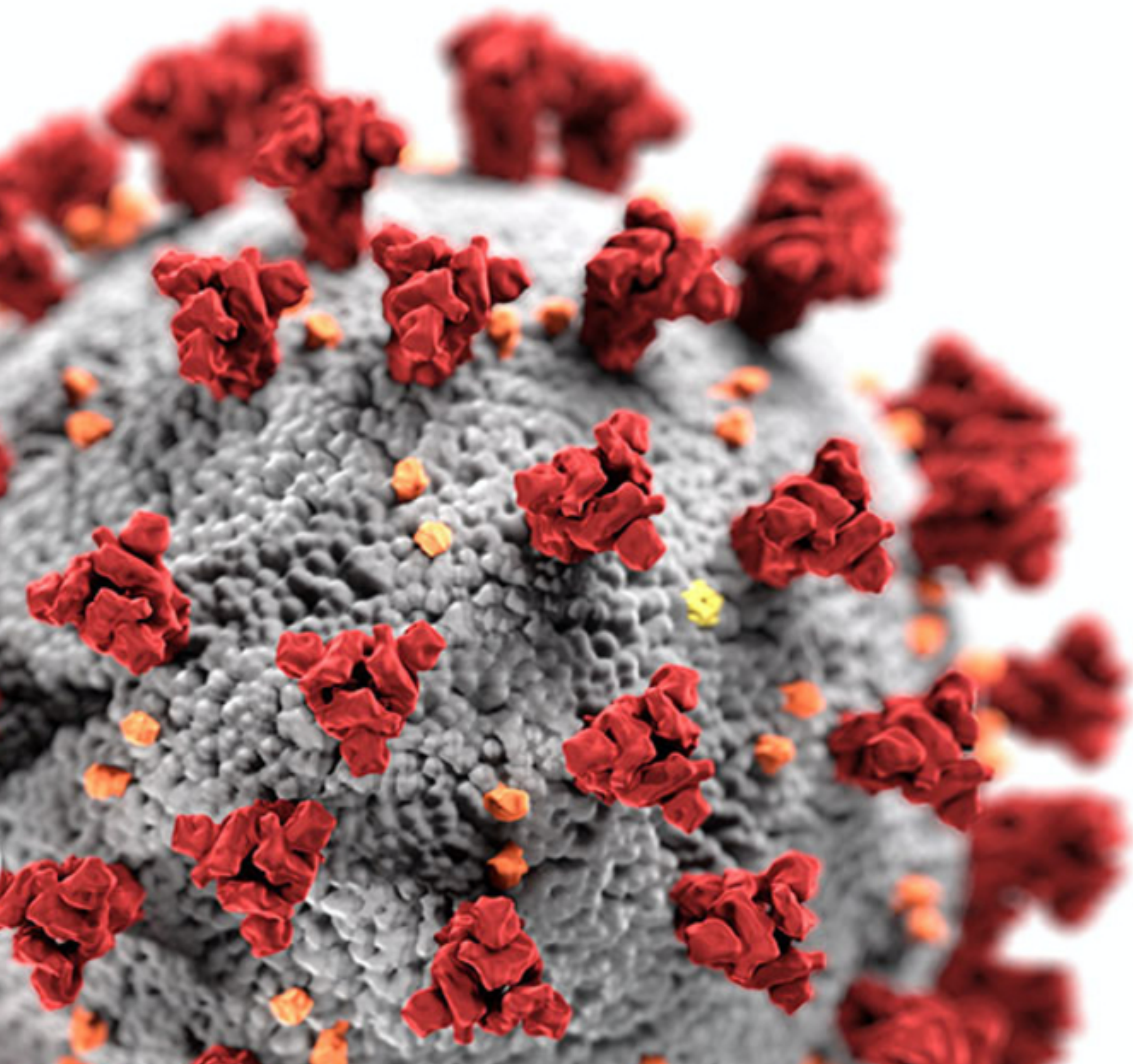
CONSTRAINING EFTs

WITH MACHINE LEARNING

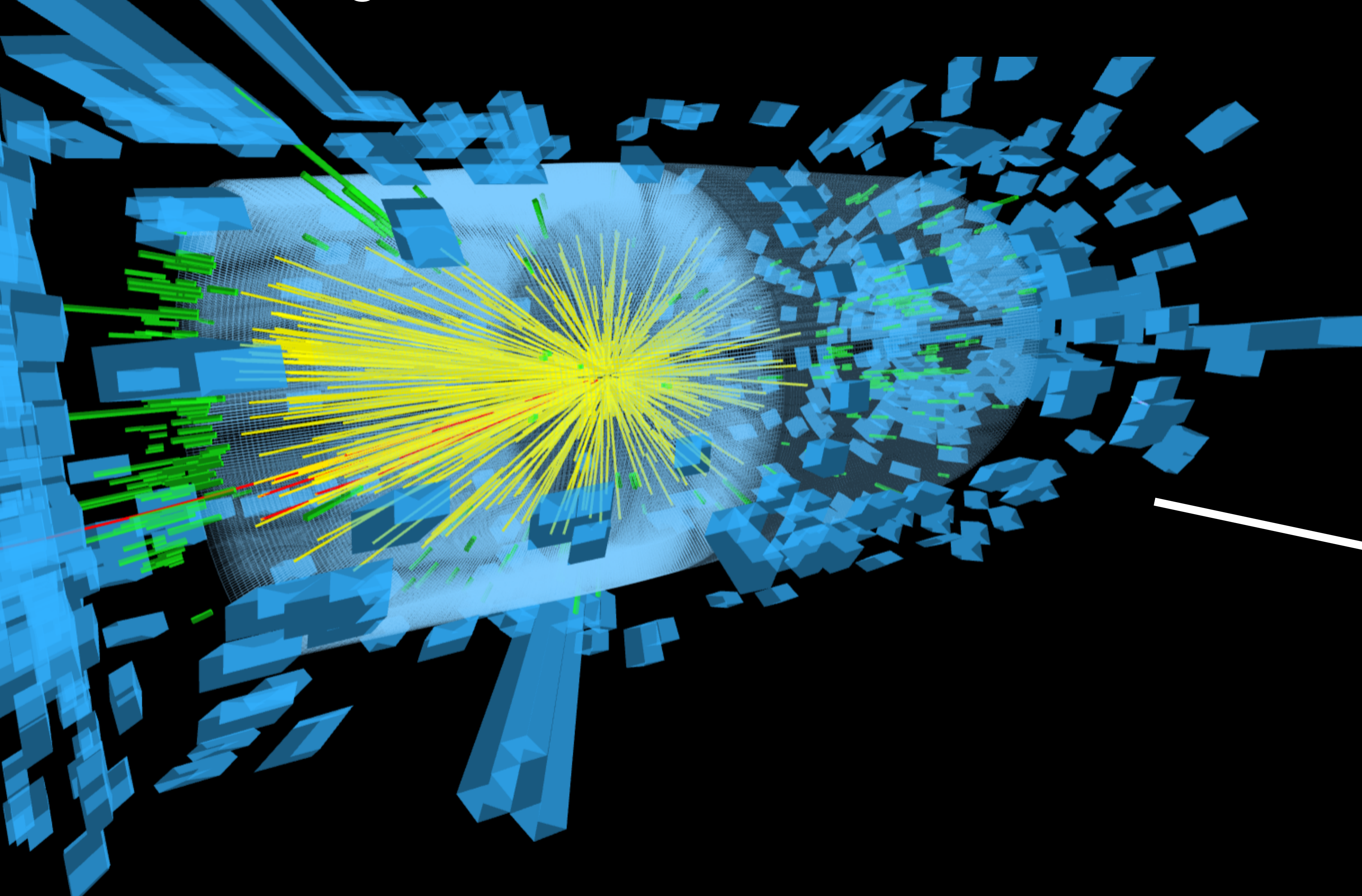


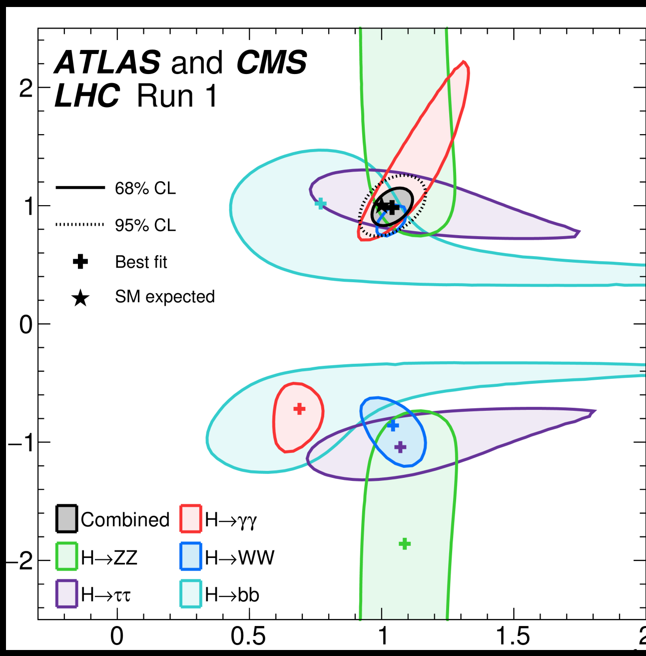
@KyleCranmer
New York University
Department of Physics
Center for Data Science
CILVR Lab

I’m sorry I couldn’t be there in person,
I was looking forward to it.



High-dimensional event data

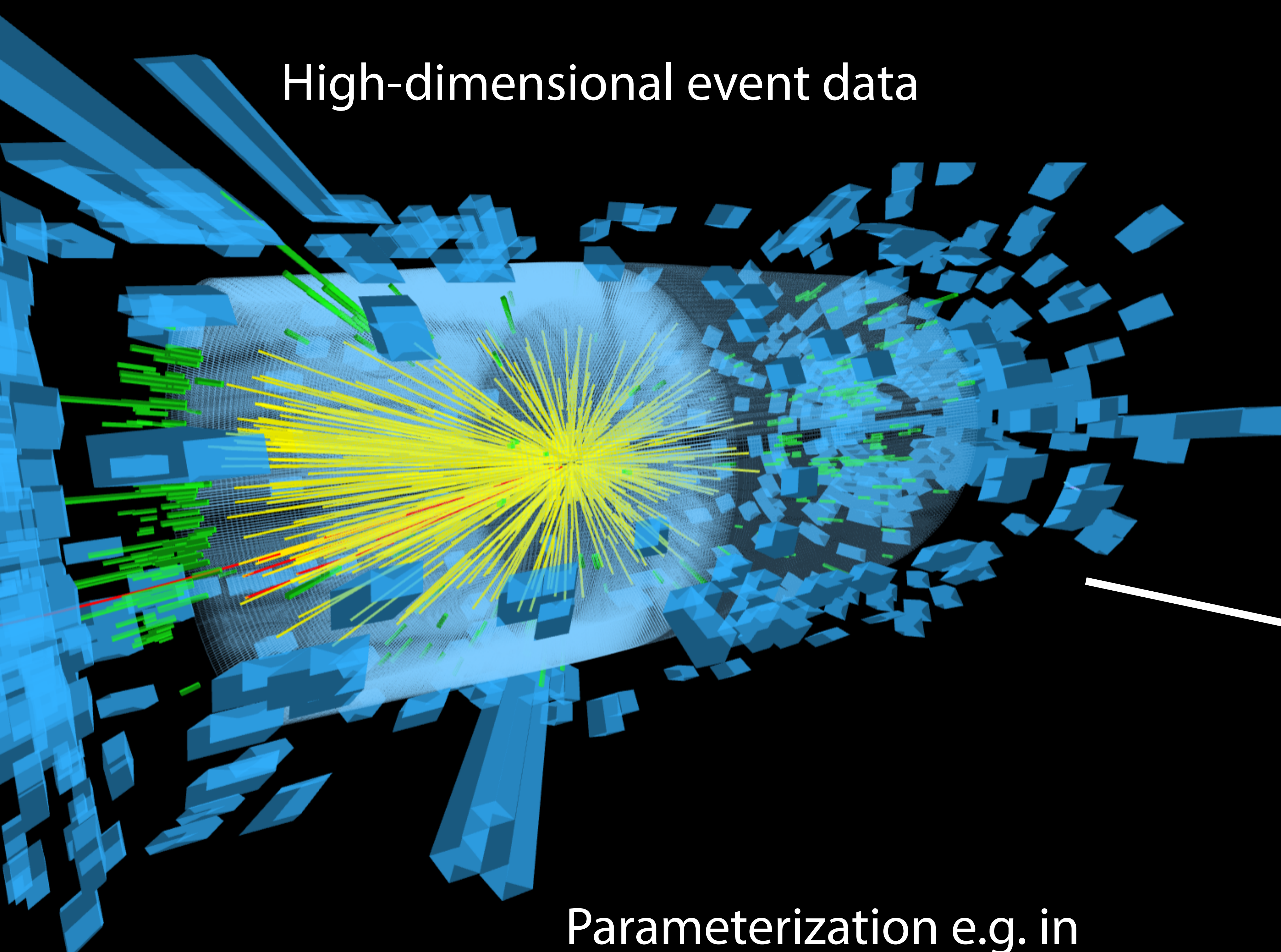


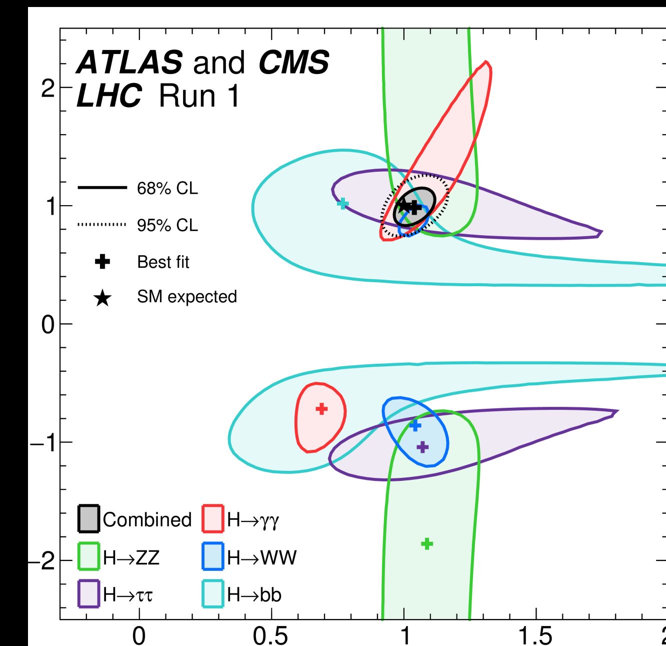


[ATLAS, CMS 1606.02266]

Precision constraints on new physics

High-dimensional event data





[ATLAS, CMS 1606.02266]

Precision constraints on new physics

Parameterization e.g. in Effective Field Theory:

systematic expansion of new physics around Standard Model

$$\mathcal{L}_{\mathrm{EFT}} = \mathcal{L}_{\mathrm{SM}} + \sum_i \frac{f_i}{\Lambda^2} \mathcal{O}_i + \dots$$

10s to 100s "universal" parameters to measure

SMEFT (Standard Model Effective Field Theory)

$$\mathcal{L}_{\mathrm{EFT}} = \mathcal{L}_{\mathrm{SM}} + \sum_i \frac{f_i}{\Lambda^2}\,\mathcal{O}_i + \ldots$$

Operators

- all possible interactions between SM particles mediated by new physics
- fixed by SM particles + SM symmetries + expansion in $1/\Lambda$, independent of high-energy physics
- affect rates + kinematics

$$\begin{aligned}
\mathcal{O}_{\phi,1} &= (D_\mu\phi)^\dagger \phi\, \phi^\dagger D^\mu \phi & \mathcal{O}_{GG} &= (\phi^\dagger\phi)\, G^a_{\mu\nu}\, G^{\mu\nu\, a} \\
\mathcal{O}_{\phi,2} &= \frac{1}{2}\,\partial_\mu(\phi^\dagger\phi)\,\partial^\mu(\phi^\dagger\phi) & \mathcal{O}_{BB} &= -\frac{g'^2}{4}\,(\phi^\dagger\phi)\, B_{\mu\nu}\, B^{\mu\nu} \\
\mathcal{O}_{\phi,3} &= \frac{1}{3}(\phi^\dagger\phi)^3 & \mathcal{O}_{WW} &= -\frac{g^2}{4}\,(\phi^\dagger\phi)\, W^a_{\mu\nu}\, W^{\mu\nu\, a} \\
\mathcal{O}_{\phi,4} &= (\phi^\dagger\phi)\,(D_\mu\phi)^\dagger D^\mu\phi & \mathcal{O}_{BW} &= -\frac{g\,g'}{4}\,(\phi^\dagger\sigma^a\phi)\, B_{\mu\nu}\, W^{\mu\nu\, a} \\
& & \mathcal{O}_{B} &= \frac{\mathrm{i}g'}{2}\,(D^\mu\phi)^\dagger D^\nu\phi\, B_{\mu\nu} \\
& & \mathcal{O}_{W} &= \frac{\mathrm{i}g}{2}\,(D^\mu\phi)^\dagger \sigma^a D^\nu\phi\, W^a_{\mu\nu}
\end{aligned}$$

Wilson coefficients

- precise measurement of these parameters is one of the most important goals of the LHC
- can be translated to high-energy physics parameters

Higher-order terms

- suppressed by additional factors of $E^2\,/\,\Lambda^2$

[W. Buchmuller, D. Wyler 85;
B. Grzadkowski, M. Iskrzynski, M. Misiak, J. Rosiek 1008.4884; …]

OUTLINE

Lecture 1:

- Inferring EFT parameters from a statistical perspective
- Information Geometry (with results at parton-level)

Lecture 2:

- Why it is hard beyond parton-level
- How Machine Learning enters (likelihood-free inference)

Lecture 3:

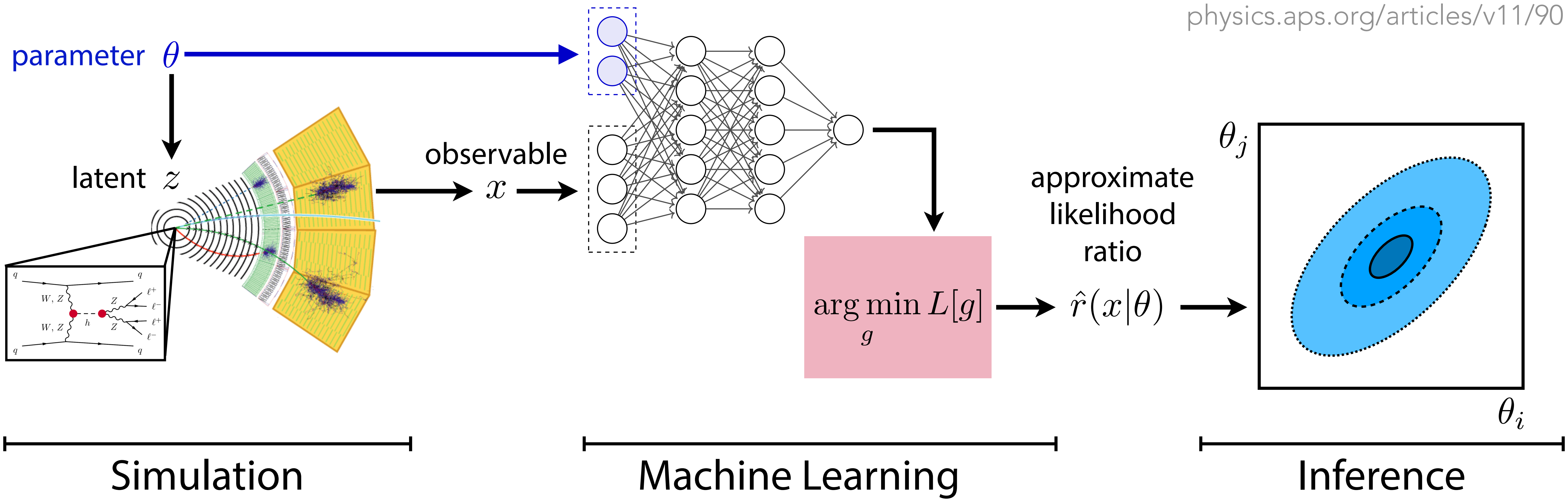
- Hands-on tutorial with MadMiner

Lecture 4:

- Other highlights of Machine Learning

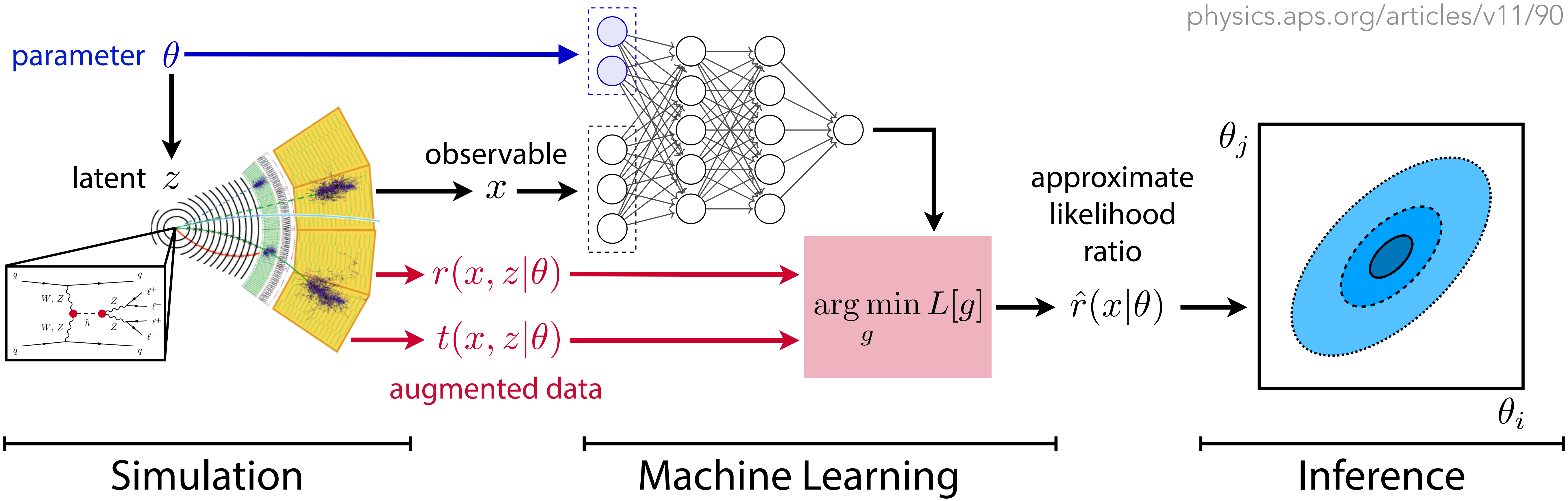
LEARNING FROM SIMULATED DATA

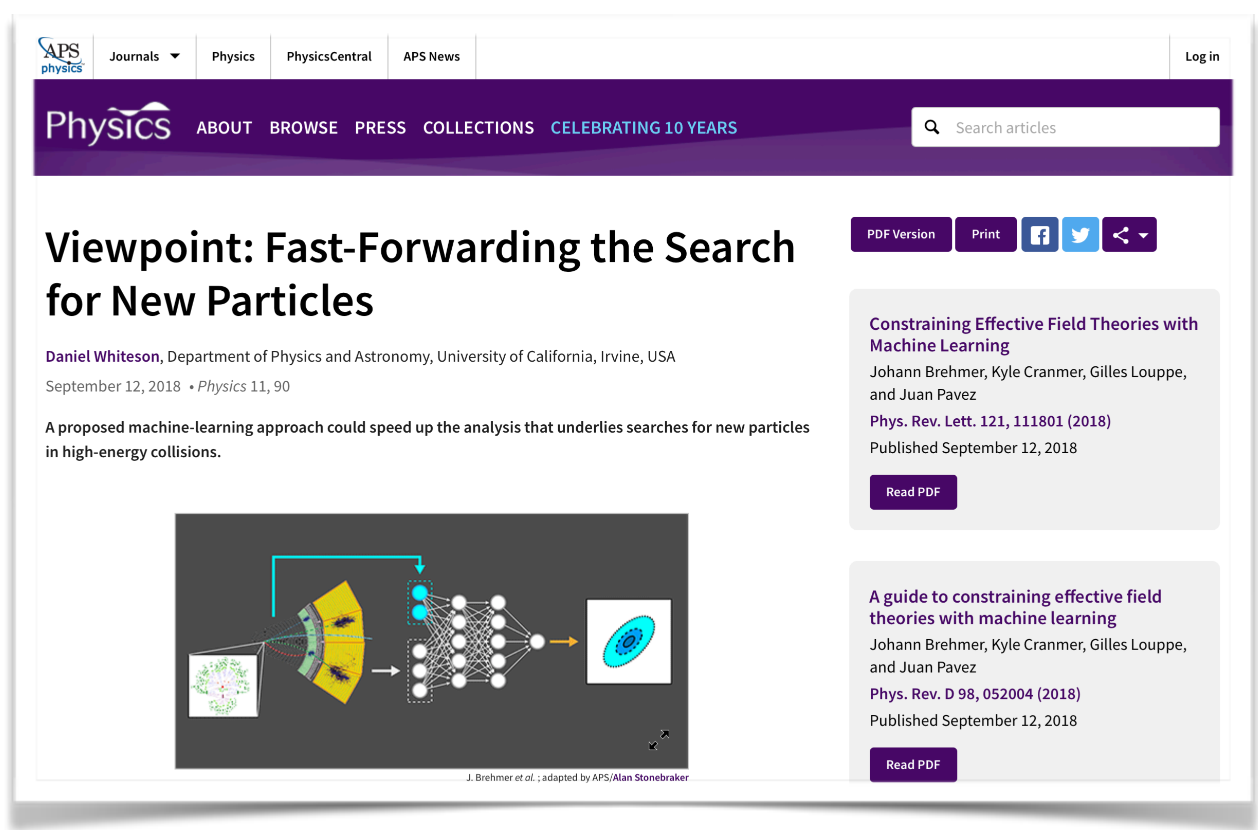
arXiv:1805.12244
PRL, arXiv:1805.00013
PRD, arXiv:1805.00020
physics.aps.org/articles/v11/90



LEARNING WITH AUGMENTED DATA

arXiv:1805.12244
PRL, arXiv:1805.00013
PRD, arXiv:1805.00020
physics.aps.org/articles/v11/90





APS physics | Journals | Physics | PhysicsCentral | APS News | Log in

Physics ABOUT BROWSE PRESS COLLECTIONS CELEBRATING 10 YEARS | Search articles

PDF Version | Print

Viewpoint: Fast-Forwarding the Search for New Particles

Daniel Whiteson, Department of Physics and Astronomy, University of California, Irvine, USA

September 12, 2018 • *Physics* 11, 90

A proposed machine-learning approach could speed up the analysis that underlies searches for new particles in high-energy collisions.

J. Brehmer *et al.*; adapted by APS/Alan Stonebraker

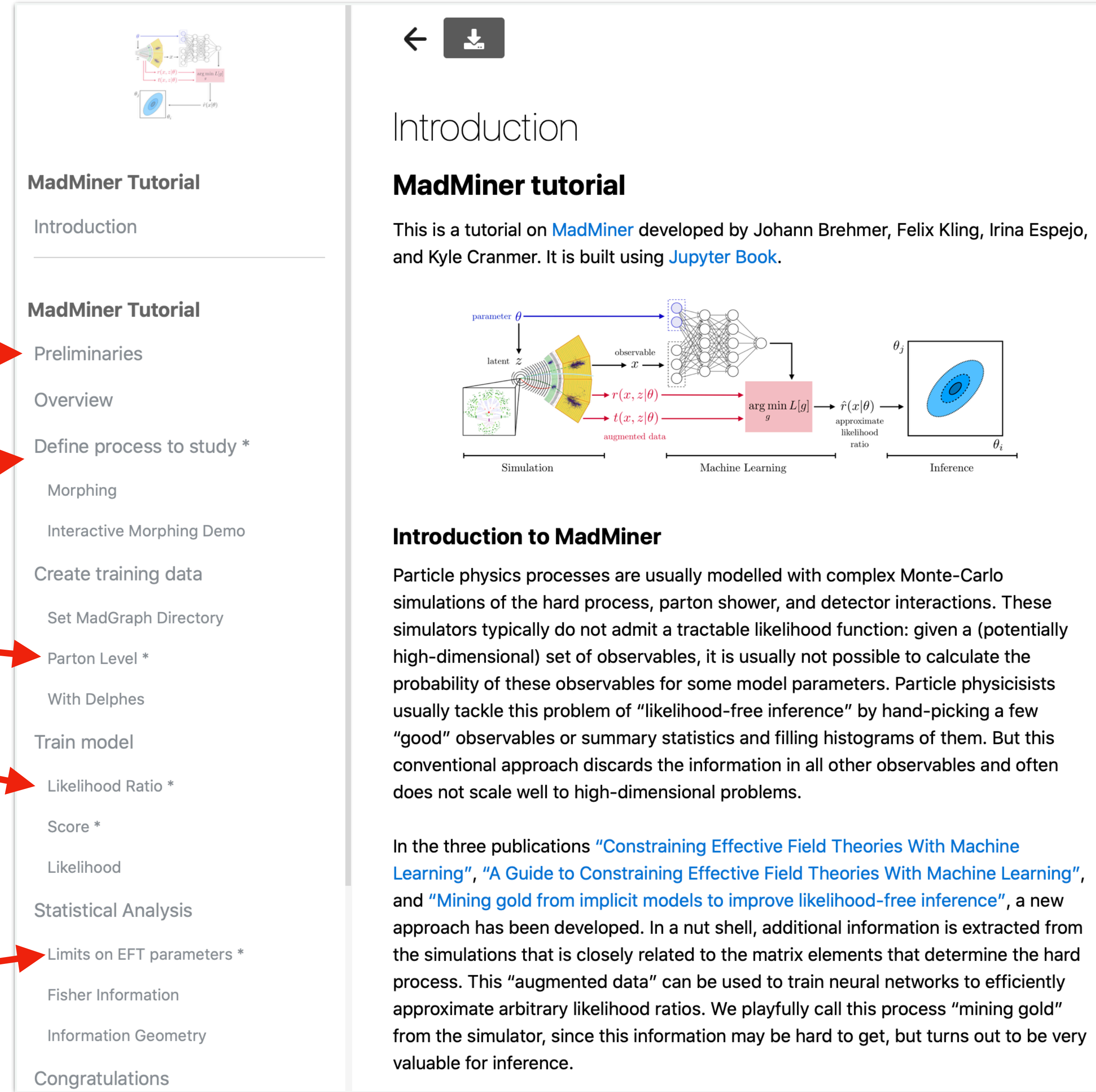
Constraining Effective Field Theories with Machine Learning
Johann Brehmer, Kyle Cranmer, Gilles Louppe, and Juan Pavez
Phys. Rev. Lett. 121, 111801 (2018)
Published September 12, 2018
Read PDF

A guide to constraining effective field theories with machine learning
Johann Brehmer, Kyle Cranmer, Gilles Louppe, and Juan Pavez
Phys. Rev. D 98, 052004 (2018)
Published September 12, 2018
Read PDF

Hands-on Tutorial

https://cranmer.github.io/madminer-tutorial/

- Step-by-step instructions
- Do Preliminaries
  - Need to install Docker
  - And then start Jupyter
- Step 1 is fast
- Step 2 takes ~25 min
- Step 3 takes ~20 min
- While they are running I will lecture
- Then we will finish with results



MadMiner Tutorial

Introduction

MadMiner Tutorial

Preliminaries

Overview

Define process to study *

Morphing

Interactive Morphing Demo

Create training data

Set MadGraph Directory

Parton Level *

With Delphes

Train model

Likelihood Ratio *

Score *

Likelihood

Statistical Analysis

Limits on EFT parameters *

Fisher Information

Information Geometry

Congratulations

Introduction

MadMiner tutorial

This is a tutorial on MadMiner developed by Johann Brehmer, Felix Kling, Irina Espejo, and Kyle Cranmer. It is built using Jupyter Book.

Introduction to MadMiner

Particle physics processes are usually modelled with complex Monte-Carlo simulations of the hard process, parton shower, and detector interactions. These simulators typically do not admit a tractable likelihood function: given a (potentially high-dimensional) set of observables, it is usually not possible to calculate the probability of these observables for some model parameters. Particle physicisists usually tackle this problem of "likelihood-free inference" by hand-picking a few "good" observables or summary statistics and filling histograms of them. But this conventional approach discards the information in all other observables and often does not scale well to high-dimensional problems.

In the three publications "Constraining Effective Field Theories With Machine Learning", "A Guide to Constraining Effective Field Theories With Machine Learning", and "Mining gold from implicit models to improve likelihood-free inference", a new approach has been developed. In a nut shell, additional information is extracted from the simulations that is closely related to the matrix elements that determine the hard process. This "augmented data" can be used to train neural networks to efficiently approximate arbitrary likelihood ratios. We playfully call this process "mining gold" from the simulator, since this information may be hard to get, but turns out to be very valuable for inference.

References

Information Geometry for Higgs EFT (no machine learning)
Brehmer, Kling, Plehn, Cranmer
"Better Higgs Measurements Through Information Geometry"
[PRD, 1612.05261]

Brehmer, Kling, Plehn, Tait
"Better Higgs-CP Tests Through Information Geometry"
[PRD, 1712.02350]

Core References for ML-based approach
JB, K. Cranmer, G. Louppe, J. Pavez:
"Constraining Effective Field Theories with machine learning"
[PRL, 1805.00013]

JB, K. Cranmer, G. Louppe, J. Pavez:
"A guide to constraining Effective Field Theories with machine learning"
[PRD, 1805.00020]

JB, G. Louppe, J. Pavez, K. Cranmer:
"Mining gold from implicit models to improve likelihood-free inference"
[PNAS, 1805.12244]

Follow-up with incremental improvements
M. Stoye, JB, K. Cranmer, G. Louppe, J. Pavez:
"Likelihood-free inference with an improved cross-entropy estimator"
[NeurIPS workshop, 1808.00973]

Opinionated review of simulation-based inference
K. Cranmer, JB, G. Louppe:
"The frontier of simulation-based inference"
[1911.01429]

Do It Yourself (for LHC physics)
JB, F. Kling, I. Espejo, K. Cranmer:
"MadMiner: Machine learning—based inference for particle physics"
[CSBS, 1907.10621, https://github.com/diana-hep/madminer]

Strong lensing
JB, S. Mishra-Sharma, J. Hermans, G. Louppe, K. Cranmer
"Mining for Dark Matter Substructure: Inferring subhalo population properties from strong lenses with machine learning"
[ApJ, 1909.02005]

LHC HXSWG YR4 STXS
JB, S. Dawson, S. Homiller, F. Kling, T. Plehn:
"Benchmarking simplified template cross sections in WH production"
[JHEP, 1908.06980]

ACKNOWLEDGEMENTS


Johann Brehmer


Gilles Louppe

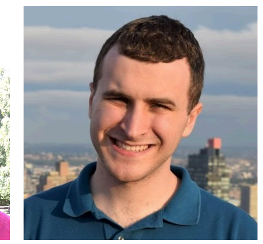

Juan Pavez

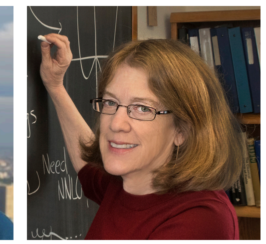

Markus Stoye


Felix Kling


Irina Espejo


Tilman Plehn


Sam Homiller


Sally Dawson


Sid Mishra-Sharma


Zubair Bhatti

NYU

MSDSE
Moore - Sloan
Data Science Environments

dianahep

iris hep
Institute for Research and Innovation
in Software for High Energy Physics

The SCAILFIN Project
scailfin.github.io

INFERRING EFT PARAMETERS FROM A STATISTICAL PERSPECTIVE

SOME STATISTICAL FOUNDATIONS

NOTATION / TERMINOLOGY

forward modeling
generation
simulation

PREDICTION

θ
parameters of interest

$p(\mathbf{x}, \mathbf{z} \mid \theta, \nu)$

x
observed data
simulated data

z
latent variables

ν
nuisance parameters

INFERENCE

inverse problem
measurement
parameter estimation

Terms & Notation

- Random variables / "observables" x
- Unobserved "latent" variables z
- Probability mass and probability density function (pdf) $p(x)$ or $f(x)$
- Parametrized Family of pdfs / "model" $p(x|\alpha)$ *or* $p(x|\theta)$
- Parameter α or θ or μ or ν
- Likelihood $L(\alpha)$
- Estimate (of a parameter) $\hat{\alpha}(x)$ or $\hat{\theta}(x)$

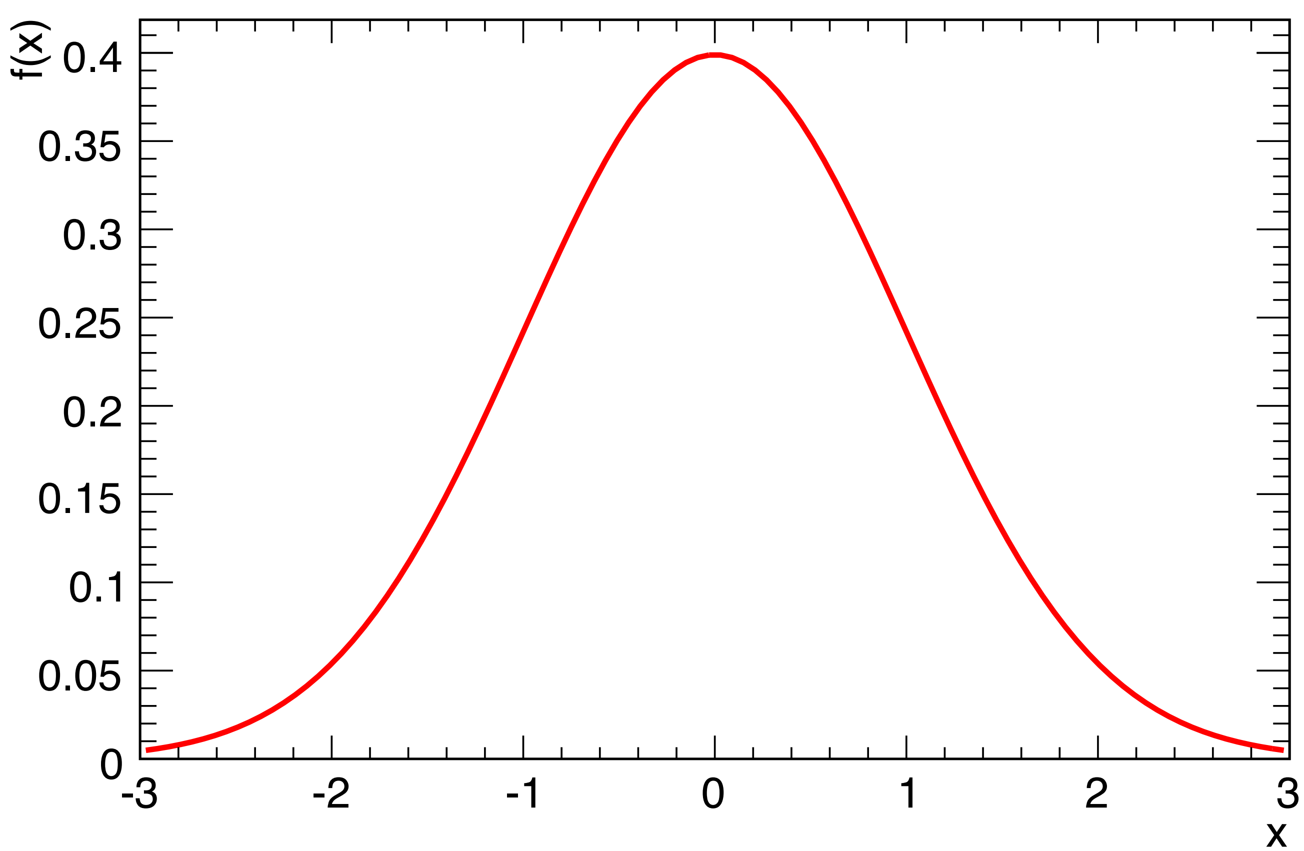
When dealing with continuous random variables, need to introduce the notion of a **Probability Density Function**

$$P(x \in [x, x + dx]) = f(x)dx$$

Note, $f(x)$ is NOT a probability

PDFs are always normalized t

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

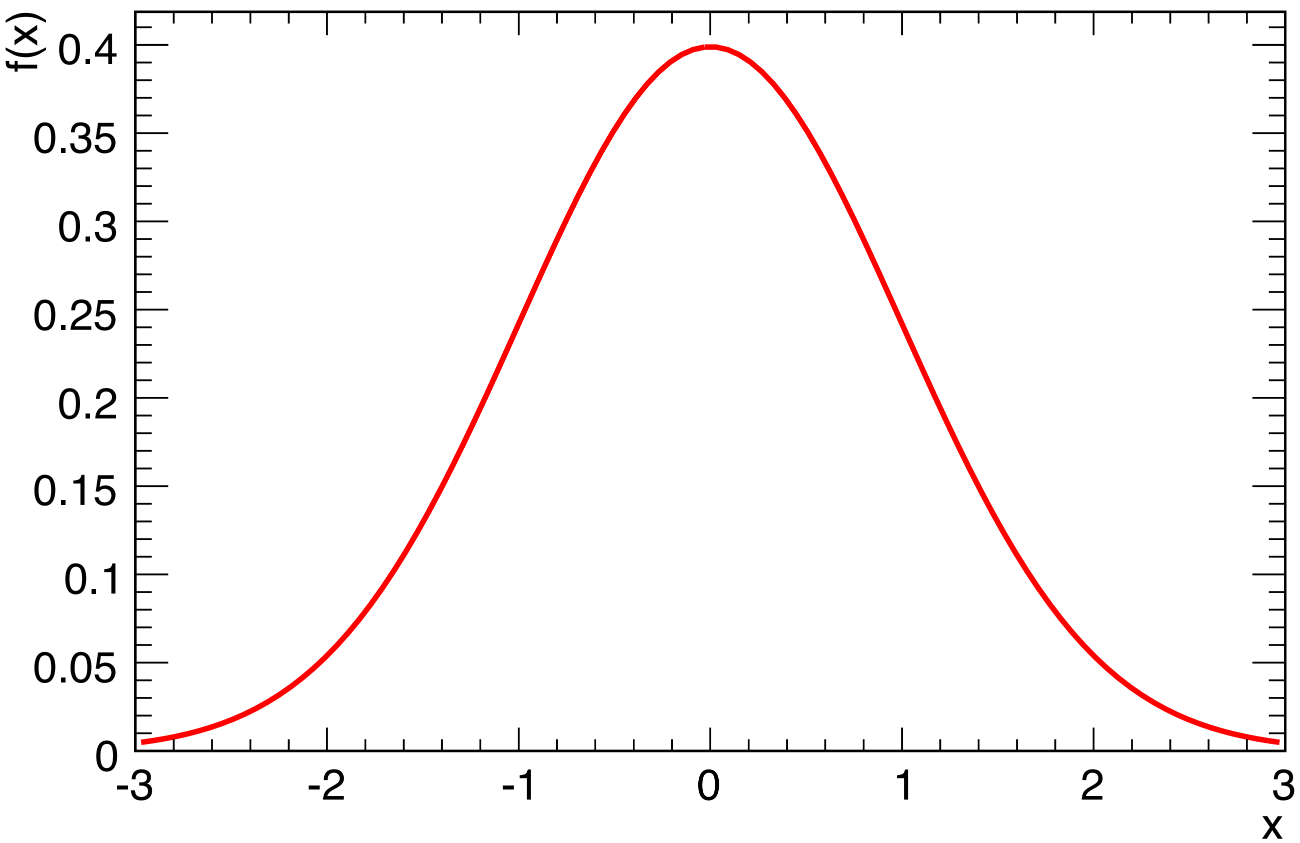


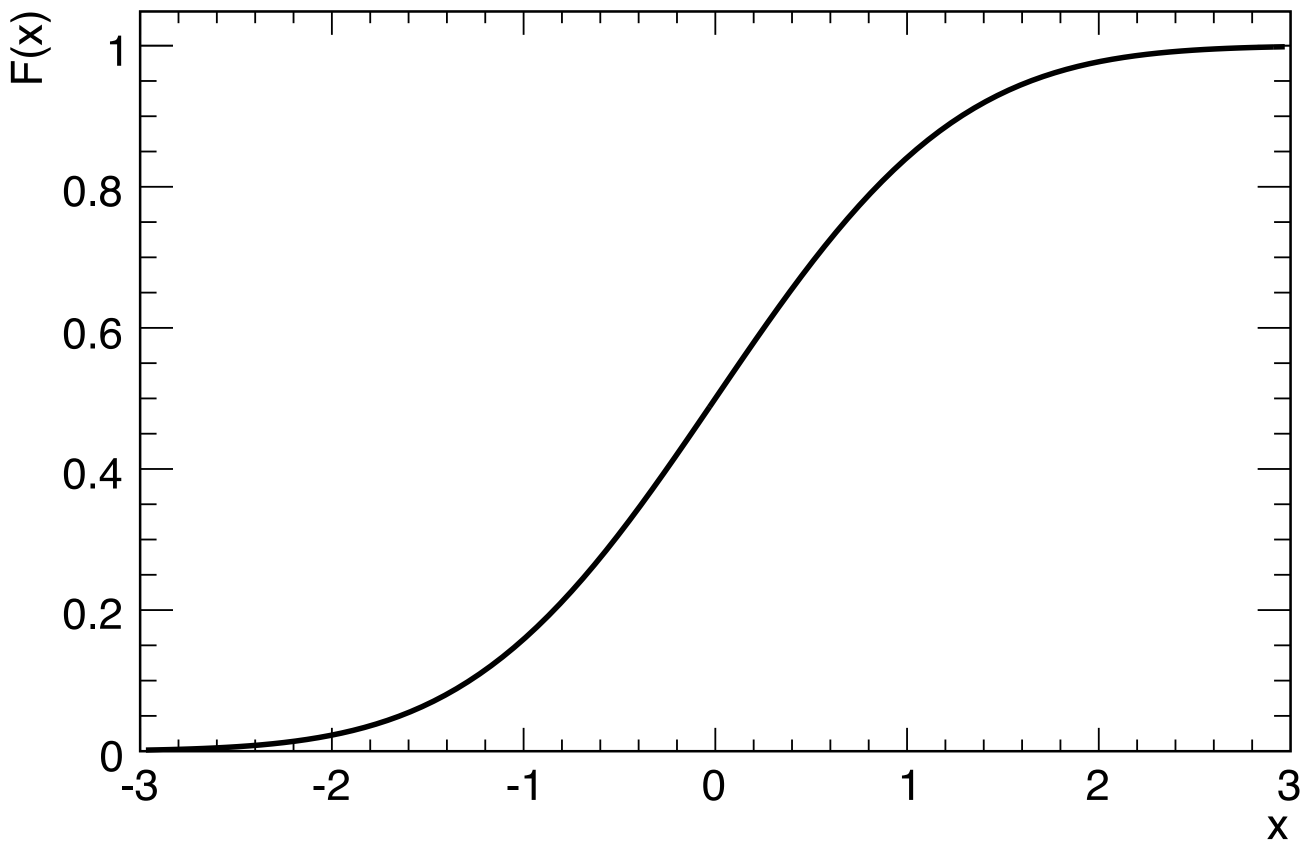
Cumulative Density Functions

Often useful to use a cumulative distribution:

- in 1-dimension:

$$\int_{-\infty}^{x} f(x')dx' = F(x)$$



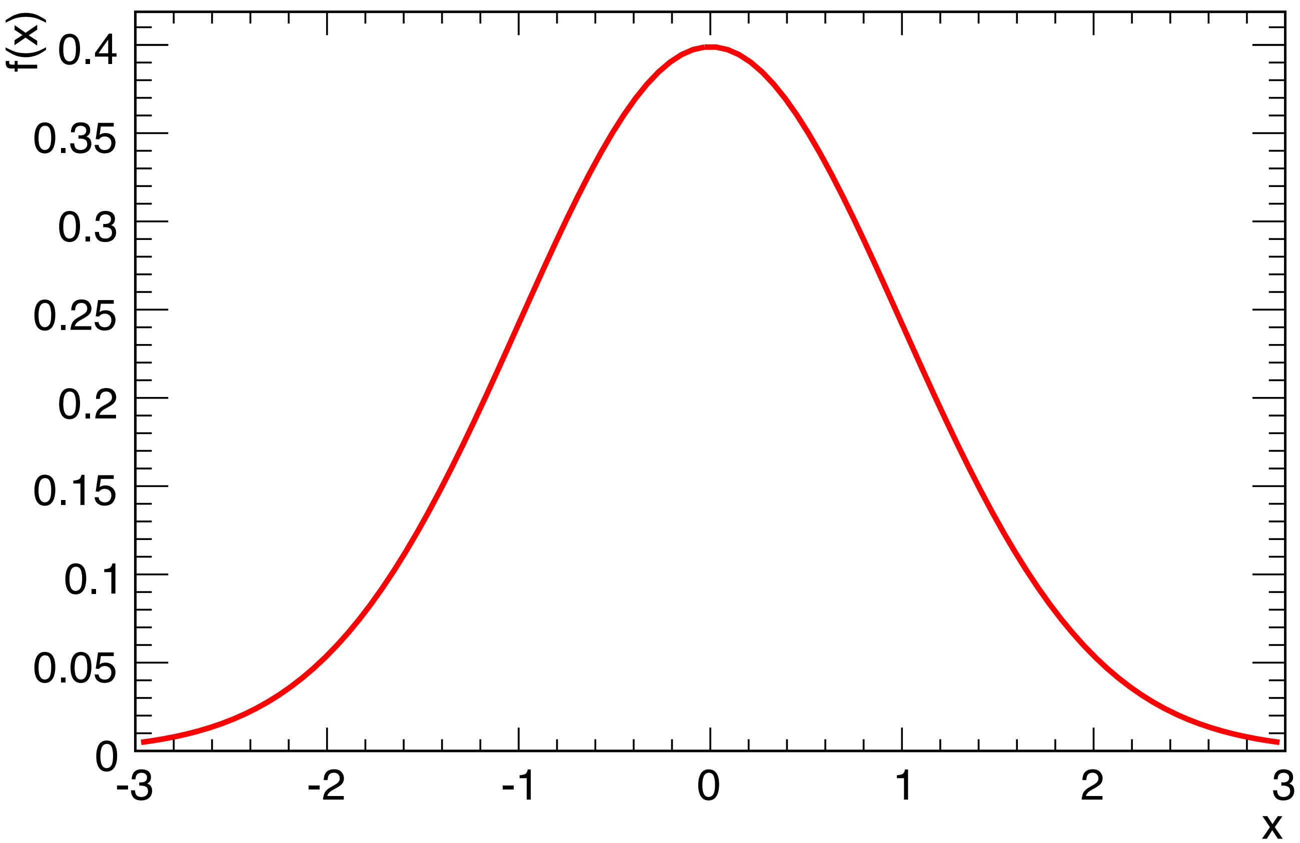


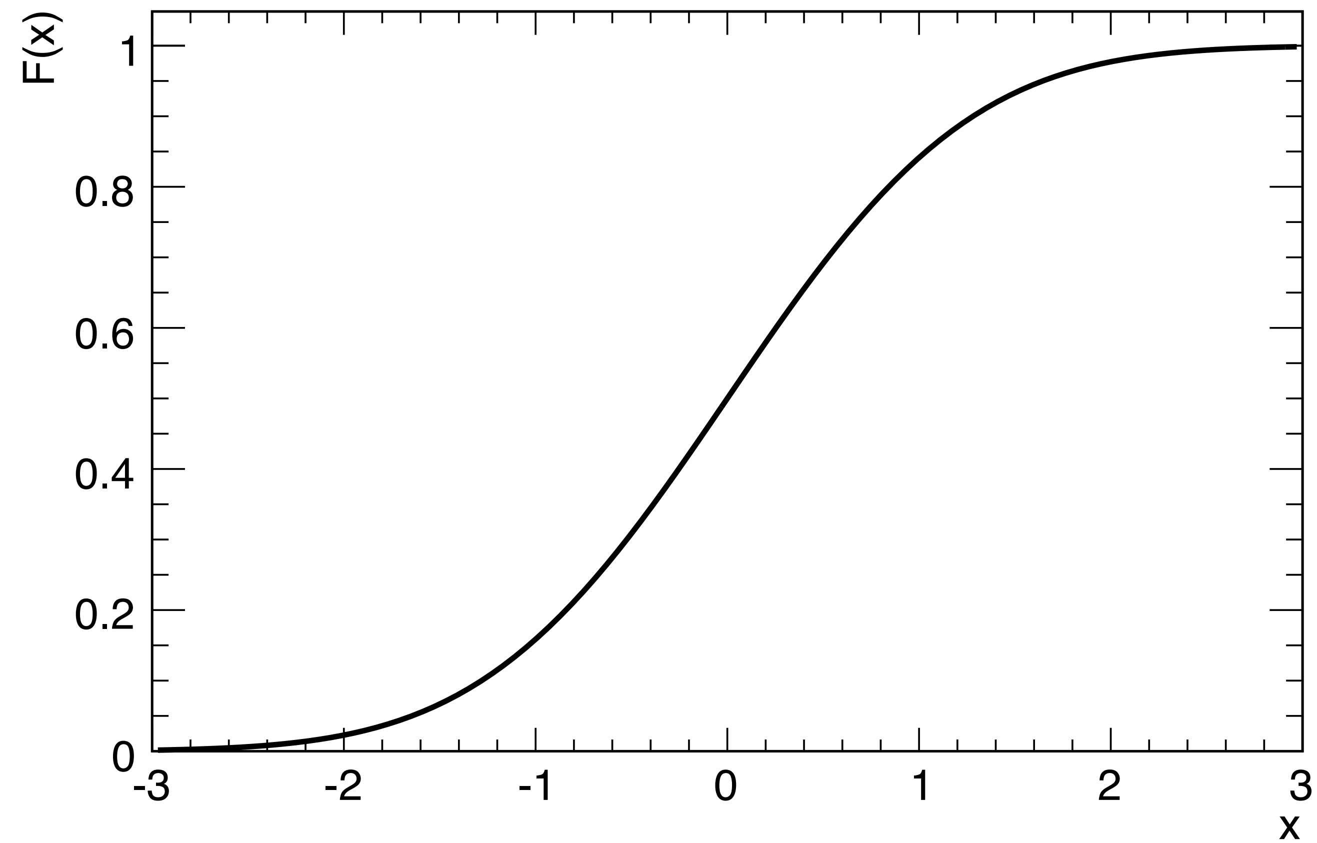
CUMULATIVE DENSITY FUNCTIONS

Often useful to use a cumulative distribution:

- in 1-dimension:

$$\int_{-\infty}^{x} f(x')dx' = F(x)$$





- alternatively, define density as partial of cumulative:

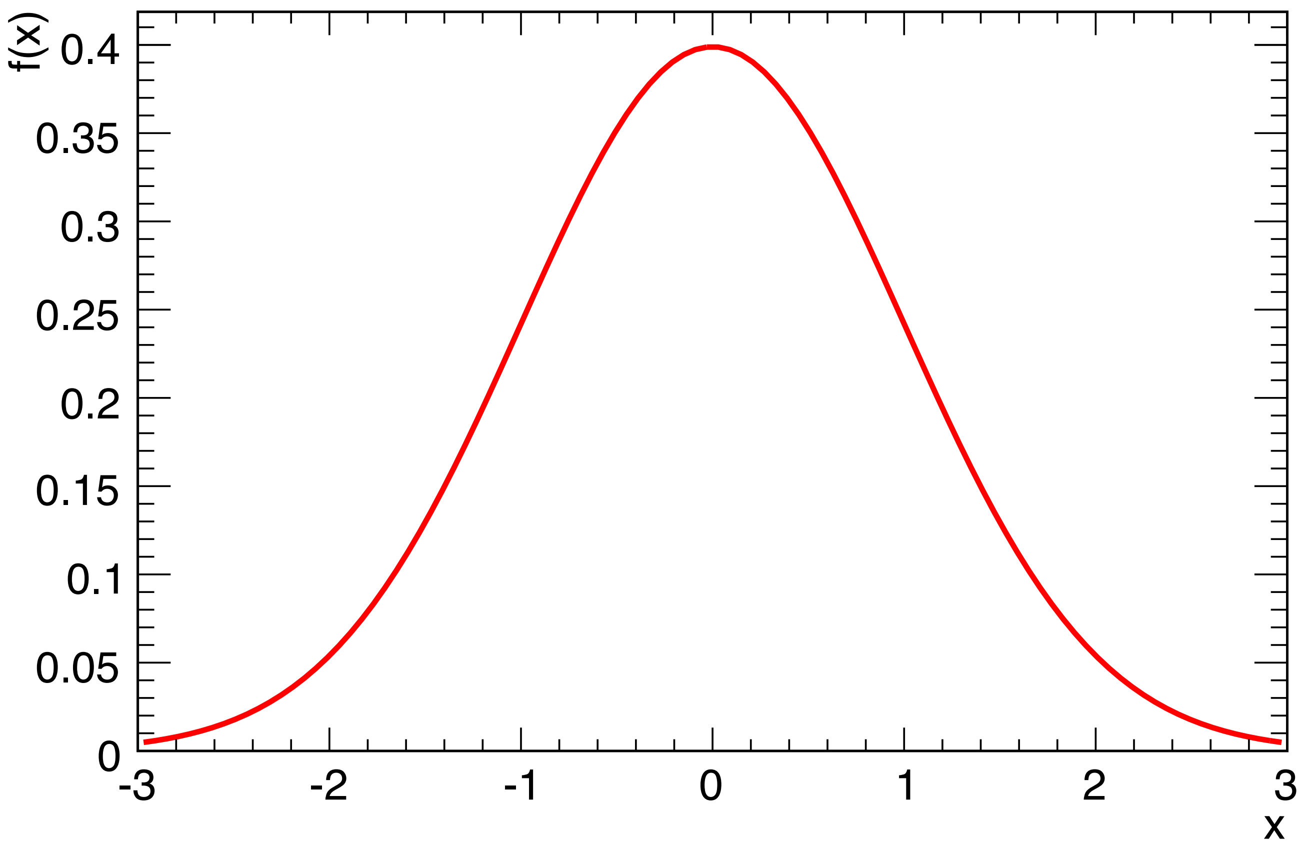
$$f(x) = \frac{\partial F(x)}{\partial x}$$

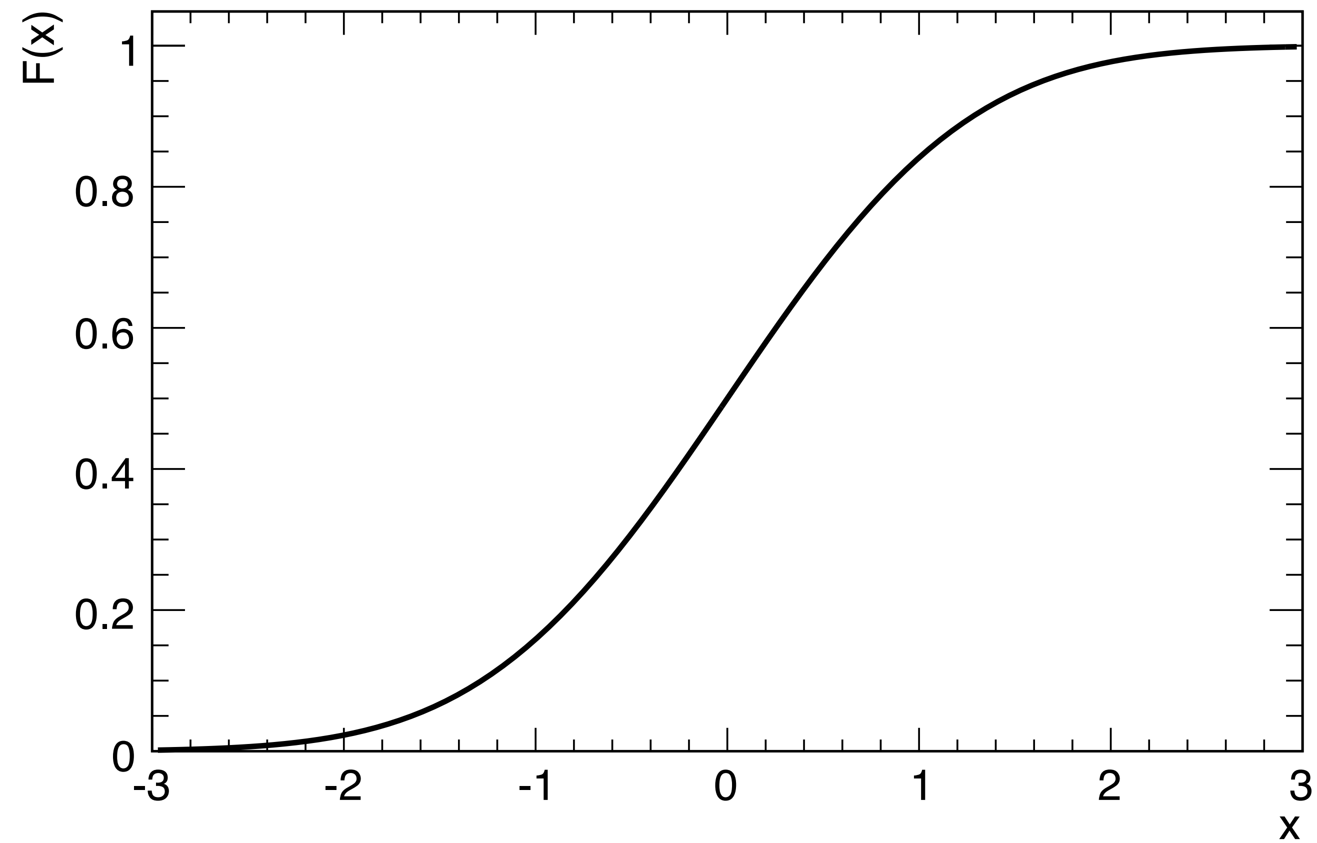
Cumulative Density Functions

Often useful to use a cumulative distribution:

- in 1-dimension:

$$\int_{-\infty}^{x} f(x')dx' = F(x)$$





- alternatively, define density as partial of cumulative:

$$f(x) = \frac{\partial F(x)}{\partial x}$$

- same relationship as total and differential cross section:

$$f(E) = \frac{1}{\sigma}\frac{\partial \sigma}{\partial E}$$

Parametrized Families / Models

Often we are interested in a parametried family of pdfs

- We will write these as: $f(x|\alpha)$ said "f of x given α"
  - where α are the parameters of the "model" (written in greek characters)

A discrete example:

- The Poisson distribution is a probability mass function for n, the number of events one observes, when one expects μ events

$$Pois(n|\mu) = \mu^n \frac{e^{-\mu}}{n!}$$

A continuous example

- The Gaussian distribution is a probability density function for a continuous variable x characterized by a mean μ and standard deviation σ

$$G(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The Likelihood Function

Consider the Poisson distribution describes a discrete event count n for a real-valued mean μ.

$$Pois(n|\mu) = \mu^n \frac{e^{-\mu}}{n!}$$

The **likelihood** of μ given n is the same equation evaluated as a function of μ

- Now it's a continuous function
- But it is not a pdf!

$$L(\mu) = Pois(n|\mu)$$

Common to plot the -ln L (or -2 ln L)

- helps avoid thinking of it as a PDF
- connection to χ^2 distribution

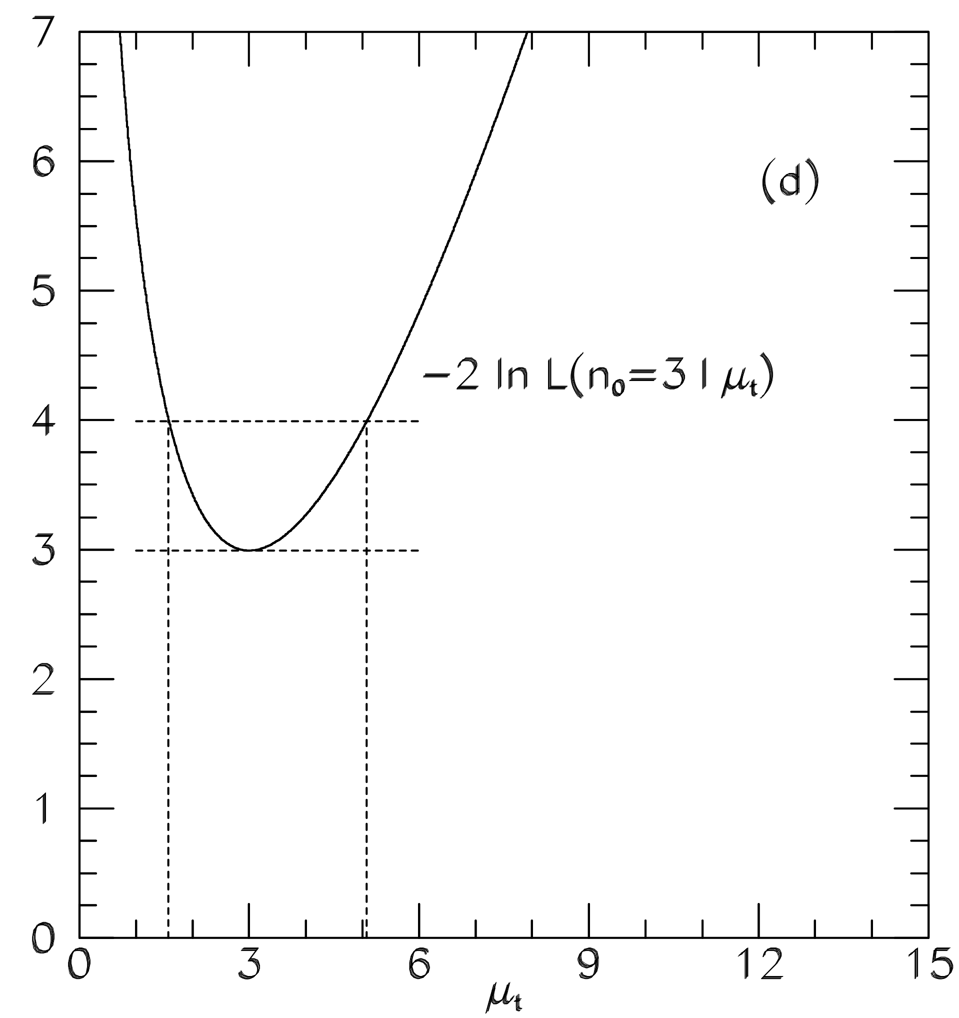


Figure from R. Cousins, Am. J. Phys. 63 398 (1995)

In particle physics we are usually able to perform repeated observations of x that are **independent & identically distributed**

- These repeated observations are written $\{x_i\}$
- and the likelihood in that case is

$$L(\alpha) = \prod_i f(x_i|\alpha)$$

- and the log-likelihood is

$$\log L(\alpha) = \sum_i \log f(x_i|\alpha)$$

HYPOTHESIS TESTS,
LIMITS, & CONFIDENCE INTERVALS

The Neyman-Pearson Lemma

In 1928-1938 Neyman & Pearson developed a theory in which one must consider competing Hypotheses:

- the Null Hypothesis H_0 (background only)
- the Alternate Hypothesis H_1 (signal-plus-background)

Given some probability that we wrongly reject the Null Hypothesis

$$\alpha = P(x \notin W | H_0)$$

(Convention: if data falls in W then we accept H_0)

Find the region W such that we minimize the probability of wrongly accepting the H_0 (when H_1 is true)

$$\beta = P(x \in W | H_1)$$

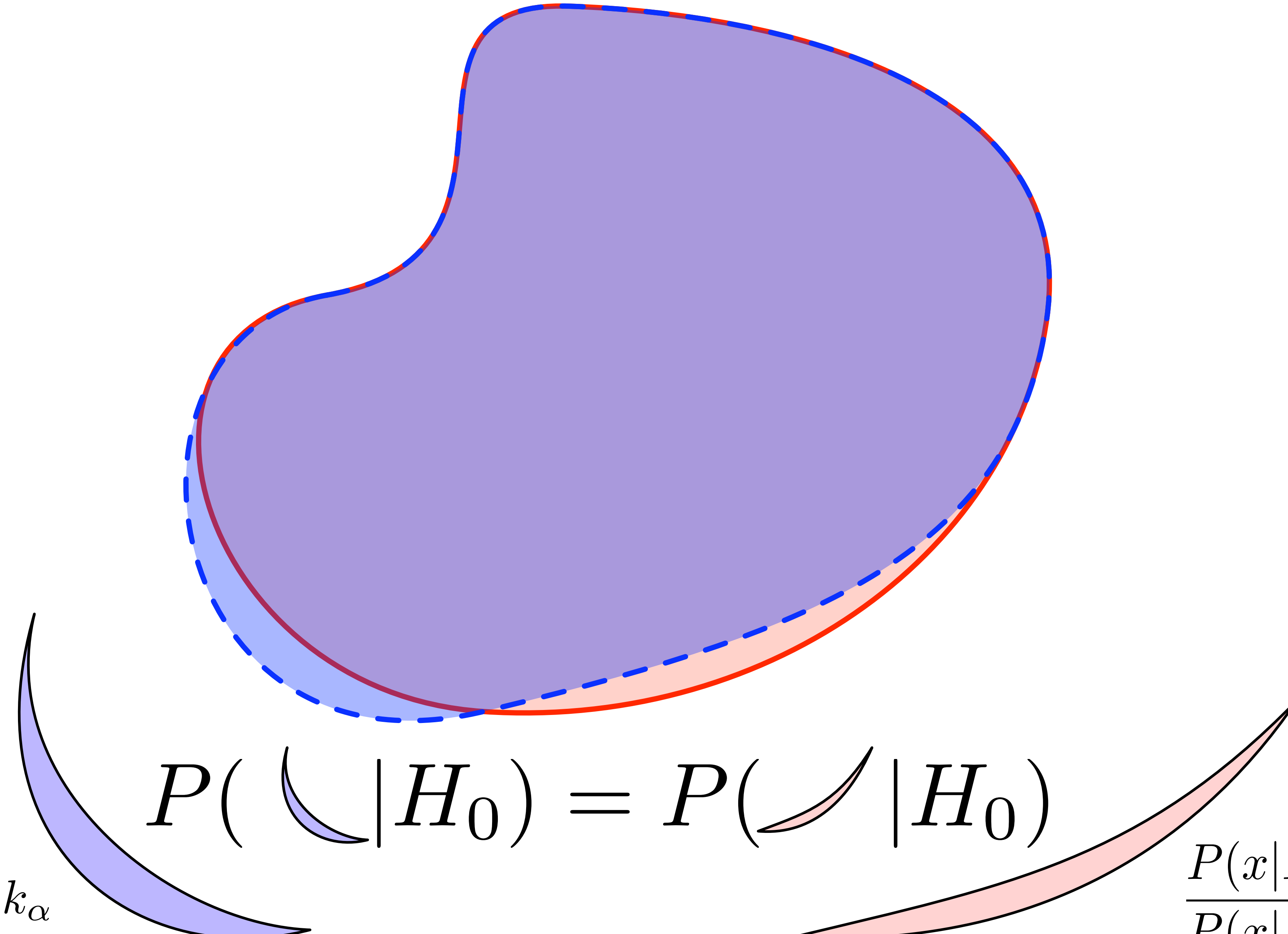
The region W that minimizes the probability of wrongly accepting H_0 is just a contour of the Likelihood Ratio

$$\frac{P(x|H_1)}{P(x|H_0)} > k_\alpha$$

Any other region of the same size will have less power

The likelihood ratio is an example of a Test Statistic, eg. a real-valued function that summarizes the data in a way relevant to the hypotheses that are being tested

A SHORT PROOF OF NEYMAN-PEARSON



$$P(\;|H_0) = P(\;|H_0)$$

$\frac{P(x|H_1)}{P(x|H_0)} < k_\alpha$

$\frac{P(x|H_1)}{P(x|H_0)} > k_\alpha$

$P(\;|H_1) < P(\;|H_0)k_\alpha$

$P(\;|H_1) > P(\;|H_0)k_\alpha$

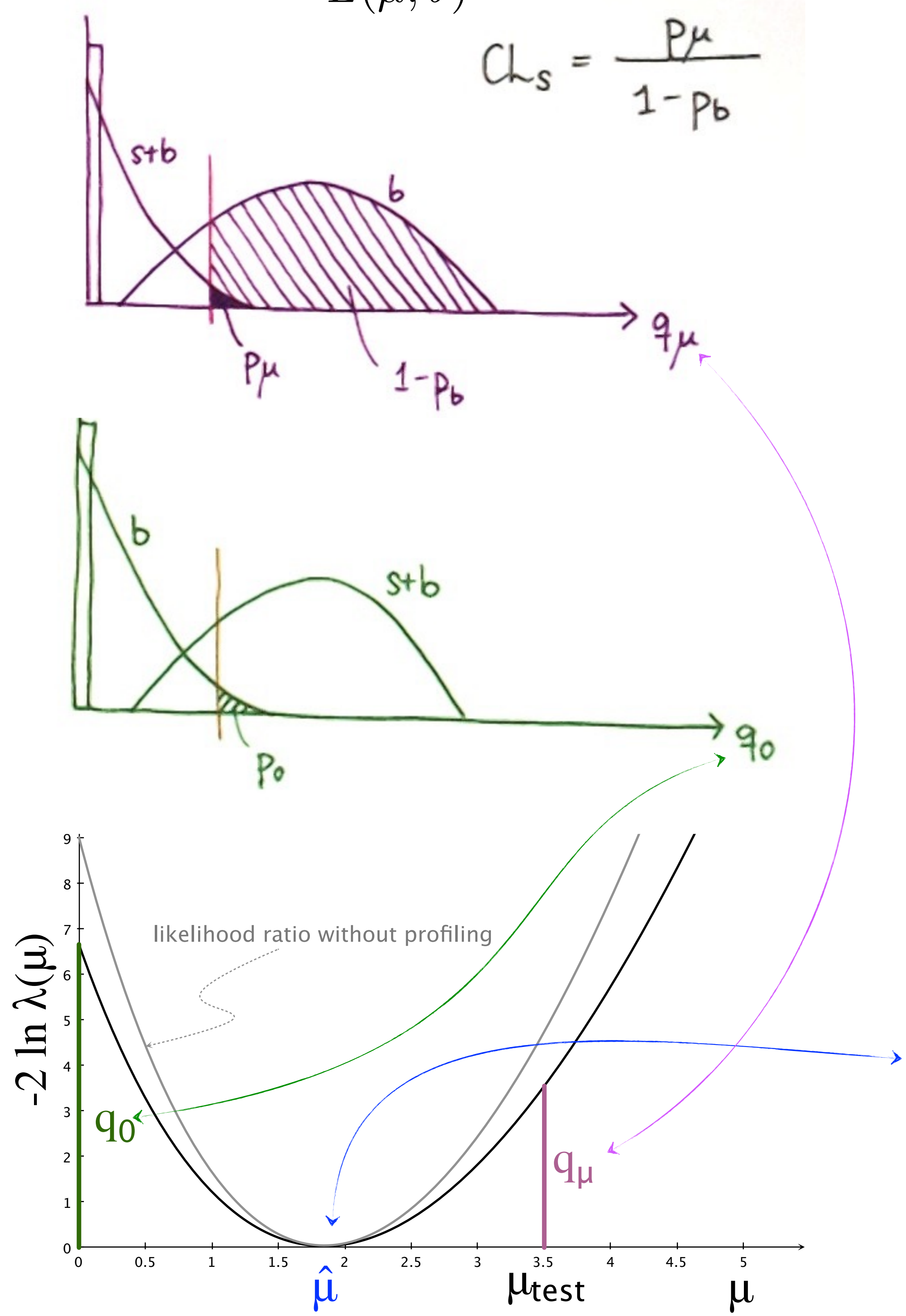
$$\boxed{P(\;|H_1) < P(\;|H_1)}$$

The new region region has less power.

THUMBNAIL OF THE LHC STATISTICAL PROCEDURE

Follow LHC-HCG Combination Procedures

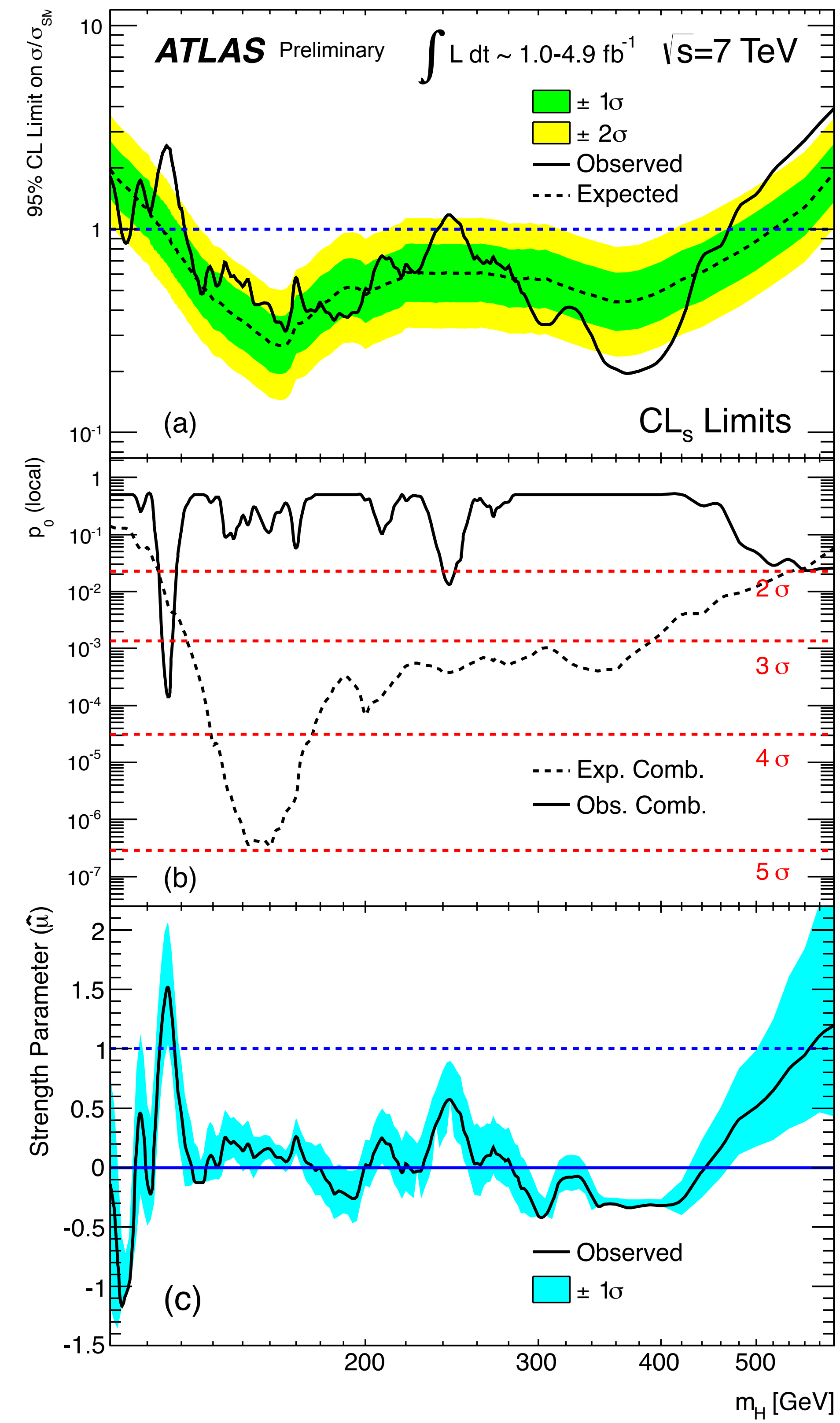
$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(\hat{\mu}, \hat{\theta})}$$



CL_s to test signal hypothesis

p_0 to test background hypothesis

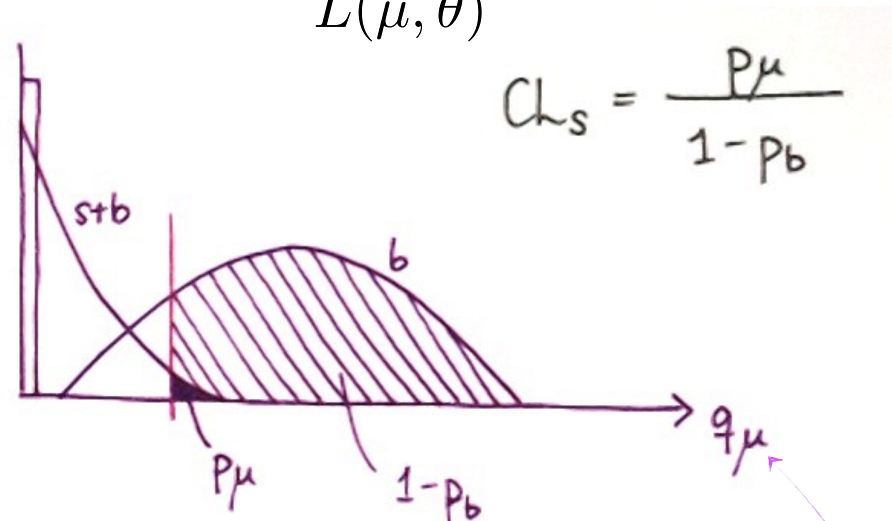
$\hat{\mu}$ to estimate signal strength



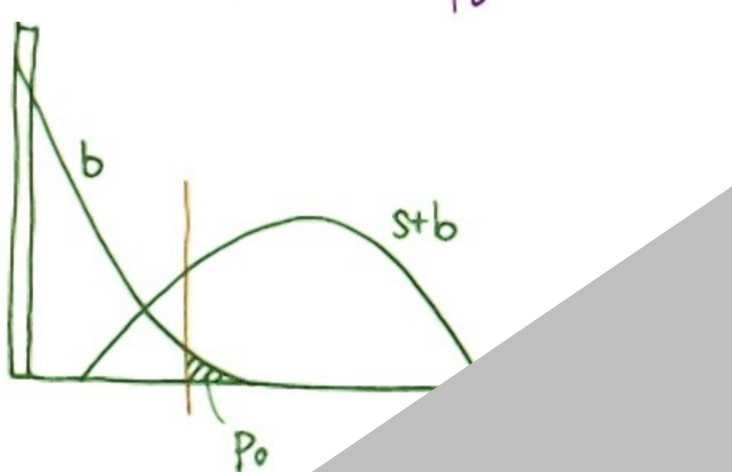
THUMBNAIL OF THE LHC STATISTICAL PROCEDURE

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(\hat{\mu}, \hat{\theta})}$$

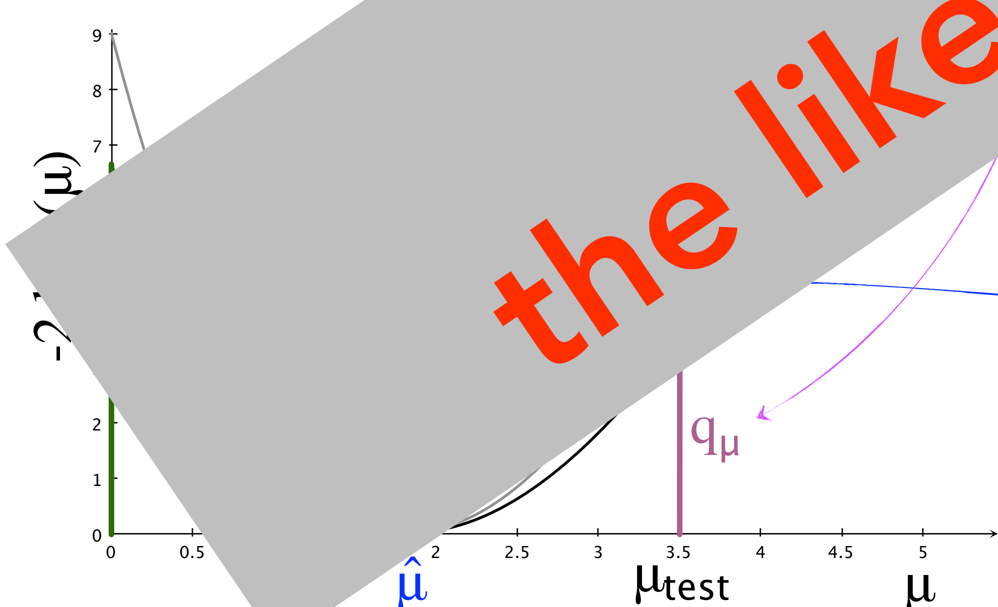
Follow LHC-HCG Combination P



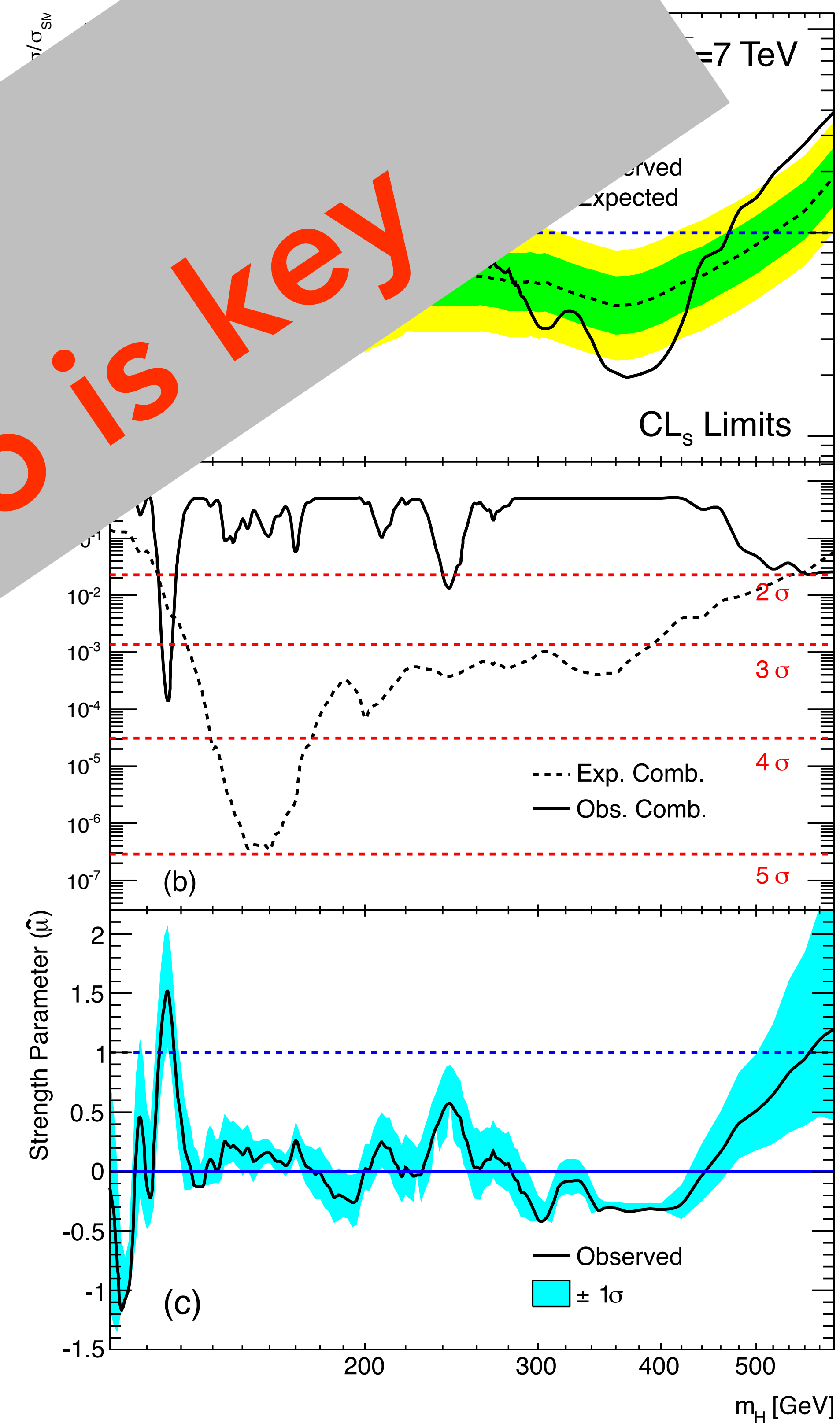
CL_s to test signal hyp



bis



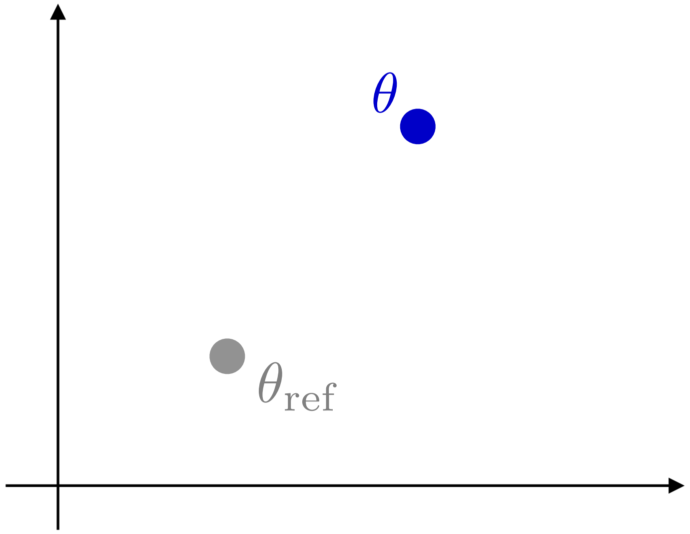
$\hat{\mu}$ to estimate signal strength



Takeaway: the likelihood ratio is key

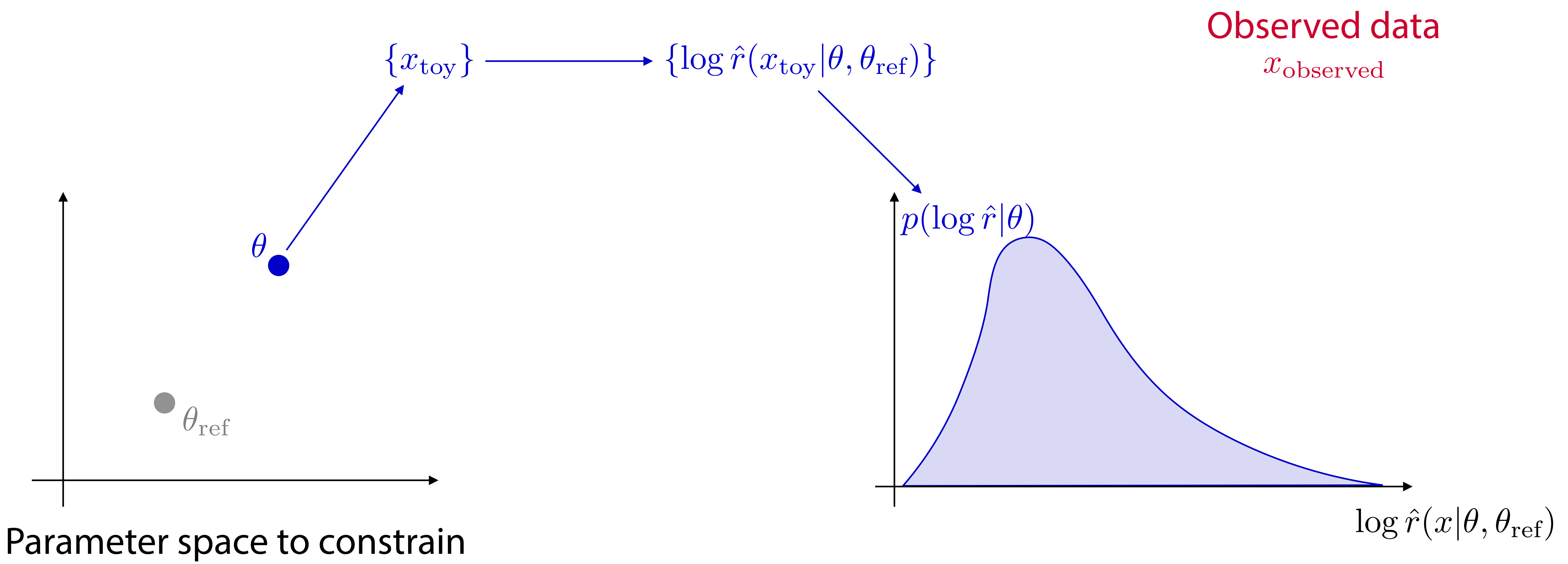
Limit setting (frequentist, standard ATLAS / CMS practice)

Observed data
x_{observed}

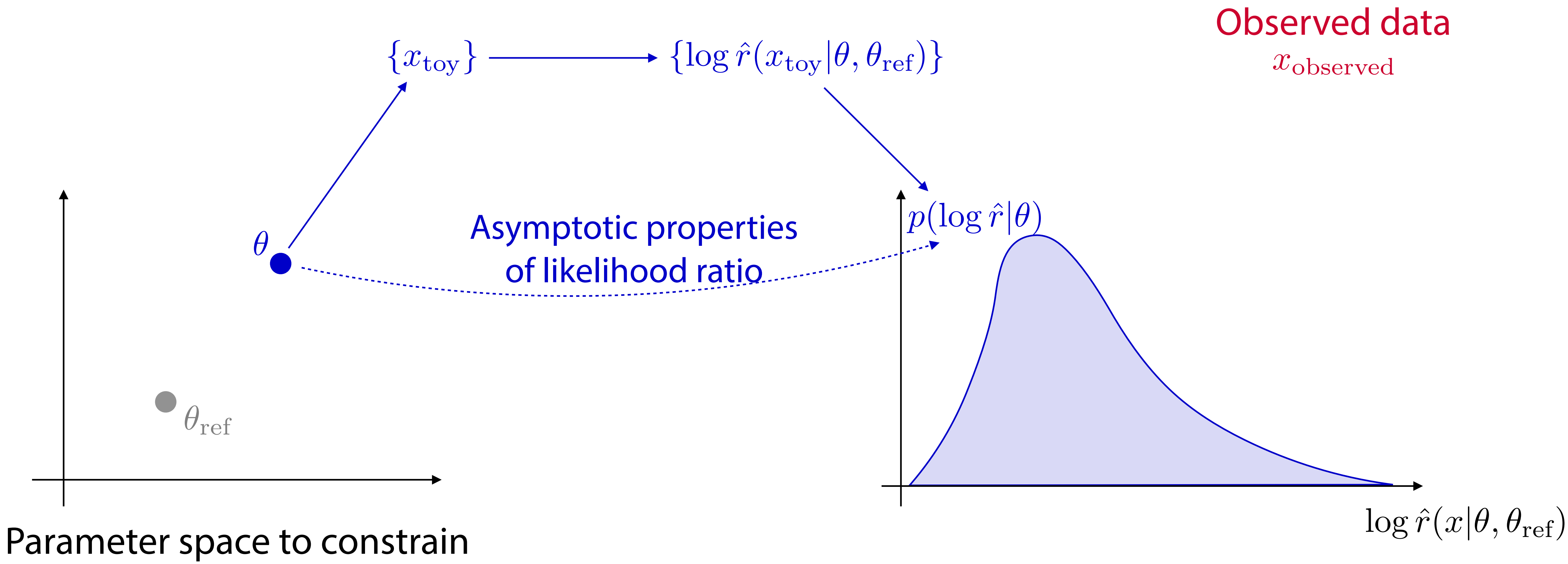


Parameter space to constrain

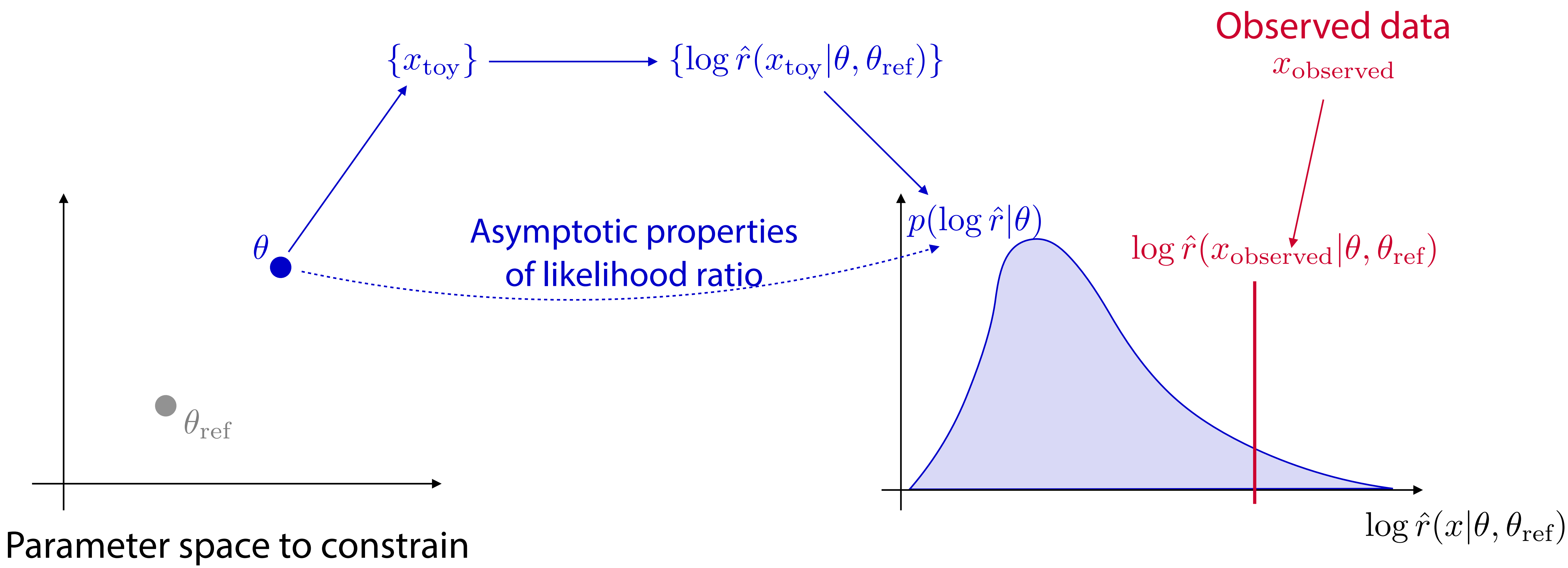
Limit setting (frequentist, standard ATLAS / CMS practice)



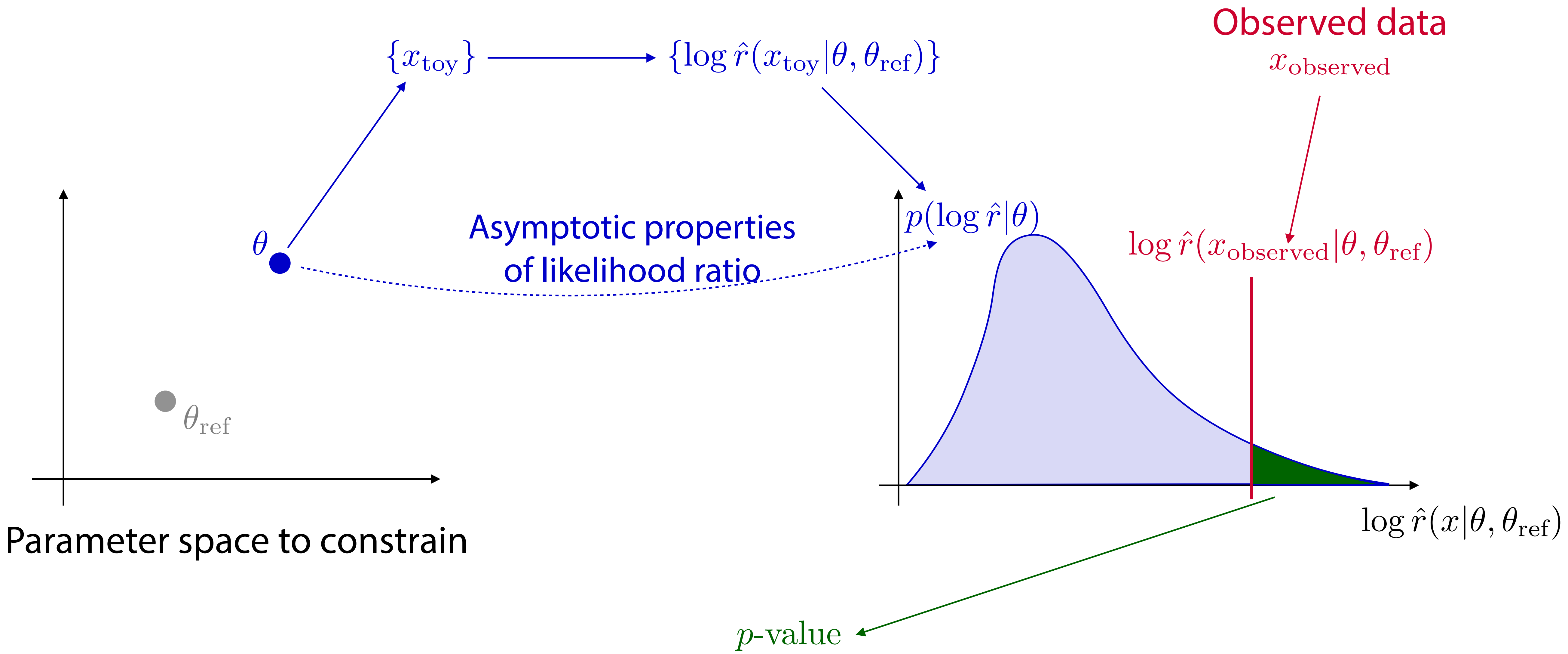
Limit setting (frequentist, standard ATLAS / CMS practice)



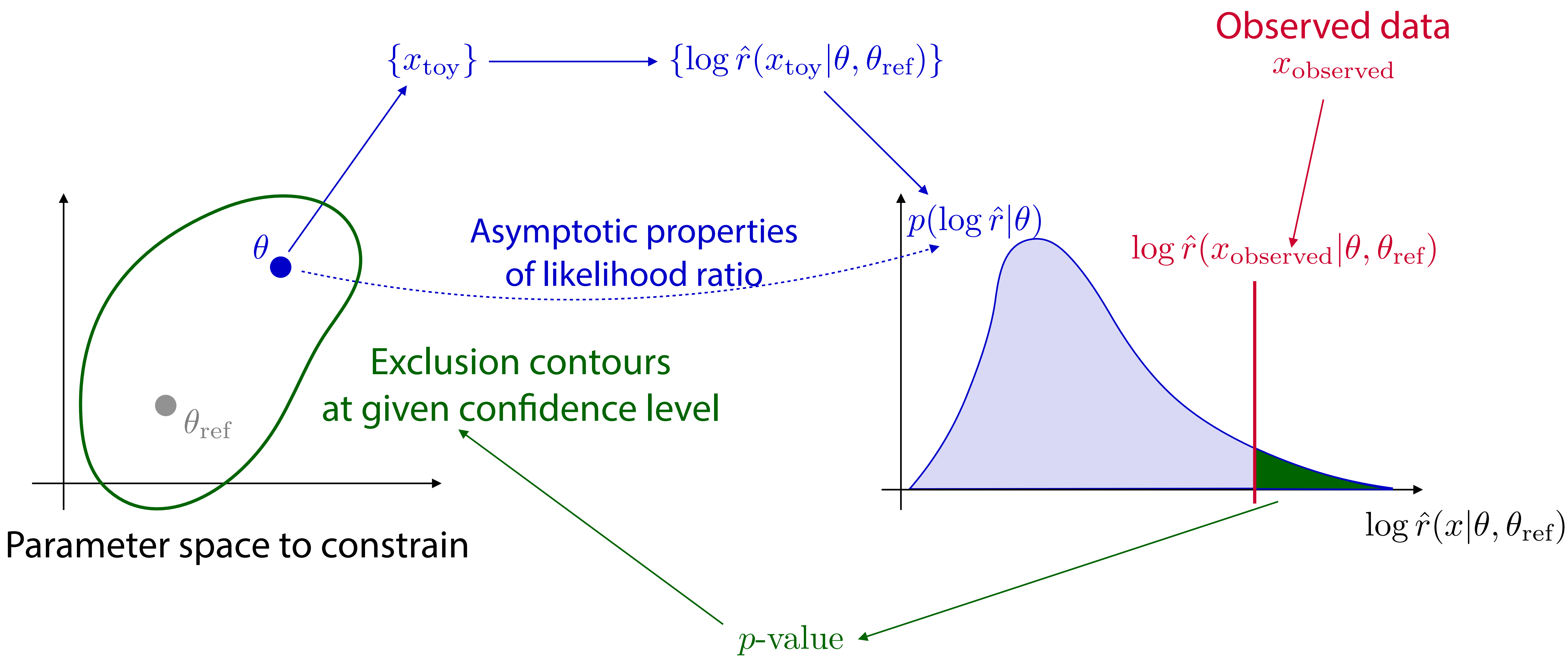
Limit setting (frequentist, standard ATLAS / CMS practice)



Limit setting (frequentist, standard ATLAS / CMS practice)

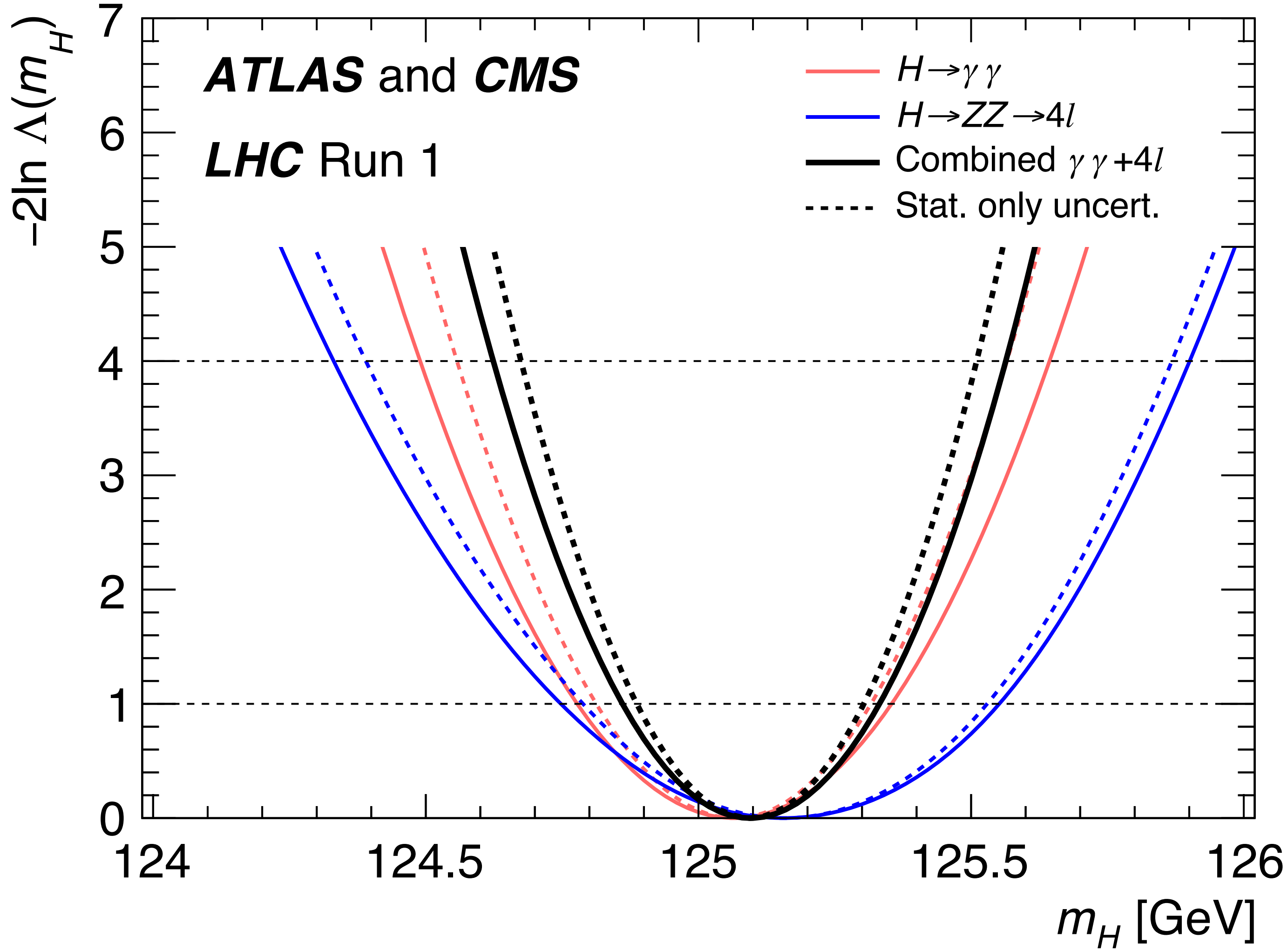


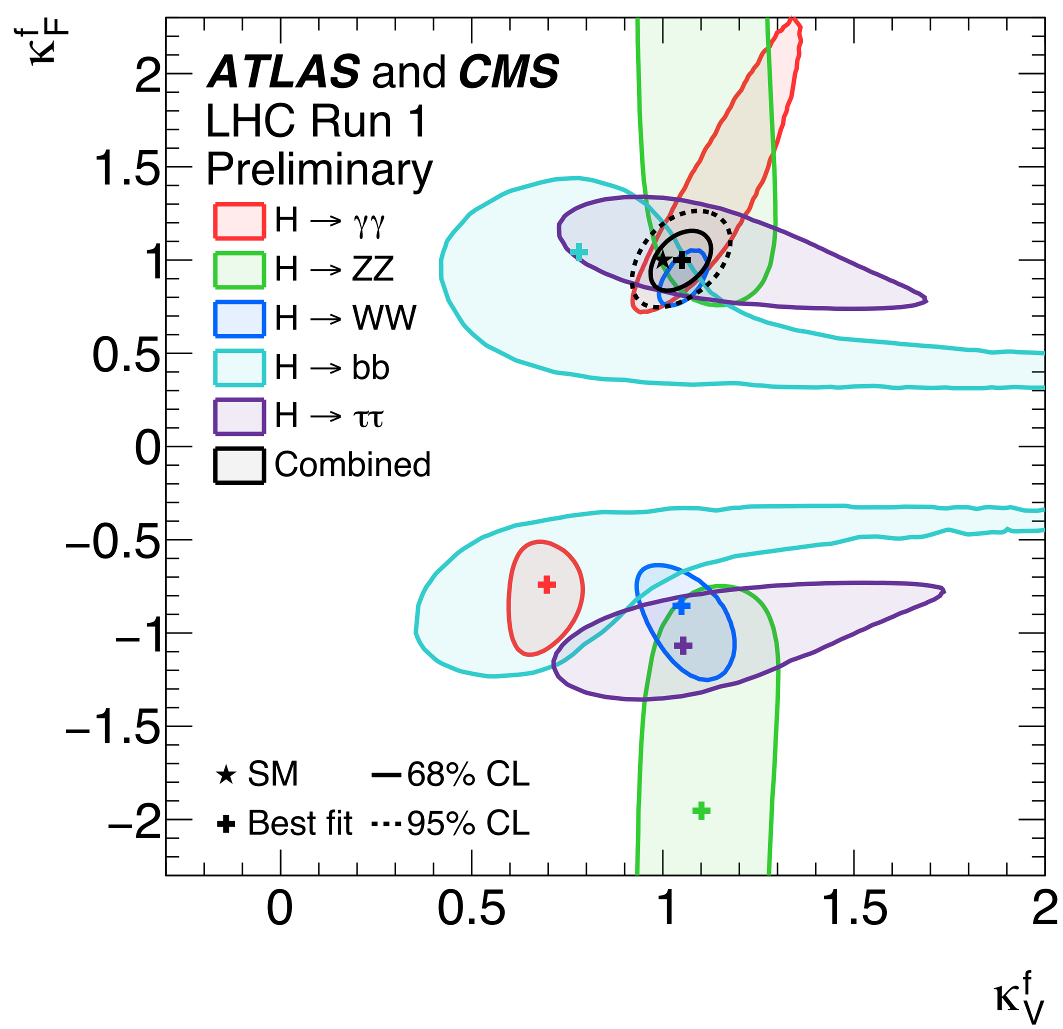
Limit setting (frequentist, standard ATLAS / CMS practice)



My goal is to approximate the likelihood p(x|**θ**)

or a likelihood ratio $r(x|\theta_0, \theta_1) \equiv \dfrac{p(x|\theta_0)}{p(x|\theta_1)}$ for high dimensional feature x





MEASUREMENT / ESTIMATORS

ESTIMATORS

Given some model $f(x|\alpha)$ and a set of observations $\{x_i\}$ often one wants to estimate the true value of α (assuming the model is true).

An **estimator** is function of the data written $\hat{\alpha}(x_1, \ldots x_n)$

- Since the data are random, so is the resulting estimate
- often it is just written $\hat{\alpha}$, where the x-dependence is implicit
- one can compute expectation of the estimator

$$E[\hat{\alpha}(x)|\alpha] = \int \hat{\alpha}(x) f(x|\alpha) dx$$

Properties of estimators:

- **bias** $E[\hat{\alpha}(x)|\alpha] - \alpha$ (unbiased means bias=0)
- **variance** $E[(\hat{\alpha}(x) - \bar{\alpha})^2|\alpha] = \int (\hat{\alpha}(x) - \bar{\alpha})^2 f(x|\alpha) dx$
- **asymptotic bias** limit of bias with infinite observations

Maximum likelihood estimators

There are many different possible estimators, but the most well-known and well-studied is the maximum likelihood estimator (MLE)

$$\hat{\alpha}(x) = \text{argmax}_{\alpha} L(\alpha) = \text{argmax}_{\alpha} f(x|\alpha)$$

This is just the value of α that maximizes the likelihood

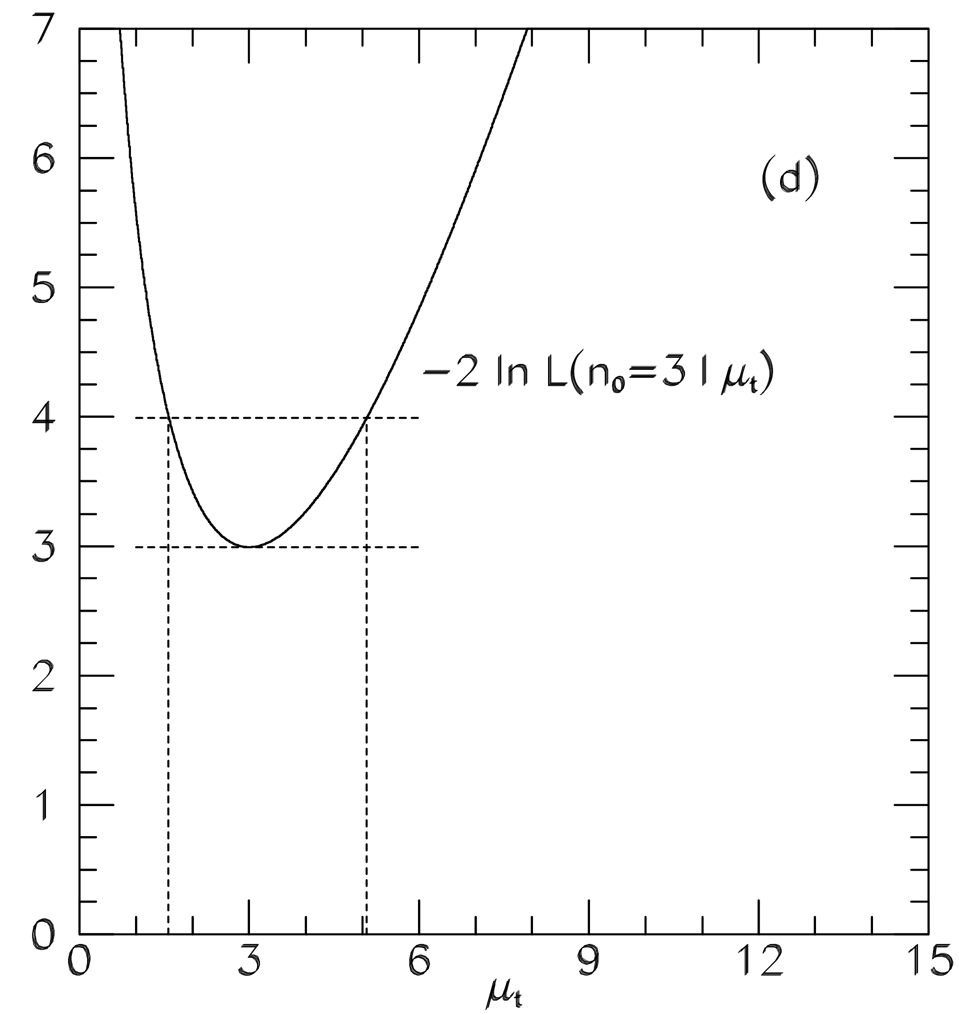


Figure from R. Cousins, Am. J. Phys. 63 398 (1995)

Example: the Poisson distribution

$$Pois(n|\mu) = \mu^n \frac{e^{-\mu}}{n!}$$

Maximizing $L(\mu)$ is the same as minimizing -ln $L(\mu)$

$$-\frac{d}{d\mu} \ln L(\mu)\Big|_{\hat{\mu}} = 0 = \frac{d}{d\mu}\left(\mu - n\ln\mu + \underbrace{\ln n!}_{\text{const}}\right) = 1 - \frac{n}{\mu}$$

$$\Rightarrow \hat{\mu} = n$$

In this case, the MLE is unbiased b/c E[n]=μ

A SECOND EXAMPLE

Consider a set of observations $\{x_i\}$ and we want to estimate the mean of a Gaussian with known σ

which gives

$$G(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$-\frac{d}{d\mu} \ln L(\mu)\Big|_{\hat{\mu}} = 0 = \frac{d}{d\mu}\left(\sum_i \frac{(x_i-\mu)^2}{2\sigma^2} + \underbrace{\ln\sqrt{2\pi}\sigma}_{\text{const}}\right) = \sum_i \frac{(x_i-\mu)}{\sigma^2}$$

$$\Rightarrow \hat{\mu} = \frac{1}{N}\sum_i x_i$$ (an unbiased estimator) .

However, the MLE $\hat{\sigma}^2 = \frac{1}{N}\sum_i (x_i-\mu)^2$ is biased

It can be shown that $\hat{\sigma}^2 = \frac{1}{N-1}\sum_i (x_i-\mu)^2$ is unbiased

Thus, the MLE is **asymptotially unbiased .**

Note: if $\hat{\sigma}^2$ is an unbiased estimate of σ^2, then $\sqrt{\{\hat{\sigma}^2\}}$ is a biased estimate of σ.

The minimum variance bound on an estimator is given by the Cramér-Rao inequality:

$$\mathrm{cov}[\hat{\theta}|\theta_0]_{ij} \geq I^{-1}_{ij}(\theta_0)$$

Expected error of best-fit parameter

Inverse of Fisher information

Fisher information matrix (is also a Riemannian metric!)

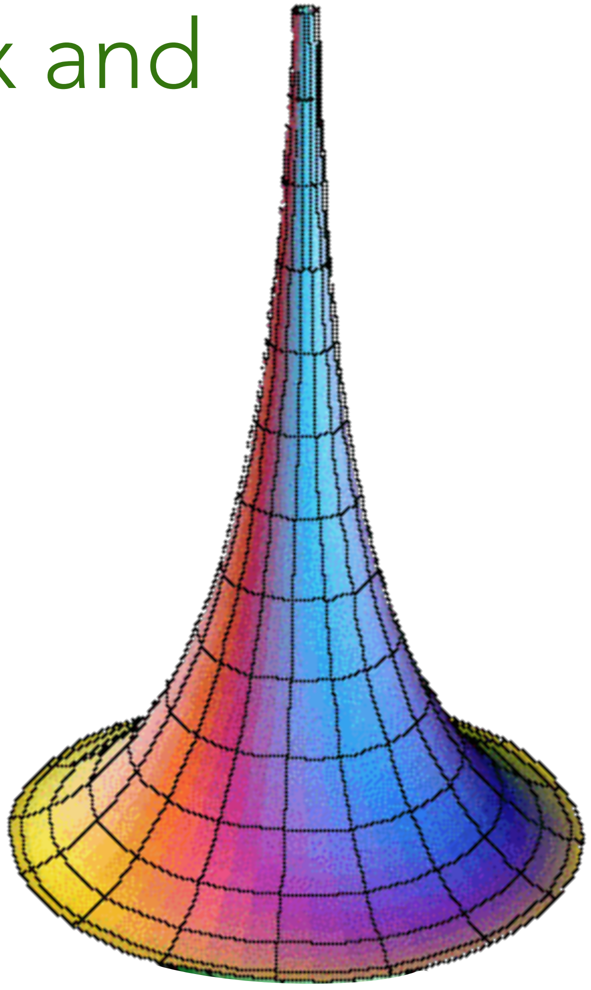
$$I_{ij}[\theta] = -\,\mathbb{E}\left[\frac{\partial^2 \log p(x|\theta)}{\partial\theta_i\,\partial\theta_j}\middle|\theta\right]$$

Maximum Likelihood Estimators *asymptotically* reach this bound

INFORMATION GEOMETRY

AN EXAMPLE

Consider a Gaussian model with observable x and parameters μ,**σ**… eg. G(x| μ,**σ**)



- 2-dim model space
- geometry is surface of constant negative curvature
- result generalized to any distribution in the exponential family

INVARIANCE & COVARIANCE

Physicists like invariance properties.

What if I measure x'=g(x) instead of x?

- Fisher information is **invariant** to reparametrization of observable

What if I wrote down my theory in terms of β instead of α?

- Fisher information is **covariant**, and transforms according to Jacobian of $\alpha(\beta)$

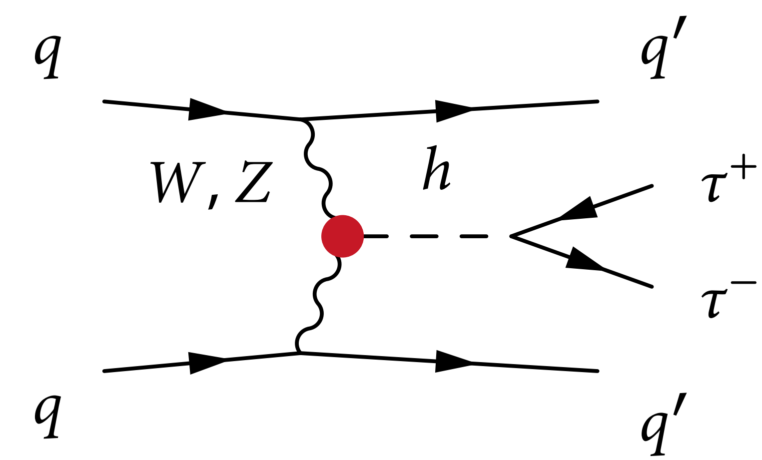
$$g_{\kappa\lambda}(\boldsymbol{\beta}) = \left[\frac{\partial \alpha_\mu}{\partial \beta_\kappa}\right] g_{\mu\nu}(\boldsymbol{\alpha}) \left[\frac{\partial \alpha_\nu}{\partial \beta_\lambda}\right] .$$

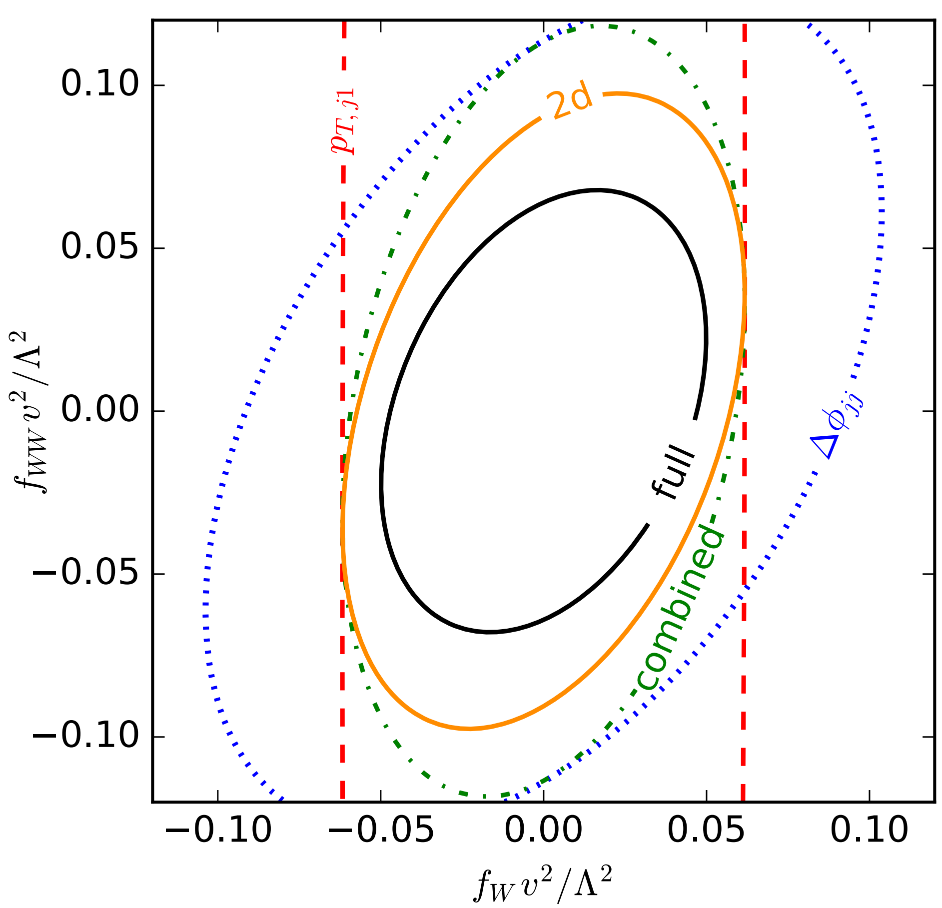
Basically I don't need to worry about these arbitrary choices!

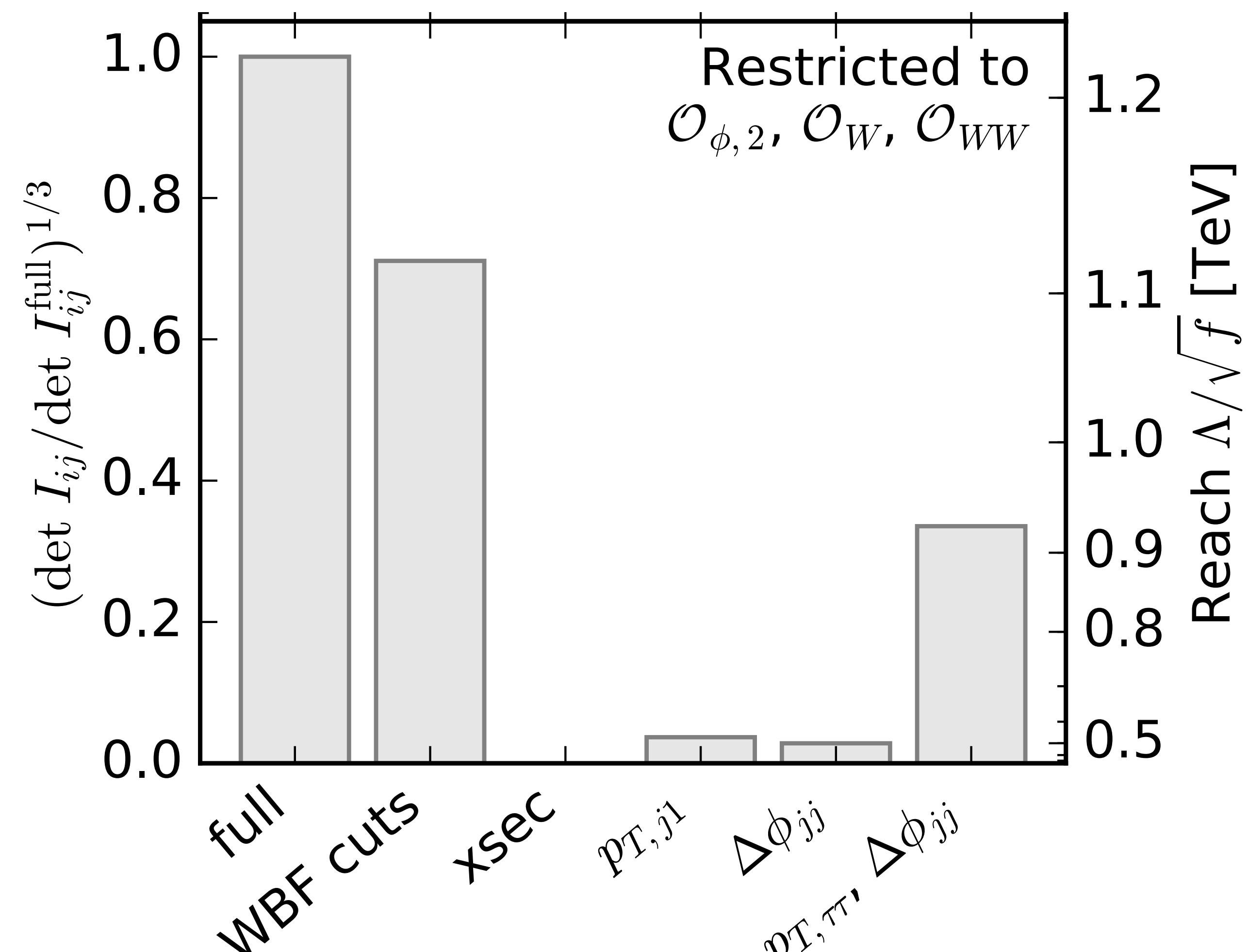
FISHER INFORMATION

[JB, K. Cranmer, F. Kling, T. Plehn 1612.05261;
JB, F. Kling, T. Plehn, T. Tait 1712.02350]

- Theory language: dimension-6 operators of SM EFT, $\mathcal{L} \supset \sum_i \frac{f_i}{\Lambda^2} \mathcal{O}_i$
  [W. Buchmuller, D. Wyler 85; K. Hagiwara, S. Ishihara, S. R. Szalapski, D. Zeppenfeld 93;
  B. Grzadkowski, M. Iskrzynski, M. Misiak, J. Rosiek 1008.4884; …]
  - Total rate: $\mathcal{O}_{\phi,2} = \frac{1}{2}\,\partial^\mu(\phi^\dagger\phi)\,\partial_\mu(\phi^\dagger\phi)$
  - New kinematic structures:
    $$\mathcal{O}_B = \mathrm{i}\frac{g}{2}\,(D^\mu\phi^\dagger)(D^\nu\phi)\,B_{\mu\nu} \qquad \mathcal{O}_W = \mathrm{i}\frac{g}{2}\,(D^\mu\phi)^\dagger\sigma^k(D^\nu\phi)\,W^k_{\mu\nu}$$
    $$\mathcal{O}_{BB} = -\frac{g'^2}{4}\,(\phi^\dagger\phi)\,B_{\mu\nu}\,B^{\mu\nu} \qquad \mathcal{O}_{WW} = -\frac{g^2}{4}\,(\phi^\dagger\phi)\,W^k_{\mu\nu}\,W^{\mu\nu\,k}$$
  - CP violation: $\mathcal{O}_{W\widetilde{W}} = -\frac{g^2}{4}\,(\phi^\dagger\phi)\,W^k_{\mu\nu}\,\widetilde{W}^{\mu\nu\,k}$
  - Others strongly constrained by EWPD or redundant

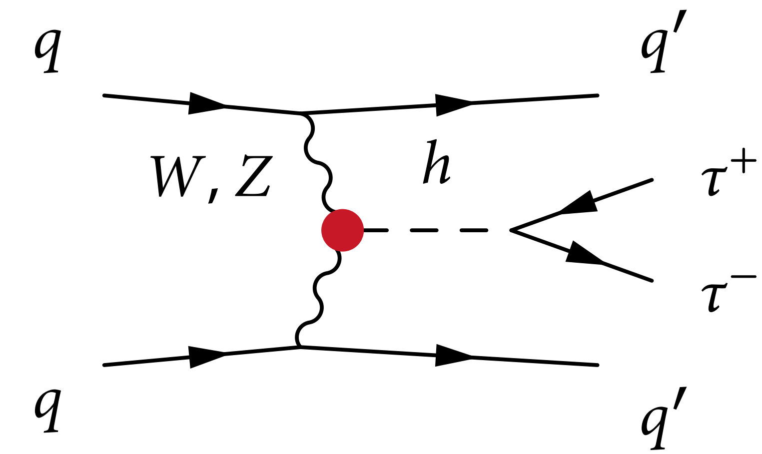


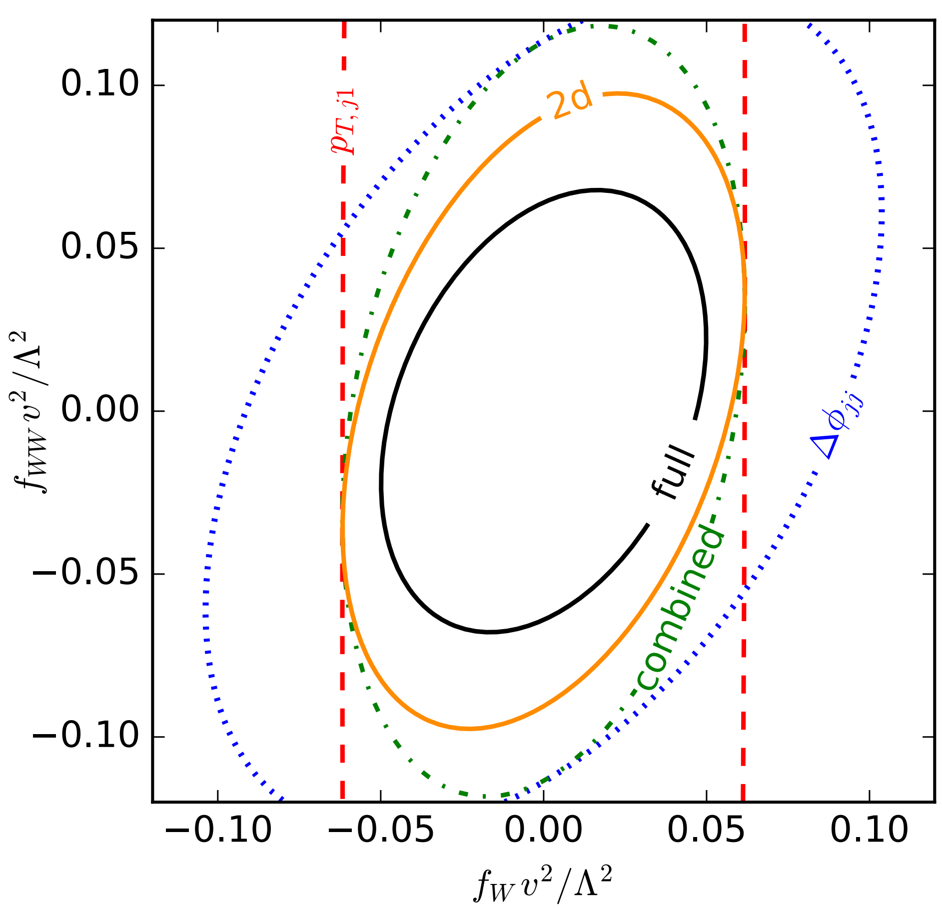


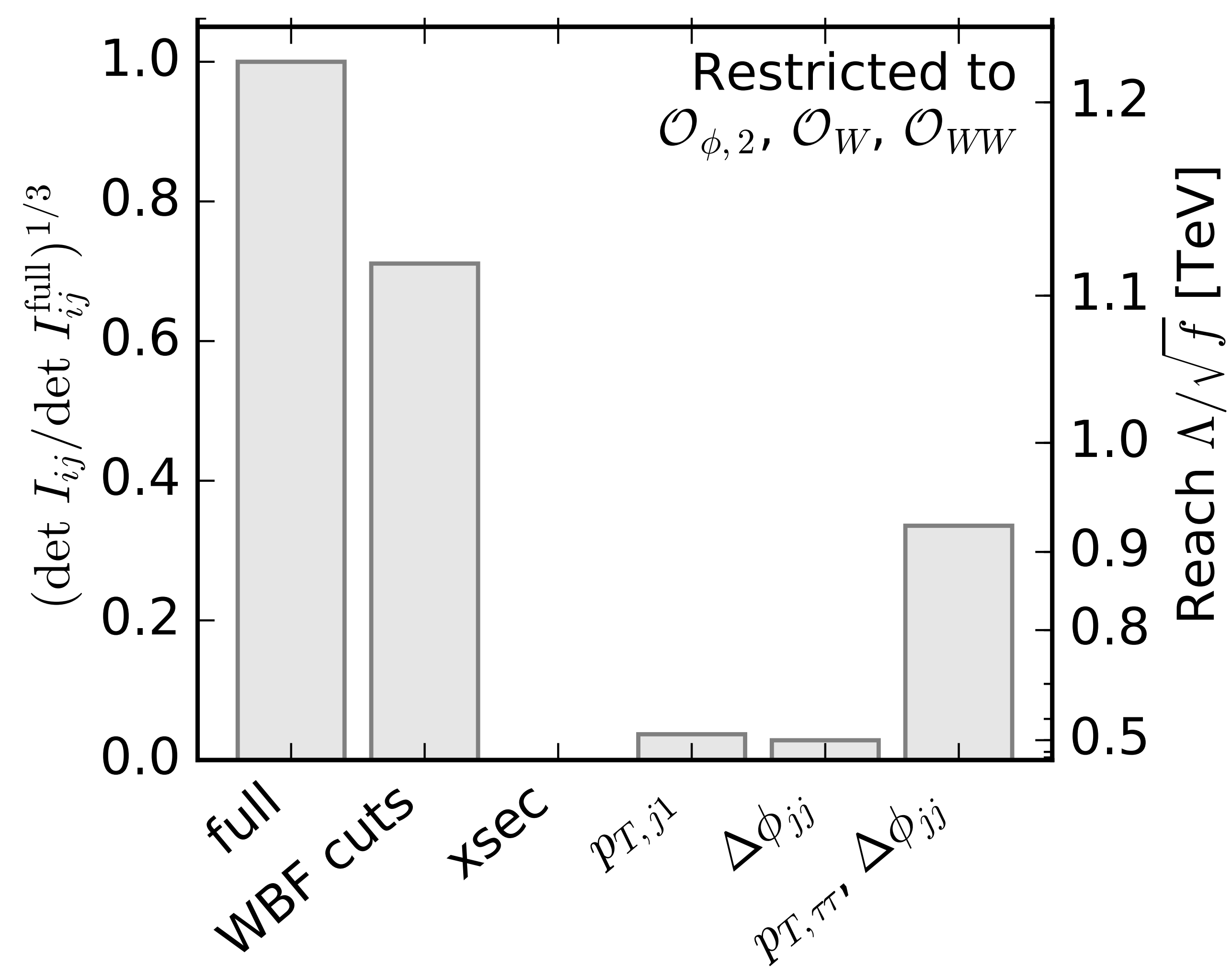


[JB, K. Cranmer, F. Kling, T. Plehn 1612.05261;
JB, F. Kling, T. Plehn, T. Tait 1712.02350]

Compared to just using standard kinematic variables, the fully differential cross-section has the potential to dramatically improve sensitivity

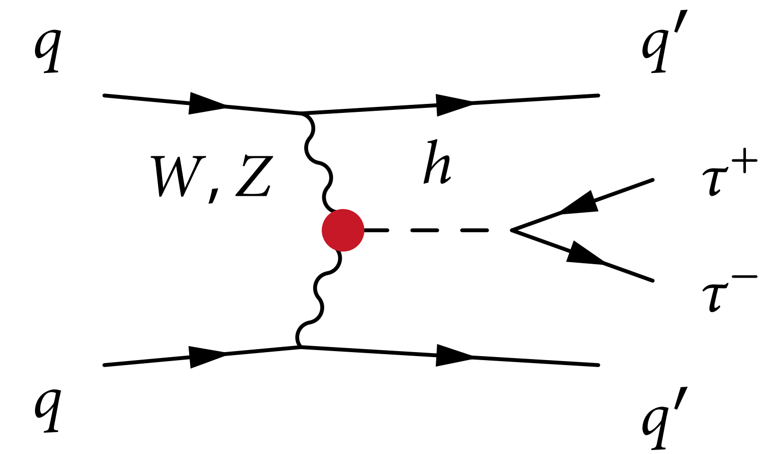


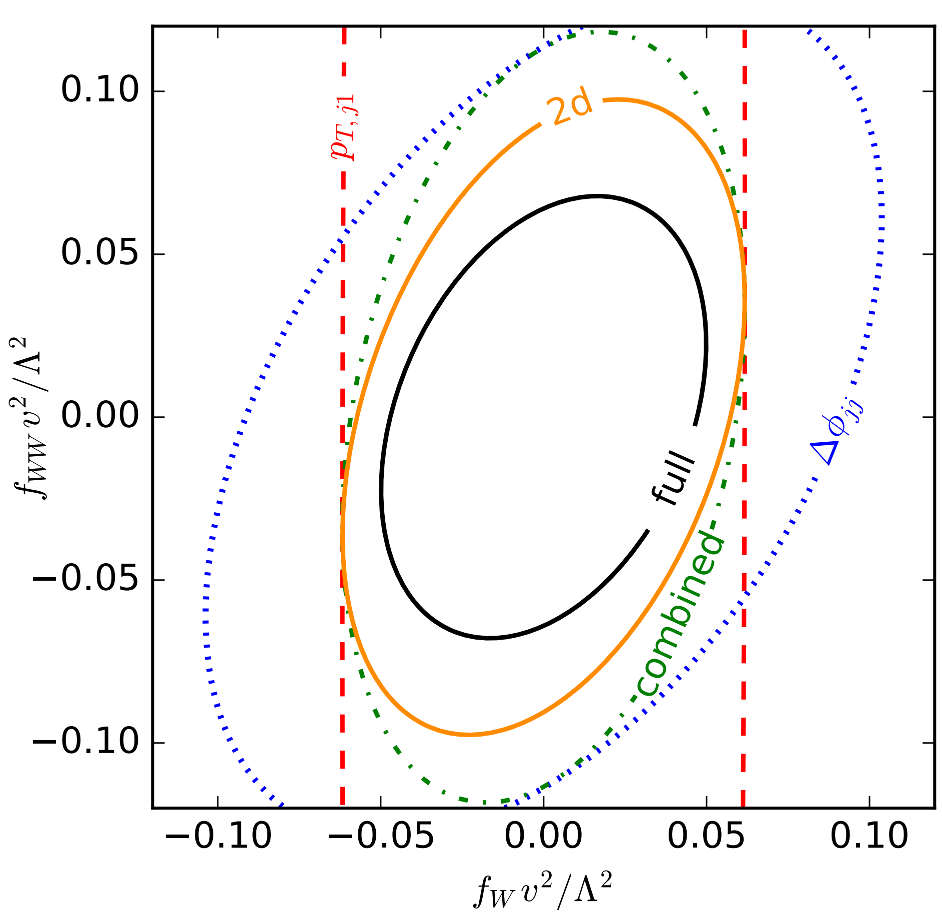


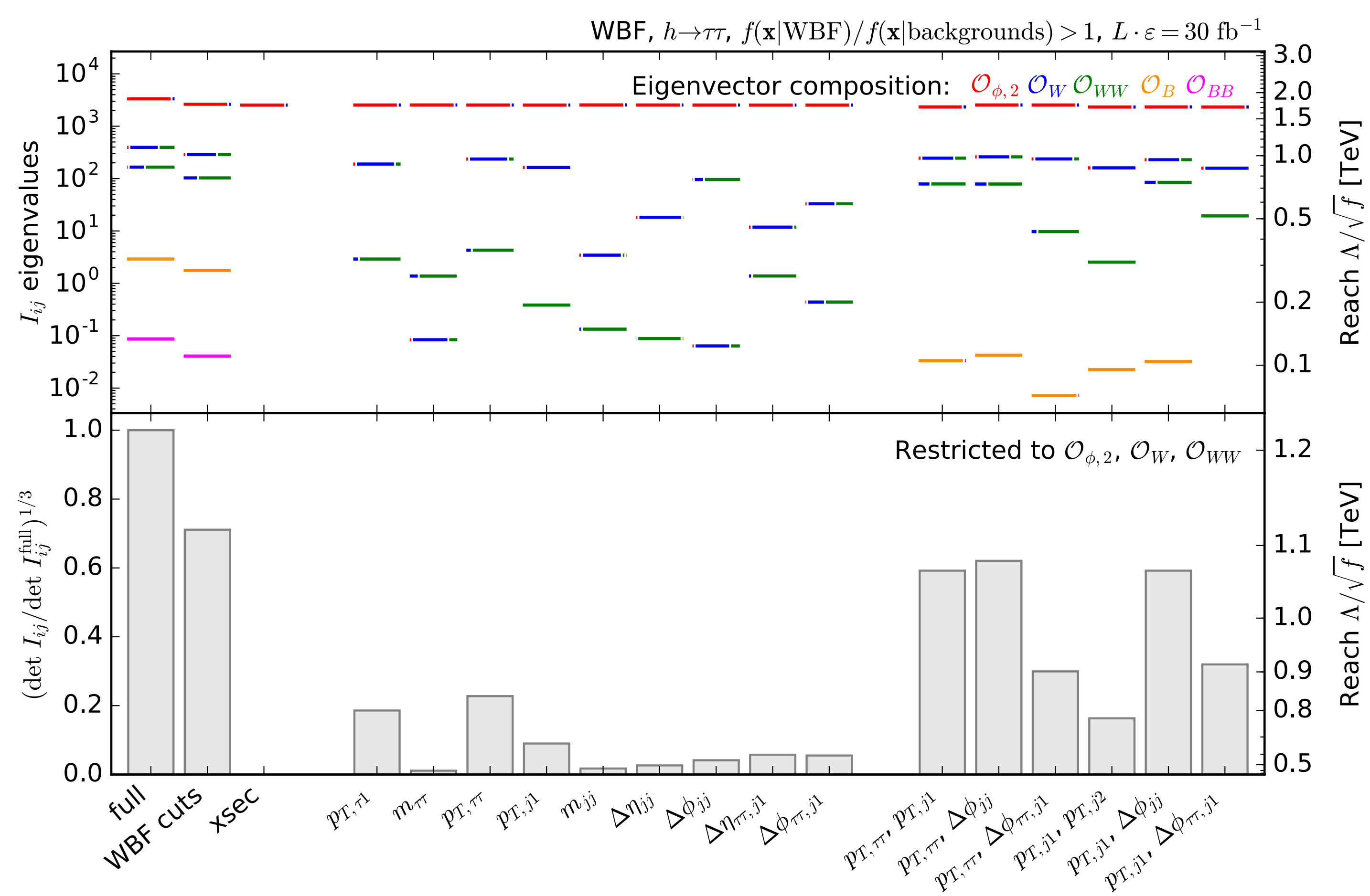


[JB, K. Cranmer, F. Kling, T. Plehn 1612.05261;
JB, F. Kling, T. Plehn, T. Tait 1712.02350]

Compared to just using standard kinematic variables, the fully differential cross-section has the potential to dramatically improve sensitivity







FISHER INFORMATION

Information geometry provides a very powerful tool for phenomenology of EFT

- formal bounds on how well parameters can be measured
- exploit fully differential cross-section
- Global fit (eg. 13 parameters) & can profile/marginalize parameters you aren't interested in (eg. CP violating vs. CP conserving)

For an effective Higgs-gauge Lagrangian truncated at mass dimension six,

$$\mathcal{L} = \mathcal{L}_{\rm SM} + \frac{f_i}{\Lambda^2}\mathcal{O}_i \tag{9}$$

our CP-even reference scenario consists of the renormalizable Standard Model Lagrangian combined with the five CP-even dimension-six operators in the HISZ basis [6, 7, 35],

$$\begin{aligned} \mathcal{O}_B &= i\frac{g}{2}\,(D^\mu\phi^\dagger)(D^\nu\phi)\,B_{\mu\nu} & \mathcal{O}_W &= i\frac{g}{2}\,(D^\mu\phi)^\dagger\sigma^k(D^\nu\phi)\,W^k_{\mu\nu} \\ \mathcal{O}_{BB} &= -\frac{g'^2}{4}\,(\phi^\dagger\phi)\,B_{\mu\nu}\,B^{\mu\nu} & \mathcal{O}_{WW} &= -\frac{g^2}{4}\,(\phi^\dagger\phi)\,W^k_{\mu\nu}\,W^{\mu\nu\,k} \\ \mathcal{O}_{\phi,2} &= \frac{1}{2}\,\partial^\mu(\phi^\dagger\phi)\,\partial_\mu(\phi^\dagger\phi)\,. \end{aligned} \tag{10}$$

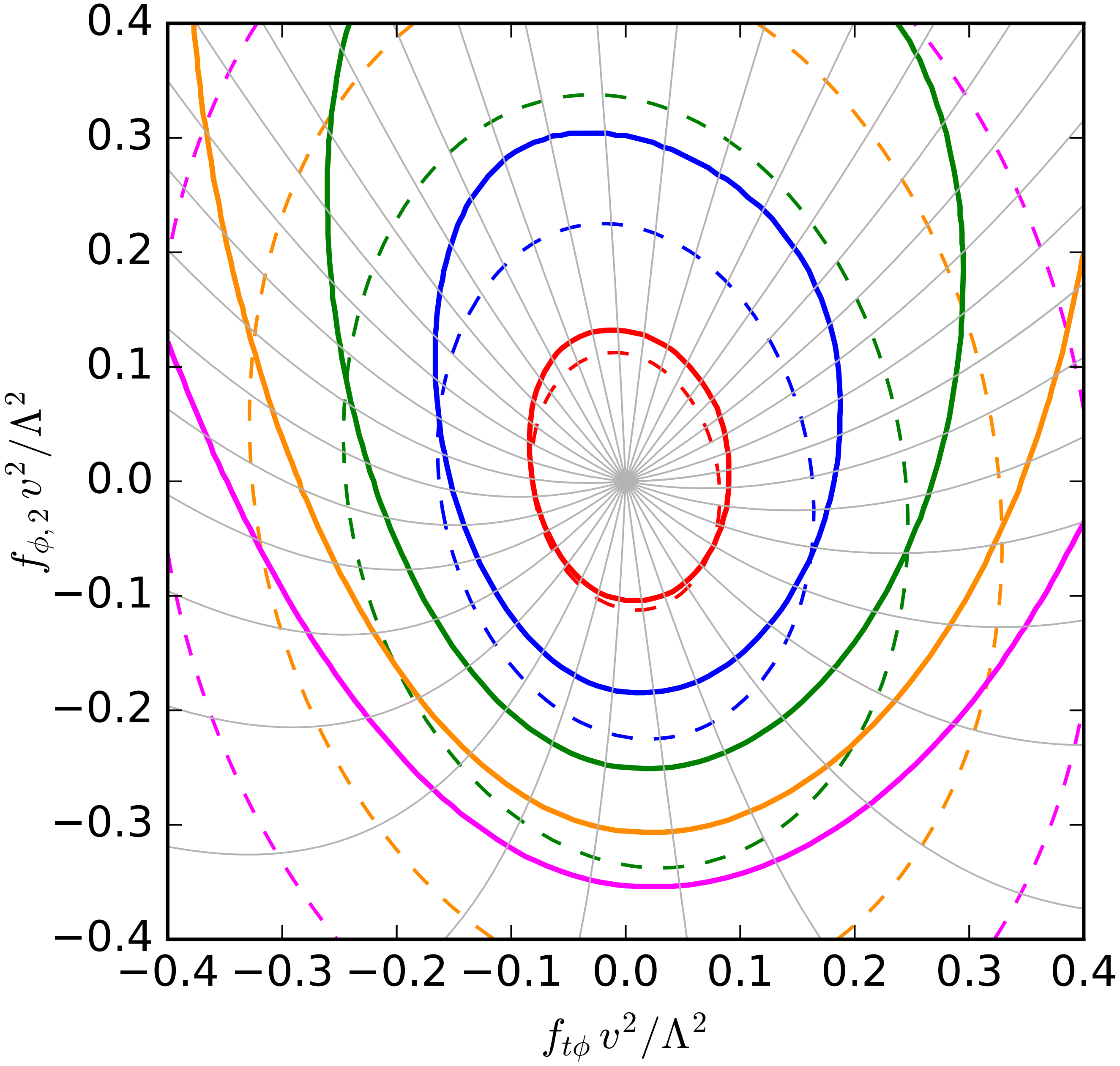
At the same mass dimension, CP-odd couplings are described by operators

$$\begin{aligned} \mathcal{O}_{B\tilde{B}} &= -\frac{g'^2}{4}\,(\phi^\dagger\phi)\,\widetilde{B}_{\mu\nu}\,B^{\mu\nu} \equiv -\frac{g'^2}{4}\,(\phi^\dagger\phi)\,\epsilon_{\mu\nu\rho\sigma}B^{\rho\sigma}\,B^{\mu\nu} \\ \mathcal{O}_{W\widetilde{W}} &= -\frac{g^2}{4}\,(\phi^\dagger\phi)\,\widetilde{W}^k_{\mu\nu}\,W^{\mu\nu\,k} \equiv -\frac{g^2}{4}\,(\phi^\dagger\phi)\,\epsilon_{\mu\nu\rho\sigma}W^{\rho\sigma\,k}\,W^{\mu\nu\,k}\,. \end{aligned} \tag{11}$$

With the Levi-Civita tensor, these operators break down as C-conserving and P-violating.

$$I_{ij} = \begin{array}{c} \begin{array}{ccccccccccccc} f_{\phi,2} & f_W & f_B & f_{WW} & f_{BB} & f_{W\widetilde{W}} & f_{B\tilde{B}} & \mathrm{Im}\,f_W & \mathrm{Im}\,f_B & \mathrm{Im}\,f_{WW} & \mathrm{Im}\,f_{BB} & \mathrm{Im}\,f_{W\widetilde{W}} & \mathrm{Im}\,f_{B\tilde{B}} \end{array} \\ \begin{pmatrix} 4942 & -968 & -50 & 54 & 2 & -7 & 0 & -1 & 0 & 2 & 0 & 36 & 0 \\ -968 & 715 & 35 & -191 & -3 & 1 & 0 & 0 & 0 & 0 & 0 & -55 & -1 \\ -50 & 35 & 6 & -9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 0 \\ 54 & -191 & -9 & 321 & 3 & -1 & 0 & 0 & 0 & 1 & 0 & 72 & 1 \\ 2 & -3 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ -7 & 1 & 0 & -1 & 0 & 359 & 4 & 41 & 1 & -81 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 41 & 0 & 6 & 0 & -12 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 1 & 0 & -81 & -1 & -12 & 0 & 23 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 36 & -55 & -2 & 72 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 21 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \begin{array}{c} \\ f_{\phi,2} \\ f_W \\ f_B \\ f_{WW} \\ f_{BB} \\ f_{W\widetilde{W}} \\ f_{B\tilde{B}} \\ \mathrm{Im}\,f_W \\ \mathrm{Im}\,f_B \\ \mathrm{Im}\,f_{WW} \\ \mathrm{Im}\,f_{BB} \\ \mathrm{Im}\,f_{W\widetilde{W}} \\ \mathrm{Im}\,f_{B\tilde{B}} \end{array}, \tag{30}$$

- With a metric, it now makes sense to talk about how far two points in theory space are from one another. And we can go beyond the ellipse and calculate geodesics!



- Basis-agnostic geometric description of sensitivity based on Fisher information

[JB, K. Cranmer, F. Kling, T. Plehn 1612.05261;
JB, F. Kling, T. Plehn , T. Tait 1712.02350]

RECAP ON LIKELIHOODS

$$\frac{P(x|H_1)}{P(x|H_0)} > k_\alpha$$

For signal vs. background searches:

- **Neyman-Pearson Lemma**: optimal hypothesis test given by **likelihood ratio** (basis of Higgs search)
- Likelihood ratio $\frac{p(x|\theta_0)}{p(x|\theta_1)}$ also used for exclusion contours

For estimates of parameters $\hat{\theta}$

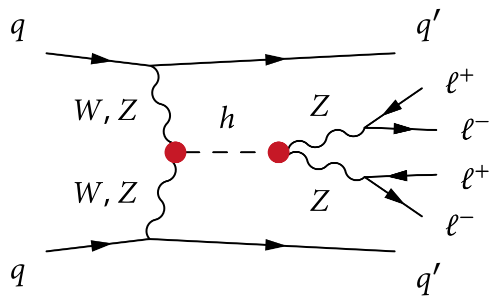
- **Cramér-Rao bound** states $\mathrm{cov}[\hat{\theta}|\theta_0]_{ij} \geq I^{-1}_{ij}(\theta_0)$ where I_{ij} is the **Fisher-information matrix** (Hessian of log-likelihood)
- Motivates **Information Geometry** as a phenomenological tool
- Maximum-likelihood (asymptotically) saturates the bound

Note: ∇_θ log p(x|θ) acts like a likelihood ratio locally

CHALLENGE FOR EFT

Let θ denote the coefficients of higher dimensional operators in the Lagrangian and **x** be high-dimensional data associated to an event

- we want to compare any two points in EFT parameter space
- evaluate the likelihood ratio $r(x|\theta_0, \theta_1) \equiv \dfrac{p(x|\theta_0)}{p(x|\theta_1)}$
- or score $t(x|\theta_0) \equiv \nabla_\theta \log p(x|\theta) \Big|_{\theta_0}$



Difficulty is that one changes the parameters of the EFT, the distributions p(x|θ) change due to interference.

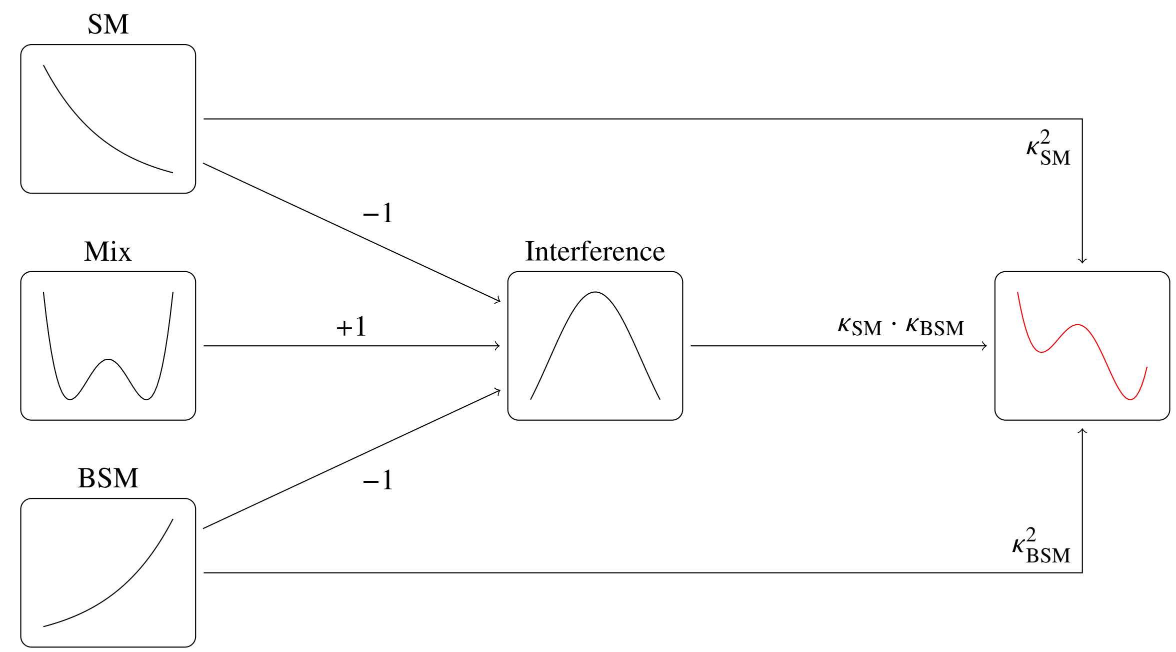
- It would be very computationally expensive to generate samples for every value of θ and estimate p(x|θ) with histograms

MORPHING

EFT EMBEDDED IN A VECTOR SPACE

ATL-PHYS-PUB-2015-047

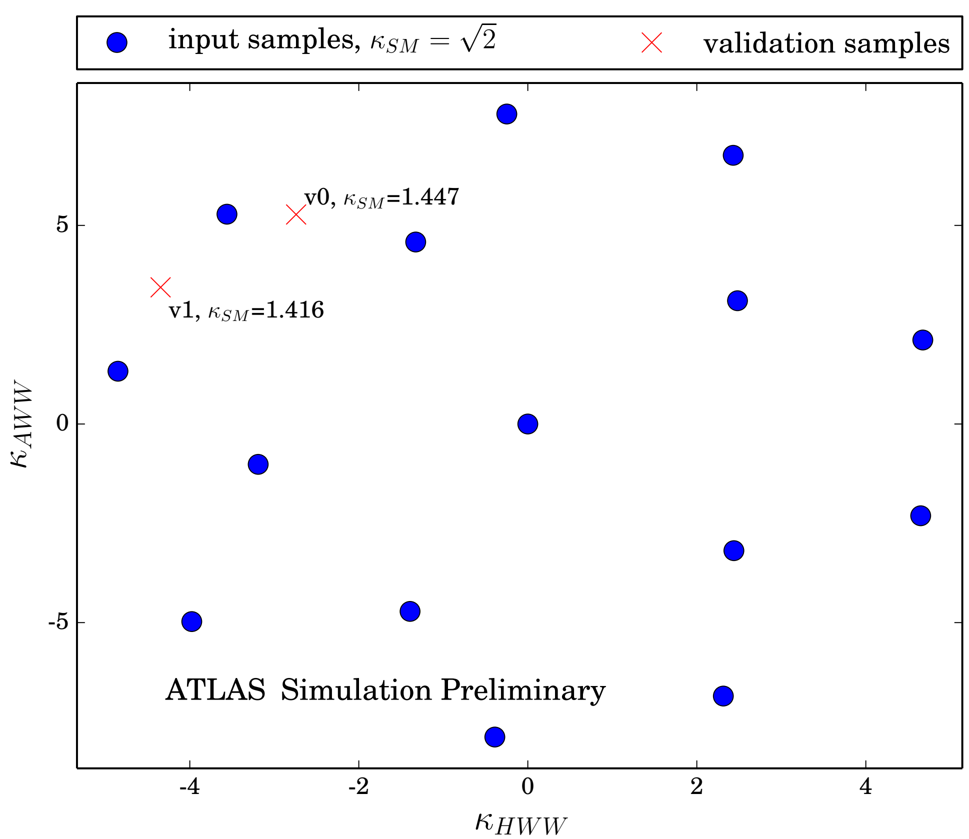
Difficulty is that one changes the parameters of the EFT, the distributions p(x|θ) change due to interference. But there is a trick:

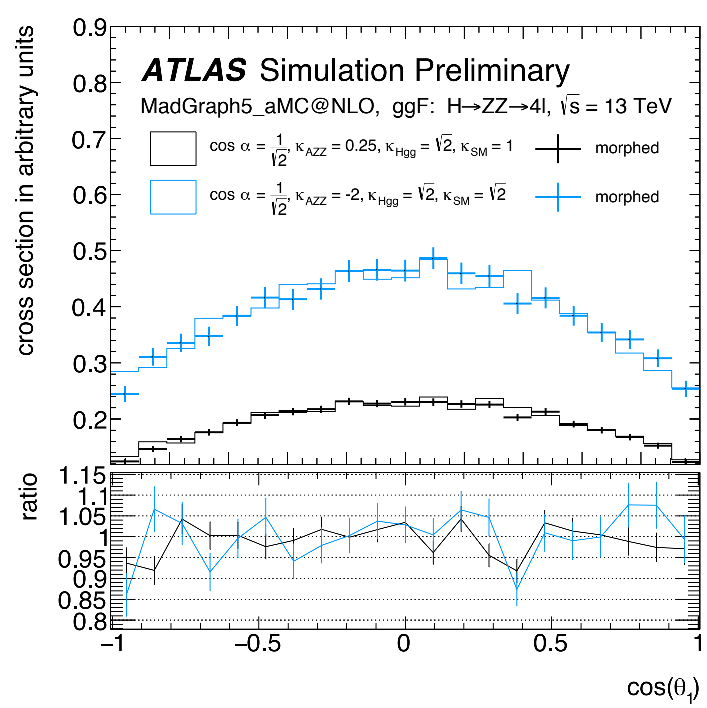


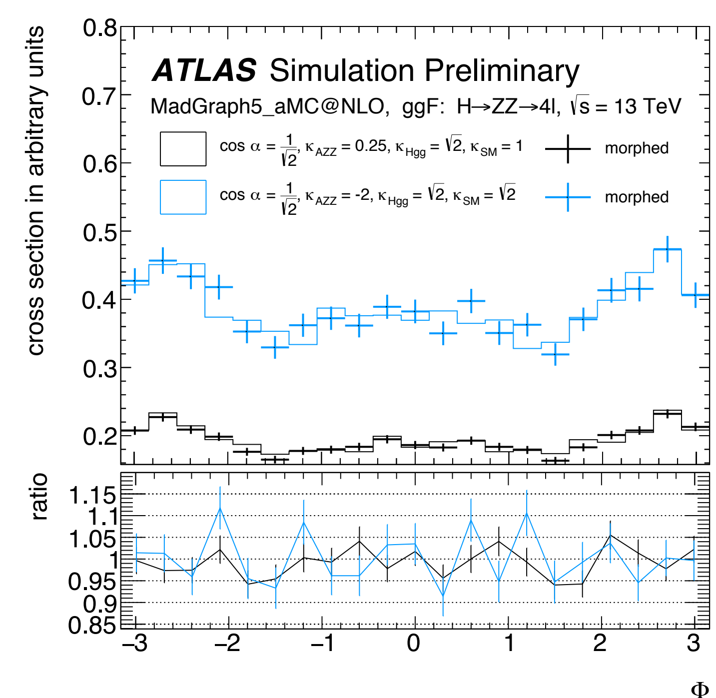
Simple example: $|g_1 M_{SM} + g_2 M_{BSM}|^2 = g_1^2 |M_{SM}|^2 + 2 g_1 g_2 Re\left[M^*_{SM} M_{BSM}\right] + g_2^2 |M_{BSM}|^2$

3-d vector space, any point in this space is linear mixture of 3 basis samples!

(real examples need more basis samples)







EFT DECOMPOSITION

$$d\sigma \propto \left| \overbrace{\left(\mathcal{M}^{p}_{\mathrm{SM}} + \sum_i \frac{f_i}{\Lambda^2} \mathcal{M}^{p}_{i}\right)}^{\text{production}} \overbrace{\left(\mathcal{M}^{d}_{\mathrm{SM}} + \sum_j \frac{f_j}{\Lambda^2} \mathcal{M}^{d}_{j}\right)}^{\text{decay}} \right|^2$$

Express EFT as a mixture:

$$p(x|\theta) = \sum_c w_c(\theta) p_c(x)$$

$w_c(\theta)$ are polynomials

$\nabla_\theta p(x|\theta)$ is now trivial!

Process	Number of components for n operators					
	$\mathcal{O}(\Lambda^0)$	$\mathcal{O}(\Lambda^{-2})$	$\mathcal{O}(\Lambda^{-4})$	$\mathcal{O}(\Lambda^{-6})$	$\mathcal{O}(\Lambda^{-8})$	$\sum$
hV / WBF production	1	n	$\frac{n(n+1)}{2}$			$\frac{(n+1)(n+2)}{2}$
$h \to VV$ decay	1	n	$\frac{n(n+1)}{2}$			$\frac{(n+1)(n+2)}{2}$
Production + decay	1	n	$\frac{n(n+1)}{2}$	$\binom{n+2}{3}$	$\binom{n+3}{4}$	$\binom{n+4}{4}$

Table 1: Number of components c as given in Eq. (6) for different processes, sorted by their suppression by the EFT cutoff scale Λ.

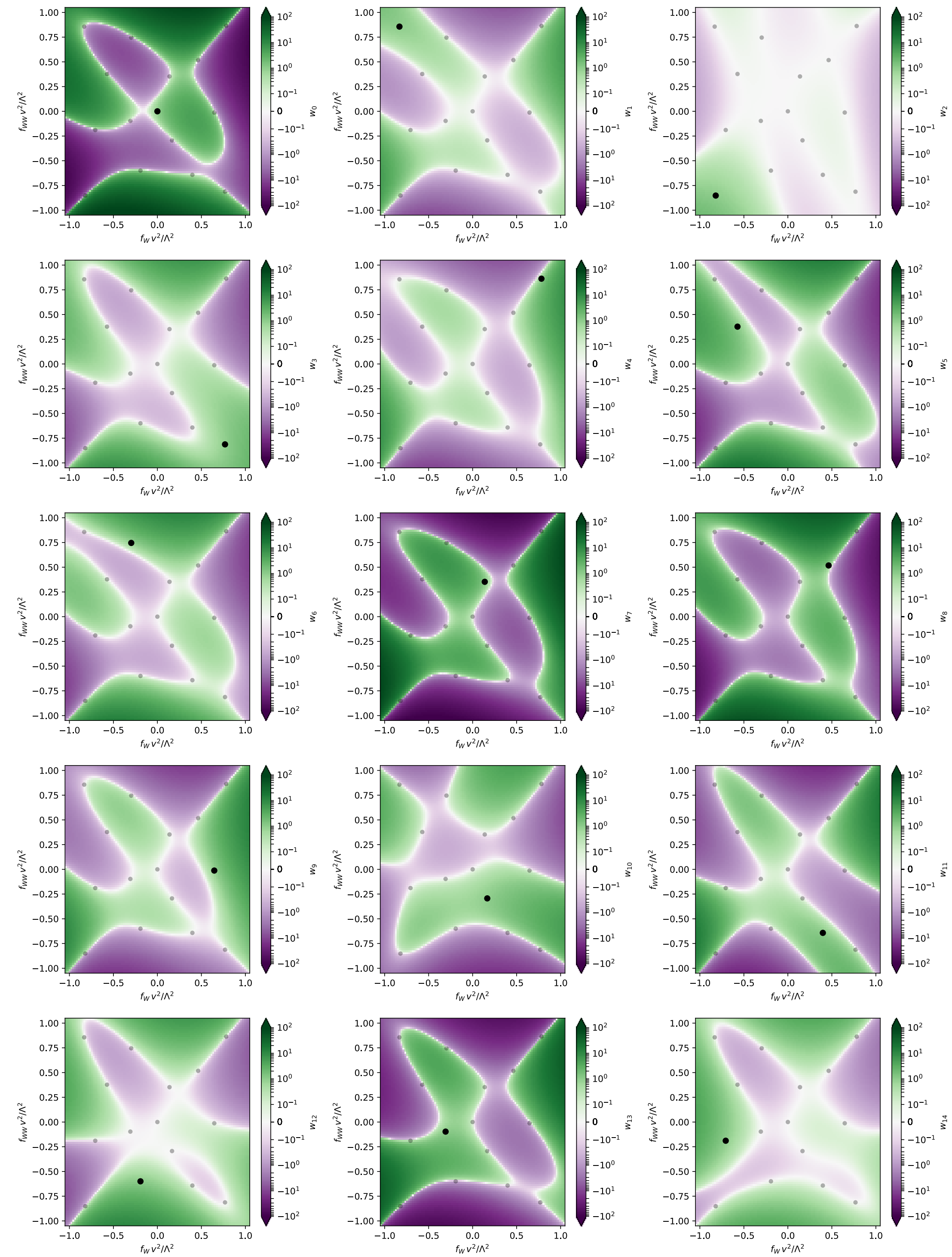
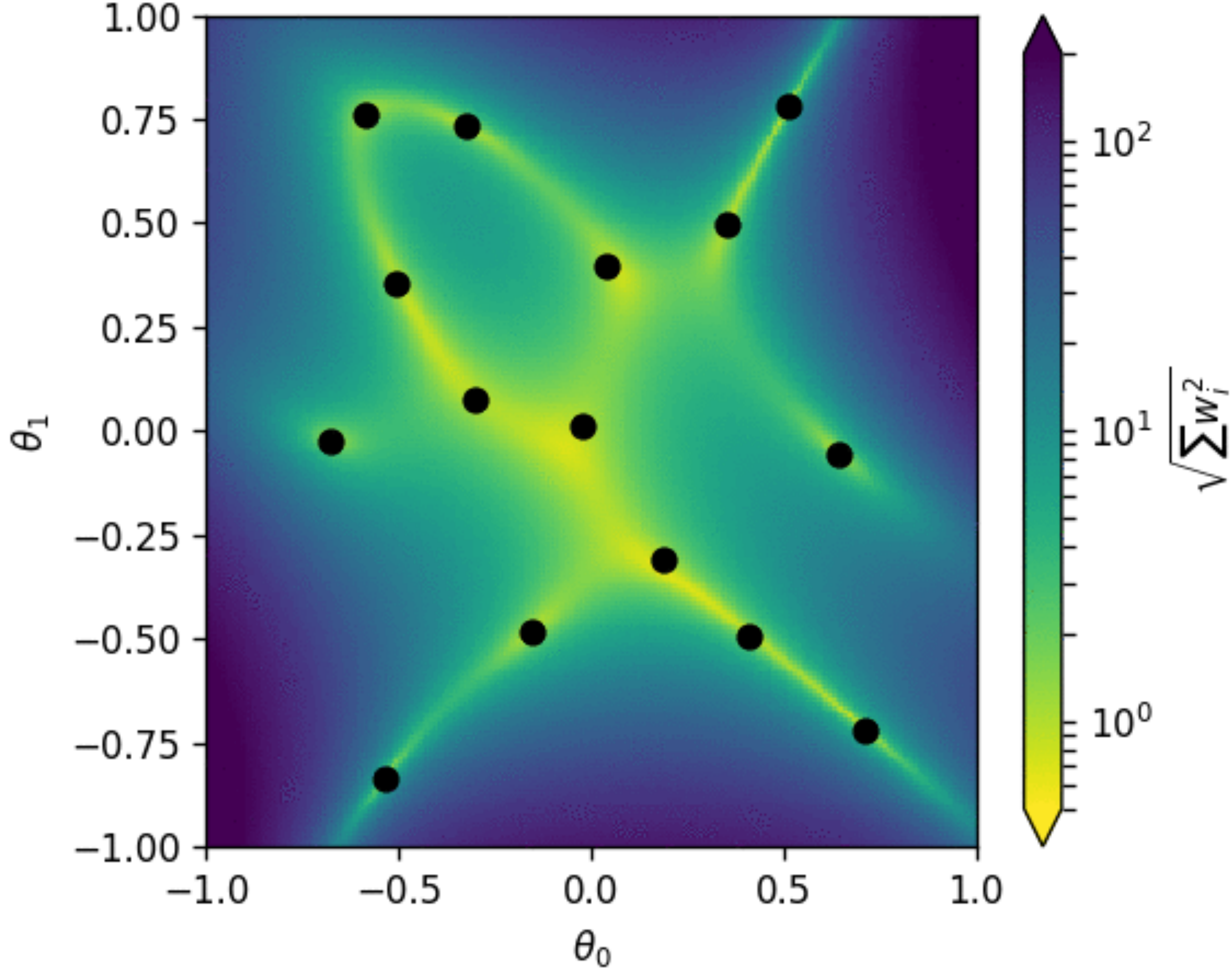


Figure 13: Morphing weights $w_i(\boldsymbol{\theta})$ for basis points distributed over the full relevant parameter space.

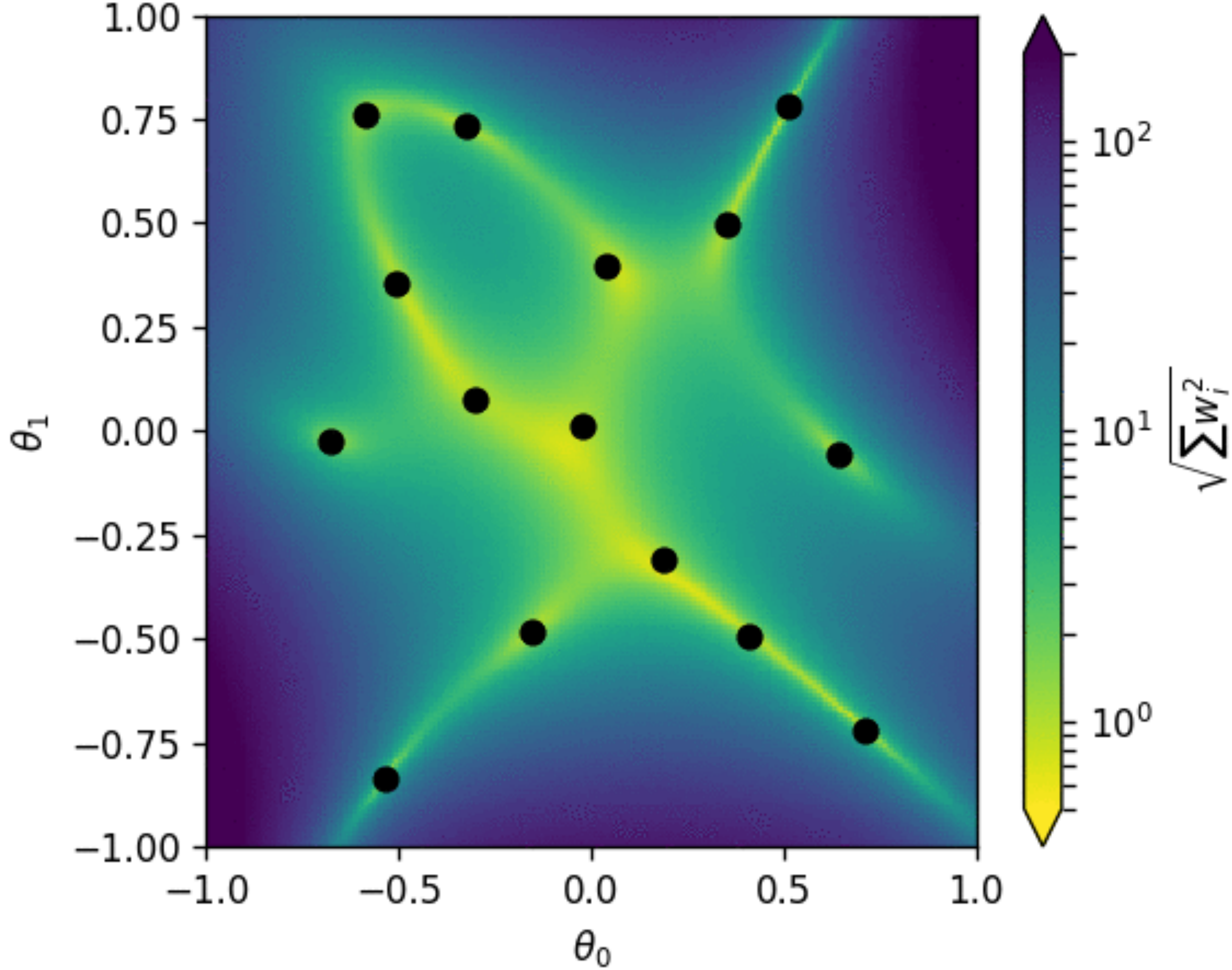
For 2 BSM operators affecting VBF Higgs production and decay, we need a 15-D vector space

For 5 BSM operators we need 126-D vector space

CHOICE OF BASIS & EFFECTIVE SAMPLE SIZE

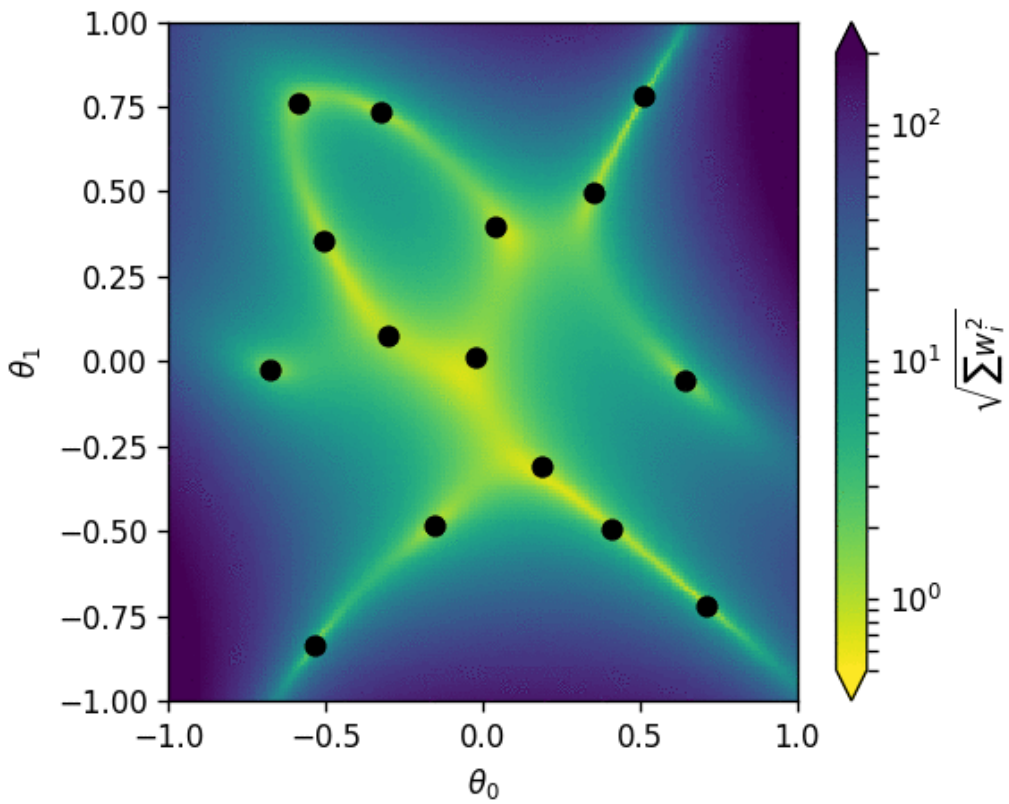


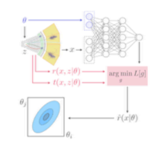
CHOICE OF BASIS & EFFECTIVE SAMPLE SIZE



Hands-on Tutorial

https://cranmer.github.io/madminer-tutorial/

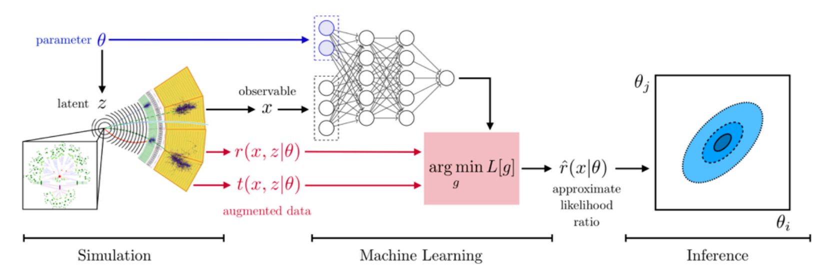




Introduction

MadMiner tutorial

This is a tutorial on MadMiner developed by Johann Brehmer, Felix Kling, Irina Espejo, and Kyle Cranmer. It is built using Jupyter Book.



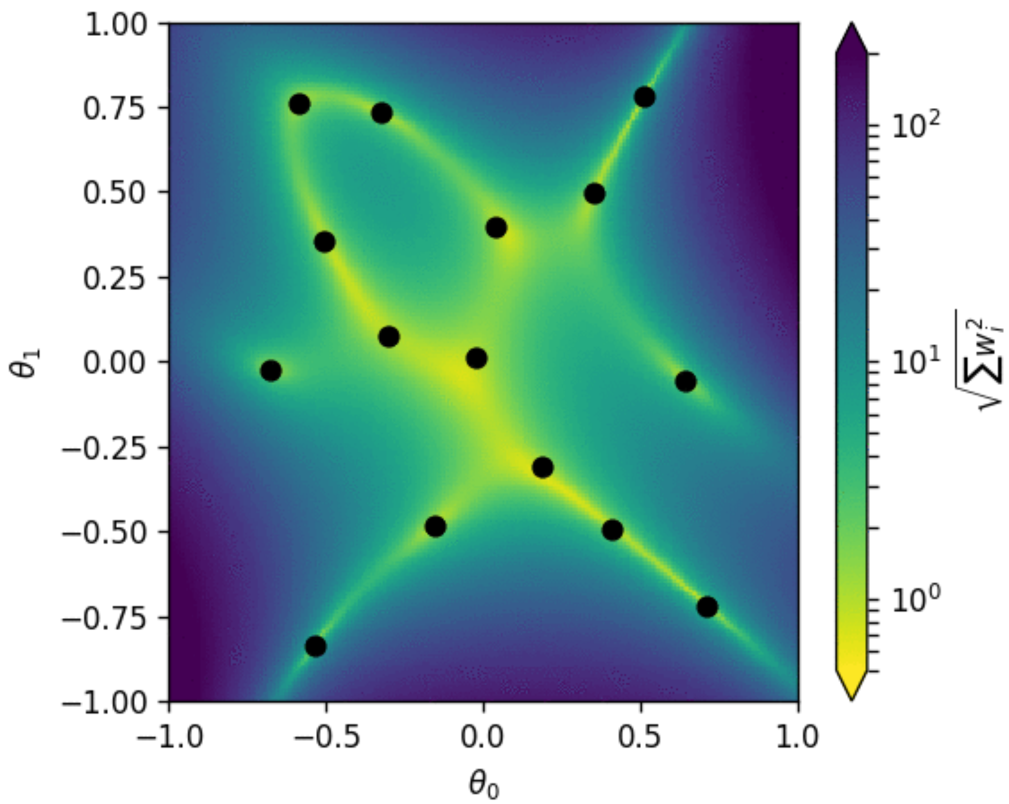
Introduction to MadMiner

Particle physics processes are usually modelled with complex Monte-Carlo simulations of the hard process, parton shower, and detector interactions. These simulators typically do not admit a tractable likelihood function: given a (potentially high-dimensional) set of observables, it is usually not possible to calculate the probability of these observables for some model parameters. Particle physicisists usually tackle this problem of "likelihood-free inference" by hand-picking a few "good" observables or summary statistics and filling histograms of them. But this conventional approach discards the information in all other observables and often does not scale well to high-dimensional problems.

In the three publications "Constraining Effective Field Theories With Machine Learning", "A Guide to Constraining Effective Field Theories With Machine Learning", and "Mining gold from implicit models to improve likelihood-free inference", a new approach has been developed. In a nut shell, additional information is extracted from the simulations that is closely related to the matrix elements that determine the hard process. This "augmented data" can be used to train neural networks to efficiently approximate arbitrary likelihood ratios. We playfully call this process "mining gold" from the simulator, since this information may be hard to get, but turns out to be very valuable for inference.

Hands-on Tutorial

https://cranmer.github.io/madminer-tutorial/



MadMiner Tutorial

Introduction

MadMiner Tutorial

Preliminaries

Overview

Define process to study *

Morphing

Interactive Morphing Demo

Create training data

Set MadGraph Directory

Parton Level *

With Delphes

Train model

Likelihood Ratio *

Score *

Likelihood

Statistical Analysis

Limits on EFT parameters *

Fisher Information

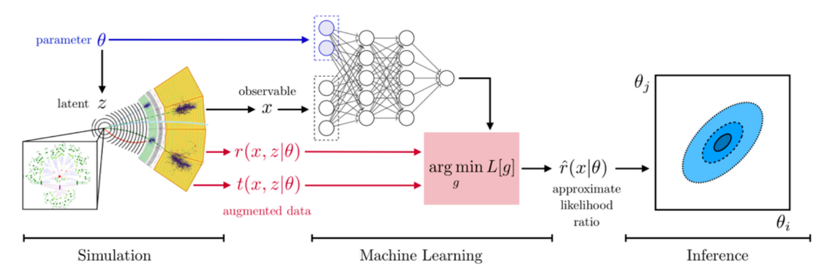
Information Geometry

Congratulations

Introduction

MadMiner tutorial

This is a tutorial on MadMiner developed by Johann Brehmer, Felix Kling, Irina Espejo, and Kyle Cranmer. It is built using Jupyter Book.



Introduction to MadMiner

Particle physics processes are usually modelled with complex Monte-Carlo simulations of the hard process, parton shower, and detector interactions. These simulators typically do not admit a tractable likelihood function: given a (potentially high-dimensional) set of observables, it is usually not possible to calculate the probability of these observables for some model parameters. Particle physicisists usually tackle this problem of "likelihood-free inference" by hand-picking a few "good" observables or summary statistics and filling histograms of them. But this conventional approach discards the information in all other observables and often does not scale well to high-dimensional problems.

In the three publications "Constraining Effective Field Theories With Machine Learning", "A Guide to Constraining Effective Field Theories With Machine Learning", and "Mining gold from implicit models to improve likelihood-free inference", a new approach has been developed. In a nut shell, additional information is extracted from the simulations that is closely related to the matrix elements that determine the hard process. This "augmented data" can be used to train neural networks to efficiently approximate arbitrary likelihood ratios. We playfully call this process "mining gold" from the simulator, since this information may be hard to get, but turns out to be very valuable for inference.