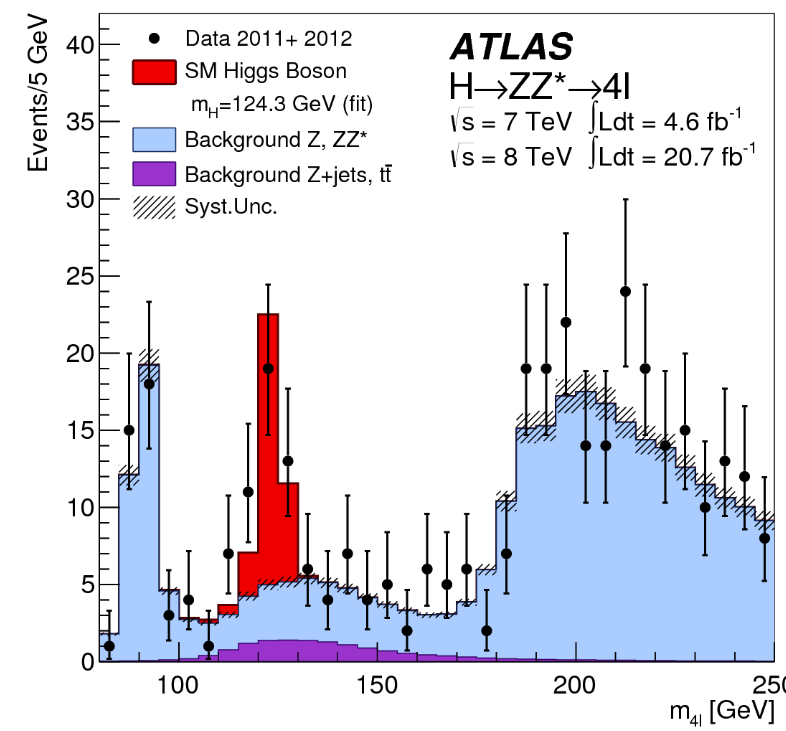
Anjishnu Bandyopadhyay
University of Bonn

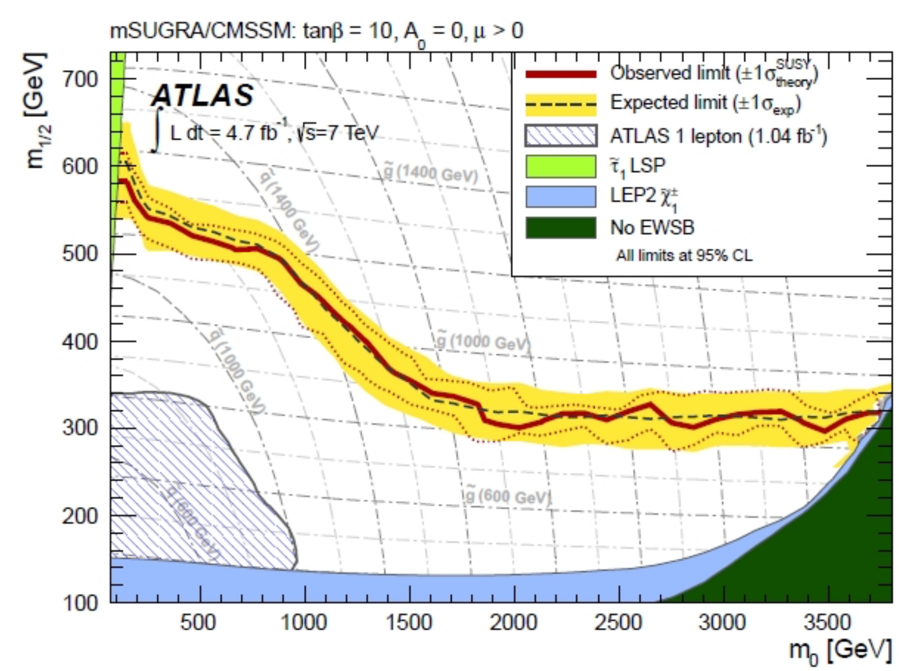
Development of morphing algorithms for Histfactory using information geometry

DPG Frühjahrstagung
Münster
27.03.17

Introduction

- Most analyses in high-energy physics use likelihood fits.
- From setting limits to measuring predicted particles.
- A Poisson likelihood is used in most cases.
- MC distributions of kinematic variables for both signal and background are used as input.





Motivation

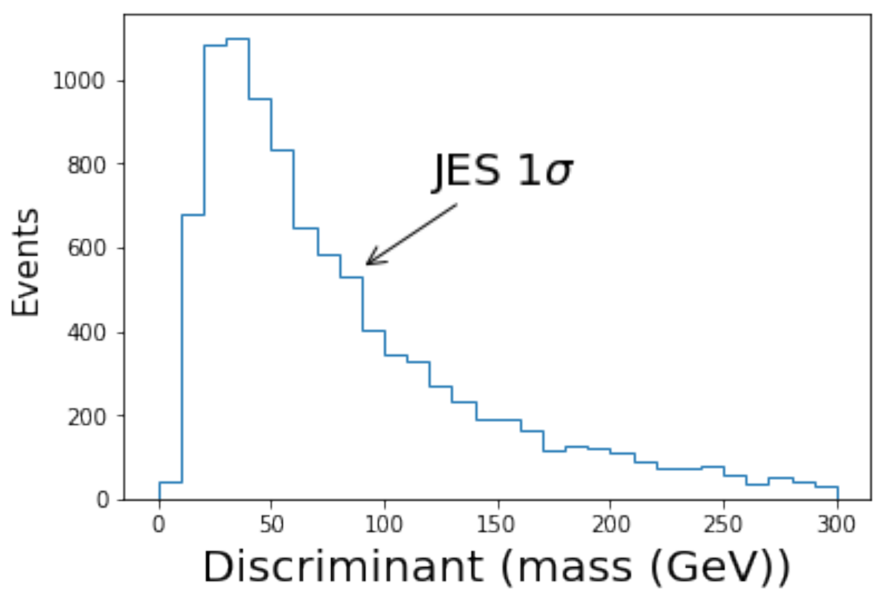
$$\mathcal{L}(\mu, \alpha) = \prod_{k}^{\text{bins}} \frac{{E_k}^{N_k} \mathrm{e}^{-E_k}}{N_k!}$$

$$E_k \equiv E_k(\mu, \alpha) = \mu E_k^{\text{signal}}(\alpha) + \sum_{j}^{\text{bkgs.}} E_k^j(\alpha)$$

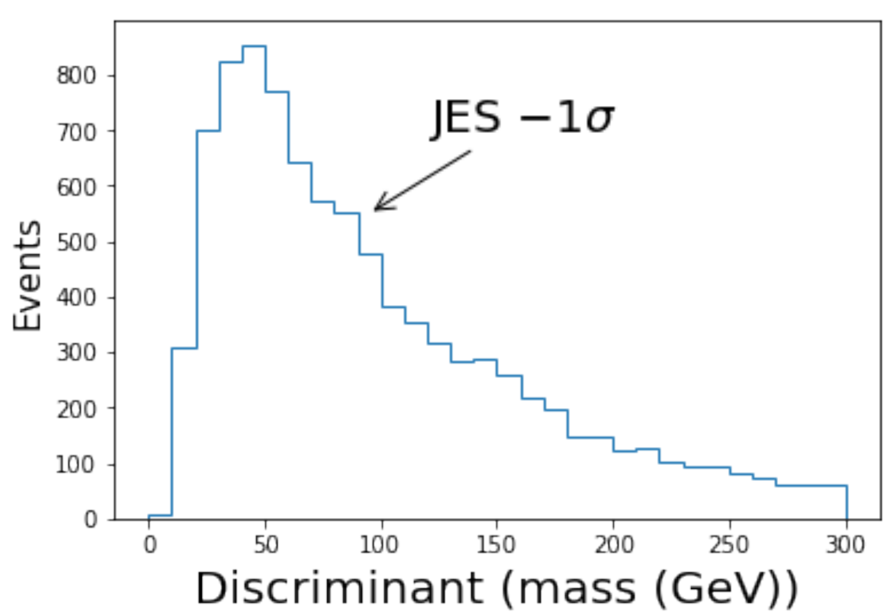
E_k : Expected number of events in the bin k.
N_k : Observed number of events in the bin k.

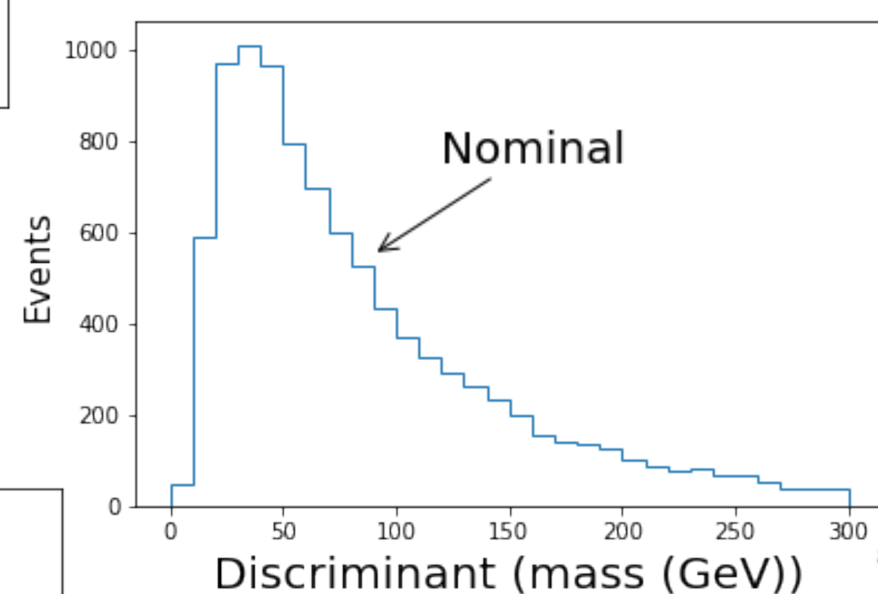
α is the nuisance parameter (JES, b-tagging unc.) vector.
We only have E_k for $\alpha = 0, \pm 1\sigma$.
We would like to compute E_k for arbitrary α

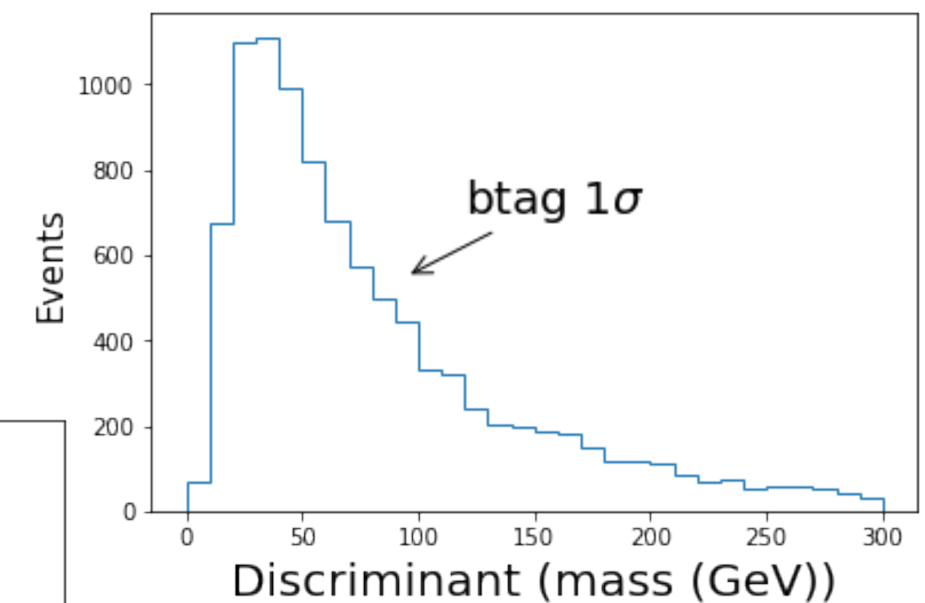
A Toy example



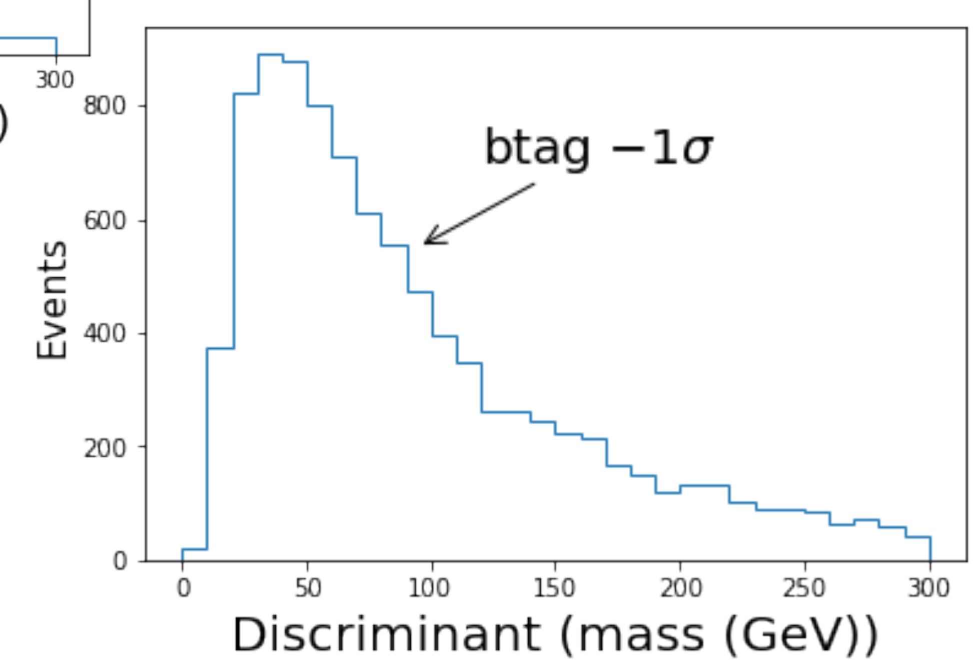
?





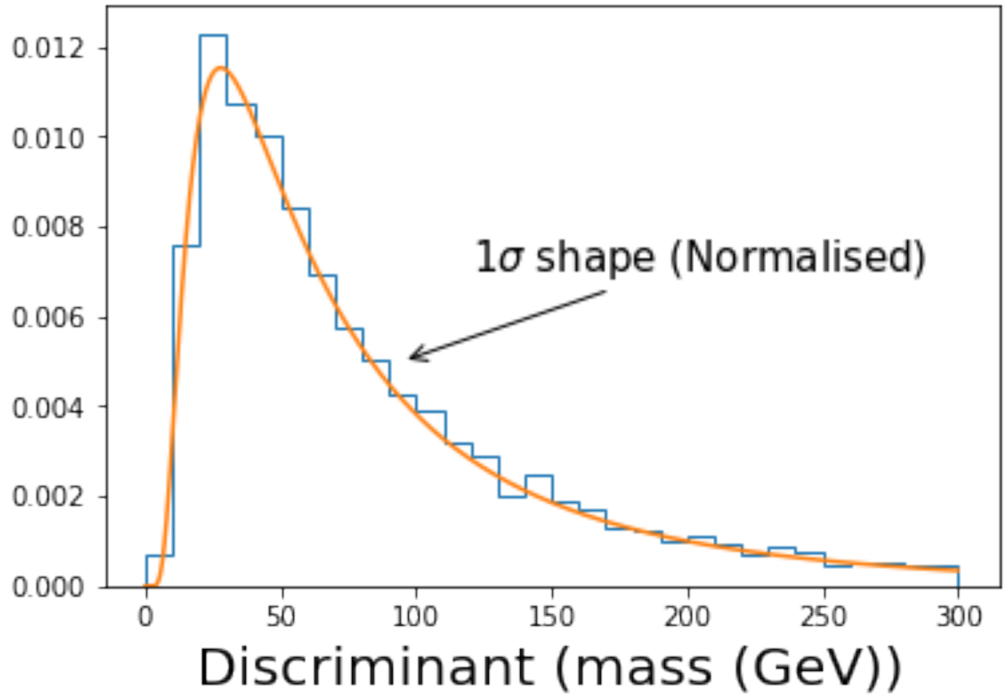


?

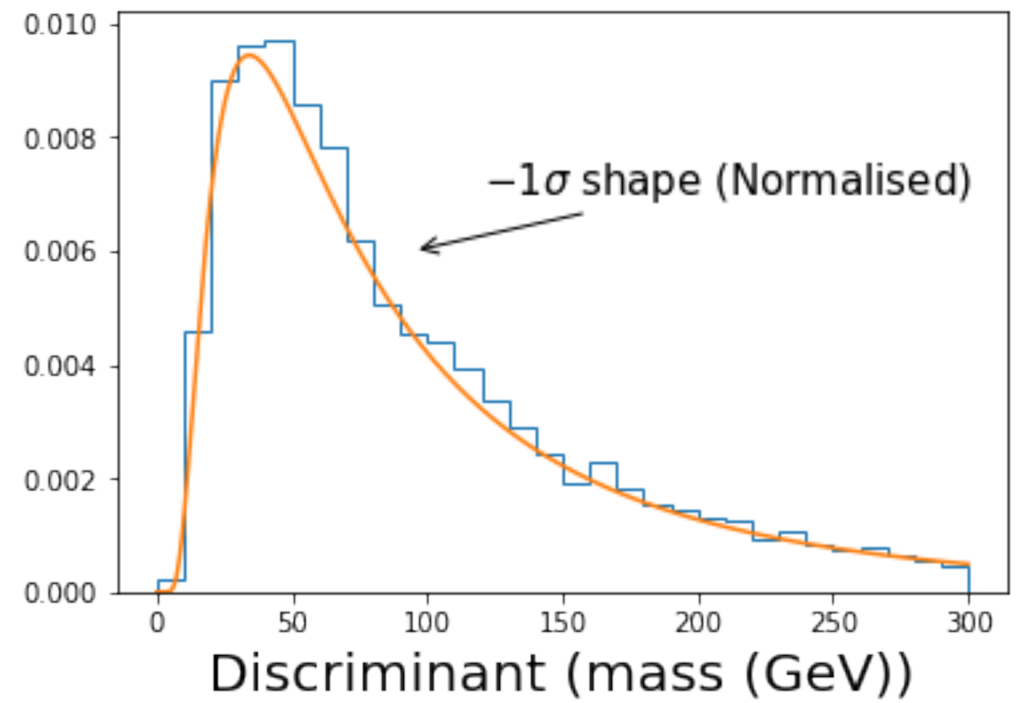


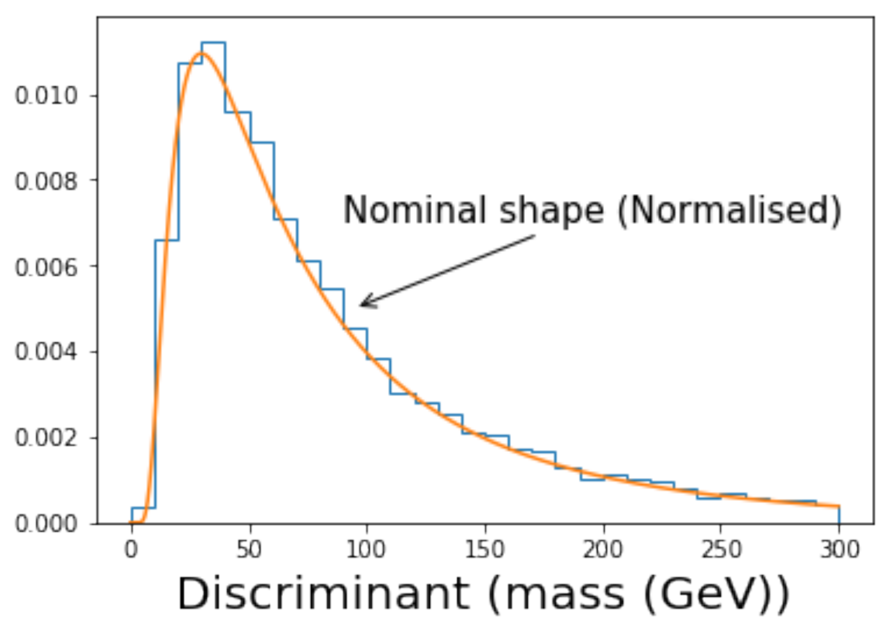
Interpolating shape uncertainties

Both uncertainties are combined quadratically. (Prior knowledge of independence)

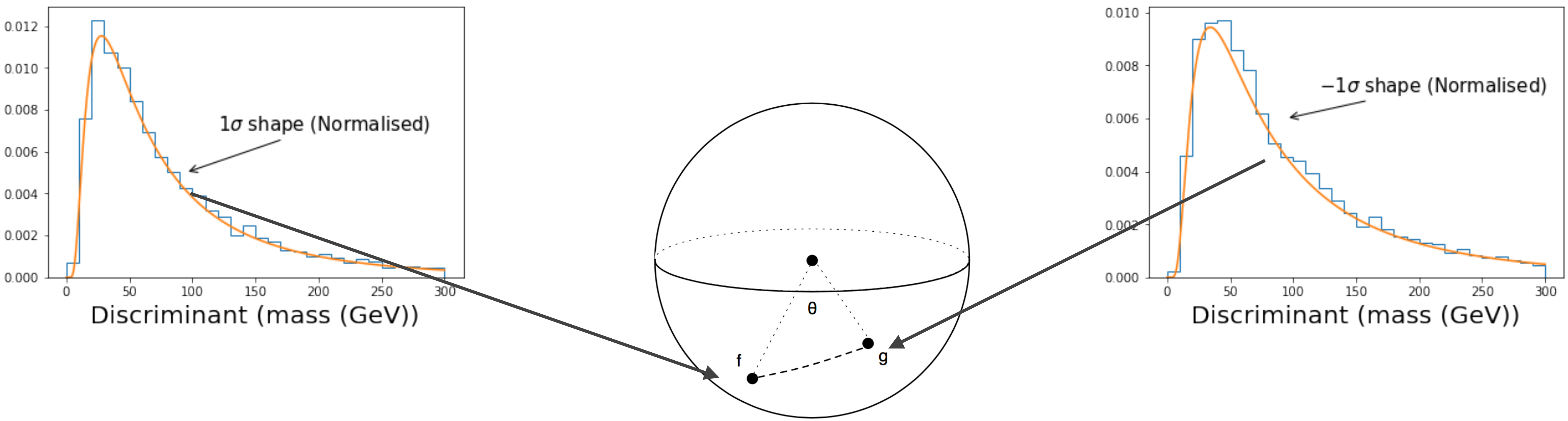


?





Ambient Fisher distance

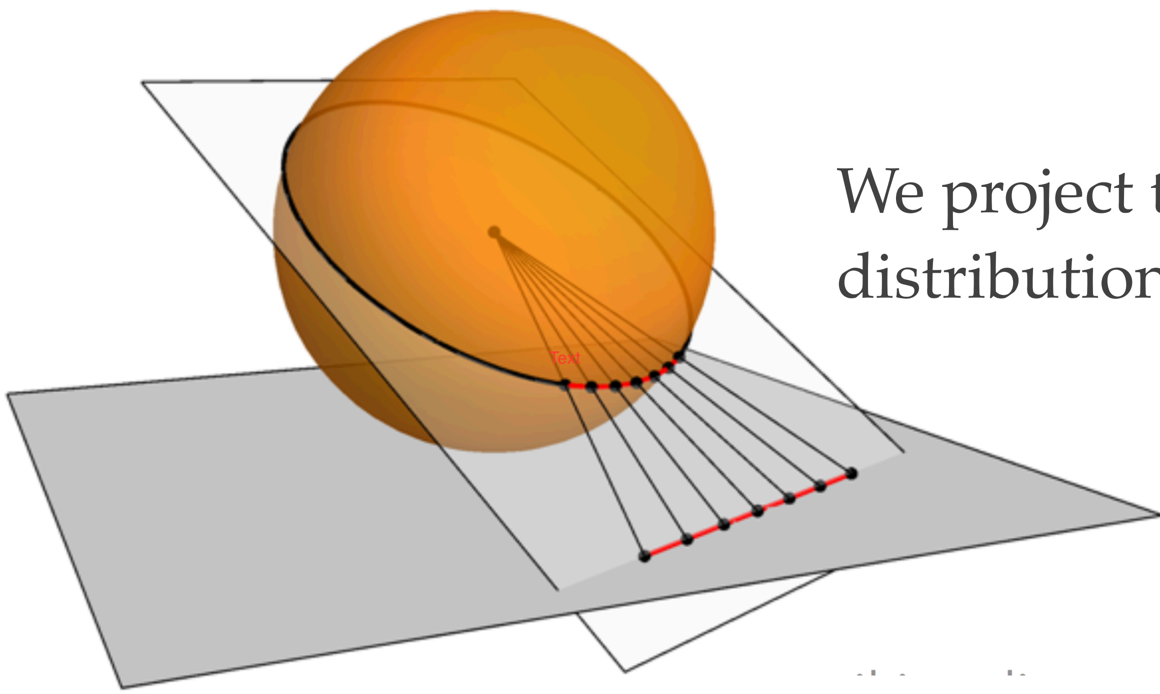


$$\langle f, g \rangle = \int \sqrt{f(x)g(x)}dx = \cos\theta$$

$$D_F(f,g) = \arccos\left(\int \sqrt{f(x)g(x)}dx\right)$$

A recursive algorithm used to embed the root distributions and $D_F(\sqrt{f},\sqrt{g})$ is the distance between these distributions

Gnomonic projection

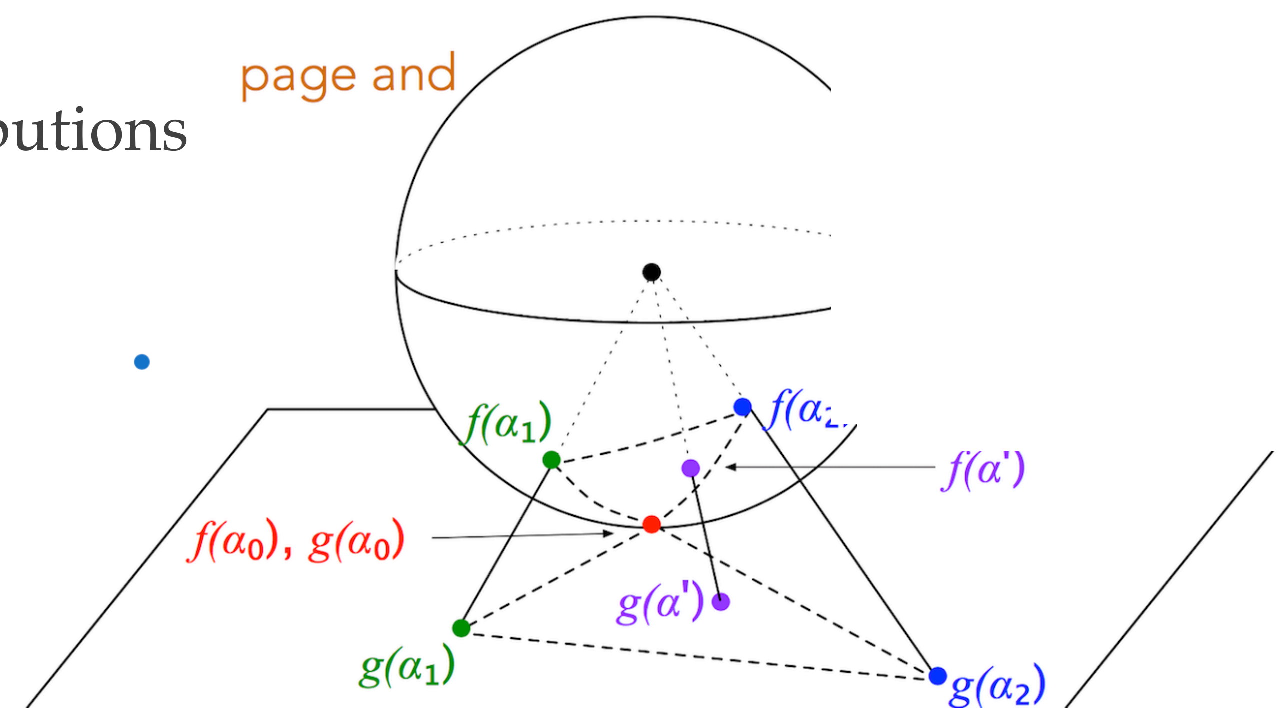


We project the variational distributions, nominal and $\pm 1\sigma$ distributions. The nominal is projected to the south pole.

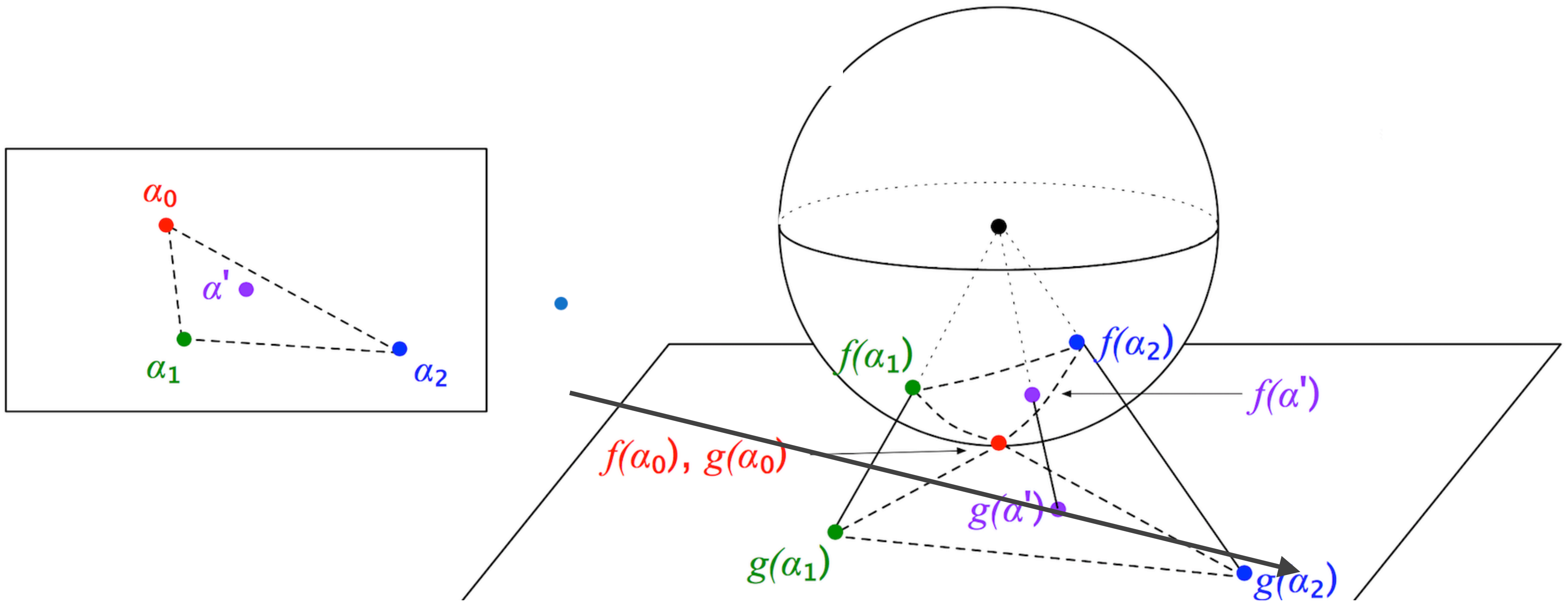
The variational distributions are projected accordingly to the (n-1) dim. hyperplane.

In our test case n=2. So we have a 2-sphere. The two root variational distributions are $f(\alpha_1)$ and $f(\alpha_2)$

We want to compute $f(\alpha')$



Barycentric coordinates

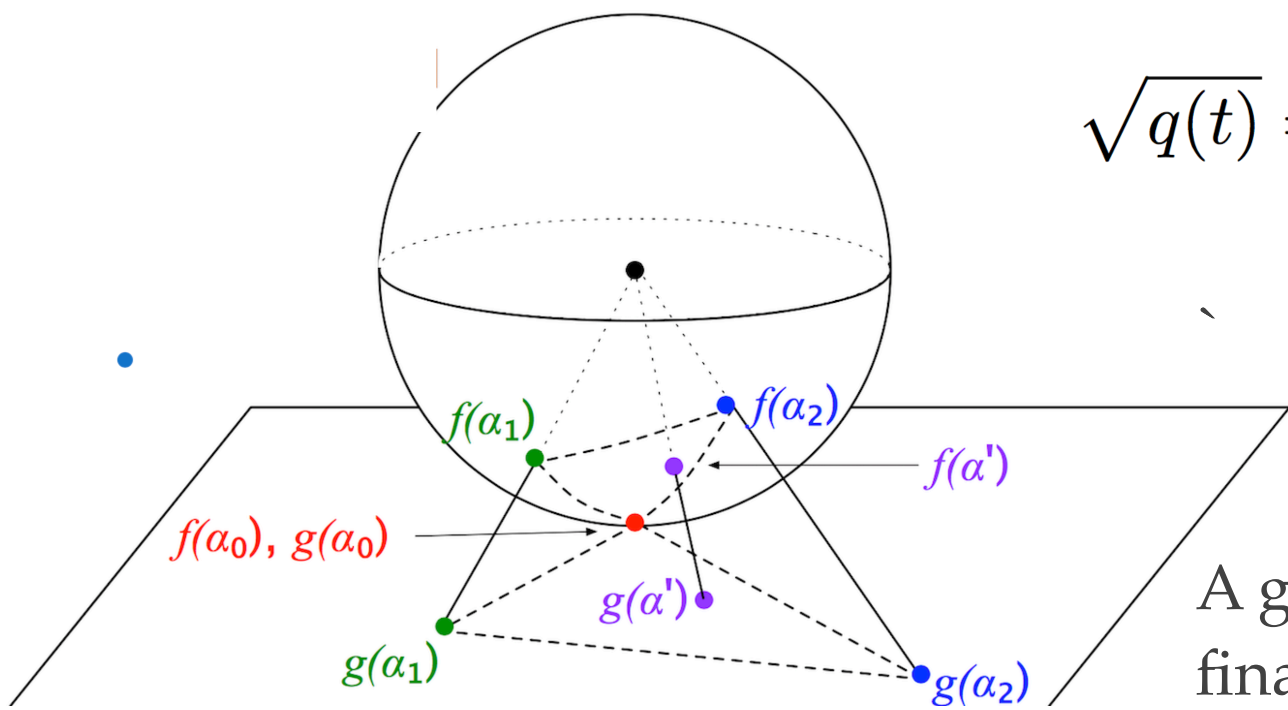


α_i are the input variational points.

$g(\alpha_i)$ are the gnomonic projections of the variational distributions.

Affine transformation to map barycentric to hyperplane

Interpolation



$$\sqrt{q(t)} = \sqrt{f}\cos t + \frac{\sqrt{g} - \langle f, g\rangle\sqrt{f}}{|\sqrt{g} - \langle f, g\rangle\sqrt{f}|}\sin t$$

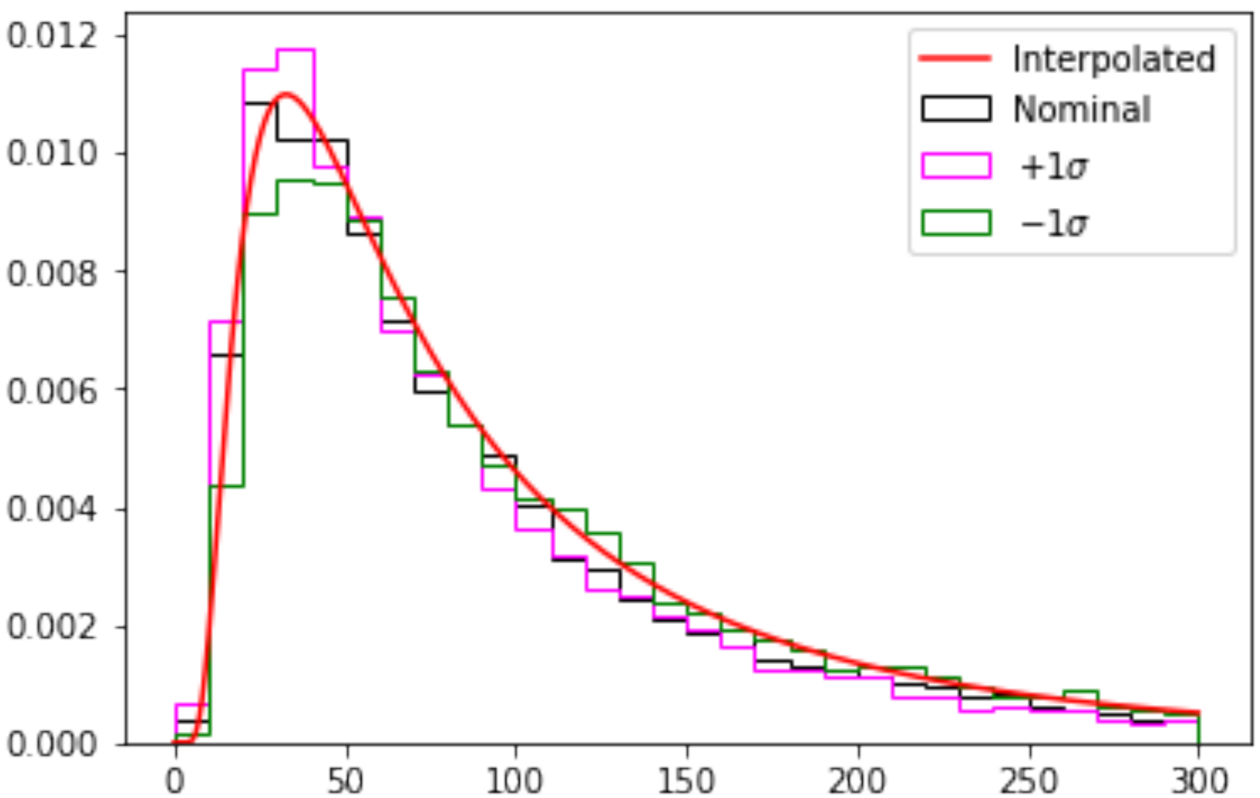
A geodesic equation is used to compute the final distribution $q(t)$

We want to compute $f(\alpha')$

We do an inverse transformation from $g(\alpha')$ to $f(\alpha')$

Results

We use this algorithm to compute an interpolated distribution for an arbitrary α'



$$f(\alpha_i), \alpha_i$$

$\downarrow$

Embed root distributions on n-sphere

$\downarrow$

Gnomonic projection

$\downarrow$

Find target point on sphere and use geodesic equation.

Summary and outlook

- An interpolation algorithm has been developed using information geometry.
- Geometric properties of root distributions used to develop the algorithm.
- Nuisance parameters can be varied simultaneously.
- Algorithm is fast.

- The algorithm has been developed for ROOSTATS framework.
- Results shown today from python implementation.
- Tests to verify working for ROOSTATS ongoing.