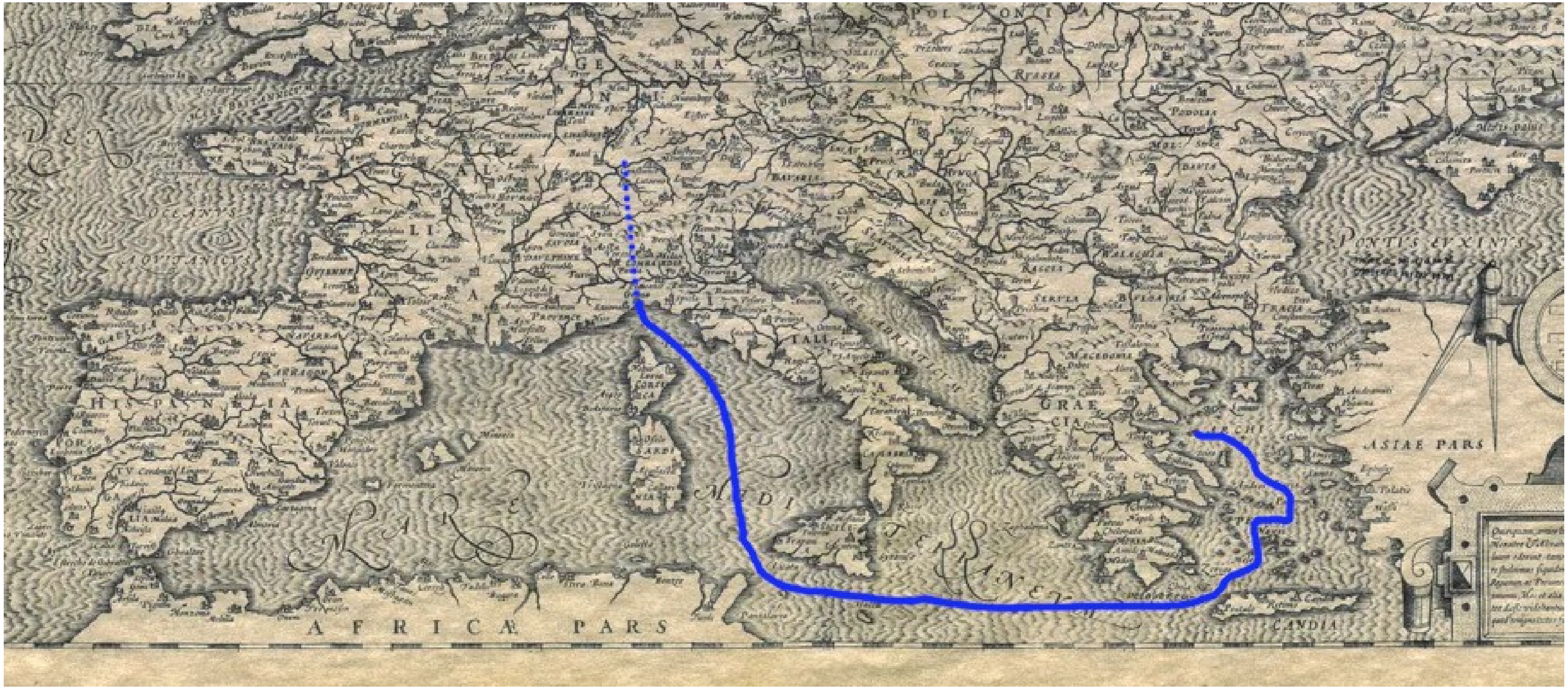
Bayesian Inference in Collider Physics

Ezequiel Alvarez
ICAS, UNSAM & Conicet (Argentina)
sequi@unsam.edu.ar

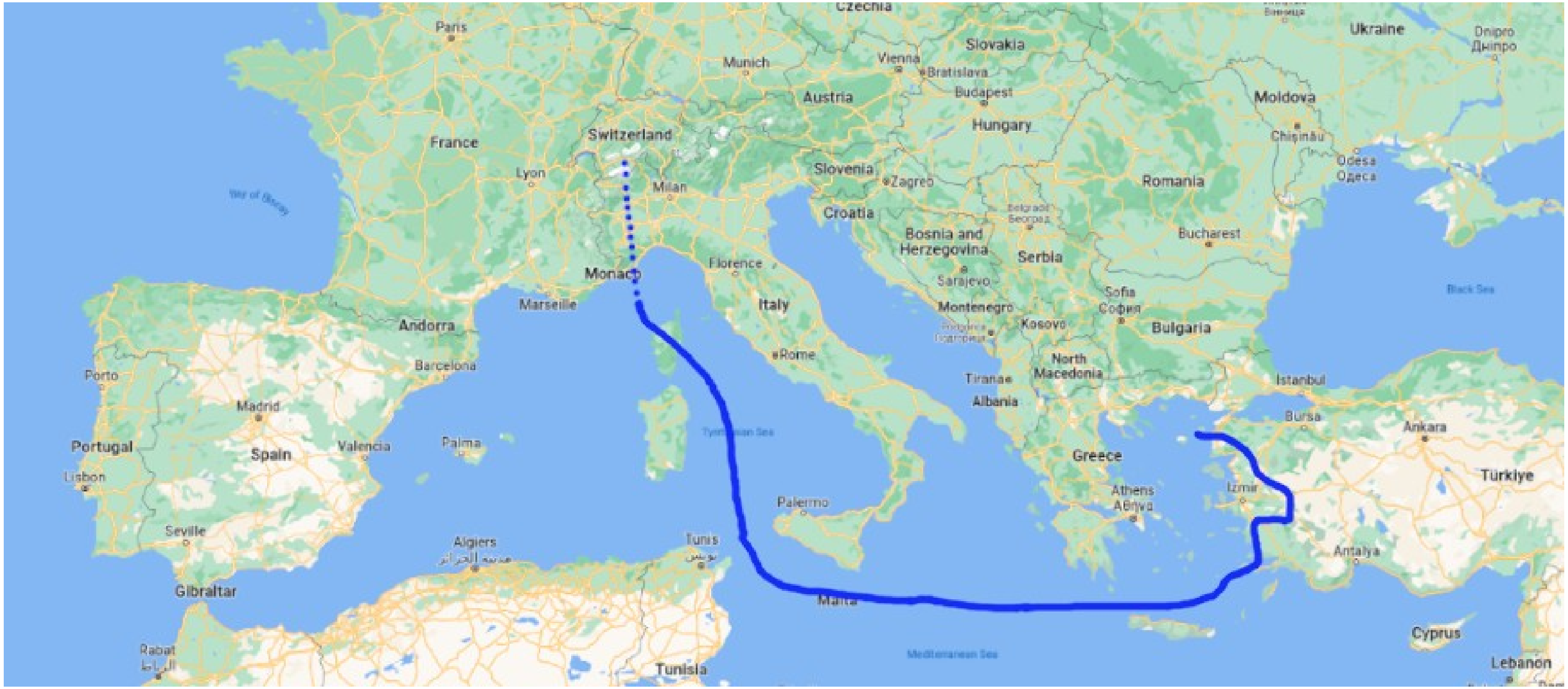
June 13th, 2023

XVIIth century map

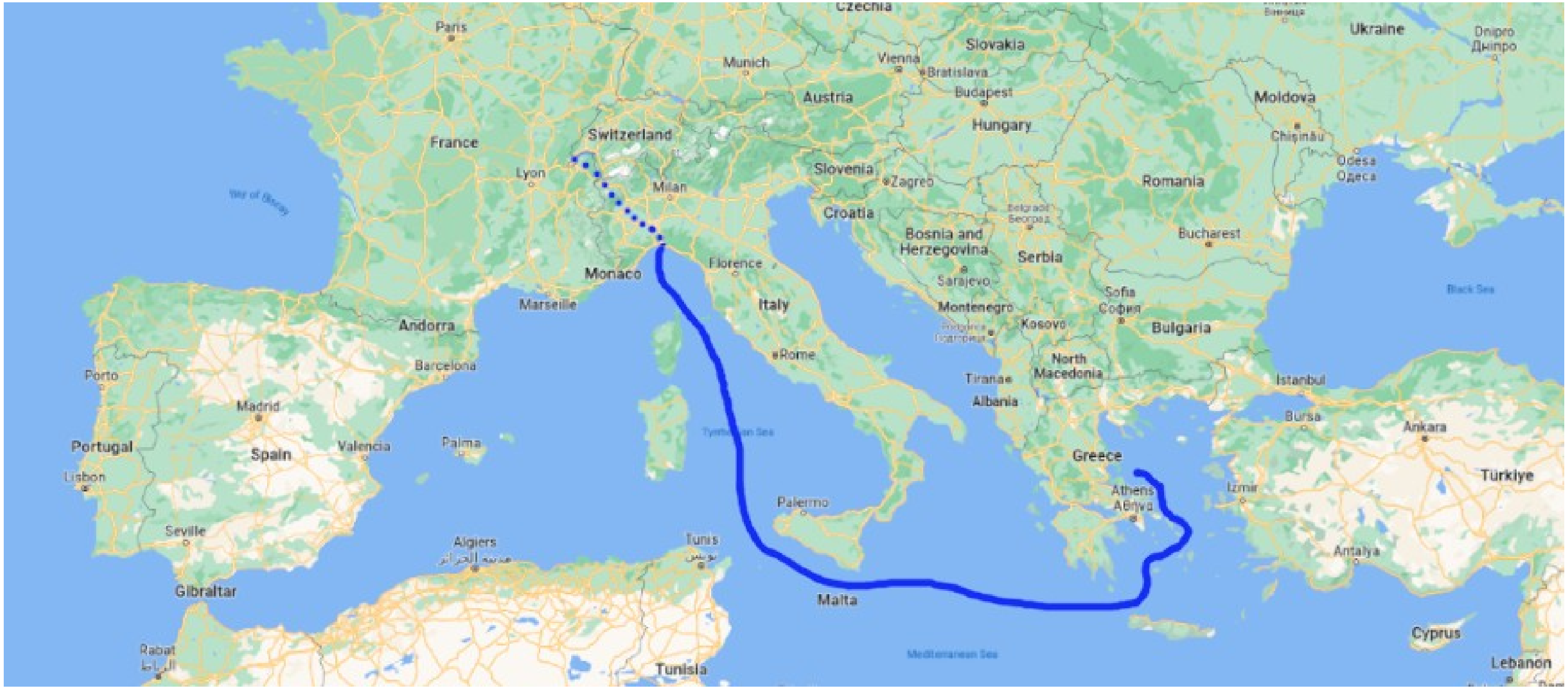
XVII[th] century map



Reality


Czechia
Paris
Munich
Vienna
Bratislava
Slovakia
Ukraine
Dnipro
Дніпро
Austria
Budapest
Hungary
Moldova
Chișinău
Switzerland
France
Lyon
Milan
Slovenia
Zagreb
Croatia
Romania
Odesa
Одеса
Bosnia and
Herzegovina
Sarajevo
Serbia
Bucharest
Florence
Monaco
Marseille
Italy
Montenegro
Kosovo
Sofia
София
Bulgaria
Black Sea
Andorra
Barcelona
Rome
Tirana
Albania
North
Macedonia
Istanbul
Porto
Madrid
Bursa
Ankara
Portugal
Spain
Valencia
Palma
Greece
Türkiye
Lisbon
Izmir
Athens
Αθήνα
Seville
Palermo
Algiers
Tunis
Antalya
Gibraltar
Malta
Cyprus
Rabat
Lebanon
Tunisia
Mediterranean Sea

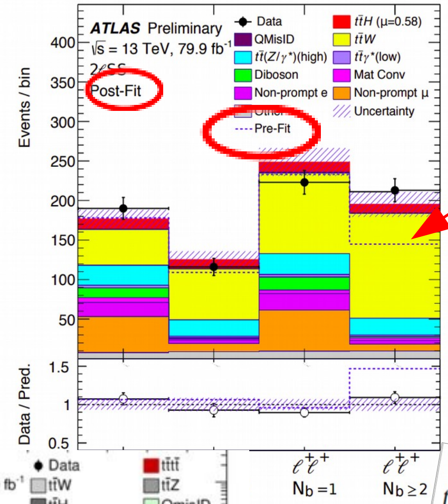
Reality



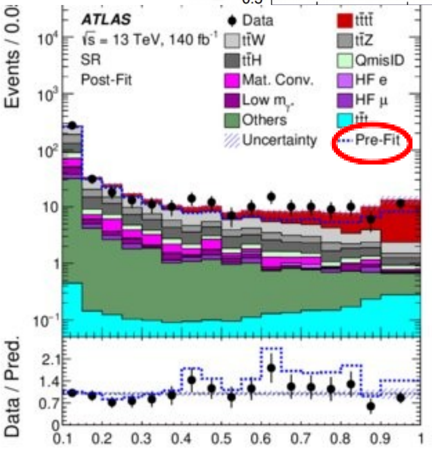
Use the inaccurate map as a guide, and then correct as you meet reality

Monte Carlo are a great guide

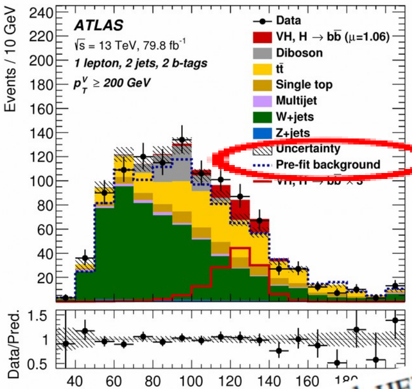
Sherpa



Herwig



wgtq3 was chosen to maximise this rate. Finally, some of the models (wgtq5, wgtq6) offer the possibility to choose sgtq·$m_{b\bar{b}}$ instead of the transverse momentum as the argument of α_S in the $g \to b\bar{b}$ vertices. Here sgtq refers to the TIMESHOWER:SCALEGLUONTOQUARK parameter, and is allowed to vary in the range $0.25 \leq$ sgtq ≤ 1, with larger values giving a smaller $g \to b\bar{b}$ rate and vice versa. For the model wgtq5, sgtq was set to 1, a combination that minimises the $g \to b\bar{b}$ rate, while for wgtq6, sgtq was set to 0.25.

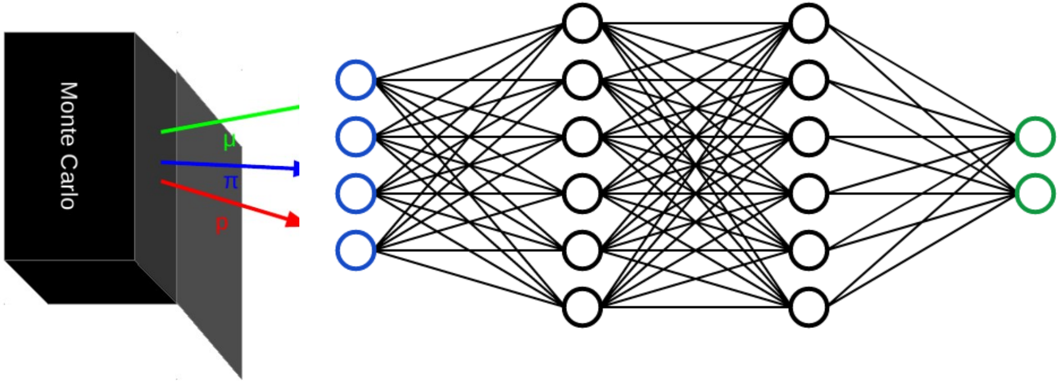


Scale factor

Pythia

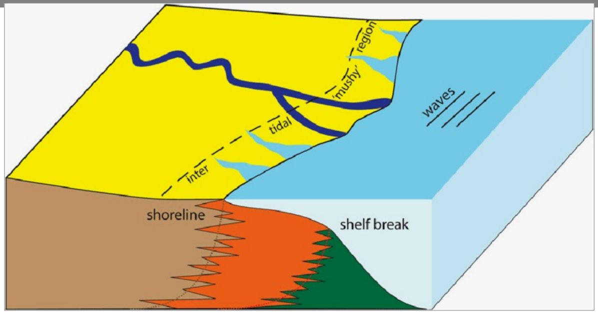
PYTHIA6 and HERWIG6. The normalization factors applied to the MADGRAPH and POWHEG predictions are found to be about 1.3 for results related to the leading additional b jet. The predictions from both generators underestimate the tt̄bb̄ cross sections by a factor 1.8, in agreement with the results from Ref. [11]. The normalization factors applied to MC@NLO are approximately 2 and 4 for the leading and subleading additional b jet quantities, respectively, reflecting the observation that the generator does not simulate sufficiently large jet multiplicities. All the predictions have slightly harder p_T spectra for the leading additional b jet than the data, while

However...



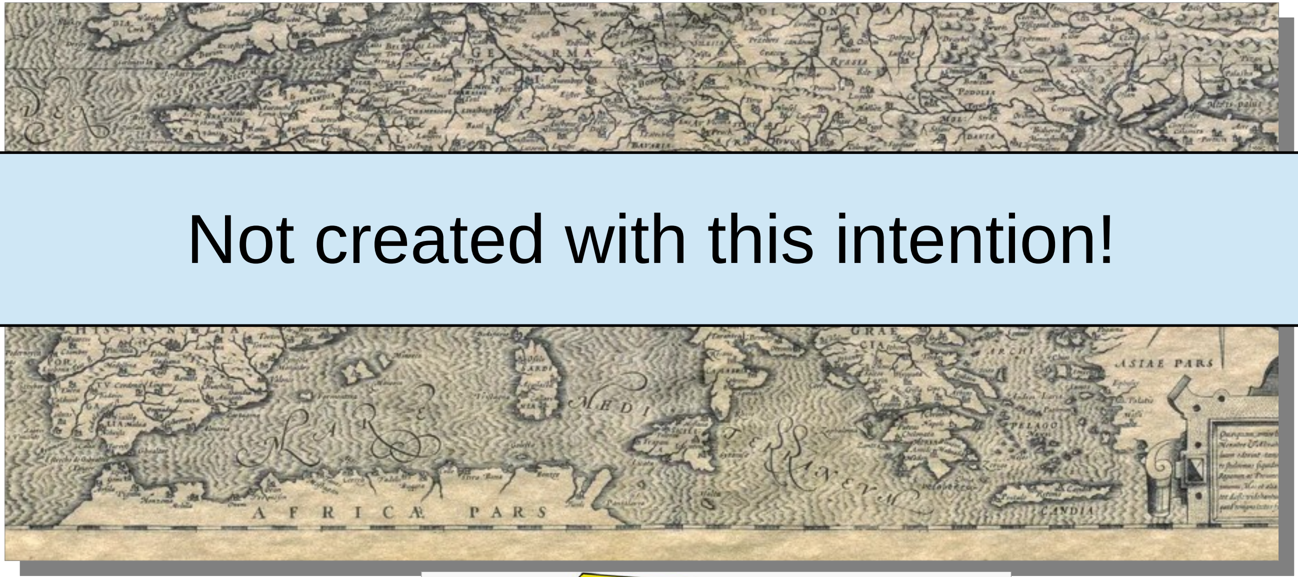
Neural Networks learning from MC:
Potential biases if learns as physics,
details, correlations, etc, that are not physics !

Just as if….



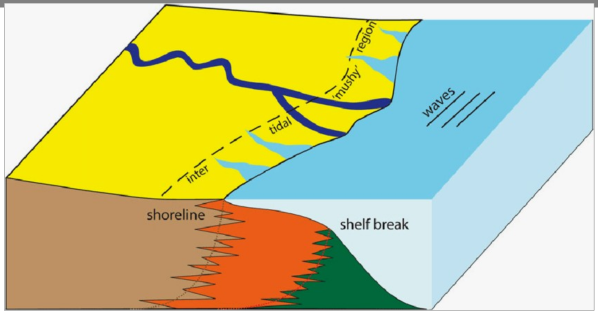
Plugging Neural Network to shore contours to learn anything

Just as if….



Not created with this intention!

Plugging Neural Network to shore contours to learn anything



Bayesian Inference

Alternative framework in which one learns from data using Monte Carlo, Theory and "more" as a guide (prior)

Bayesian Inference

Alternative framework in which one learns from data using Monte Carlo, Theory and “more” as a guide (prior)

Disclaimer: This talk has nothing to do with any Frequentist vs. Bayesian (pointless) discussion

Bayesian Inference

Alternative framework in which one learns from data using Monte Carlo, Theory and "more" as a guide (prior)

Disclaimer: This talk has nothing to do with any Frequentist vs. Bayesian (pointless) discussion
Instead: is about new tools and techniques that are more suitable within a Bayesian framework

Summary

- Intro to Bayesian framework
- Graphical Models (the Feynman diagrams in statistics!)

Applications

- q- Vs, g-jets using softdrop Poisson shapes
- Four tops: correlating N_J and N_b
- Di-Higgs: correlation and full info extraction
- Posterior predictive ← (check your model with data)
- LHC measuring techniques

Intro to

Bayesian Framework

Intro to Bayesian framework

Bayes Theorem

$$P(\theta|X) = \frac{P(X|\theta) \times P(\theta)_{prior}}{P(X)}$$

Intro to Bayesian framework

Bayes Theorem

$$P(\theta|X) = \frac{P(X|\theta) \times P(\theta)_{\text{prior}}}{P(X)}$$

Cleverness: the data is modeled to be sampled from a given PDF

X : data
θ : parameters of a PDF

Intro to Bayesian framework

Bayes Theorem

$$P(\theta|X) = \frac{P(X|\theta) \times P(\theta)_{\text{prior}}}{P(X)}$$

Cleverness: the data is modeled to be sampled from a given PDF

X : data
θ : parameters of a PDF

Hence, in Bayesian the "*probability of a probability*" is always buzzing around

Intro to Bayesian framework

Bayes Theorem

$$P(\theta|X) = \frac{P(X|\theta) \times P(\theta)_{\text{prior}}}{P(X)}$$

By seeing the data you improve your knowledge of your PDF

Cleverness: the data is modeled to be sampled from a given PDF

X : data
θ : parameters of a PDF

Hence, in Bayesian the "*probability of a probability*" is always buzzing around

Intro to Bayesian framework

Bayes Theorem

$$P(\theta|X) = \frac{P(X|\theta) \times P(\theta)_{\text{prior}}}{P(X)}$$

By seeing the data you improve your knowledge of your PDF

Crucial info about Physics!

Cleverness: the data is modeled to be sampled from a given PDF

X : data
θ : parameters of a PDF

Hence, in Bayesian the "*probability of a probability*" is always buzzing around

Intro to Bayesian framework

Bayes Theorem

Statistics
is about
Modeling!

$$P(\theta|X) = \frac{P(X|\theta) \times P(\theta)_{\text{prior}}}{P(X)}$$

By seeing the data you improve your knowledge of your PDF

Crucial info about Physics!

Cleverness: the data is modeled to be sampled from a given PDF

X : data
θ : parameters of a PDF

Hence, in Bayesian the "*probability of a probability*" is always buzzing around

Graphical Models

Graphical Models

Probabilistic model for which a graph expresses the conditional dependence structure between random variables.

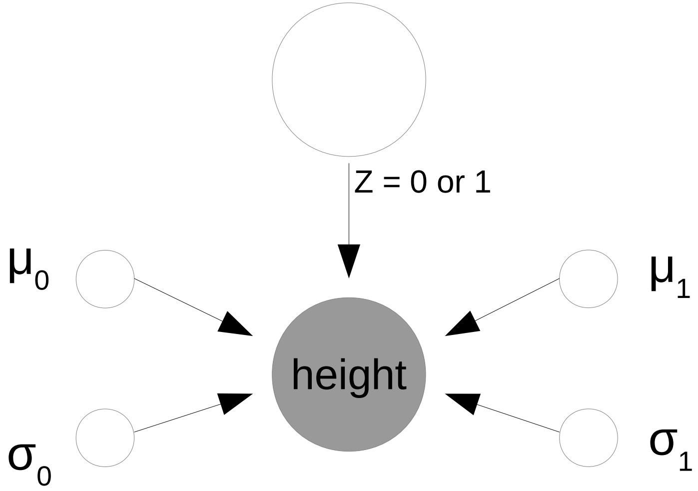
Graphical Models

Probabilistic model for which a graph expresses the conditional dependence structure between random variables.

Just a PDF, but more sophisticated than plain Gaussian, Exponential, etc

Graphical Models

Probabilistic model for which a graph expresses the conditional dependence structure between random variables.



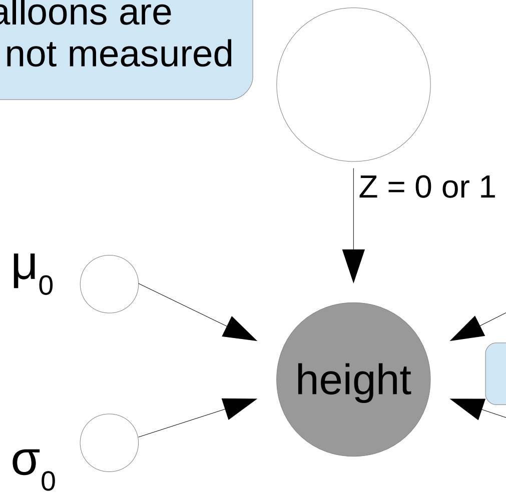
For instance, model the height of the persons in this room as coming from 2 populations

Graphical Models

Each balloon is a random variable

Empty balloons are sampled but not measured

Arrows indicate dependence

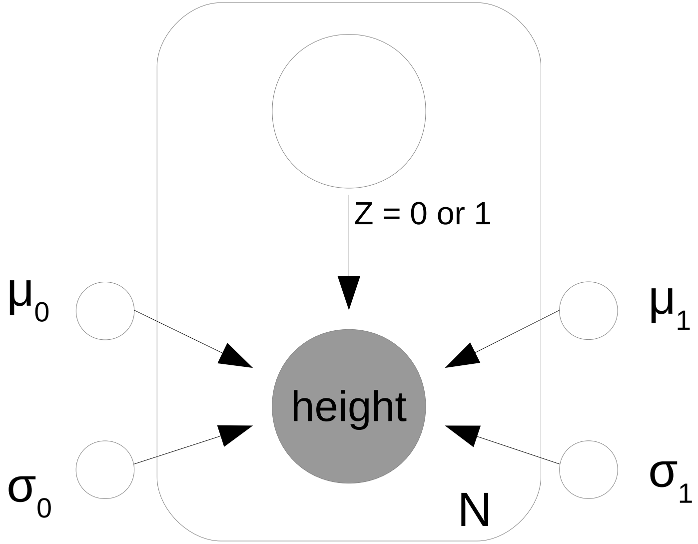


Filled balloons are sampled and measured

For instance, model the height of the persons in this room as coming from 2 populations

Graphical Models

Each balloon is a random variable

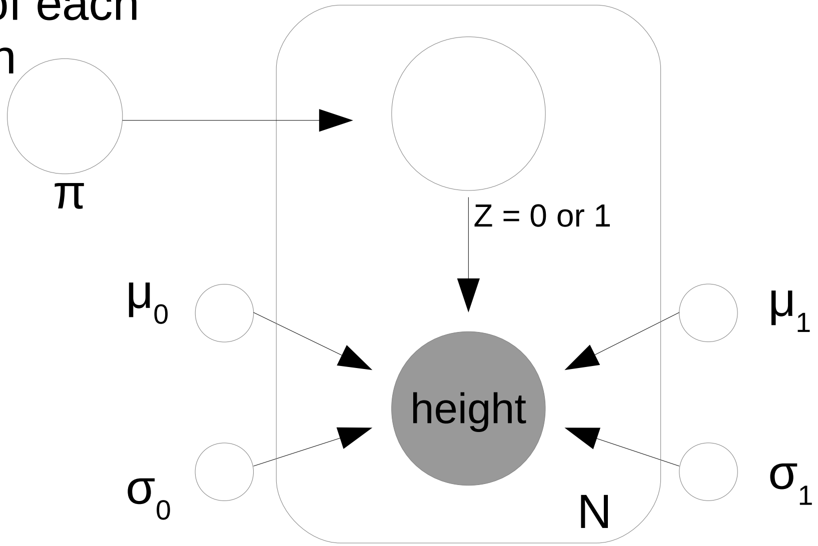


For instance, model the height of the persons in this room as coming from 2 populations

Graphical Models

Each balloon is a random variable

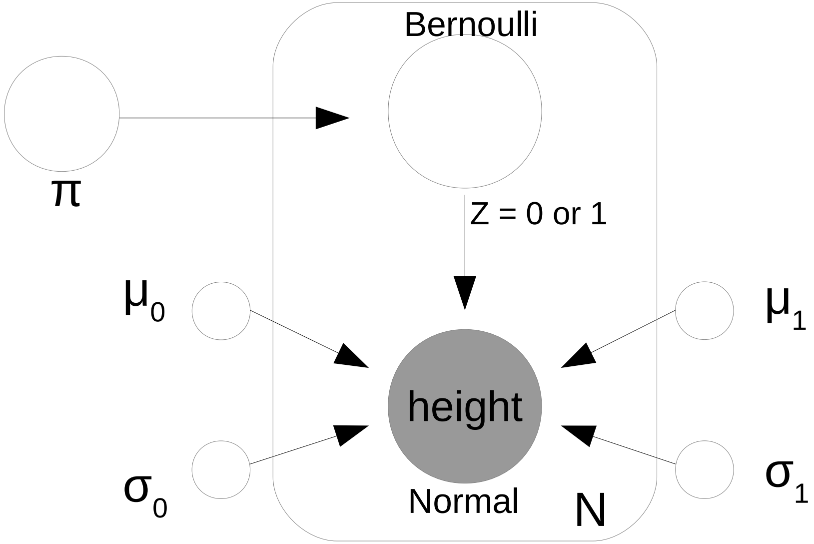
Fraction of each population



For instance, model the height of the persons in this room as coming from 2 populations

Graphical Models

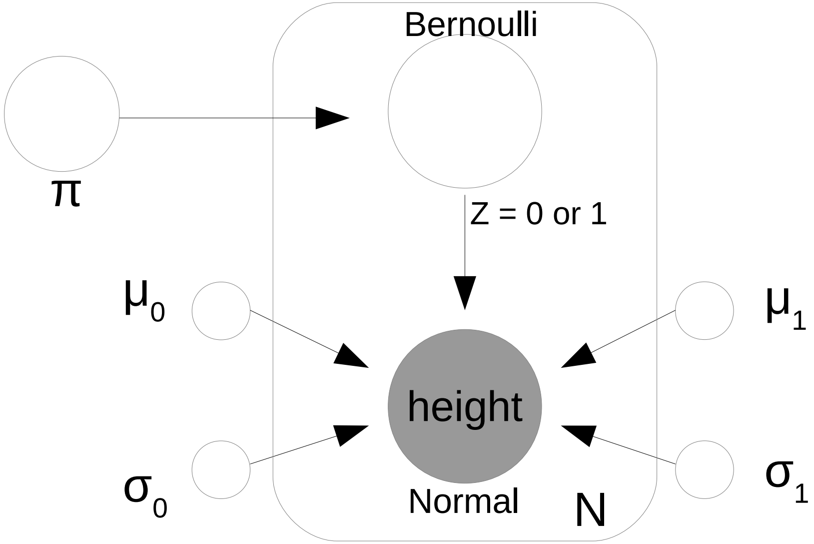
Z = 0 or 1 is drawn from a Bernoulli with parameter π.
Then the height is drawn from either of 2 Normal, depending on Z

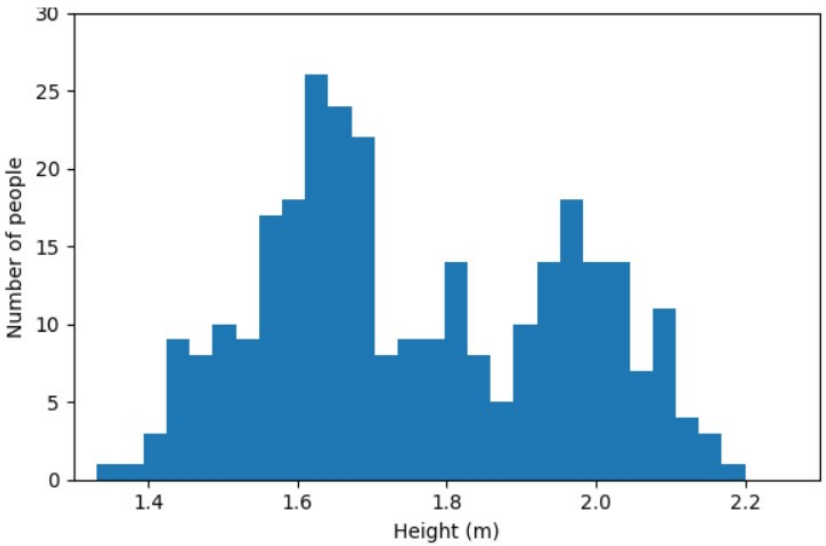


For instance, model the height of the persons in this room as coming from 2 populations

Graphical Models

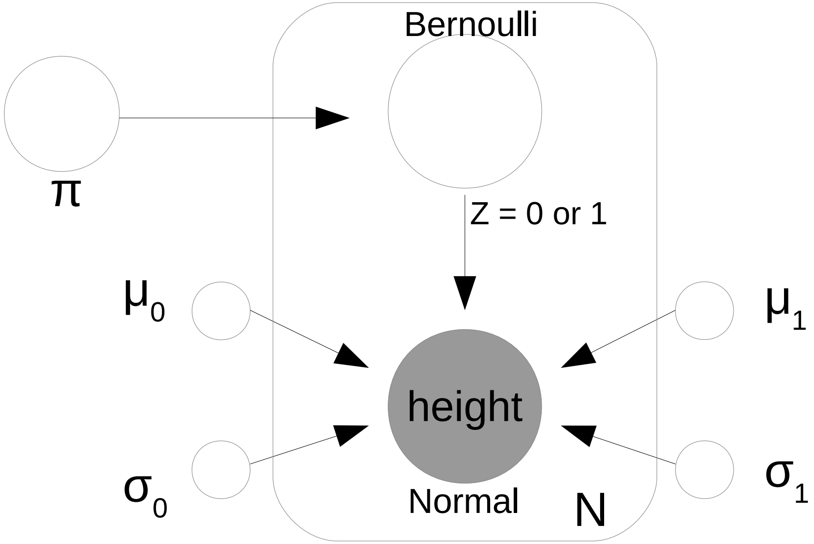
Model that data is sampled from this given PDF
and compute $P(X|\theta) = P(X \mid \mu_0 \sigma_0 \mu_1 \sigma_1 \pi)$

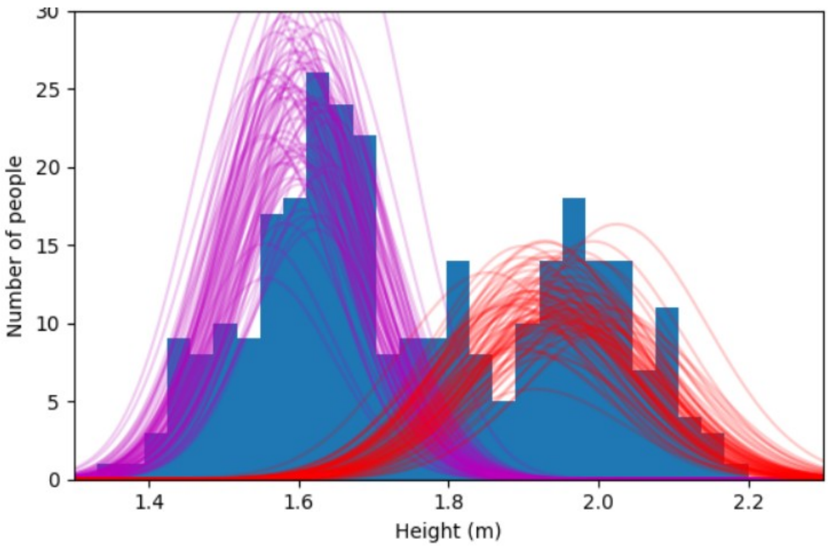




Graphical Models

Use Bayesian techniques to obtain P(θ|X)

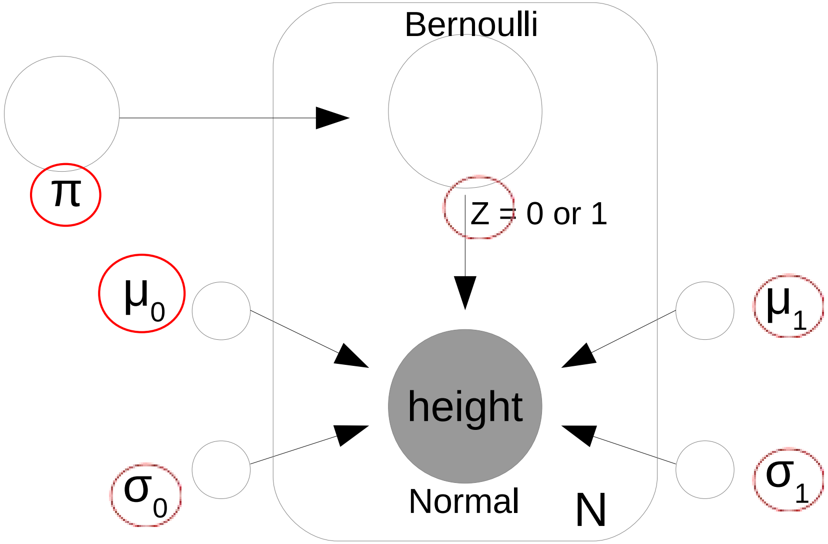


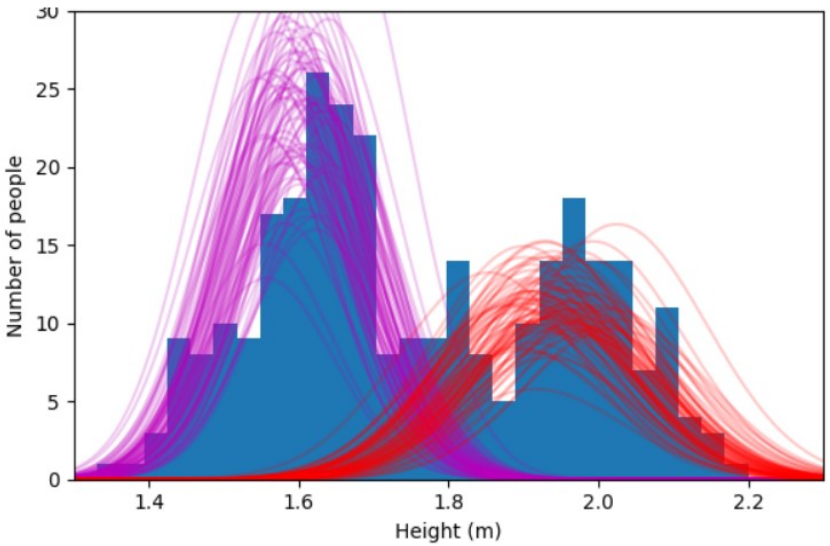


Graphical Models

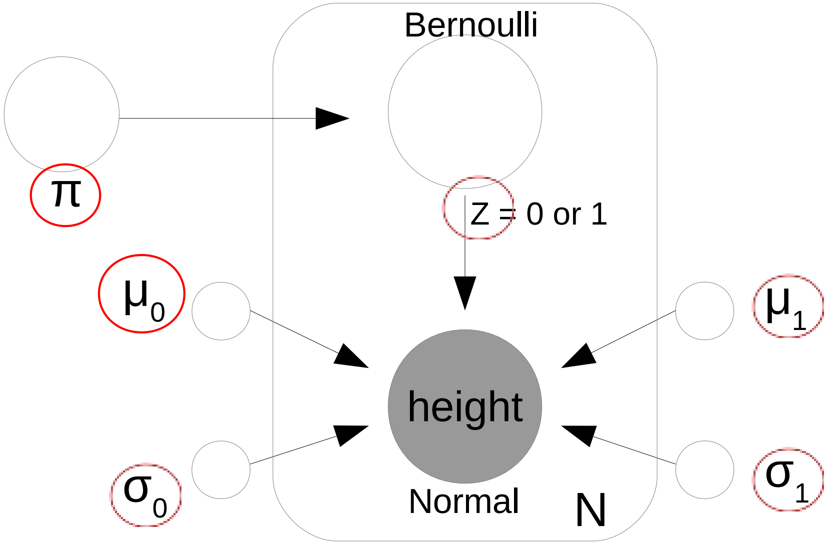
Use Bayesian techniques to obtain P(θ|X)

The posterior is a distribution
Over all latent variables of the model





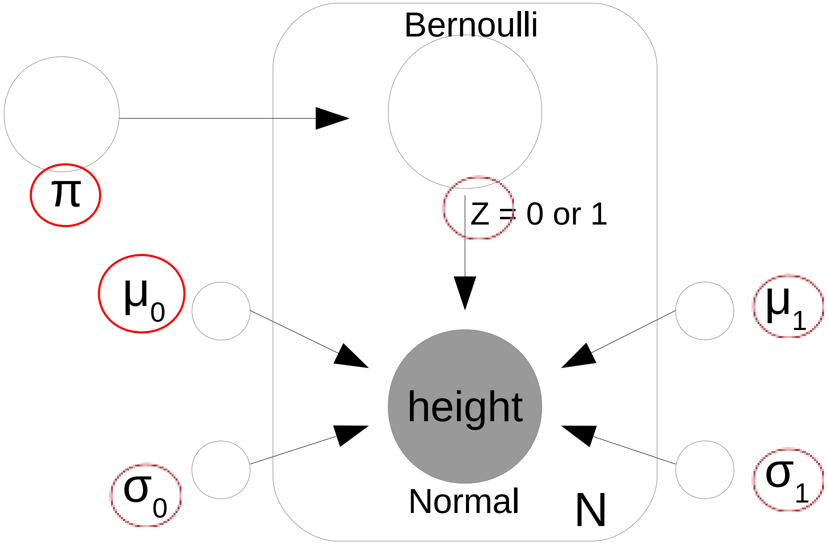
Graphical Models



Few remarks

- Access the internal structure of the data

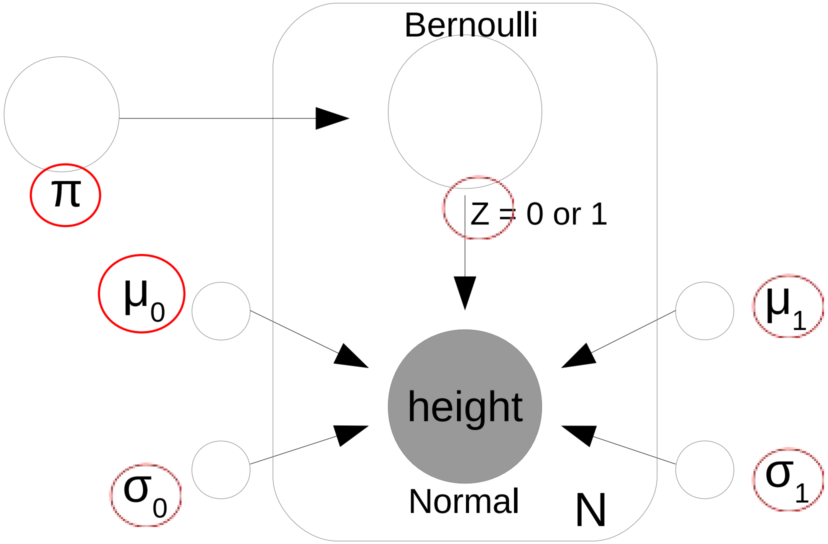
Graphical Models



Few remarks

- Access the internal structure of the data
- Very complex data can be constructed from simple PDFs

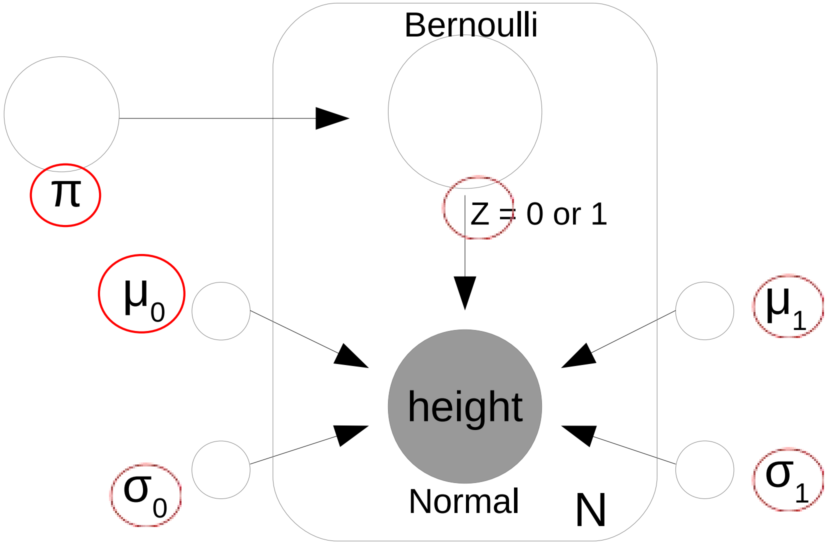
Graphical Models



Few remarks

- Access the internal structure of the data
- Very complex data can be constructed from simple PDFs
- Identify many signals just by using some prior knowledge on their shape

Graphical Models

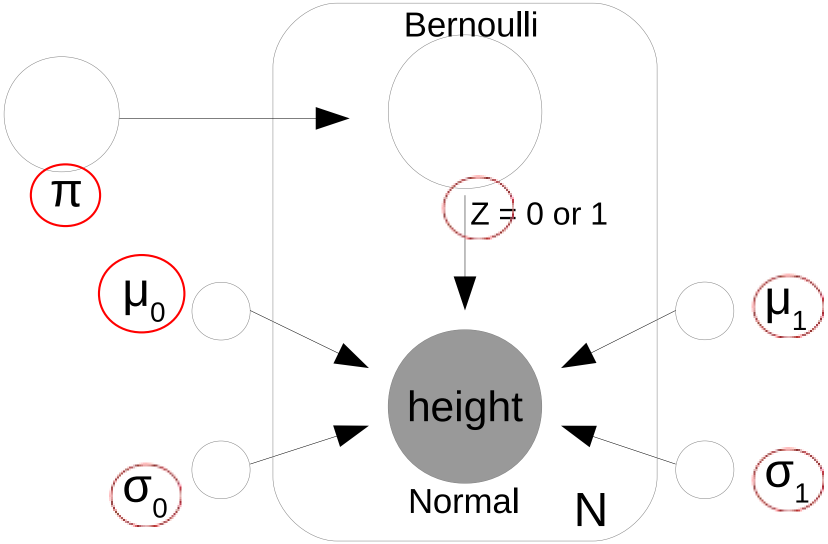


Few remarks

- Access the internal structure of the data
- Very complex data can be constructed from simple PDFs
- Identify many signals just by using some prior knowledge on their shape

This is all in signal region
(you don’t need control region!)

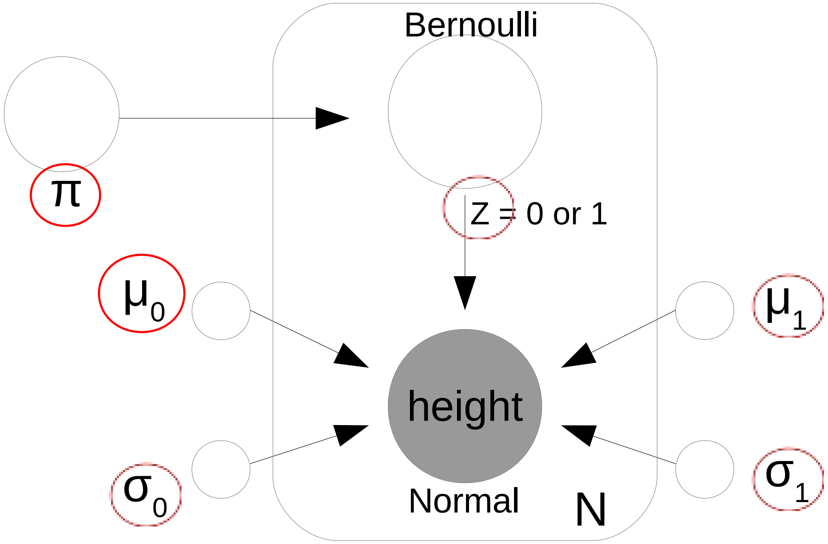
Graphical Models



Few remarks

- Access the internal structure of the data
- Very complex data can be constructed from simple PDFs
- Identify many signals just by using some prior knowledge on their shape
- Parameters pursue maximization of the probability of the data
- Recent numerical techniques, such as Stochastic Variational Inference, or Black Box Inference, etc.
- If you can construct $P(X|\theta)$, you're all set
- They are like Feynman Diagrams in Statistics

Graphical Models



Few remarks

This happens
in collider physics
much more often
than what we think!

- If you can construct P(X|θ), you're all set
- They are like Feynman Diagrams in Statistics

Applications:

quark- Vs gluon-jet

Four tops

$hh \rightarrow b b \gamma\gamma$

$hh \rightarrow bbbb$

Applications:

quark- Vs gluon-jet	Shapes
Four tops	Correlation
$hh \rightarrow bb\gamma\gamma$	Shapes + correlation
$hh \rightarrow bbbb$ (In preparation)	Shapes (arbitrary) + many correlations

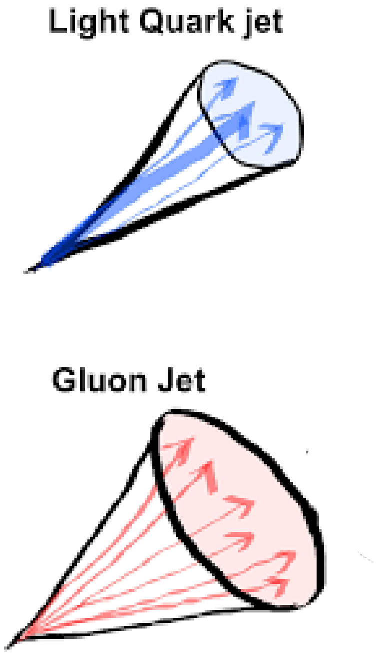
Applications:

quark- Vs gluon-jet

2112.11352
E.Alvarez
M.Spannowsky
M.Szewc

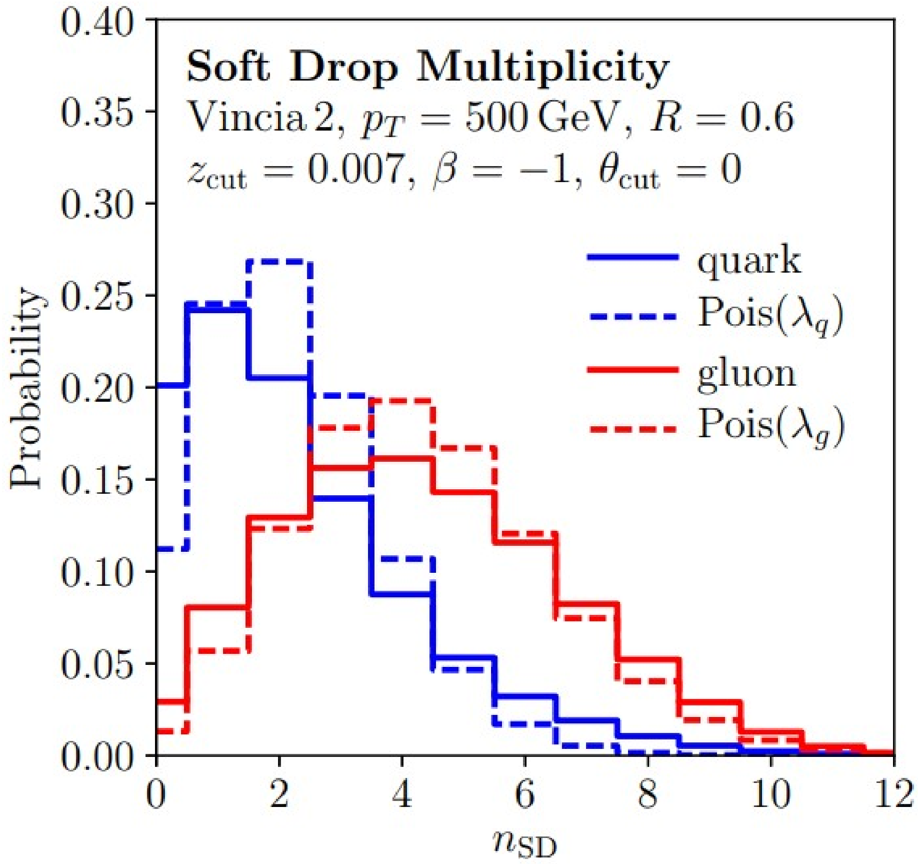
Quark and gluon jet

Quark and gluon jet

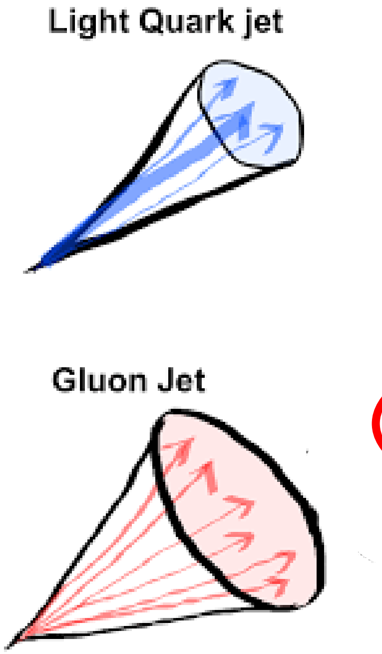


SoftDrop (n_{SD}) is an integer number for any jet. At leading-log:

$n_{SD} \sim \text{Poisson}(\lambda_{q,g})$



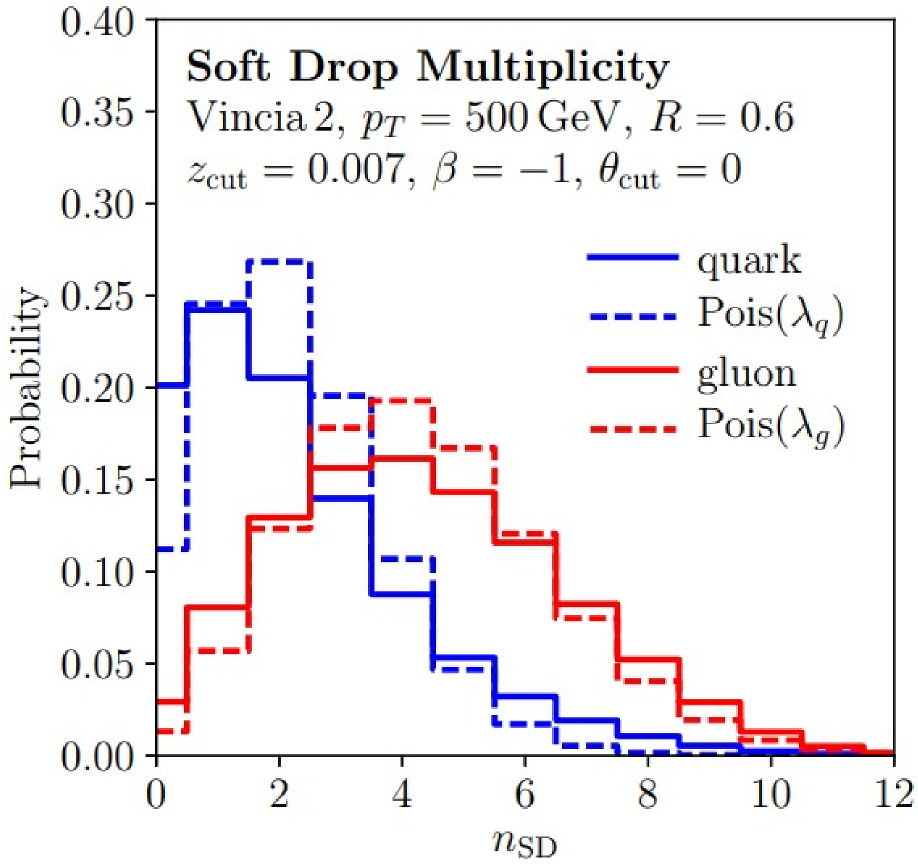
Quark and gluon jet



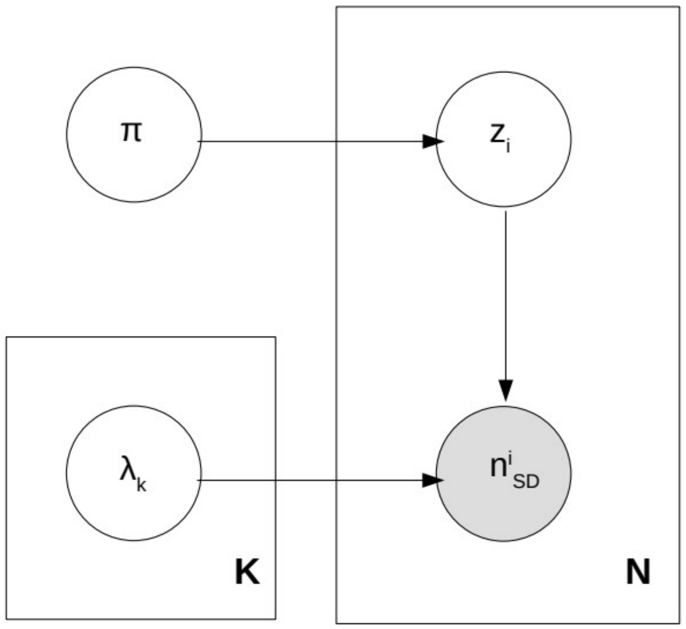
SoftDrop (n_{SD}) is an integer number for any jet. At leading-log:

n_{SD} ~ Poisson($\lambda_{q,g}$)

Very well defined shape each class!

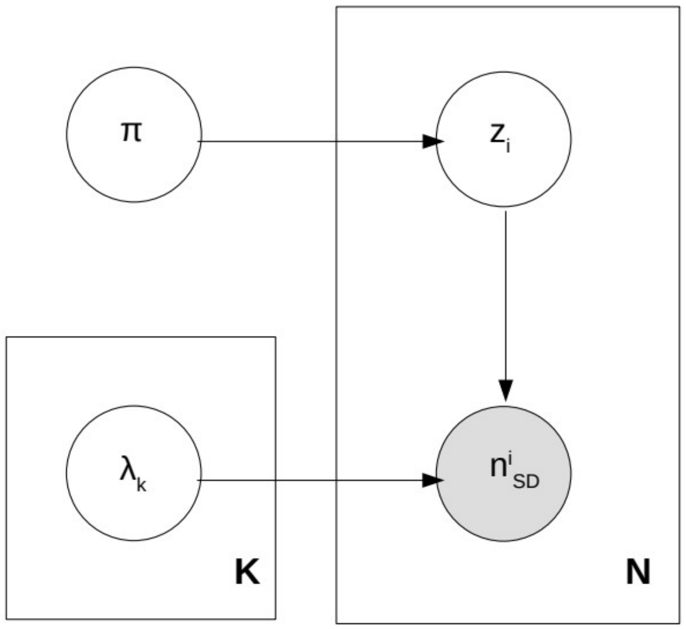


Quark and gluon jet



Graphical Model
(or PDF)

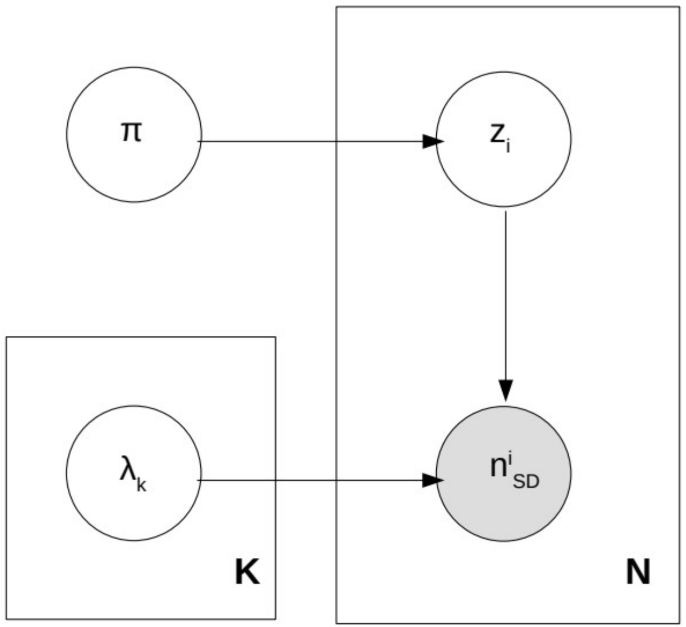
Quark and gluon jet



Graphical Model
(or PDF)

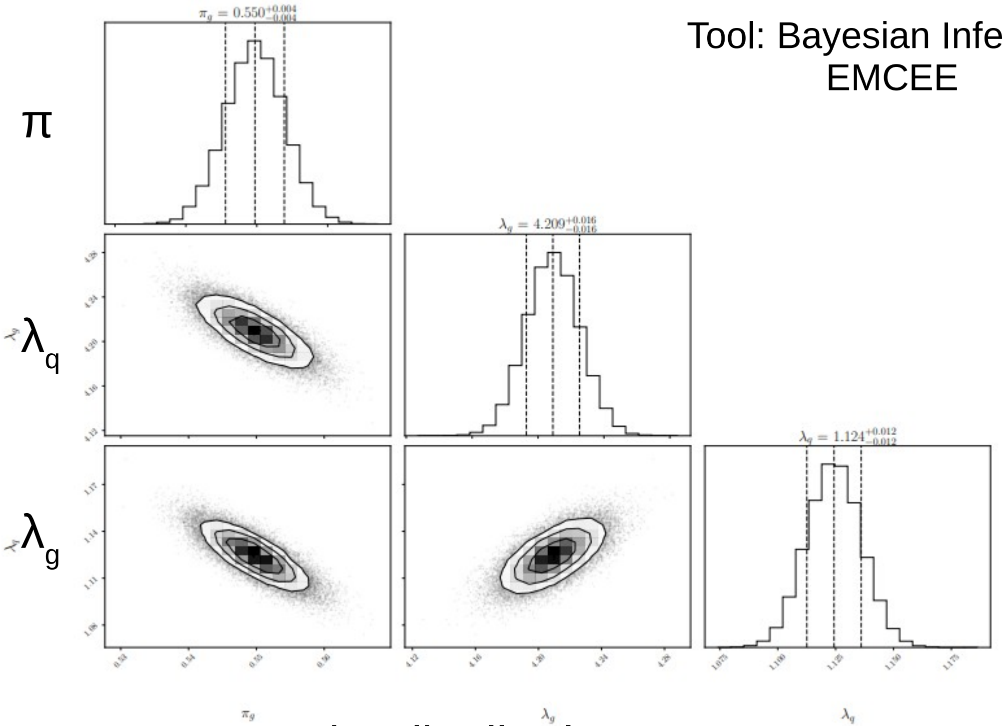
Get n_{SD} from a
simulated sample
using Pythia and/or Hergiw

Quark and gluon jet



Graphical Model
(or PDF)

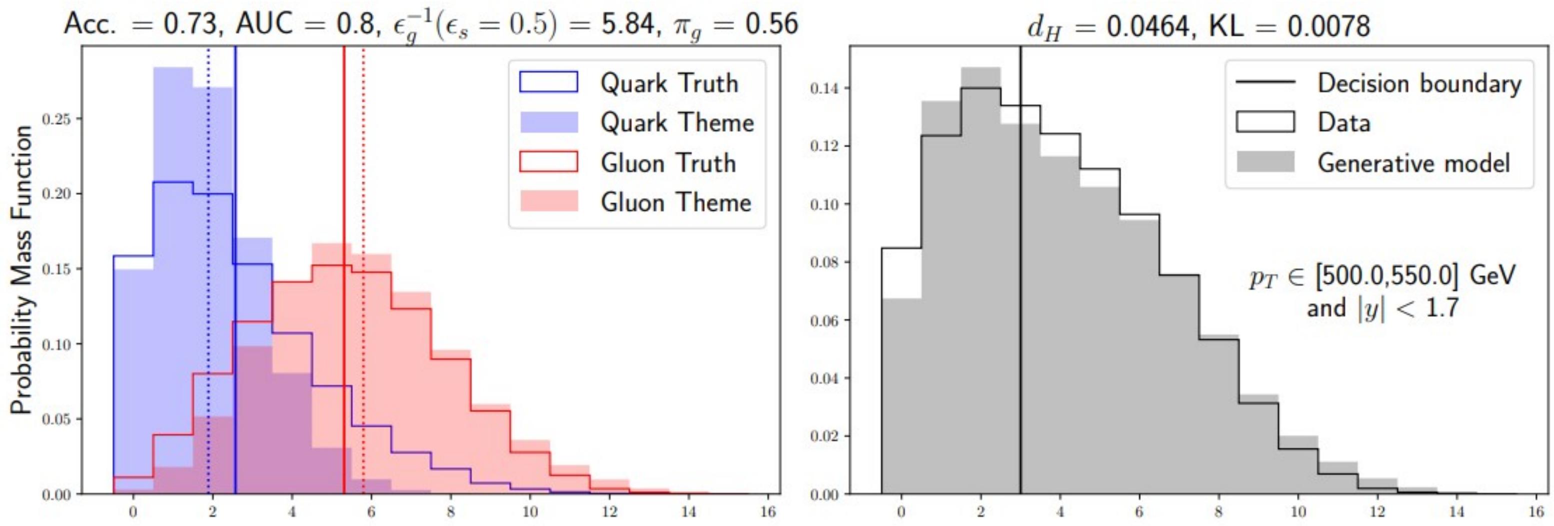
Tool: Bayesian Inference
EMCEE



Extract a posterior distribution over parameters
P(θ|X)

Quark and gluon jet

Results:

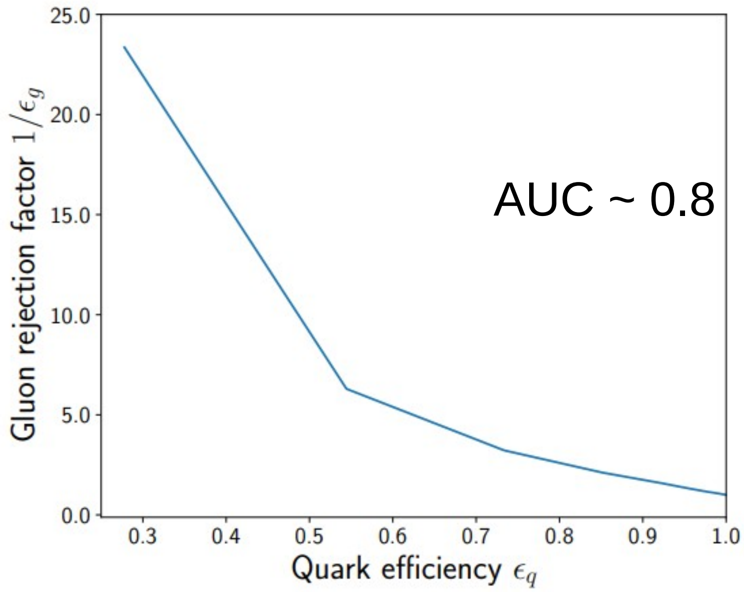


Tool: stochastic Variational Inference

Quark and gluon jet

Results:

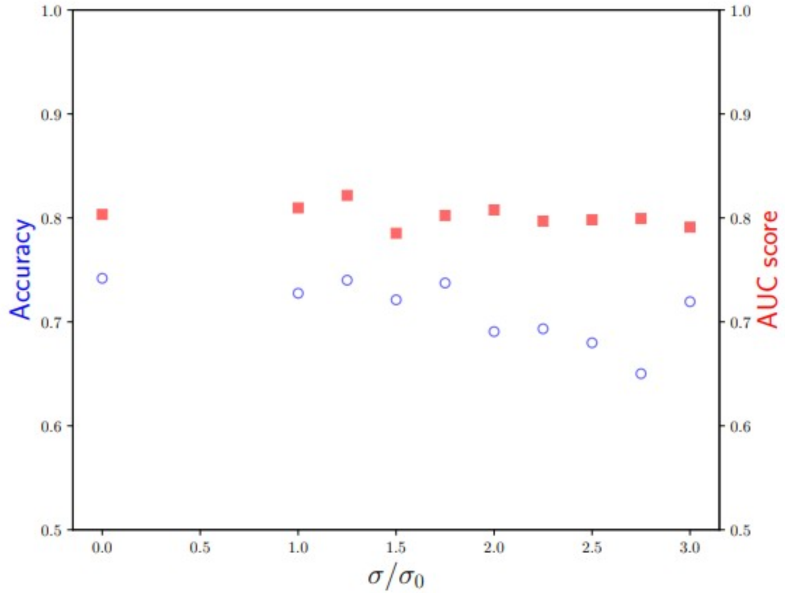
Robust to simple detector effects



AUC ~ 0.8

Accuracy ~ 0.71

Fully unsupervised

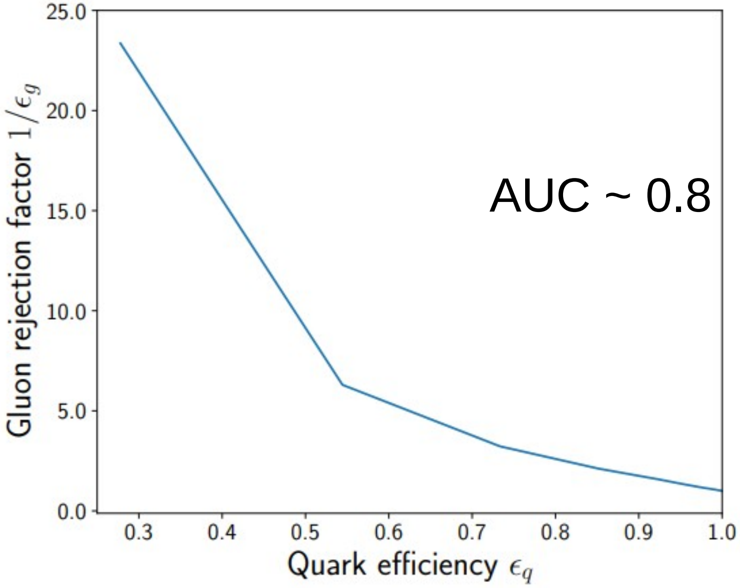


Smearing η and φ with a N(0,σ)

Quark and gluon jet

2212.10493
A.Butter
B.Dillon
T.Plehn
L.Vogel

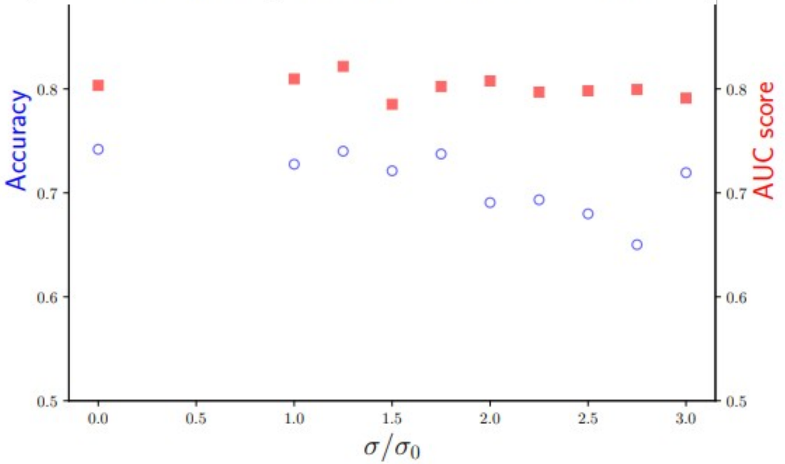
Results:



Accuracy ~ 0.71

Fully unsupervised

AUC		training		
		PYTHIA	SHERPA	HERWIG
testing	PYTHIA	0.900	0.893	0.886
	SHERPA	0.856	0.863	0.853
	HERWIG	0.804	0.820	0.833



Smearing η and φ with a N(0,σ)

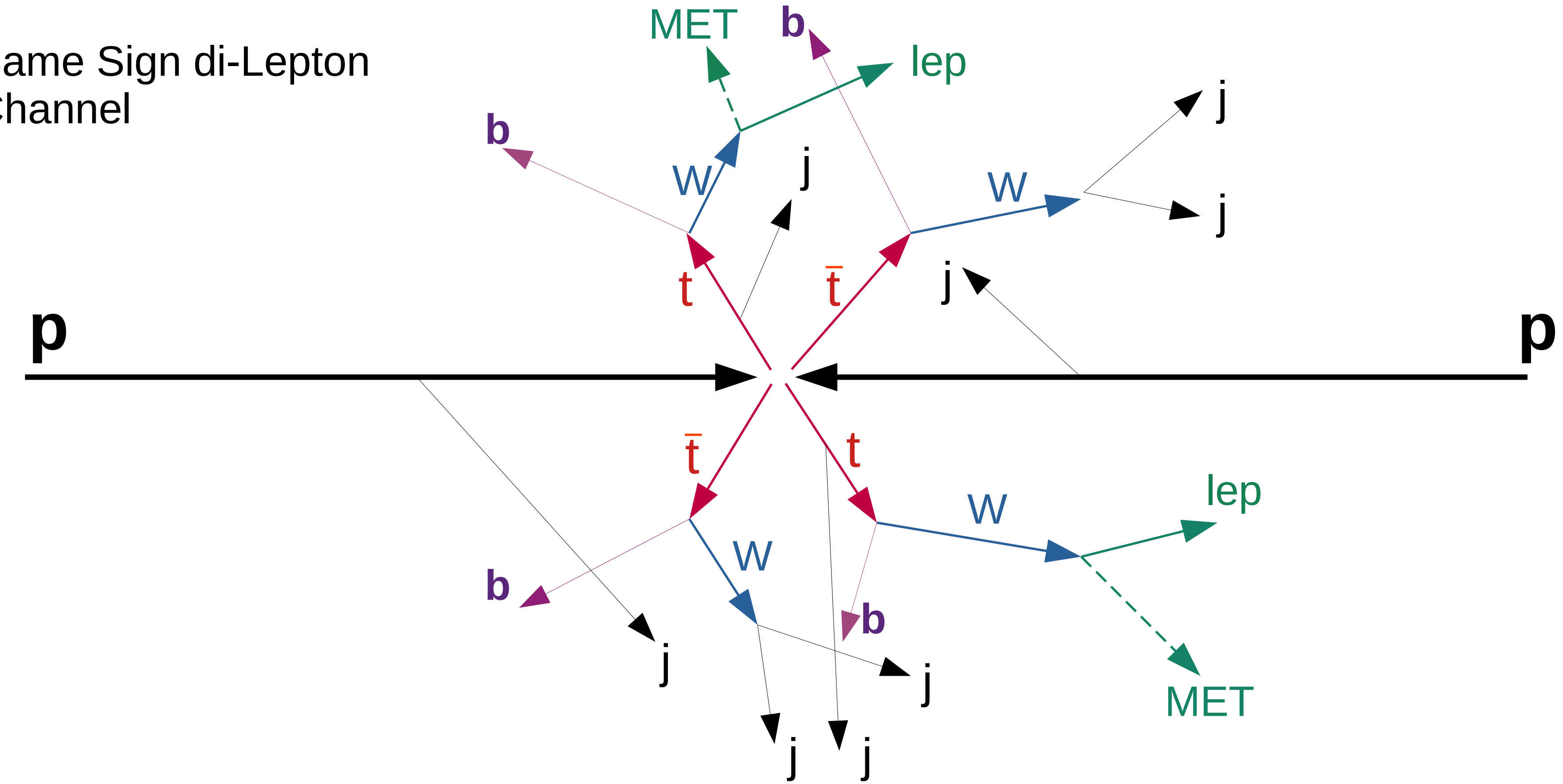
Applications:

Four tops

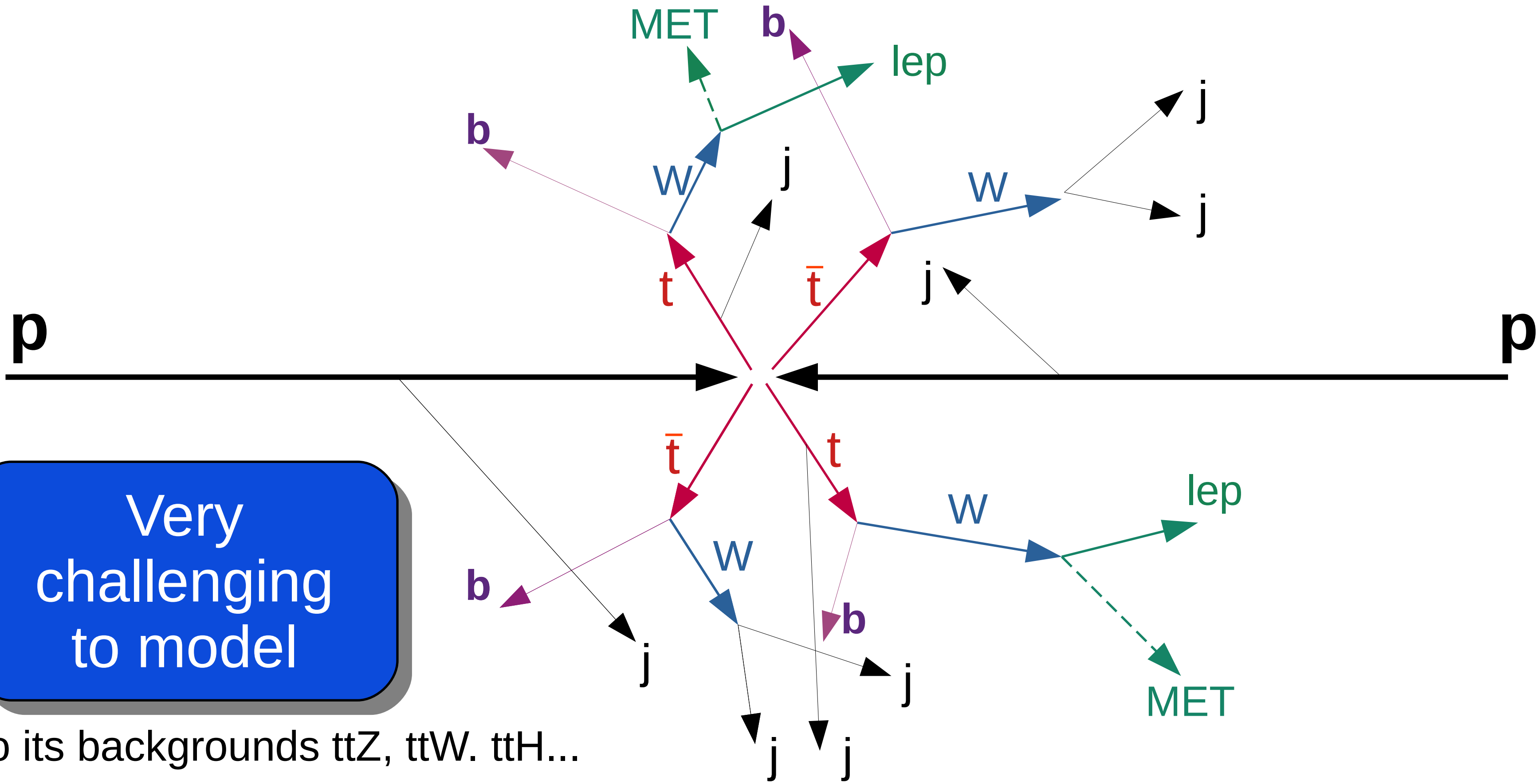
2107.00668
E.Alvarez
B.Dillon
D.Faroughy
J.Kamenik
F.Lamagna
M.Szewc

Four tops

Same Sign di-Lepton Channel



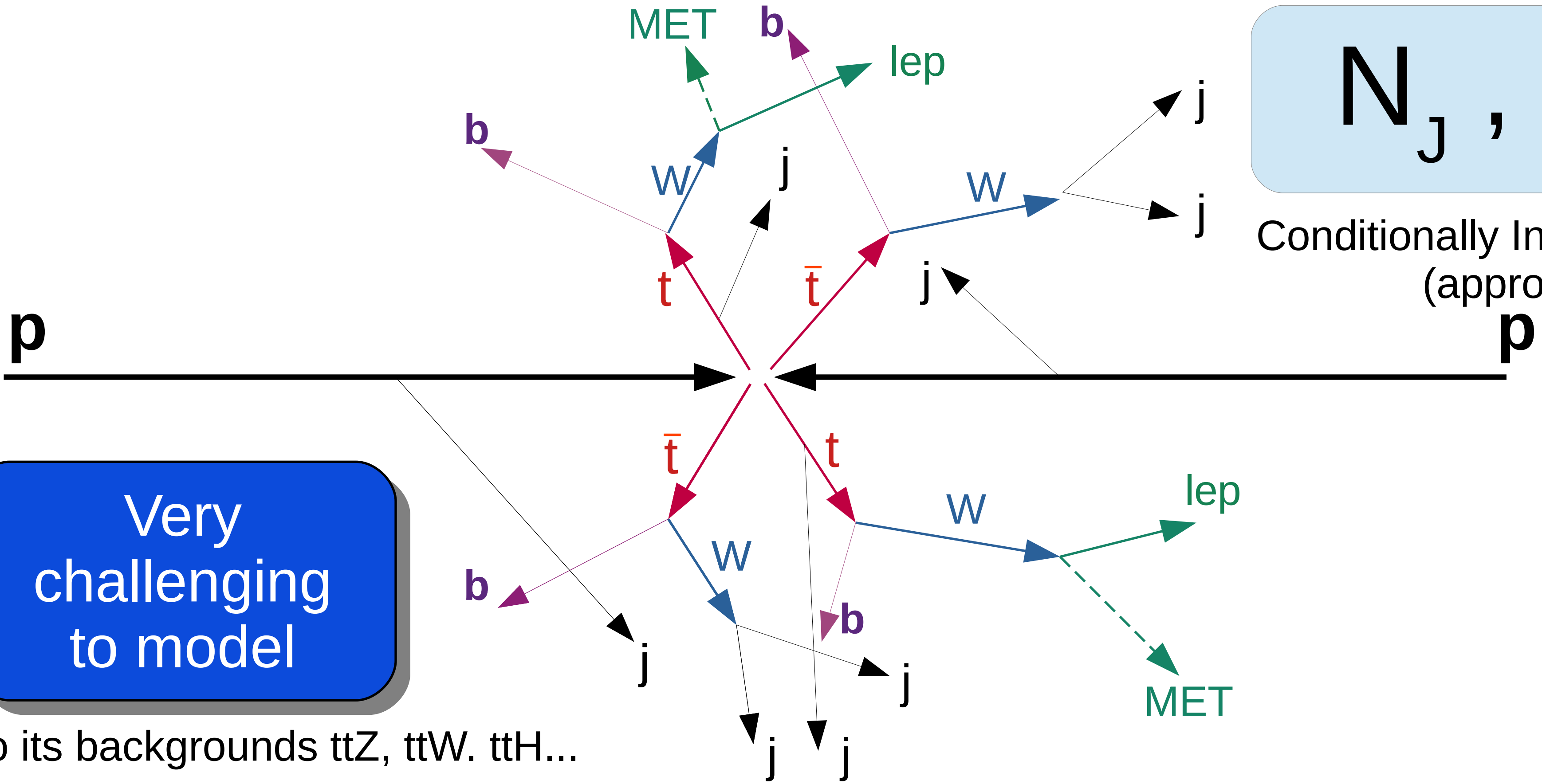
Four tops



Very challenging to model

Also its backgrounds ttZ, ttW. ttH...

Four tops



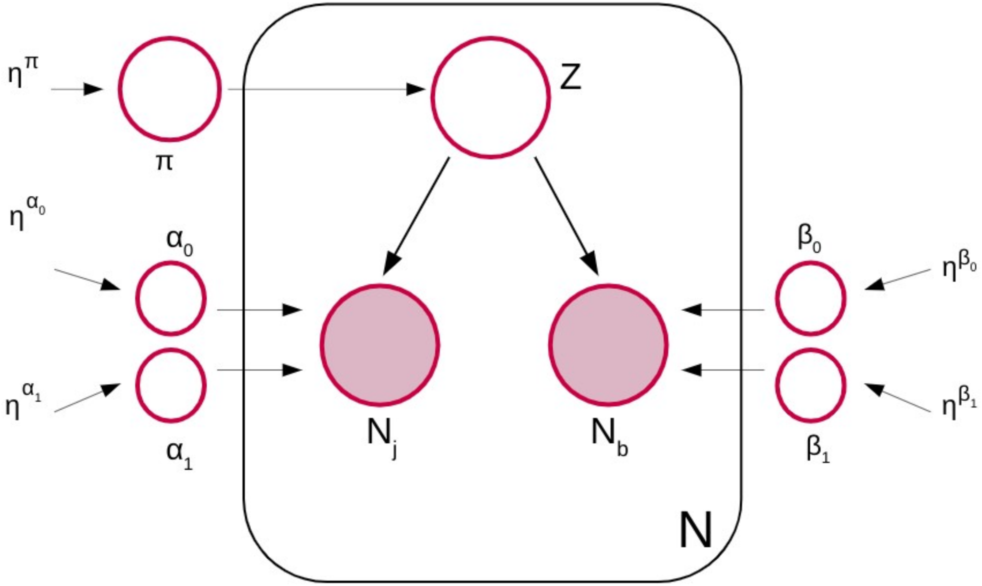
N_j , N_b

Conditionally Independent (approx)

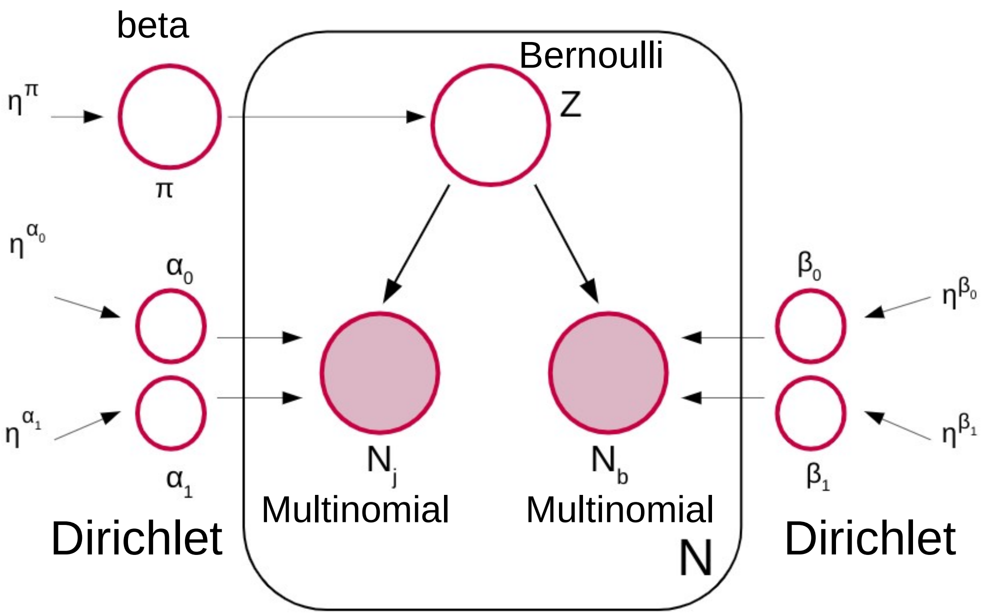
Very challenging to model

Also its backgrounds ttZ, ttW. ttH...

Four tops

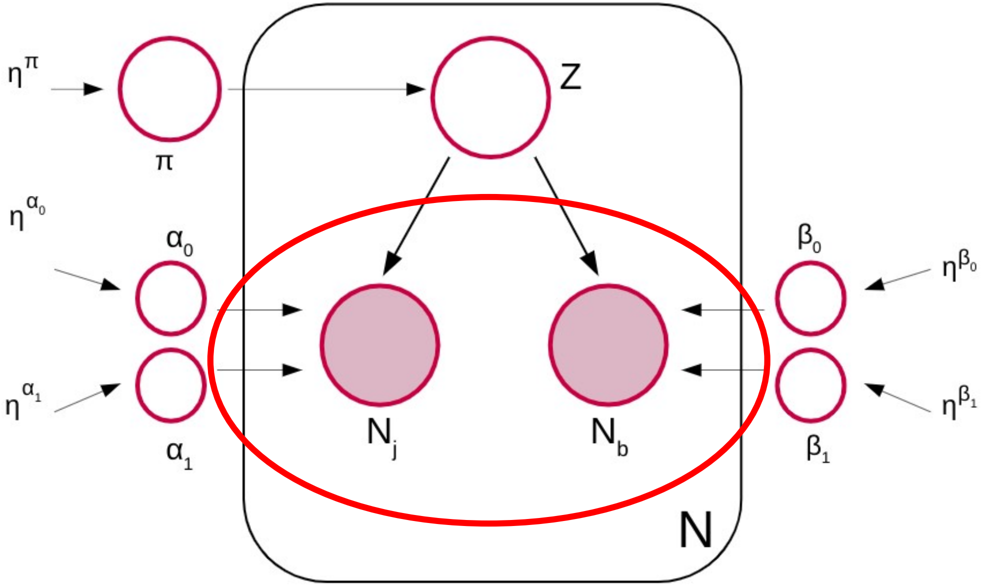

η^{π}
π
Z
η^{α_0}
α_0
η^{α_1}
α_1
N_j
N_b
β_0
η^{β_0}
β_1
η^{β_1}
N

Four tops



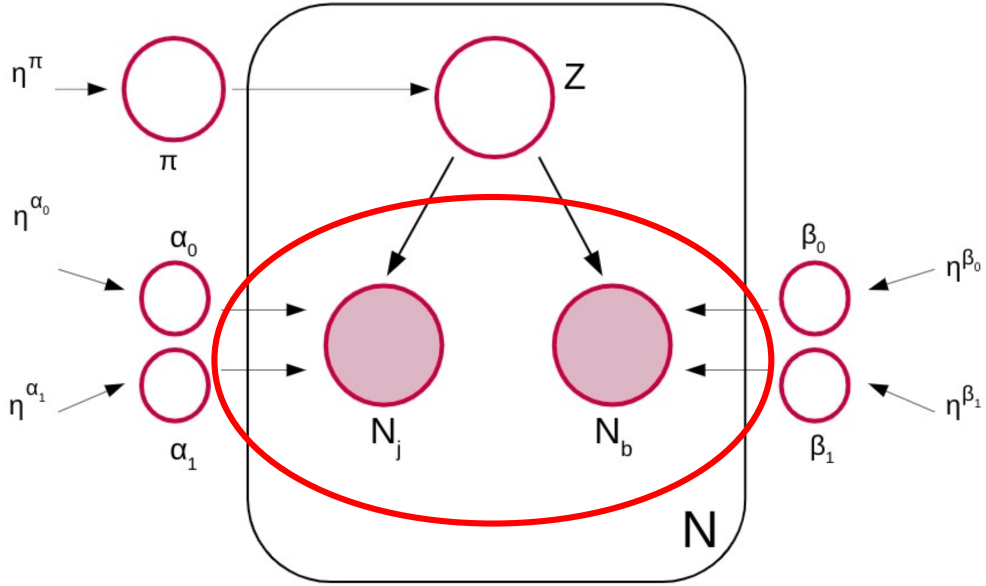
beta
Bernoulli
η^{π}
π
Z
η^{α_0}
α_0
η^{α_1}
α_1
β_0
η^{β_0}
β_1
η^{β_1}
N_j
Multinomial
N_b
Multinomial
N
Dirichlet
Dirichlet

Four tops



Multinomials are too flexible,
but N_J - N_b correlation fixes the issue

Four tops



Multinomials are too flexible,
but N_J - N_b correlation fixes the issue

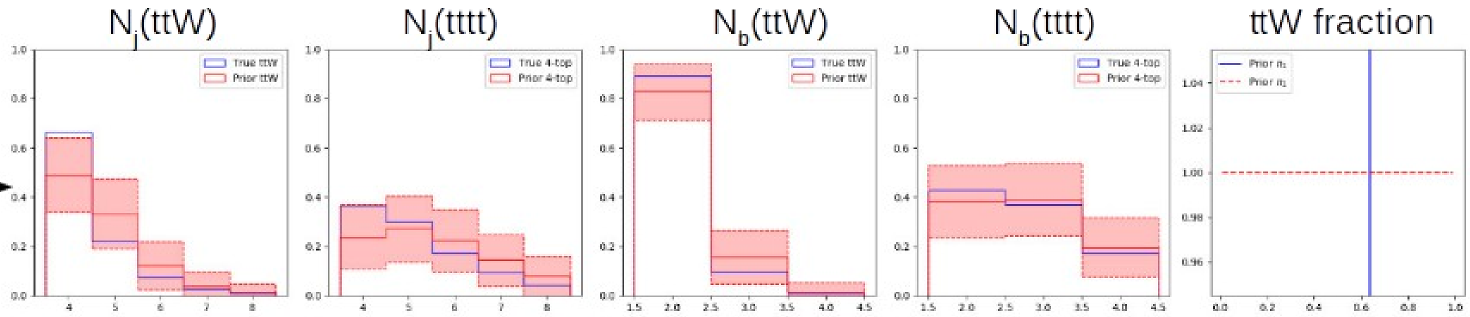
Signal (tttt) expects larger N_J and N_b

Background (ttW) expects smaller N_J and N_b

Four tops

Results:

Prior(θ)

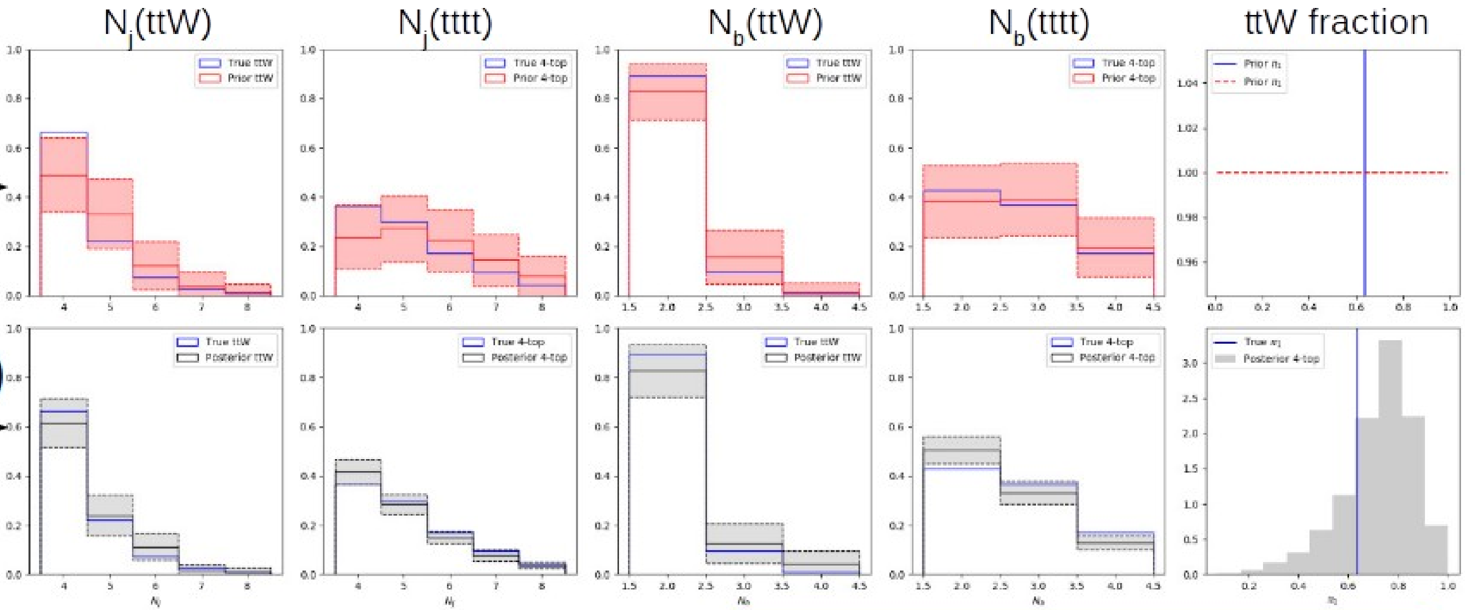


Four tops

Results:

Prior(θ)

Posterior(θ|X)

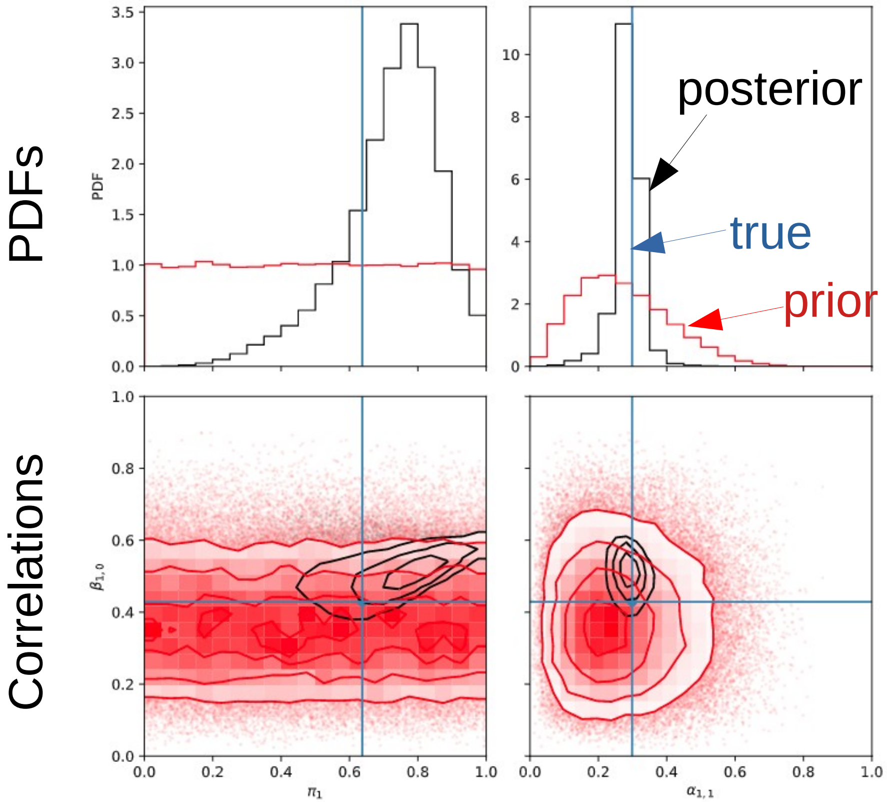


(500 fb^{-1})

$$P(\theta|X) = \frac{P(X|\theta) \times \text{Prior}(\theta)}{P(X)}$$

Inference correctly approaches true Values!

Four tops



Each parameter approaches the true values with the posterior!

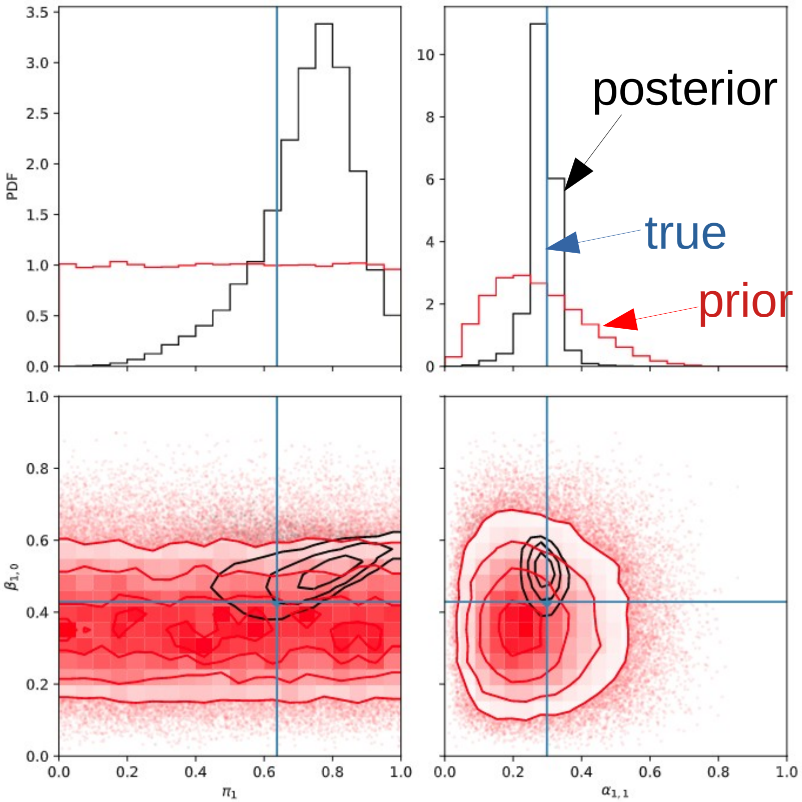
Excerpt from Corner-plot panels

(500 fb^{-1})

Four tops

PDFs

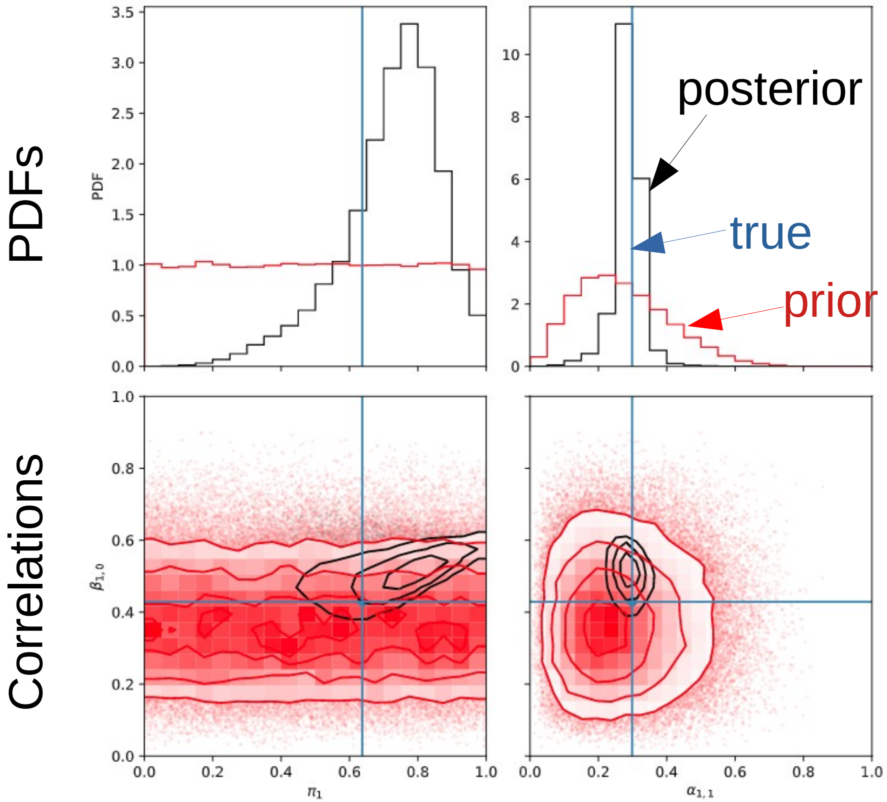
Correlations



Four tops has large discrepancies between data and MC.

We considerably reduce MC impact

Four tops



Four tops has large discrepancies between data and MC.
We considerably reduce MC impact

- N_J N_b at the event-by-event level
- Use prior info
- Bayesian Inference techniques

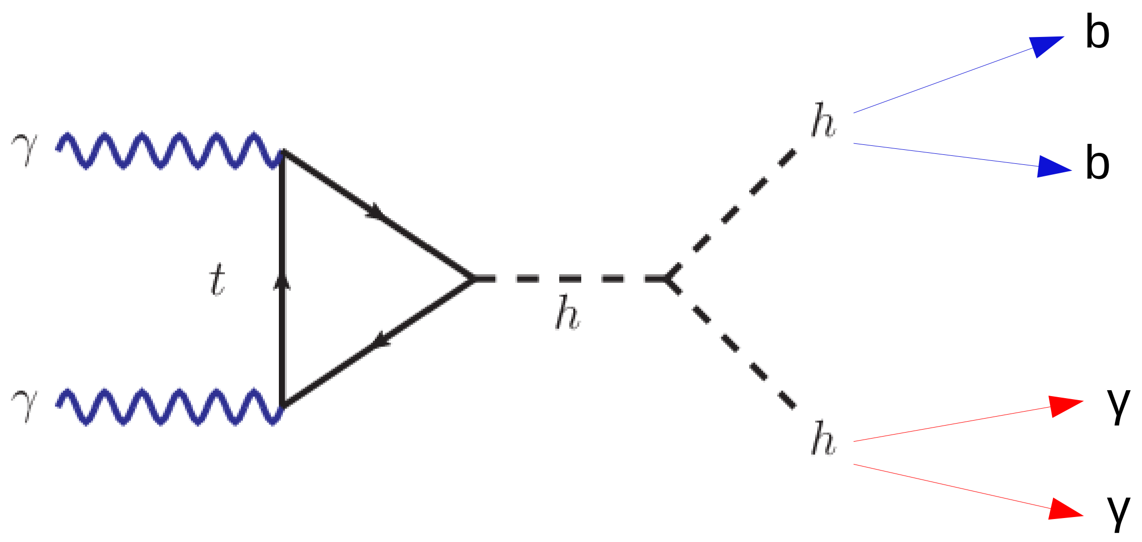
Applications:

Di-Higgs

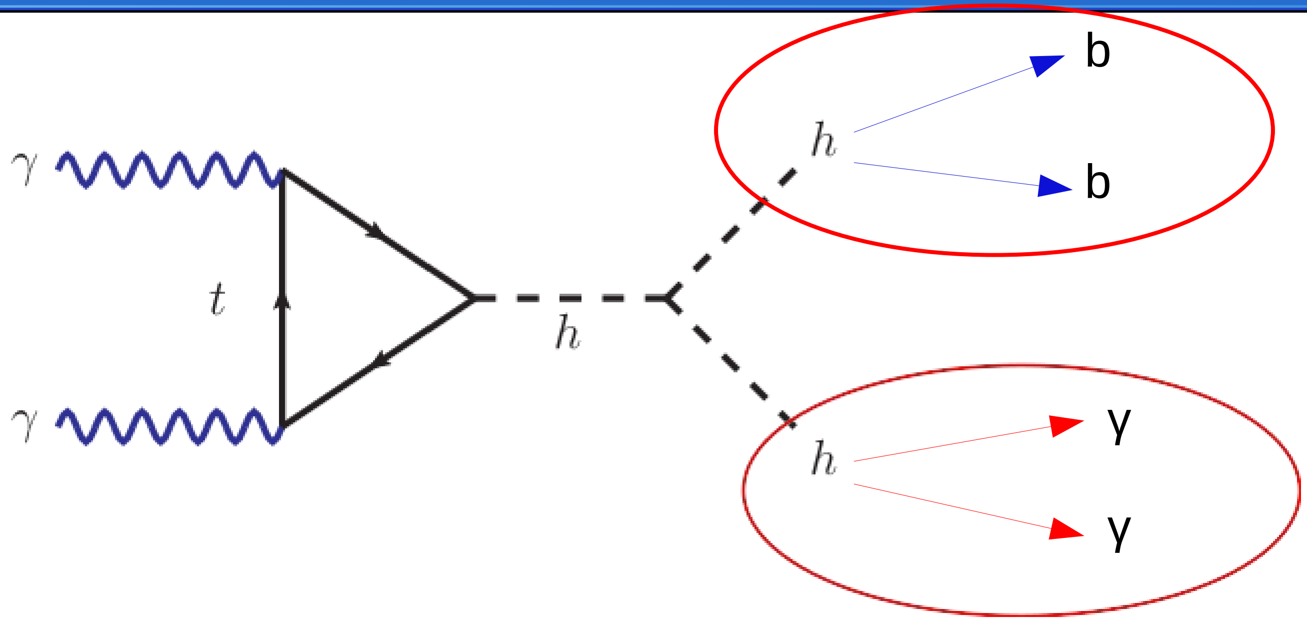
Simplified for the sake of the algorithm

2210.07358 (hh $\rightarrow$ bbγγ)
E.Alvarez
+ in preparation (hh $\rightarrow$ bbbb)
A.Alvarez,L. Da Rold,
S.Tanco, T.Tarutina,
M.Szewc, A.Szynkman

Di-Higgs: hh → bbγγ

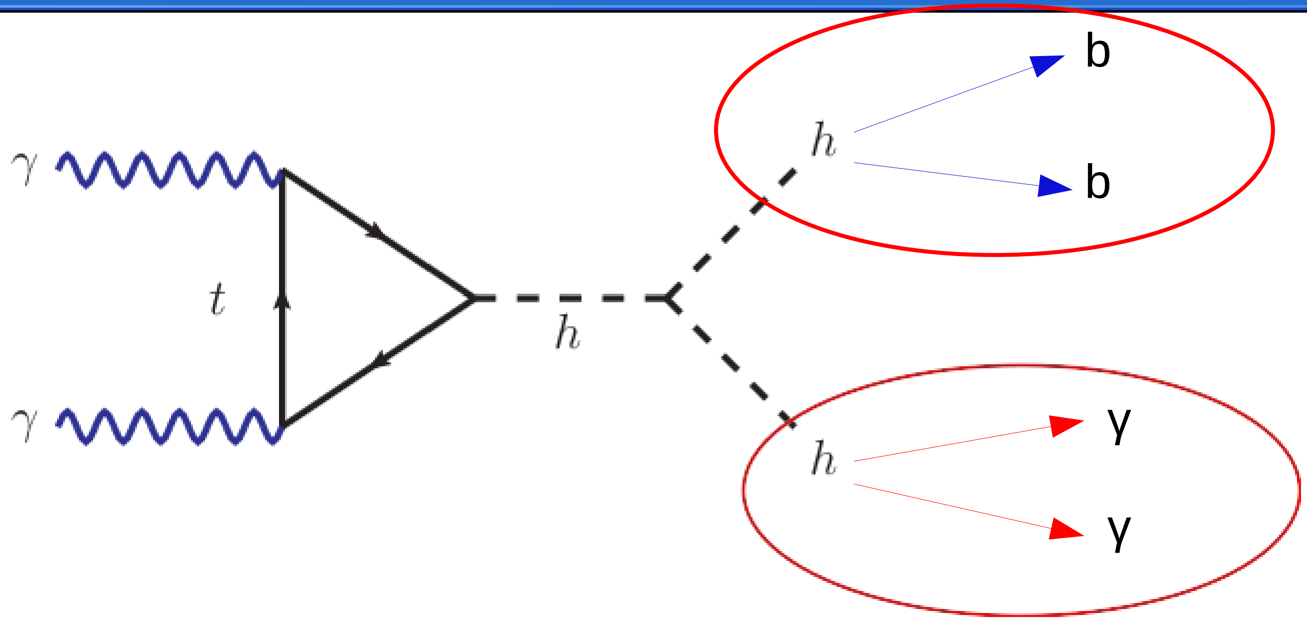


Di-Higgs: hh → bbγγ



Independent resonant decays!

Di-Higgs: hh → bbγγ

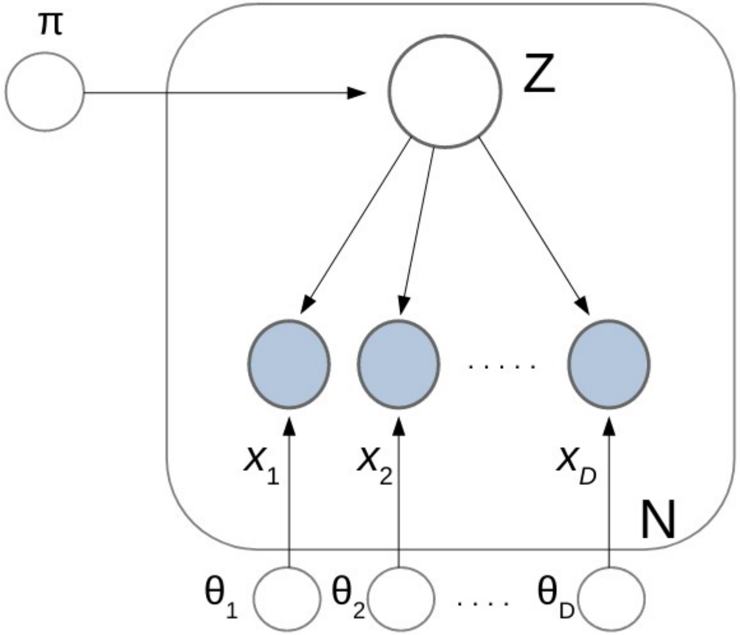


Independent resonant decays!

Versus continuum exponentially decaying background

(plus semi-resonant, and others)

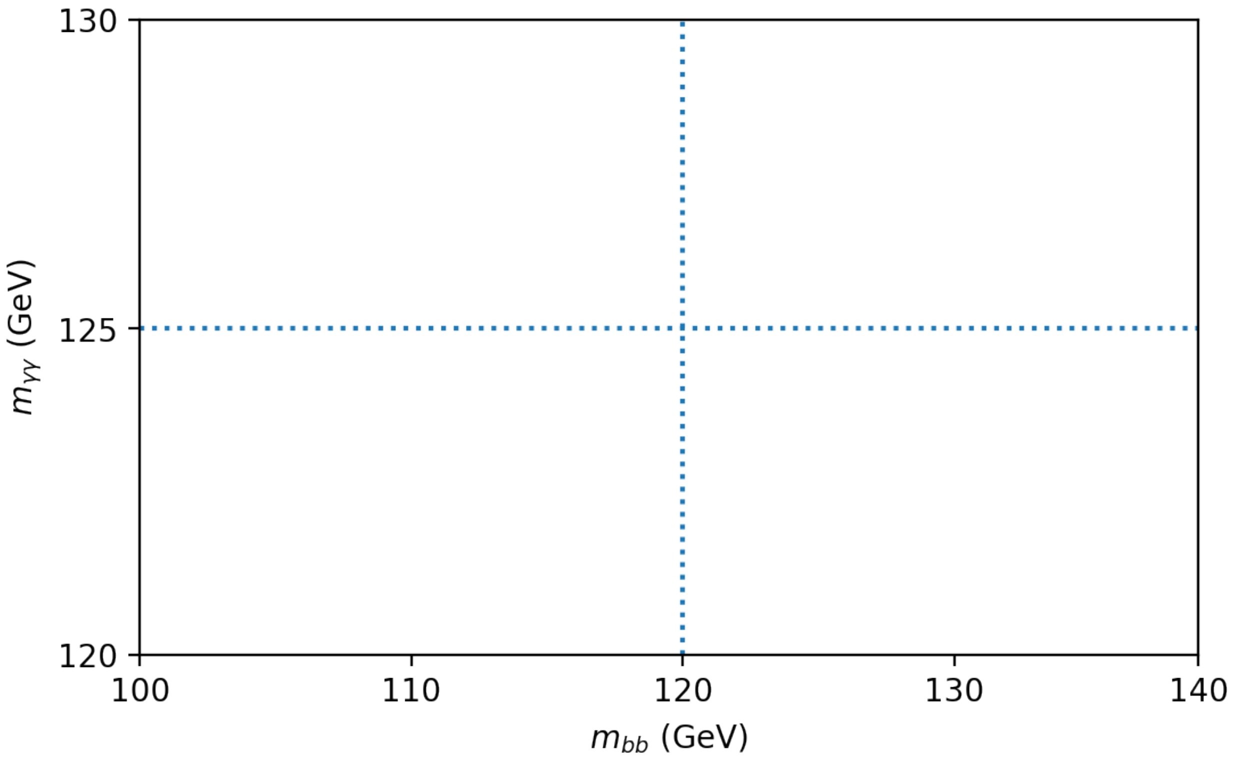
Di-Higgs: hh → bbγγ



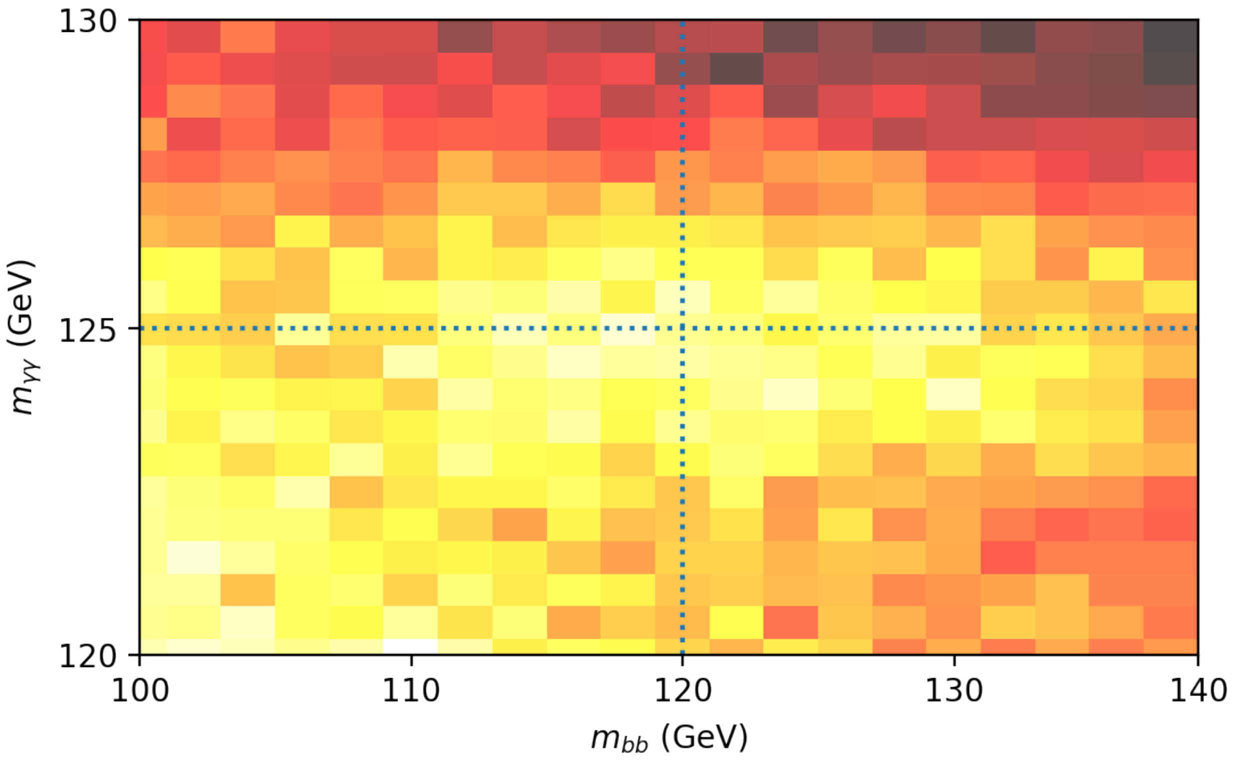
Observables are (approx) independent once they are conditioned on the class

m_{bb} and $m_{\gamma\gamma}$ correlation in the data is the key!

Di-Higgs: hh → bbγγ



Di-Higgs: hh → bbγγ

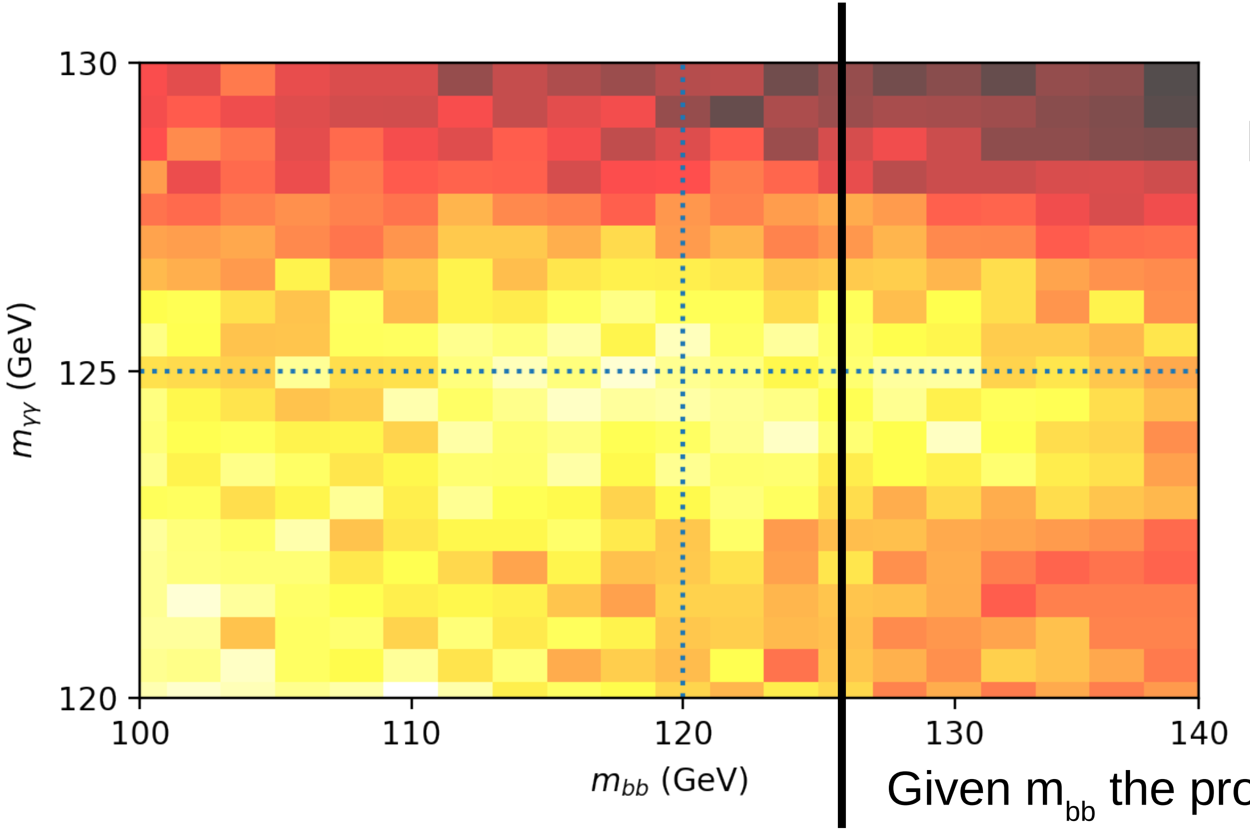


PDF

Signal (10%) + background

(MG5+Pythia+Delphes)

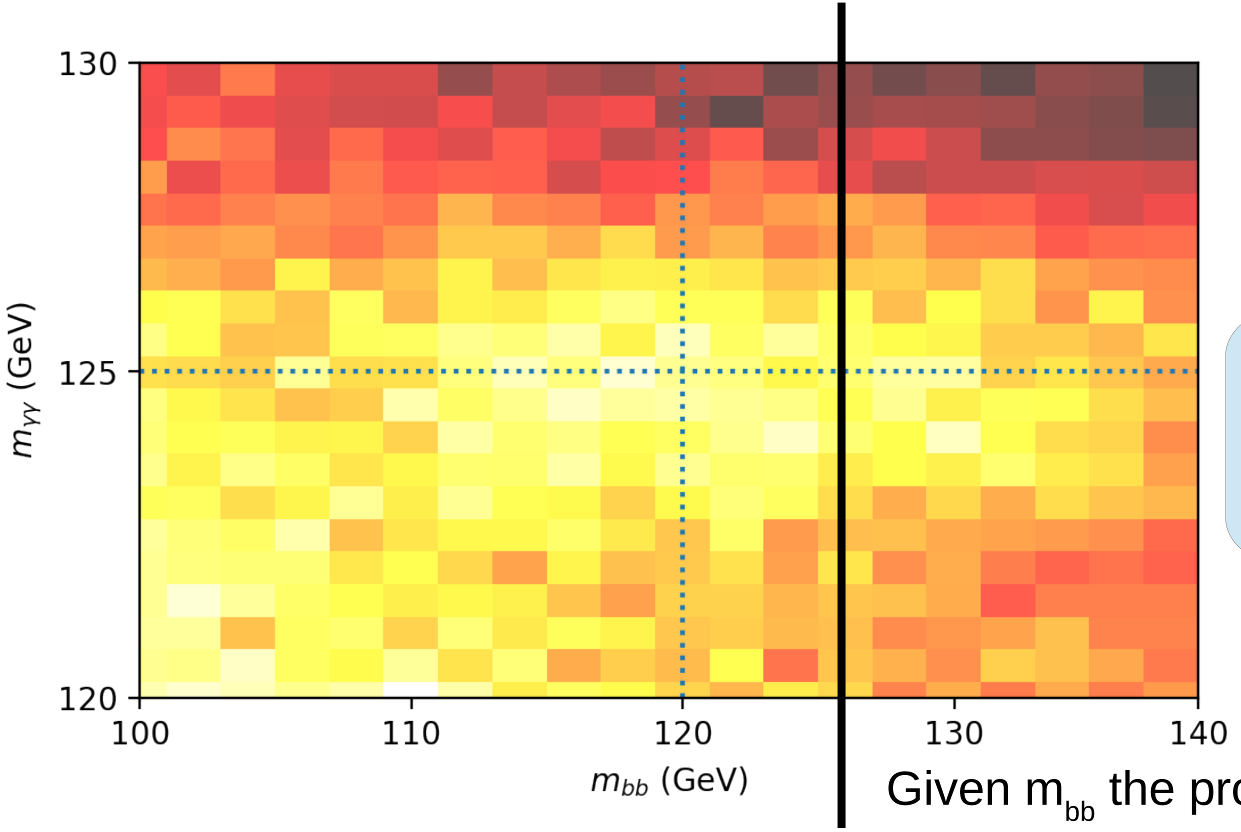
Di-Higgs: hh → bbγγ



PDF

Given m_{bb} the probability for $m_{\gamma\gamma}$ depends on all parameters, including the fraction

Di-Higgs: hh → bbγγ

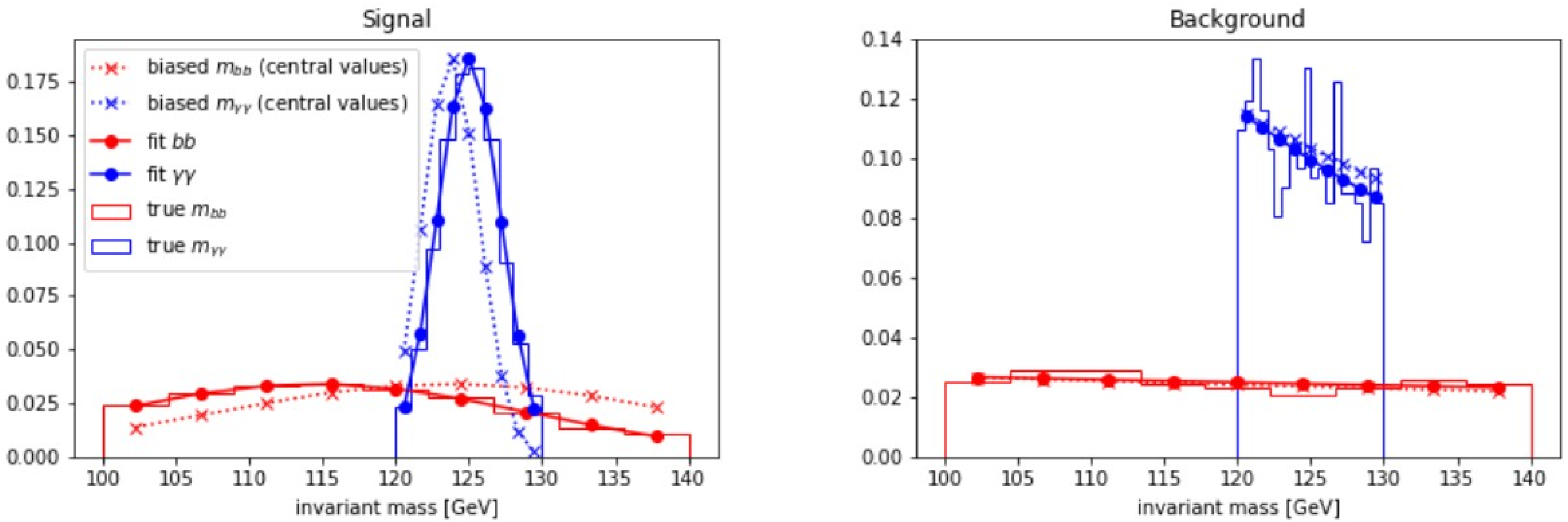


PDF

The Bayesian algorithm does all the *'thinking'* and extracts a posterior over all parameters

Given m_{bb} the probability for $m_{\gamma\gamma}$ depends on all parameters, including fraction

Di-Higgs: hh → bbγγ



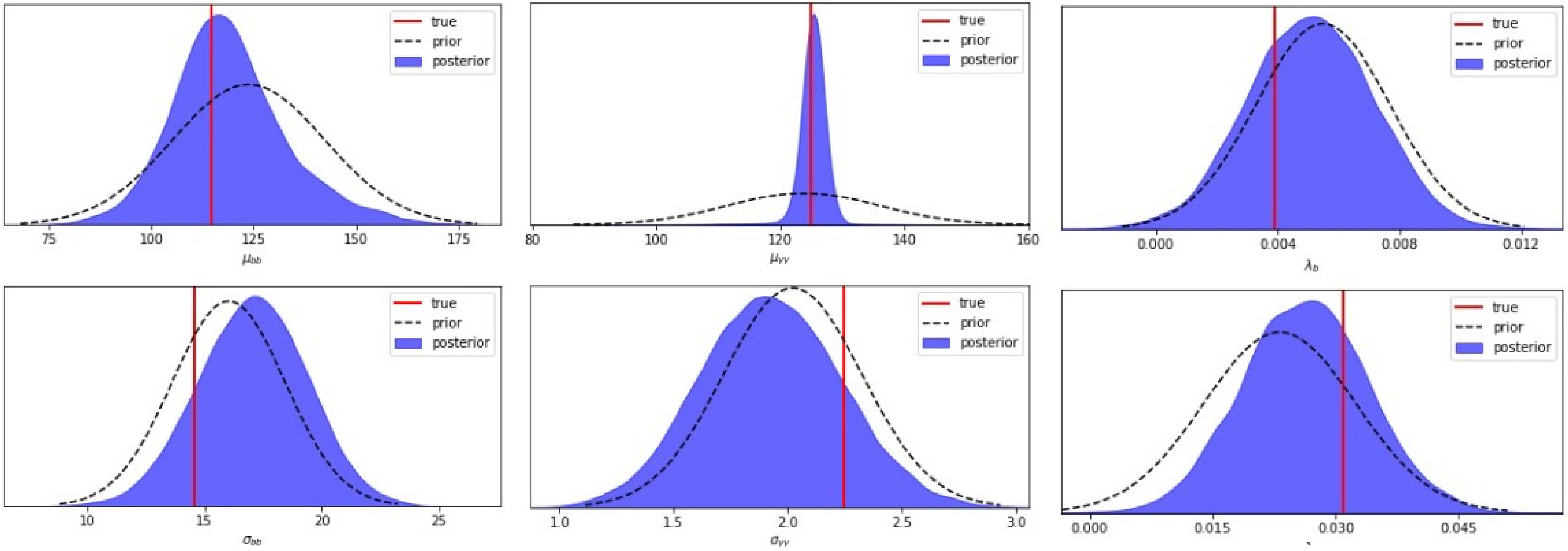
Generate 1k events (MG5+Pythia+Delphes).
Use a biased prior to emulate an inaccurate Montecarlo

Di-Higgs: hh → bbγγ

m_{bb} $m_{\gamma\gamma}$

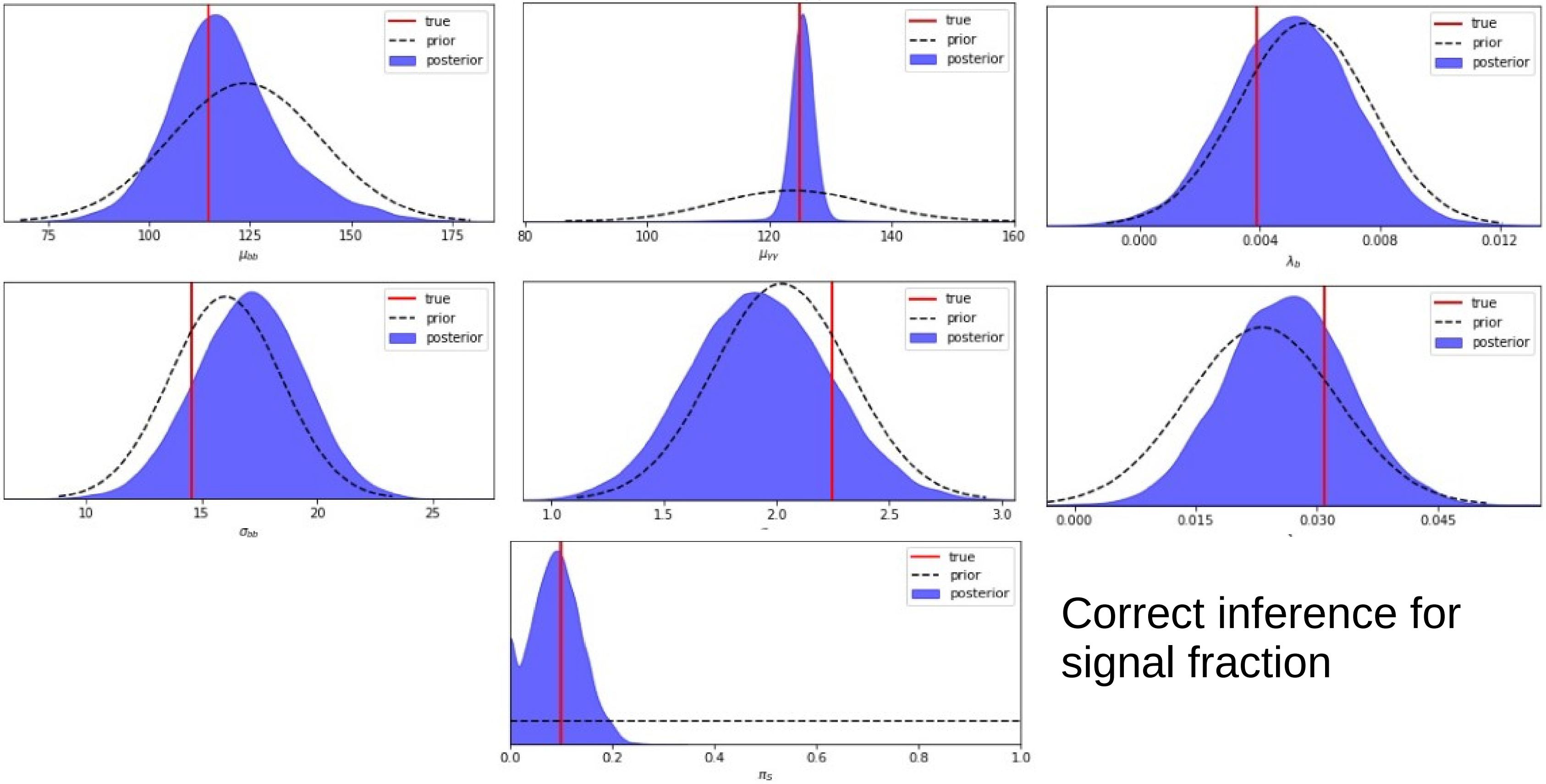
λ_b

λ_γ



Di-Higgs: hh → bbγγ

m_{bb} $m_{\gamma\gamma}$

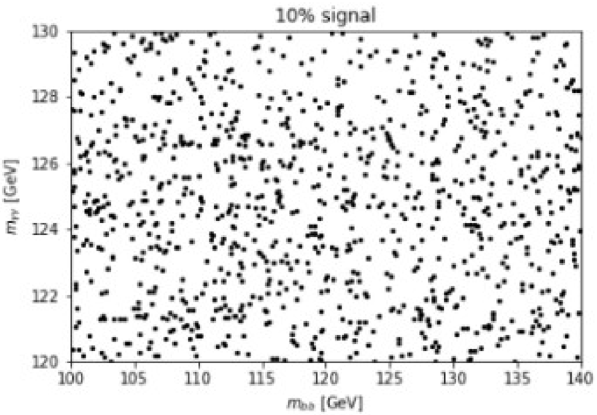


λ_b

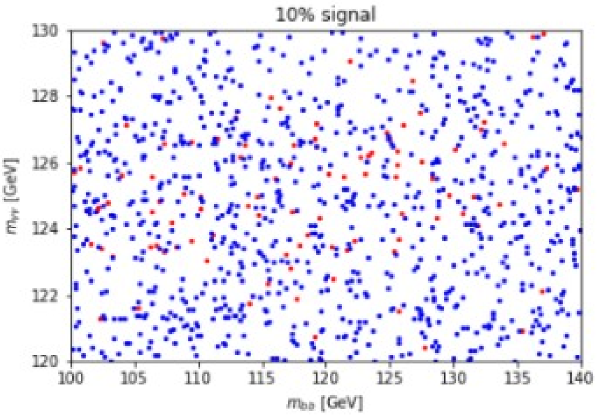
λ_γ

Correct inference for signal fraction

Di-Higgs: hh → bbγγ

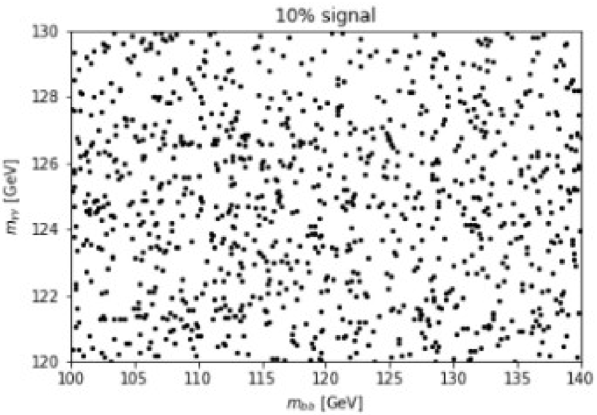


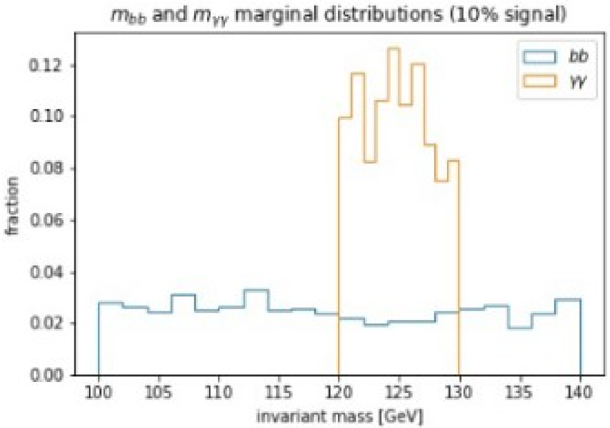
This is what we actually see



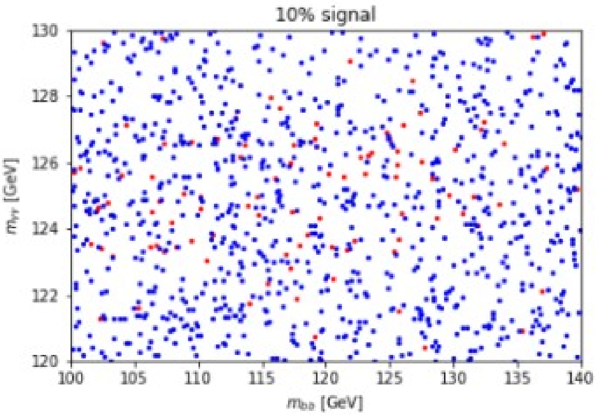
(here with labels)

Di-Higgs: hh → bbγγ

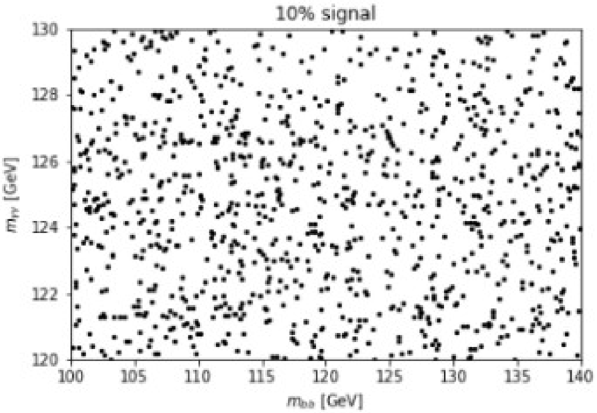


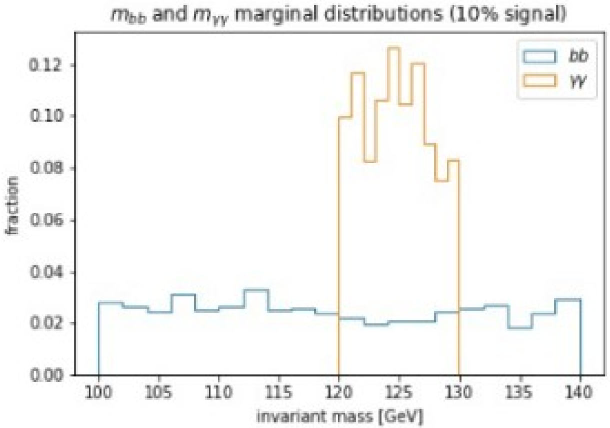


Hard to recognize something

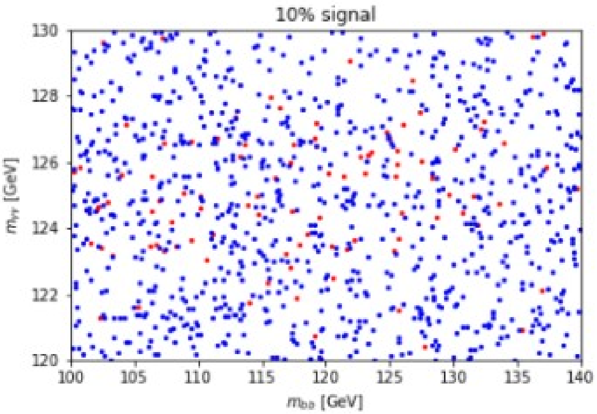


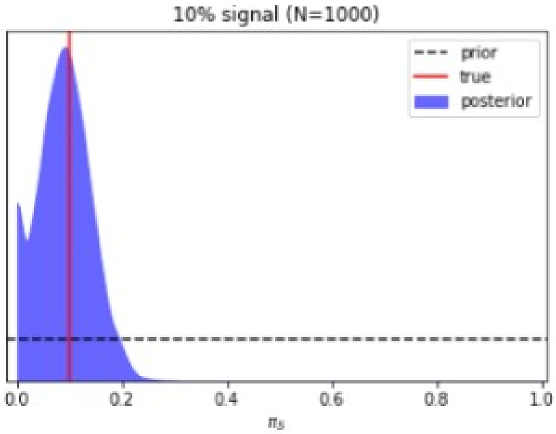
Di-Higgs: hh → bbγγ





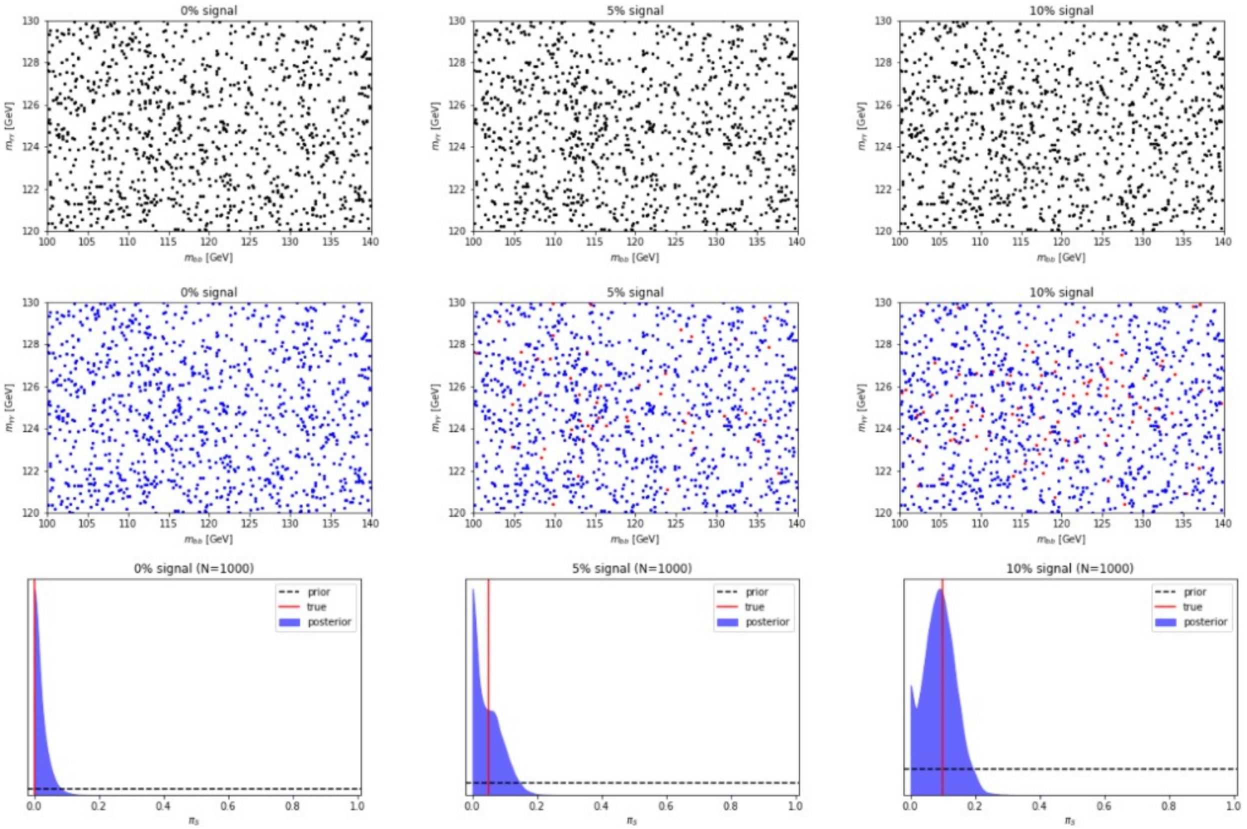
Hard to recognize something





Fraction inferred

Di-Higgs: hh → bbγγ



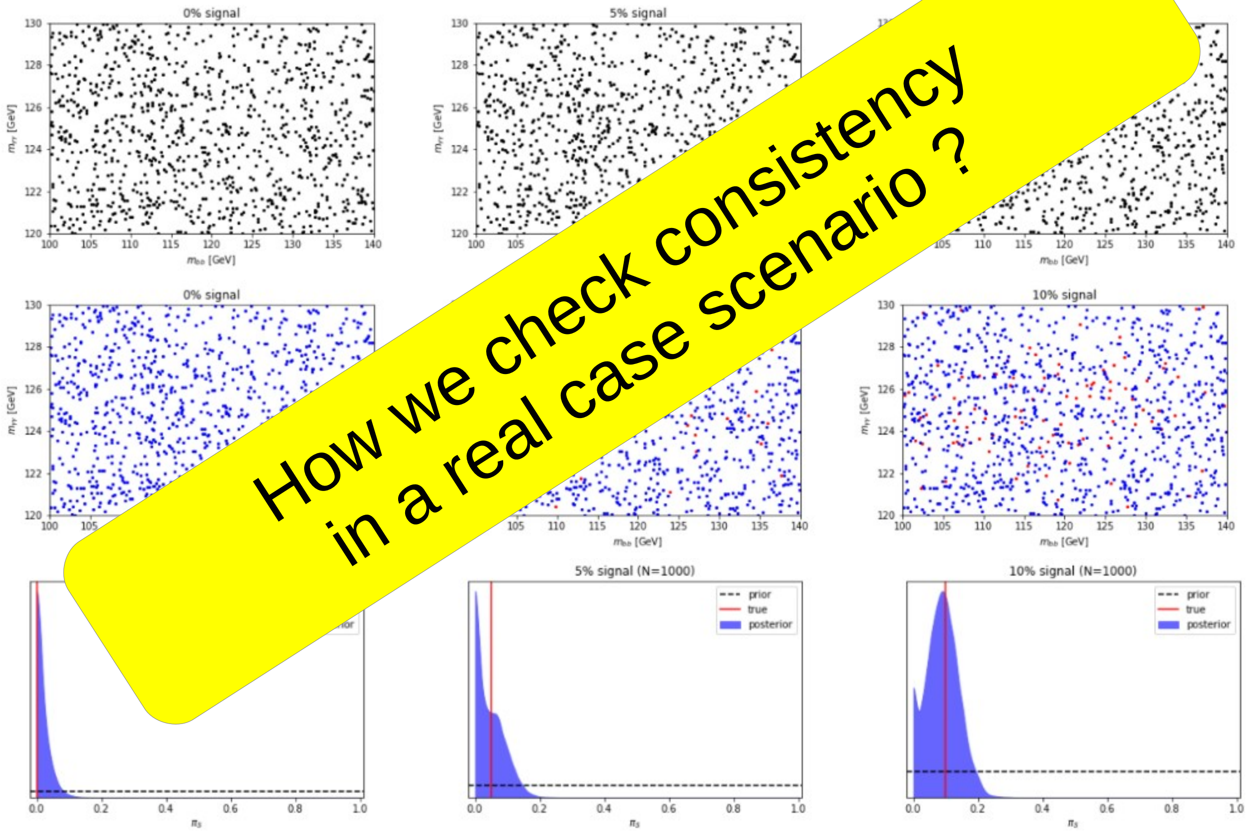
0% 5% 10%

What we see

With Labels

Posterior

Di-Higgs: hh → bbγγ



How we check consistency in a real case scenario ?

0% 5% 10%

What we see

With Labels

Posterior

Posterior predictive check

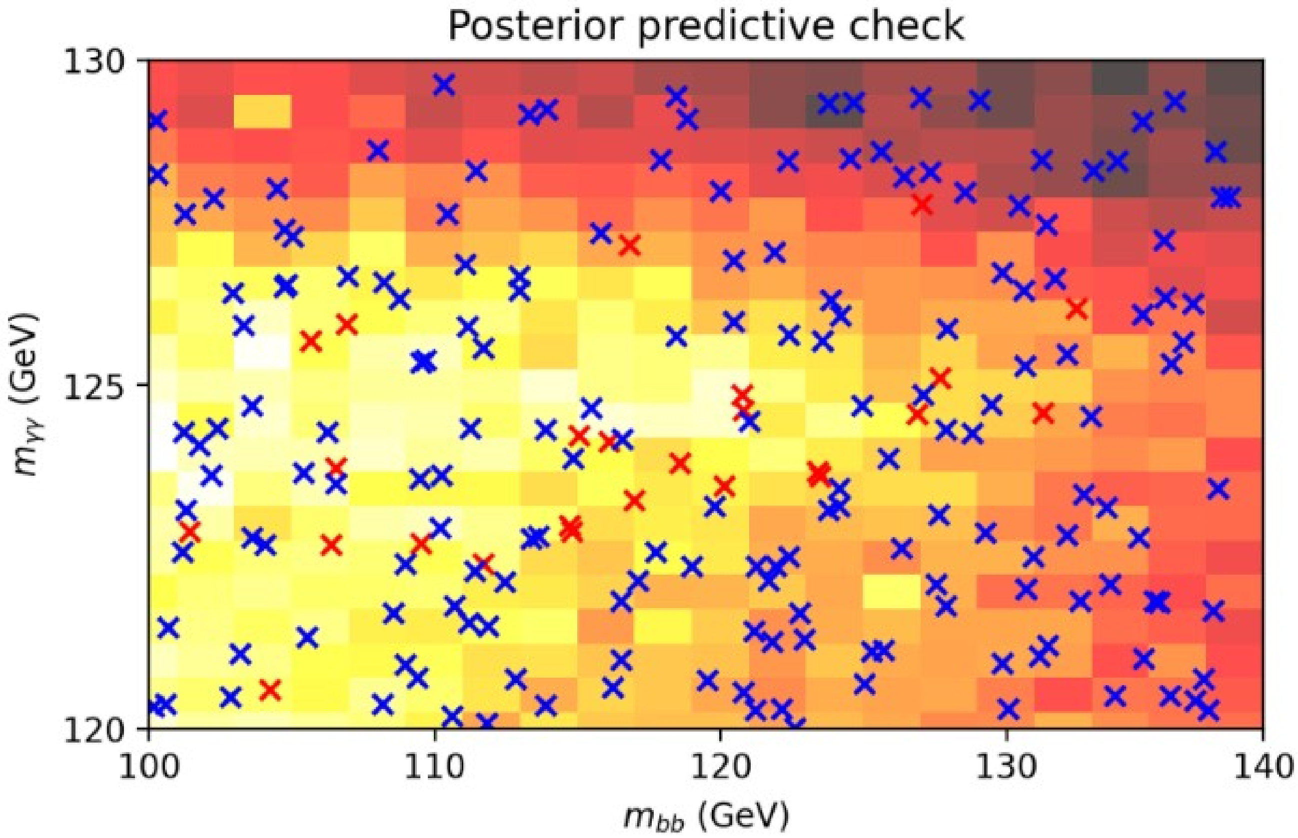
Probability of the data

→ Leave some held-out data aside, and then check the probability of having sampled those data-points with the inferred PDF

Posterior predictive check

Probability of the data

→ Leave some held-out data aside, and then check the probability of having sampled those data-points with the inferred PDF

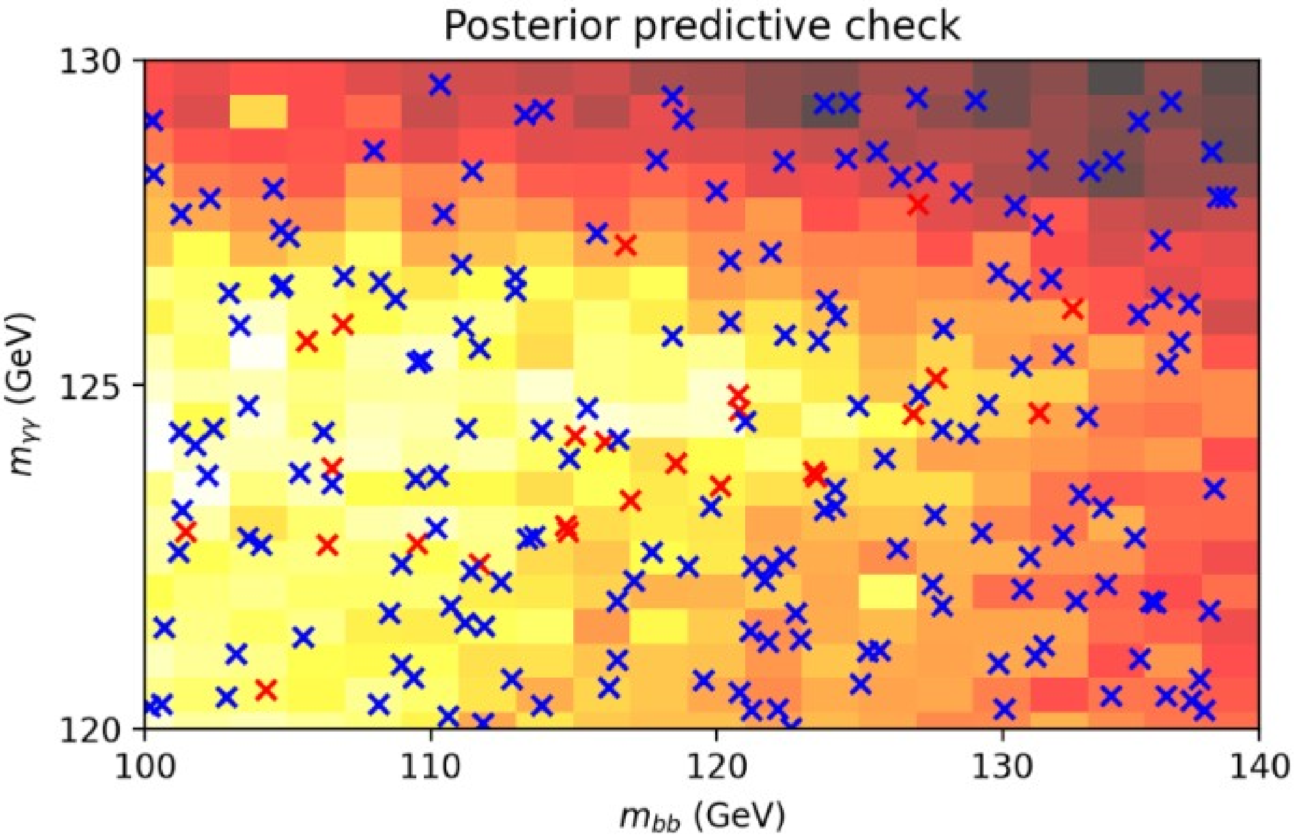


Posterior predictive check

1. Sample replicate data with the posterior
2. Compute the predictive score

$$p\left(p(X_{rep}|X_{obs}) < p(X_{held}|X_{obs})\right)$$

Probability that probability of replicate data is less than the probability of held-out data



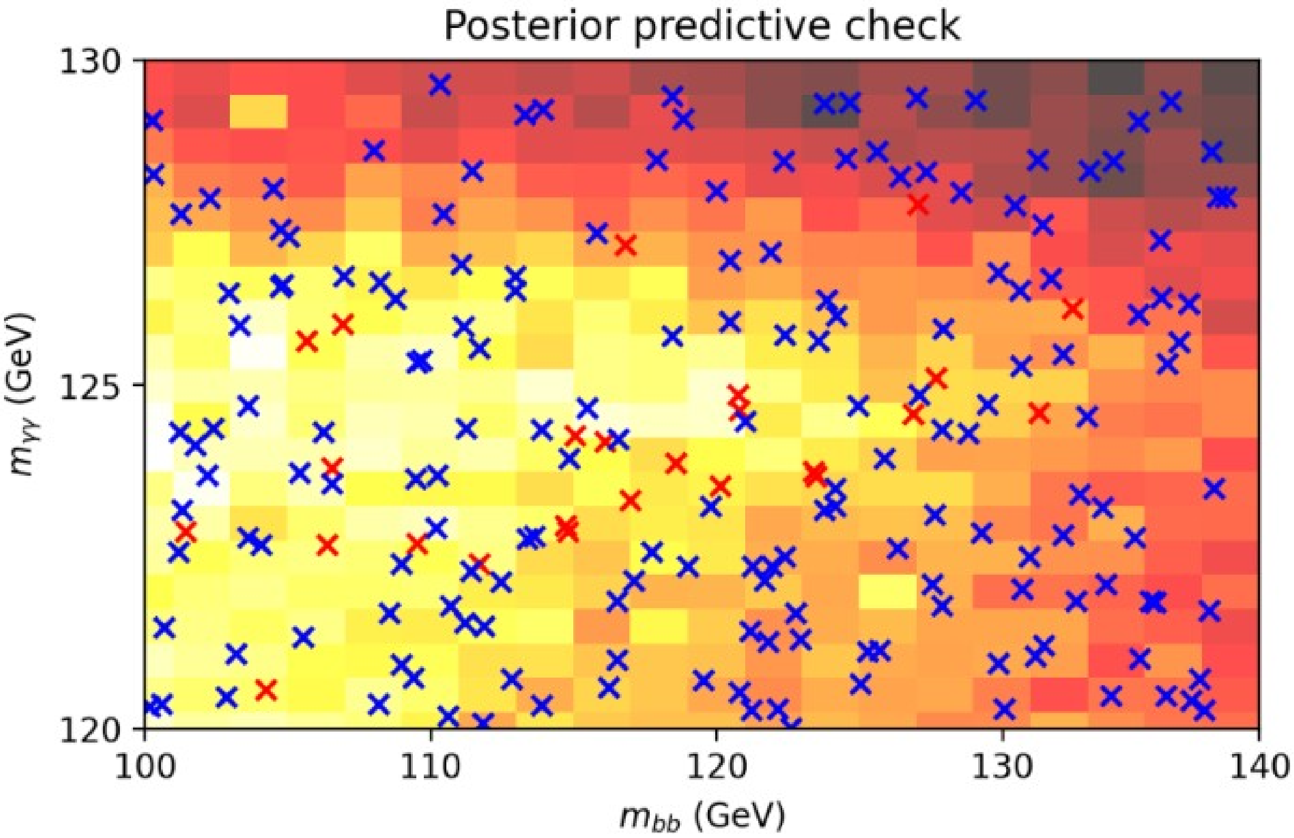
Posterior predictive check

1. Sample replicate data with the posterior
2. Compute the predictive score

$$p\left(p(X_{rep}|X_{obs}) < p(X_{held}|X_{obs})\right)$$

Probability that probability of replicate data is less than the probability of held-out data

Score: 0.5 +/- 0.03



ATLAS @ hh → bbγγ

Search for Higgs boson pair production in the two bottom quarks plus two photons final state in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector

ATLAS EXPERIMENT

Phys. Rev. D 106 (2022) 052001
DOI: 10.1103/PhysRevD.106.052001

4 Object and event selections

Variable	Definition
Photon-related kinematic variables	
$p_T/m_{\gamma\gamma}$	Transverse momentum of each of the two photons divided by the diphoton invariant mass $m_{\gamma\gamma}$
η and ϕ	Pseudorapidity and azimuthal angle of the leading and subleading photon
Jet-related kinematic variables	
b-tag status	Tightest fixed b-tag working point (60%, 70%, or 77%) that the jet passes
p_T, η and ϕ	Transverse momentum, pseudorapidity and azimuthal angle of the two jets with the highest b-tagging score
$p_T^{b\bar{b}}$, $\eta_{b\bar{b}}$ and $\phi_{b\bar{b}}$	Transverse momentum, pseudorapidity and azimuthal angle of the b-tagged jets system
$m_{b\bar{b}}$	Invariant mass of the two jets with the highest b-tagging score
H_T	Scalar sum of the p_T of the jets in the event
Single topness	For the definition, see Eq. (1)

7 Results

The statistical framework used to derive the results for both the nonresonant and resonant searches is described in the following.

7.1 Statistical framework

For both the nonresonant and resonant searches, the results of the analysis are obtained from a maximum-likelihood fit of the $m_{\gamma\gamma}$ distribution in the range $105 < m_{\gamma\gamma} < 160$ GeV, performed simultaneously over all relevant categories described in Section 4.2. The likelihood function is defined in Eq. (3):

$$\mathcal{L} = \prod_c \left(\mathrm{Pois}(n_c | N_c(\boldsymbol{\theta})) \cdot \prod_{i=1}^{n_c} f_c(m^i_{\gamma\gamma}, \boldsymbol{\theta}) \cdot G(\boldsymbol{\theta}) \right), \quad (3)$$

ATLAS: first selects using m_{bb} and then uses $m_{\gamma\gamma}$ to make the analysis.
No correlation info.

CMS @ hh → bbγγ

Search for nonresonant Higgs boson pair production in final states with two bottom quarks and two photons in proton-proton collisions at $\sqrt{s} = 13$ TeV

CMS-HIG-19-018

6 Analysis strategy

To improve the sensitivity of the search, MVA techniques are used to distinguish the ggF and VBF HH signal from the dominant nonresonant background. The output of the MVA classifiers is then used to define mutually exclusive analysis categories targeting VBF and ggF HH production. The HH signal is extracted from a fit to the invariant masses of the two Higgs boson candidates in the $(m_{\gamma\gamma}, m_{jj})$ plane simultaneously in all categories.

Table 2: Summary of the analysis categories. Two VBF- and tw
are defined based on the output of the MVA classifiers and the n
system $\widetilde{M}_X$. The VBF and ggF categories are mutually exclusive.

Category	MVA	$\widetilde{M}_X$ (GeV)
VBF CAT 0	0.52–1.00	>500
VBF CAT 1	0.86–1.00	250–500
ggF CAT 0	0.78–1.00	>600
ggF CAT 1		510–600
ggF CAT 2		385–510
ggF CAT 3		250–385
ggF CAT 4	0.62–0.78	>540
ggF CAT 5		360–540
ggF CAT 6		330–360
ggF CAT 7		250–330
ggF CAT 8	0.37–0.62	>585
ggF CAT 9		375–585
ggF CAT 10		330–375
ggF CAT 11		250–330

They do take into account correlation at the event-by-event level!

They rely on MVA over the Montecarlo

In preparation

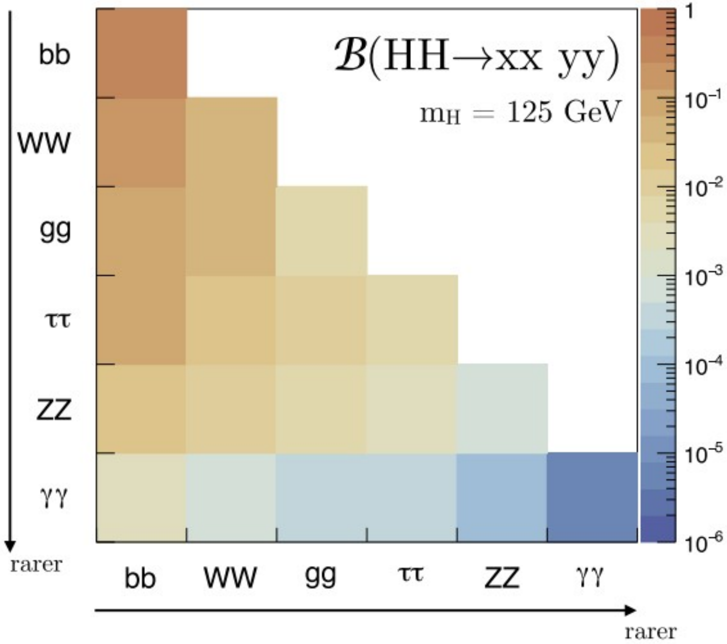
$h\,h \rightarrow bbbb$

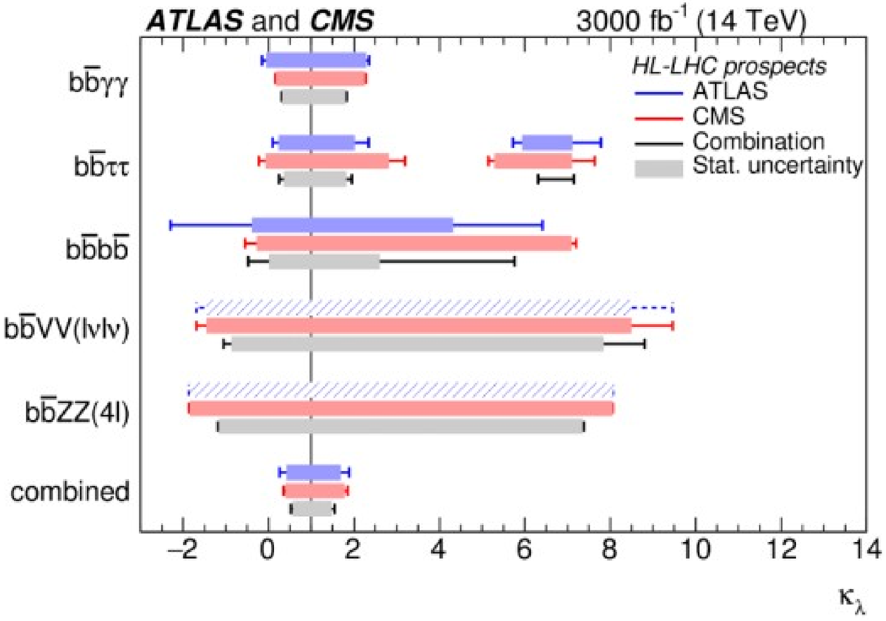
Di-Higgs: hh → bbbb

A review of Higgs boson pair production

Maxime Gouzevitch [a], Alexandra Carvalho [b]

Reviews in Physics (2020), 100039





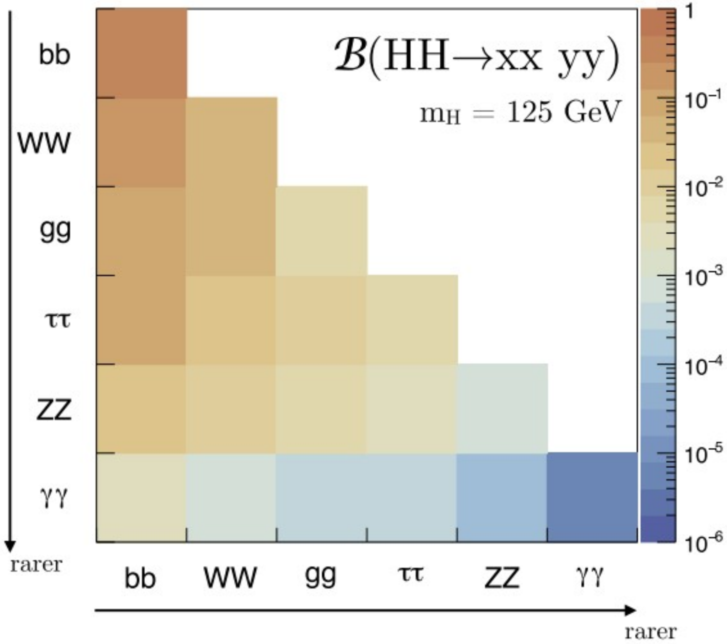
Di-Higgs: hh → bbbb

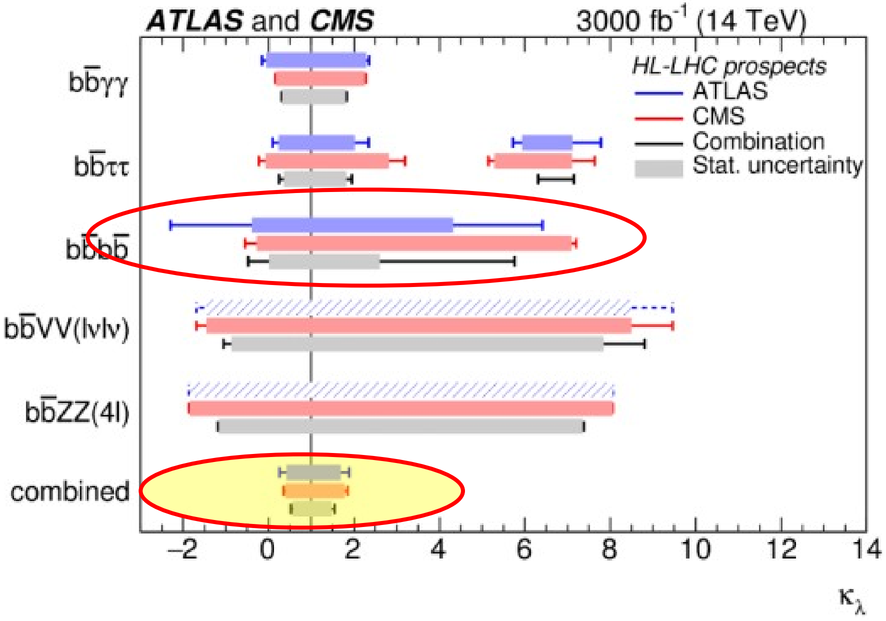
A review of Higgs boson pair production

Maxime Gouzevitch [a], Alexandra Carvalho [b]

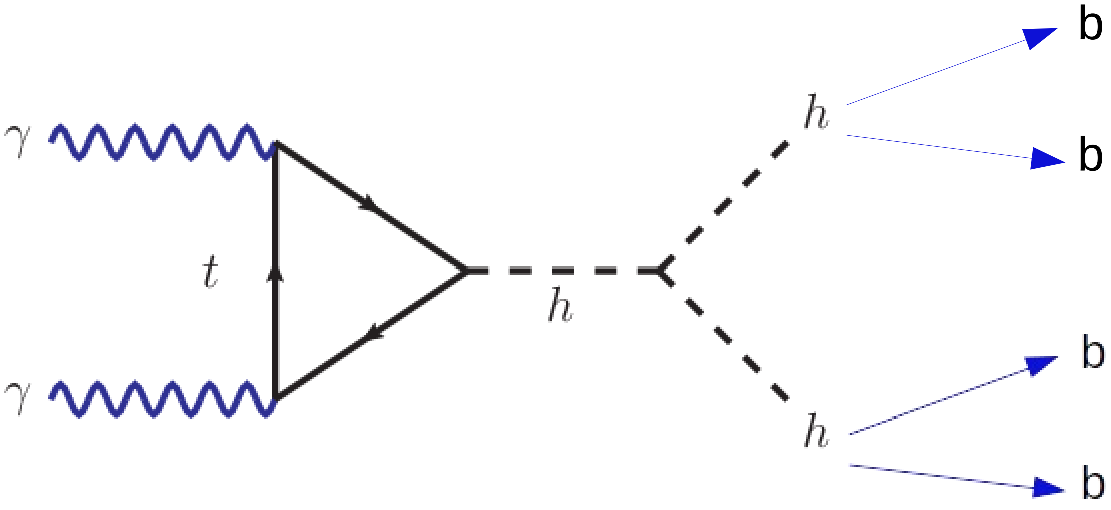
Reviews in Physics (2020), 100039

Explore improvements to bbbb is important

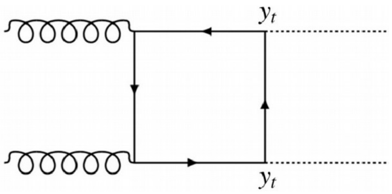




Di-Higgs: hh → bbbb

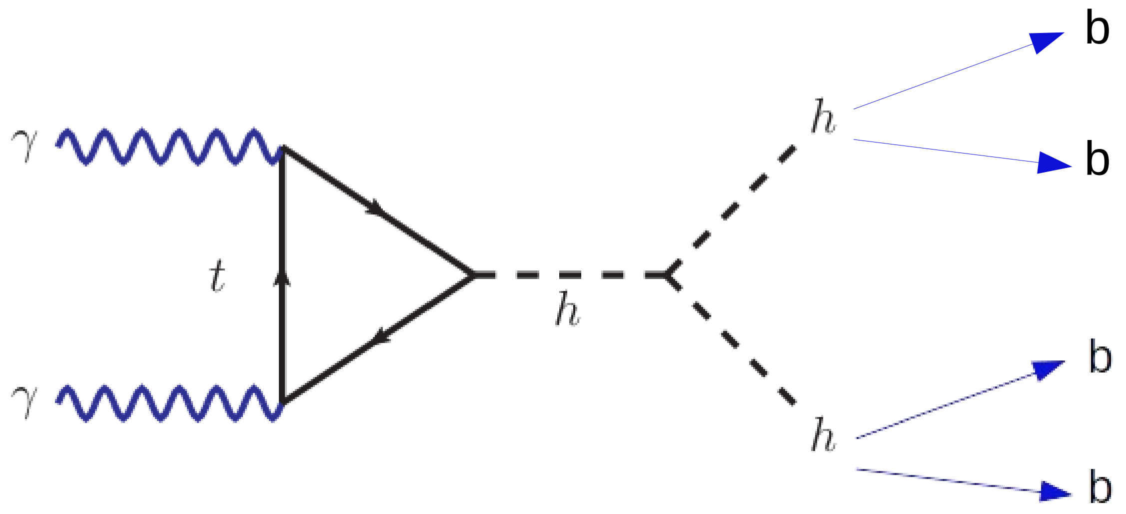


+

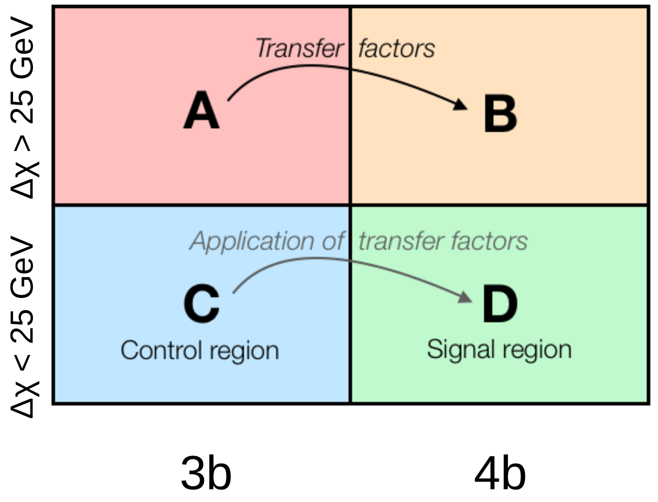


Very difficult to simulate and model,
ATLAS & CMS go data-driven.
Large backgrounds.

Di-Higgs: hh → bbbb

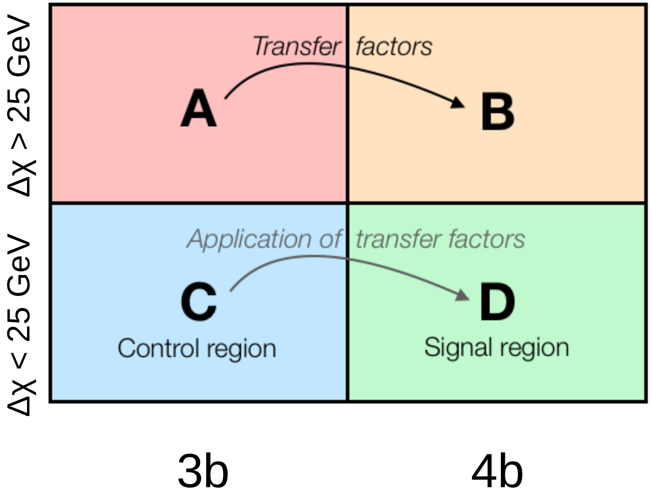


~ ABCD method



Very difficult to simulate and model,
ATLAS & CMS go data-driven.
Large backgrounds.

Di-Higgs: hh → bbbb



Search for Higgs boson pair production in the four b quark final state in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$

hep-ex/2202.09617

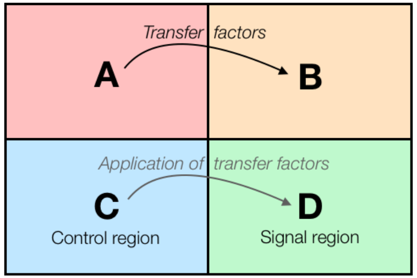
CMS-HIG-20-005

ABCD
tuned

The large multijet background that originates from QCD and $t\bar{t}$ hadronic processes is estimated from the data using background-dominated regions. Analysis signal (A_{SR}) and control (A_{CR}) regions are defined by requiring $\chi < 25\,\text{GeV}$ and $25 \leq \chi < 50\,\text{GeV}$, respectively, where χ is the distance from the expected peak position of the two Higgs boson candidates' invariant masses and is defined as $\chi = \sqrt{(m_{H_1} - c_1)^2 + (m_{H_2} - c_2)^2}$, where c_1 and c_2 are as defined for

Background events in the A^{4b}_{SR} region are modeled from events in the A^{3b}_{SR} region. The former represents the sensitive region of the analysis, while the latter provides a sample enriched in multijet background events with similar kinematic properties. Events in A^{4b}_{SR} were analyzed only after all the methods were defined and validated. The normalization is determined by scaling the observed number of events in A^{3b}_{SR} by a transfer factor computed as the ratio of the number of events in the A^{4b}_{CR} and A^{3b}_{CR} regions. Variations of the transfer factor depending on the position in the (m_{H_1}, m_{H_2}) plane are accounted for by measuring it as a function of $m_{\parallel}$, defined as the projection of the point in the plane on the line $m_{H_1} = (c_1/c_2)\, m_{H_2}$ that is used for the H candidate reconstruction. Higher values of $m_{\parallel}$ are correlated with a higher average p_T of the selected jets.

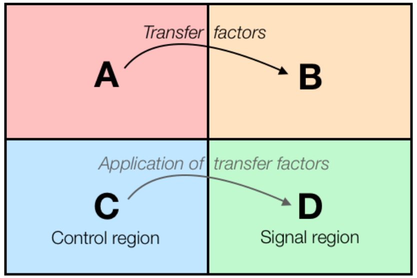
ABCD Vs Bayesian techniques

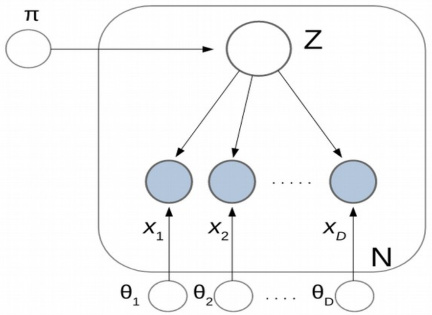


Regions should be

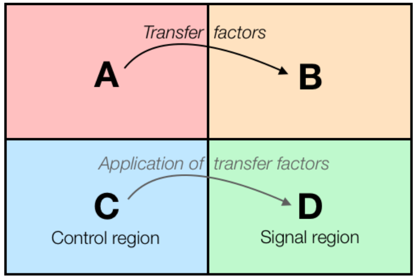
- Close by to maintain similarity
- Separated to avoid contamination

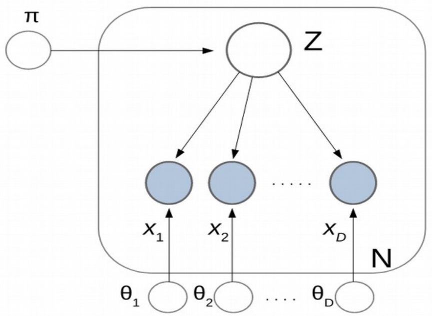
ABCD Vs Bayesian techniques





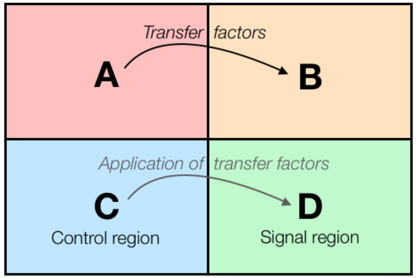
ABCD Vs Bayesian techniques

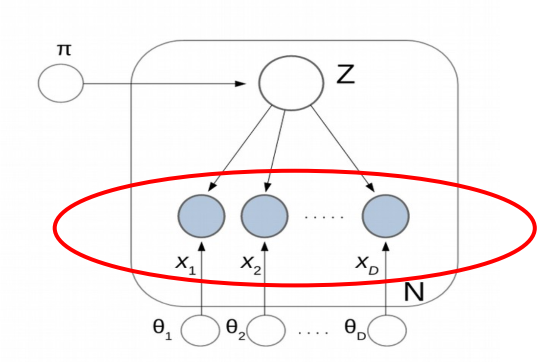




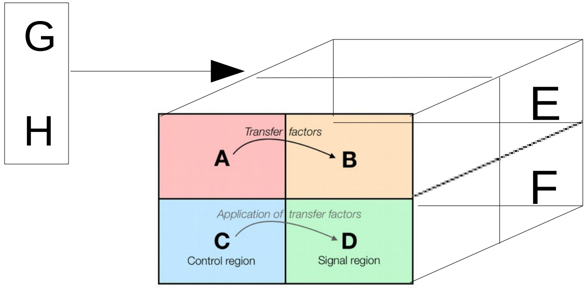
Two observables & must be independent

ABCD Vs Bayesian techniques

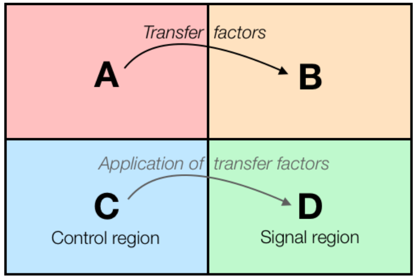


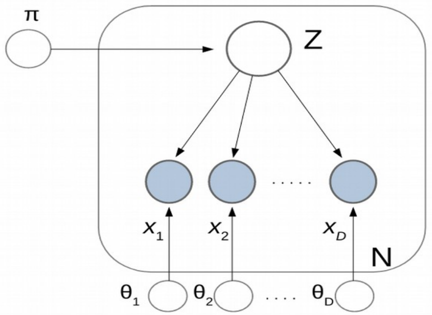


Two observables & must be independent → Any number of observables & independent



ABCD Vs Bayesian techniques

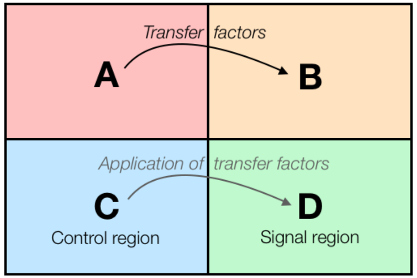


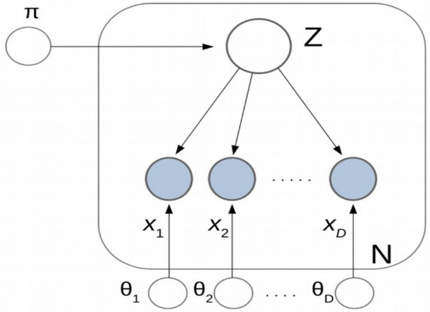


Two observables & must be independent → Any number of observables & independent

Each observable has two outcomes

ABCD Vs Bayesian techniques

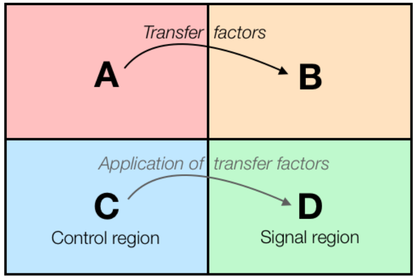


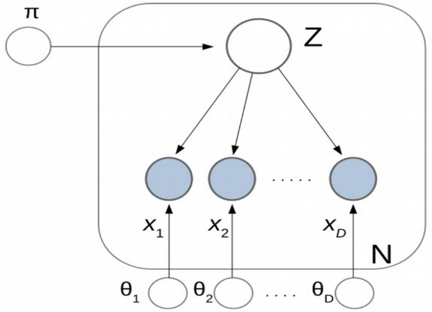


Two observables & must be independent → Any number of observables & independent

Each observable has two outcomes → Each observable can have any number of Outomes. Usually continuous is better

ABCD Vs Bayesian techniques



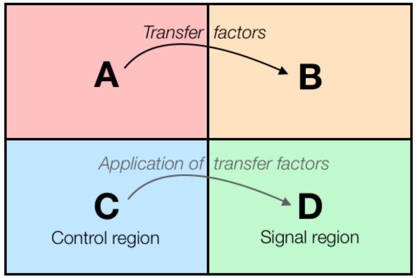


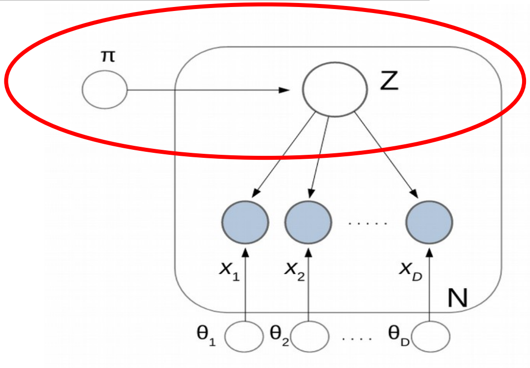
Two observables & must be independent → Any number of observables & independent

Each observable has two outcomes → Each observable can have any number of Outomes. Usually continuous is better

Two classes: signal & background

ABCD Vs Bayesian techniques



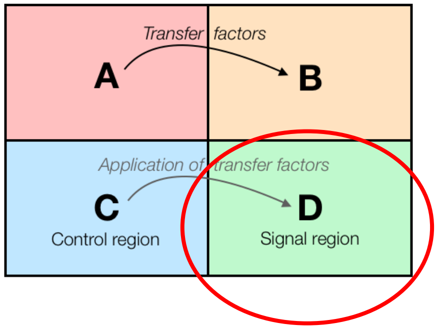


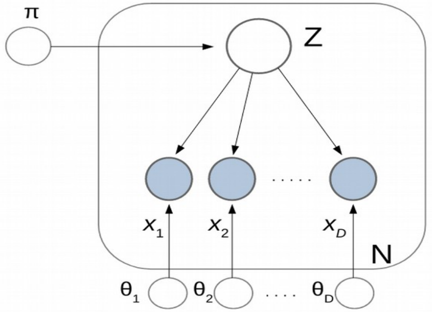
Categorical (~multinomial)

ABCD	Bayesian
Two observables & must be independent	Any number of observables & independent
Each observable has two outcomes	Each observable can have any number of Outomes. Usually continuous is better
Two classes: signal & background	It can have many classes. E.g. many back-grounds and use prior knowledge on them

ABCD Vs Bayesian techniques

Need control regions!





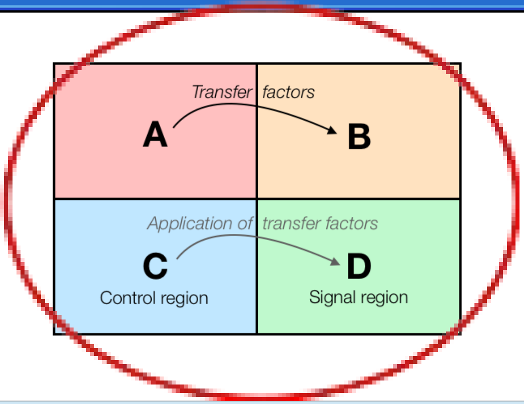
Two observables & must be independent → Any number of observables & independent

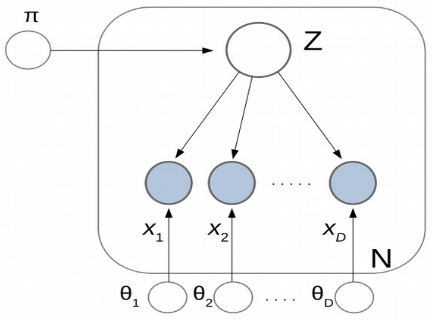
Each observable has two outcomes → Each observable can have any number of Outomes. Usually continuous is better

Two classes: signal & background → It can have many classes. E.g. many back-grounds and use prior knowledge on them

Signal should be in only one region, usually D

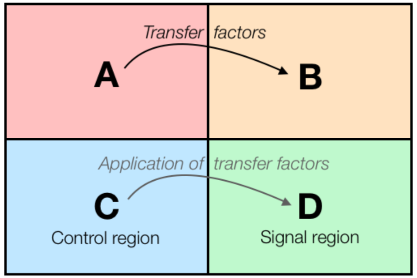
ABCD Vs Bayesian techniques



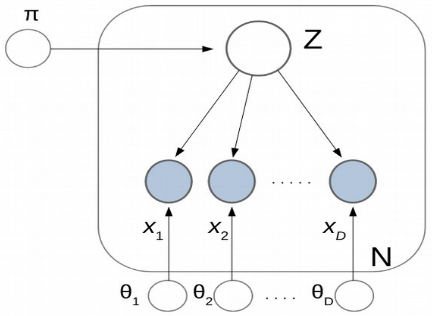


ABCD	Bayesian
Two observables & must be independent	Any number of observables & independent
Each observable has two outcomes	Each observable can have any number of Outomes. Usually continuous is better
Two classes: signal & background	It can have many classes. E.g. many back-grounds and use prior knowledge on them
Signal should be in only one region, usually D	S & B mixed in different shape and proportions everywhere. No control region. No hard cuts!

ABCD Vs Bayesian techniques

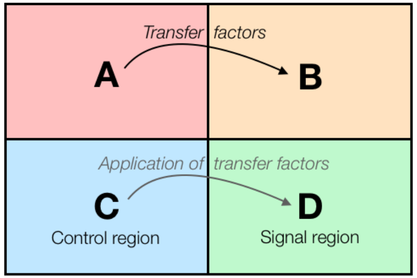


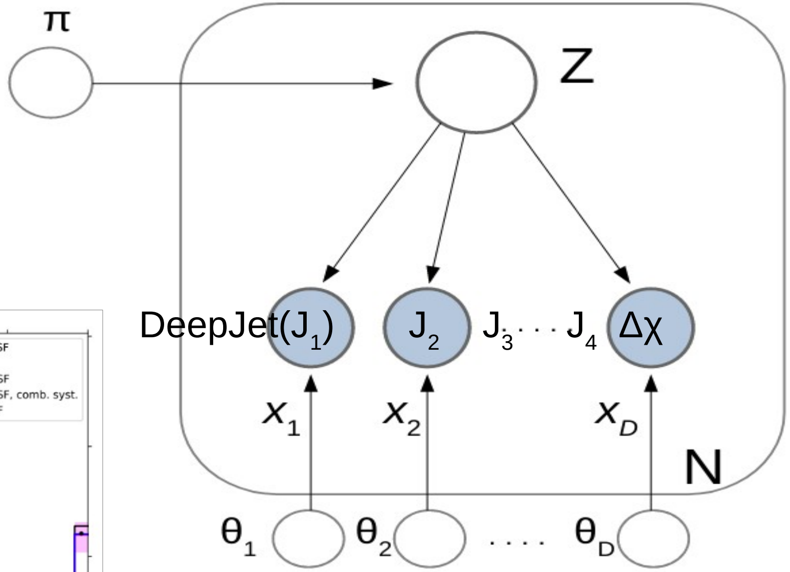
Are there real cases of many independent observable?



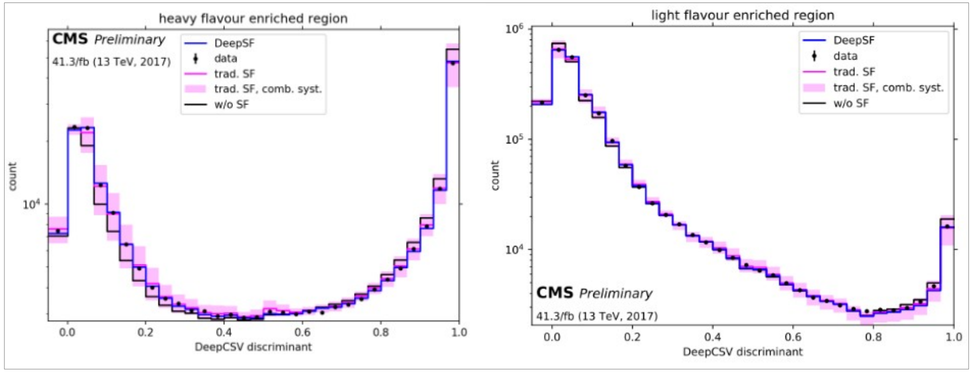
ABCD	Bayesian
Two observables & must be independent	Any number of observables & independent
Each observable has two outcomes	Each observable can have any number of Outomes. Usually continuous is better
Two classes: signal & background	It can have many classes. E.g. many back-grounds and use prior knowledge on them
Signal should be in only one region, usually D	S & B mixed in different shape and proportions everywhere. No control region. No hard cuts!

ABCD Vs Bayesian techniques





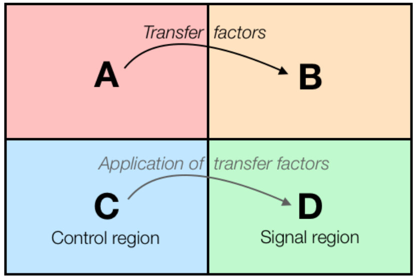
hh → bbbb

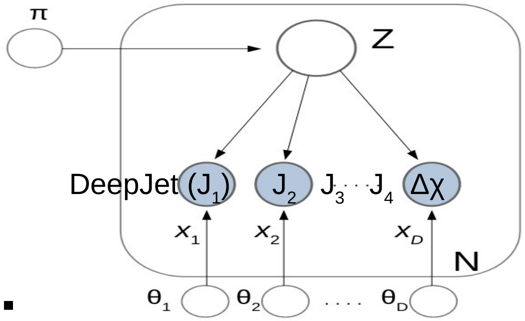


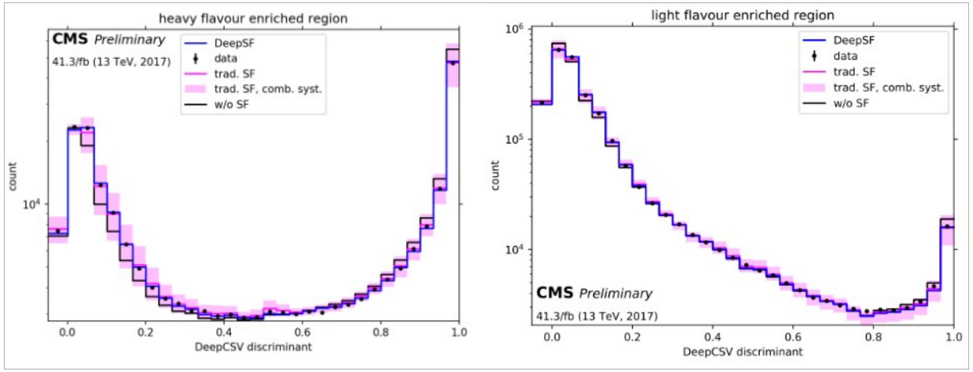
DeepJet PDFs according to class!

Instead of counting b-jets, use the four continuous Deepjet variables

ABCD Vs Bayesian techniques







DeepJet PDFs according to class!

Ultimate goal:

- Arbitrary priors with uncertainty (usually MC)
- Not arbitrarily flexible
- Model observed variables as being sampled from there
- Infere everything!
- Calibration (?)

Conclusions

Conclusions

- ML-Industry and Statistics very successful tools in Bayesian framework

Conclusions

- ML-Industry and Statistics very successful tools in Bayesian framework
- Have not yet been adequately tested @ LHC

Conclusions

- ML-Industry and Statistics very successful tools in Bayesian framework
- Have not yet been adequately tested @ LHC
- Simplified examples show good perspectives
  - ✓ q-g jet discrimination (2112.11352)
  - ✓ Four-tops (2107.00668)
  - ✓ Unsupervised top-tagging (2212.13583)
  - ✓ hh → bbγγ (2210.07358)

Conclusions

- ML-Industry and Statistics very successful tools in Bayesian framework
- Have not yet been adequately tested @ LHC
- Simplified examples show good perspectives
    - ✓ q-g jet discrimination (2112.11352)
    - ✓ Four-tops (2107.00668)
    - ✓ Unsupervised top-tagging (2212.13583)
    - ✓ hh → bbγγ (2210.07358)
- Potential enhancement hh → bbbb (in preparation)

Conclusions

- ML-Industry and Statistics very successful tools in Bayesian framework
- Have not yet been adequately tested @ LHC
- Simplified examples show good perspectives
  - ✓ q-g jet discrimination (2112.11352)
  - ✓ Four-tops (2107.00668)
  - ✓ Unsupervised top-tagging (2212.13583)
  - ✓ hh → bbγγ (2210.07358)
- Potential enhancement hh → bbbb (in preparation)
- Are some LHC analysis sub-optimal ?

Conclusions

- ML-Industry and Statistics very successful tools in Bayesian framework
- Have not yet been adequately tested @ LHC
- Simplified examples show good perspectives
  - ✓ q-g jet discrimination (2112.11352)
  - ✓ Four-tops (2107.00668)
  - ✓ Unsupervised top-tagging (2212.13583)
  - ✓ hh → bbγγ (2210.07358)
- Potential enhancement hh → bbbb (in preparation)
- Are some LHC analysis sub-optimal ?
- Bayesian ML techniques may yield improvement in observables

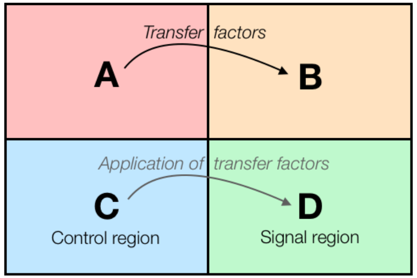
Conclusions

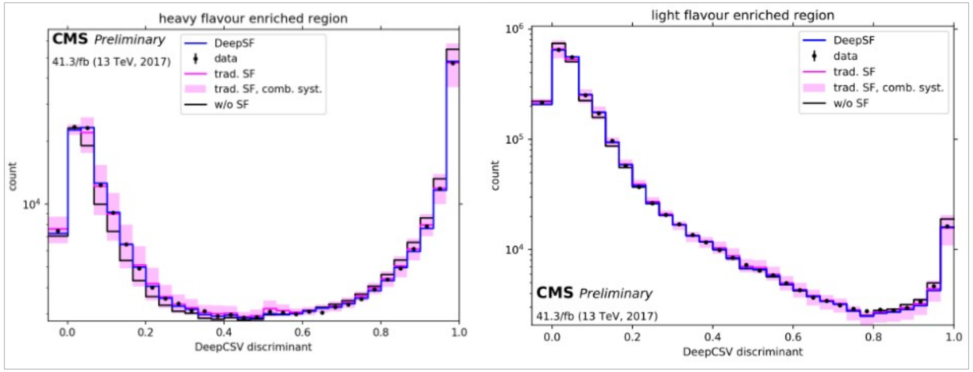
- ML-Industry and Statistics very successful tools in Bayesian framework
- Have not yet been adequately tested @ LHC
- Simplified examples show good perspectives
  - ✓ q-g jet discrimination (2112.11352)
  - ✓ Four-tops (2107.00668)
  - ✓ Unsupervised top-tagging (2212.13583)
  - ✓ hh $\rightarrow$ bbγγ (2210.07358)
- Potential enhancement hh $\rightarrow$ bbbb (in preparation)
- Are some LHC analysis sub-optimal ?
- Bayesian ML techniques may yield improvement in observables

Thank you!

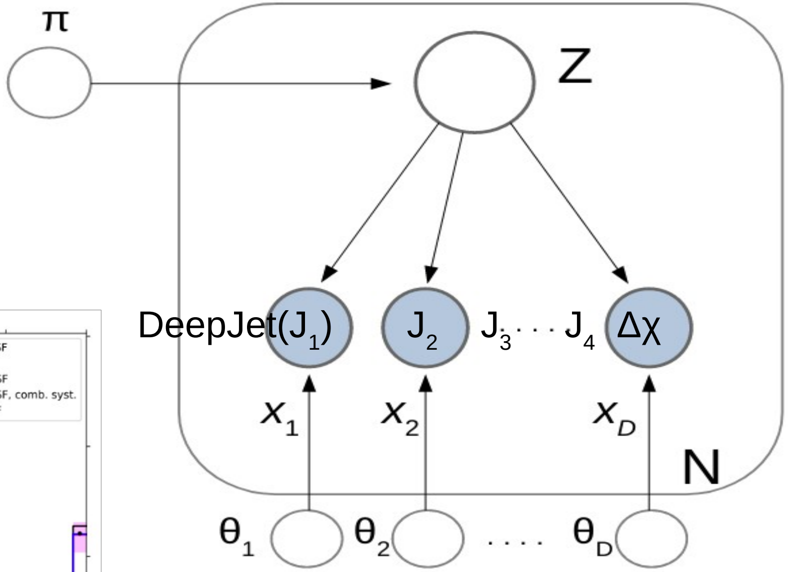
Backup Slides

ABCD Vs Bayesian techniques





DeepJet PDFs according to class!



b-jet if DJ > 0.7

DJ = 0.69, 0.99, 0.99, 0.99
DJ = 0.1, 0.71, 0.71, 0.71
DJ = 0.71, 0.71, 0.71, 0.71

Backup slides

Eur. Phys. J. C (2016) 76:11
DOI 10.1140/epjc/s10052-015-3852-4

THE EUROPEAN PHYSICAL JOURNAL C

Regular Article - Experimental Physics

Measurements of fiducial cross-sections for $t\bar{t}$ production with one or two additional b-jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector

ATLAS Collaboration*

malised to the NNLO+NNLL result [32–37]. PYTHIA 8 offers several options for modelling $g \rightarrow b\bar{b}$ splittings in the final-state parton showers, which may be accessed by varying the TIMESHOWER:WEIGHTGLUONTOQUARK (wgtq) parameter [75]. Differences between the models arise by neglecting (wgtq5) or retaining (wgtq3, wgtq6) the mass-dependent terms in the $g \rightarrow b\bar{b}$ splitting kernels. Differences also arise with respect to the treatment of the high-$m_{b\bar{b}}$ region, with specific models giving an enhanced or suppressed $g \rightarrow b\bar{b}$ rate. The model corresponding to wgtq3 was chosen to maximise this rate. Finally, some of the models (wgtq5, wgtq6) offer the possibility to choose sgtq$\cdot m_{b\bar{b}}$ instead of the transverse momentum as the argument of α_S in the $g \rightarrow b\bar{b}$ vertices. Here sgtq refers to the TIMESHOWER:SCALEGLUONTOQUARK parameter, and is allowed to vary in the range $0.25 \leq$ sgtq ≤ 1, with larger values giving a smaller $g \rightarrow b\bar{b}$ rate and vice versa. For the model wgtq5, sgtq was set to 1, a combination that minimises the $g \rightarrow b\bar{b}$ rate, while for wgtq6, sgtq was set to 0.25.

(see discussion in 1701.04427)

Backup slides

Eur. Phys. J. C (2016) 76:379
DOI 10.1140/epjc/s10052-016-4105-x

THE EUROPEAN PHYSICAL JOURNAL C

Regular Article - Experimental Physics

Measurement of $t\bar{t}$ production with additional jet activity, including b quark jets, in the dilepton decay channel using pp collisions at $\sqrt{s} = 8$ TeV

CMS Collaboration*

11. CMS Collaboration, Measurement of the cross section ratio $\sigma_{t\bar{t}b\bar{b}}/\sigma_{t\bar{t}jj}$ in pp collisions at $\sqrt{s} = 8$ TeV. Phys. Lett. B **746**, 132 (2015). doi:10.1016/j.physletb.2015.04.060. arXiv:1411.5621

PYTHIA6 and HERWIG6. The normalization factors applied to the MADGRAPH and POWHEG predictions are found to be about 1.3 for results related to the leading additional b jet. The predictions from both generators underestimate the $t\bar{t}b\bar{b}$ cross sections by a factor 1.8, in agreement with the results from Ref. [11]. The normalization factors applied to MC@NLO are approximately 2 and 4 for the leading and subleading additional b jet quantities, respectively, reflecting the observation that the generator does not simulate sufficiently large jet multiplicities. All the predictions have slightly harder p_T spectra for the leading additional b jet than the data, while they describe the behaviour of the $|\eta|$ and m_{bb} distributions within the current precision. The predictions favour smaller ΔR_{bb} values than the measurement, although the differences are in general within two standard deviations of the total uncertainty.

(see discussion in 1701.04427)

Backup slides

Four-tops
ATLAS-CONF-2021-013

7.2 Sequential kinematic reweighting

Following the flavour rescaling, a sequential reweighting is used to mitigate the kinematic mismodelling observed in $t\bar{t}$+jets MC. The reweighting corrects for the distributions of N_{jets}, the number of large-R jets ($N_{\mathrm{LR-jets}}$), the scalar sum of all jet and lepton p_{T} in the event ($H_{\mathrm{T}}^{\mathrm{all}}$), and the average ΔR between any two jets ($\Delta R_{\mathrm{avg.}}^{\mathrm{jj}}$). These variables are related to the overall jet activities in the events and are observed to be mismodelled, especially the N_{jets} and $H_{\mathrm{T}}^{\mathrm{all}}$ spectra. These variables capture the most representative global kinematics of the events, as well as kinematic properties of the individual jets such as p_{T} and their angular distributions.

The $t\bar{t}$+jets events in $\geq 3b$ regions are reweighted according to the discrepancy between data and MC in the 2b regions. The reweighting factors are derived such that the overall MC prediction matches the data in the 2b regions. This is done based on the assumption that the deficiency of the radiation modelling in the parton shower is independent of the flavour of the radiated jets. Systematic variations on the $t\bar{t}$+jets modelling cover possible deviations from such assumption.

Backup slides

Accuracy = 0.71

Another feature of Bayesian computation is that we can compute the probability of a given measurement n_{SD} belonging to class z integrated over the λ_g, λ_q and π_g posterior distribution. Using our Monte Carlo samples, we calculate

$$p(z \,|\, n_{\mathrm{SD}}, X) \approx \frac{1}{T} \sum_{t=1}^{T} p(z \,|\, n_{\mathrm{SD}}, \pi_g^{(t)}, \lambda_g^{(t)}, \lambda_q^{(t)}) \tag{4.1}$$

where X represents the training dataset and t is the posterior sample index. We show

https://arxiv.org/pdf/2112.11352.pdf

Backup slides

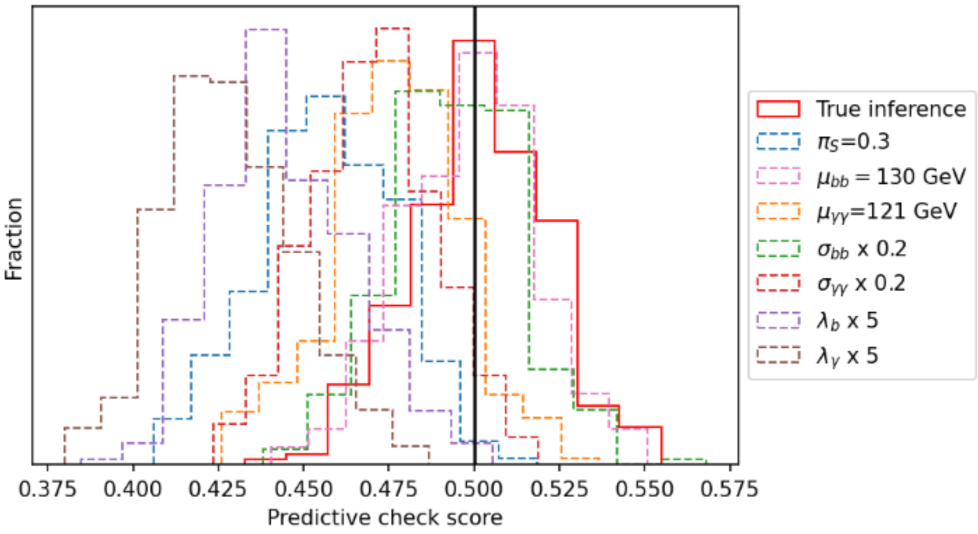


Figure 7. Predictive check score distributions for the true inference and for a few artificially shifted replicate data sets. We indicate in the right panel which parameters have been artificially fixed/shifted for each case. From the plot it can be recognized that the replicate data set from the inference process has a good agreement with the held-out data since it is centered around 0.5, as expected. We can also have a grasp on how much shift in the predictive check score is produced by different shifts in the parameters of the replicate data set PDF. As expected, in these cases the predictive score goes below 0.5, indicating the (injected) bias in the replicate PDF. From the plot it can be recognized that the data set and problem is not much sensitive to variations in $\mu_{b\bar{b}}$ and $\sigma_{b\bar{b}}$, this is in agreement with the discussion in Sect. III A and it is because of the very little variation that has the $m_{b\bar{b}}$ background in conjunction with large $\sigma_{b\bar{b}}$ and small signal fraction.

Backup slides

ICAS

A multi-dimensional search for new heavy resonances decaying to boosted WW, WZ, or ZZ boson pairs in the dijet final state at 13 TeV

CMS Collaboration*

CERN, 1211 Geneva 23, Switzerland

Abstract A search in an all-jet final state for new massive resonances decaying to WW, WZ, or ZZ boson pairs using a novel analysis method is presented. The analysis is performed on data corresponding to an integrated luminosity of 77.3 fb^{-1} recorded with the CMS experiment at the LHC at a centre-of-mass energy of 13 TeV. The search is focussed on potential narrow-width resonances with masses above 1.2 TeV, where the decay products of each W or Z boson are expected to be collimated into a single, large-radius jet. The signal is extracted using a three-dimensional maximum likelihood fit of the two jet masses and the dijet invariant mass, yielding an improvement in sensitivity of up to 30% relative to previous search methods. No excess is observed above the estimated standard model background. In a heavy vector triplet model, spin-1 Z′ and W′ resonances with masses below 3.5 and 3.8 TeV, respectively, are excluded at 95% confidence level. In a bulk graviton model, upper limits on cross sections are set between 27 and 0.2 fb for resonance masses between 1.2 and 5.2 TeV, respectively. The limits presented in this paper are the best to date in the dijet final state.

CMS

Search for new heavy resonances decaying to WW, WZ, ZZ, WH, or ZH boson pairs in the all-jets final state in proton-proton collisions at $\sqrt{s} = 13$ TeV

CMS-B2G-20-009

Existing search by CMS improving ~30% sensitivity when using the correlation of the invariant masses

Backup slides

2007.1440

Decorrelated variables based on Montecarlo simulations.

Danger: catch patterns that are from MC and expect them to be in real data

Squark pair events and multijet events are generated with PYTHIA 8.230 [97, 98] at a center-of-mass-energy of $\sqrt{s} = 13$ TeV interfaced with DELPHES 3.4.1 [99] using the default CMS run card. Jets are clustered using the anti-k_t algorithm [100] with radius parameter $R = 0.4$ implemented in FASTJET 3.2.1 [93, 101]. 1M signal events and 10M background events were generated, of which about 100k signal events and

For the Double Disco ABCD method, we use the loss function

$$\mathcal{L}[f,g] = \mathcal{L}_{\text{classifier}}[f(X),y] + \mathcal{L}_{\text{classifier}}[g(X),y] + \lambda\,\text{dCorr}^2_{y=0}[f(X),g(X)], \qquad (3.2)$$

where now f and g are two neural networks that are trained simultaneously. When $\lambda = 0$, the loss will be minimized when $f = g$ is the optimal classifier (up to degeneracies).

ABCDisCo: Automating the ABCD Method with Machine Learning

Gregor Kasieczka,[1] Benjamin Nachman,[2] Matthew D. Schwartz,[3] and David Shih,[2,4,5]

[1] *Institut für Experimentalphysik, Universität Hamburg, Luruper Chaussee 149, D-22761 Hamburg, Germany*

[2] *Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*

[3] *Department of Physics, Harvard University, Cambridge, MA 02138*

[4] *NHETC, Department of Physics and Astronomy, Rutgers University, Piscataway, NJ 08854, USA*

[5] *Berkeley Center for Theoretical Physics, University of California, Berkeley, CA 94720, USA*

E-mail: gregor.kasieczka@uni-hamburg.de, bpnachman@lbl.gov, shih@physics.rutgers.edu, schwartz@g.harvard.edu

ABSTRACT: The ABCD method is one of the most widely used data-driven background estimation techniques in high energy physics. Cuts on two statistically-independent classifiers separate signal and background into four regions, so that background in the signal region can be estimated simply using the other three control regions. Typically, the independent classifiers are chosen "by hand" to be intuitive and physically motivated variables. Here, we explore the possibility of automating the design of one or both of these classifiers using machine learning. We show how to use state-of-the-art decorrelation methods to construct powerful yet independent discriminators. Along the way, we uncover a previously unappreciated aspect of the ABCD method: its accuracy hinges on having low signal contamination in control regions not just overall, but *relative* to the signal fraction in the signal region. We demonstrate the method with three examples: a simple model consisting of three-dimensional Gaussians; boosted hadronic top jet tagging; and a recasted search for paired dijet resonances. In all cases, automating the ABCD method with machine learning significantly improves performance in terms of ABCD closure, background rejection and signal contamination.

Backup slides

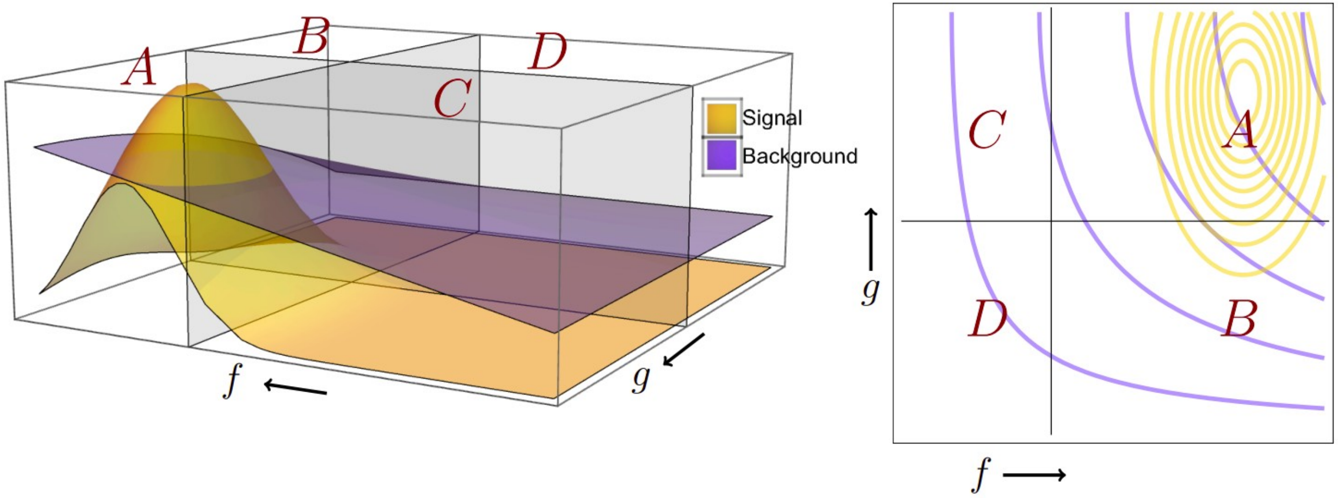


Figure 1. The ABCD method is used to estimate the background in region A as $N_A = \frac{N_B N_C}{N_D}$. It requires the signal to be relatively localized in region A and the observables to be independent on background. The shaded planes (left) or lines (right) denote thresholds which isolate the signal in region A.

Backup slides

Backup slides

Backup slides