Maximizing the potential of LHC data

A new frontier in particle physics

Università degli Studi Roma Tre
24 June 2024
Ezequiel Alvarez @ICAS @Argentina
sequi@unsam.edu.ar

Data & Information

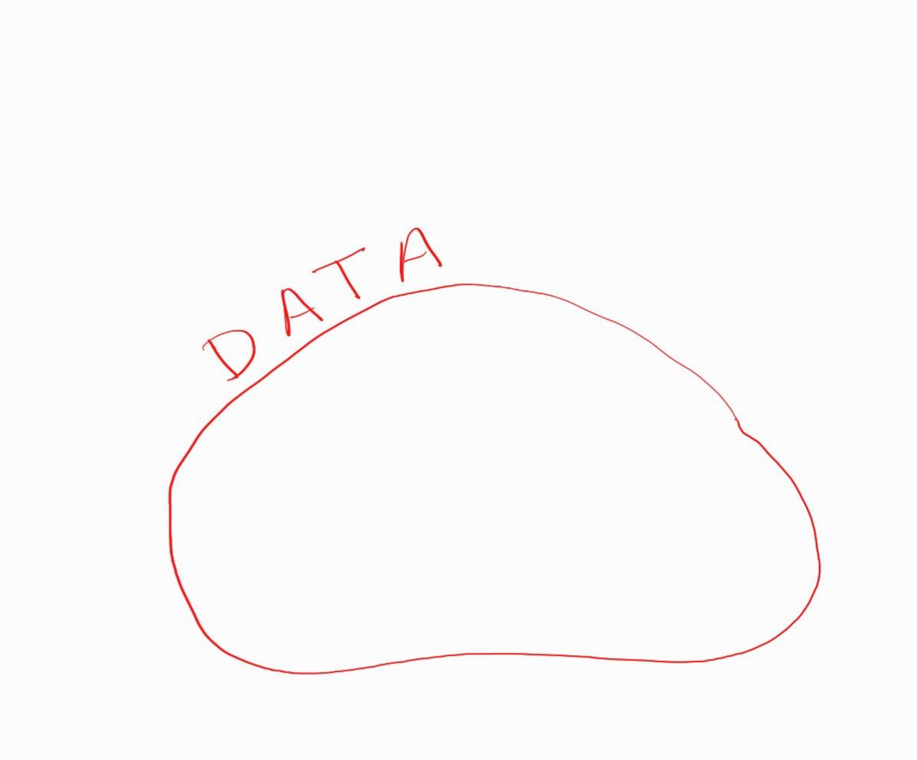
- **Data:** Measure N times the length of the table
- **Information:** An estimation on the length of the table

Data & Information

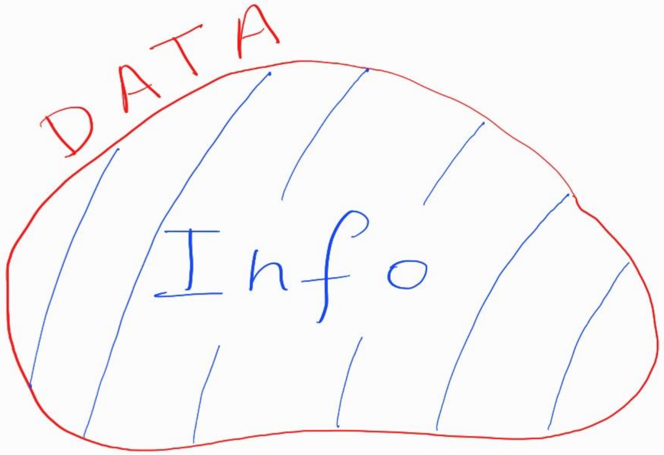
- **Data:** Measure N times the length of the table
- **Information:** An estimation on the length of the table

There is not such a thing as intrinsic information in the Data

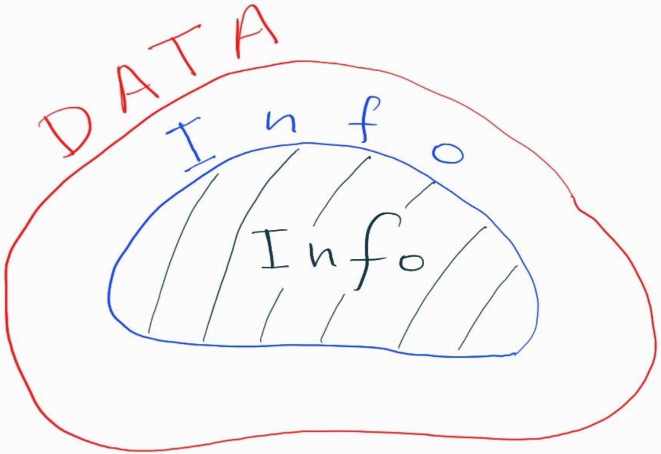
Data & Information



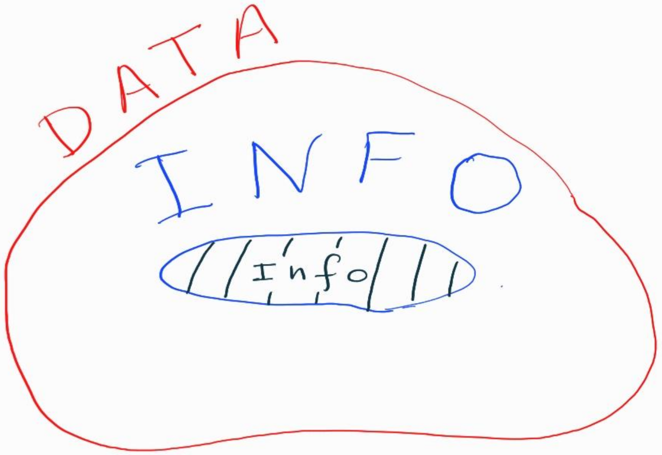
Data & Information



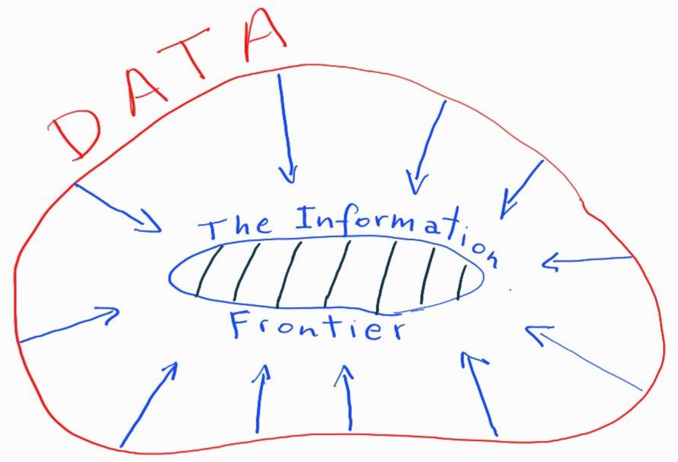
Data & Information



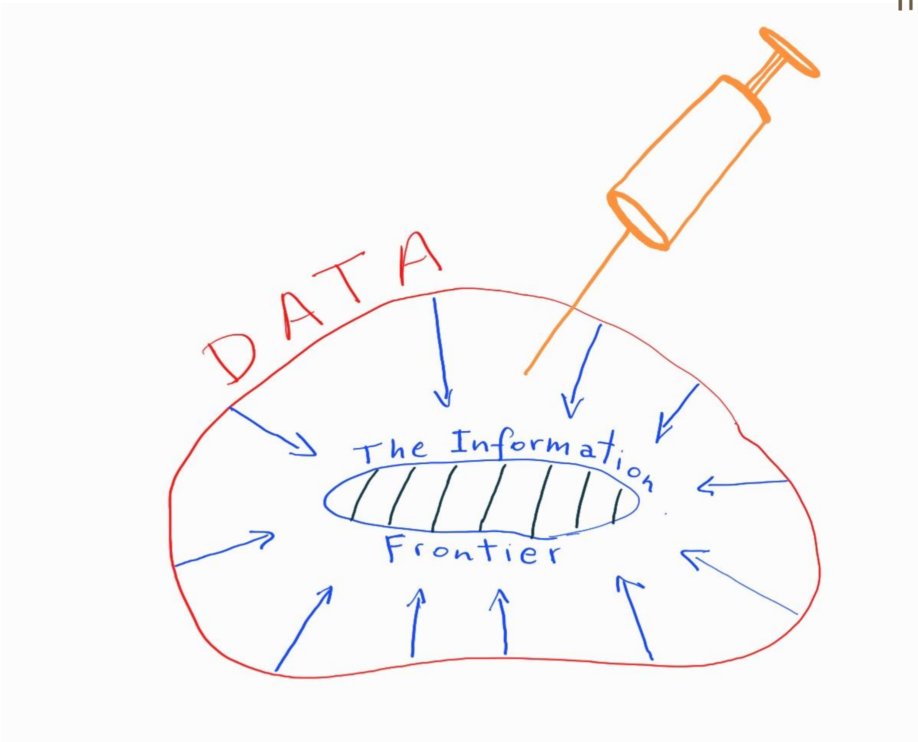
The Information Frontier



The Information Frontier



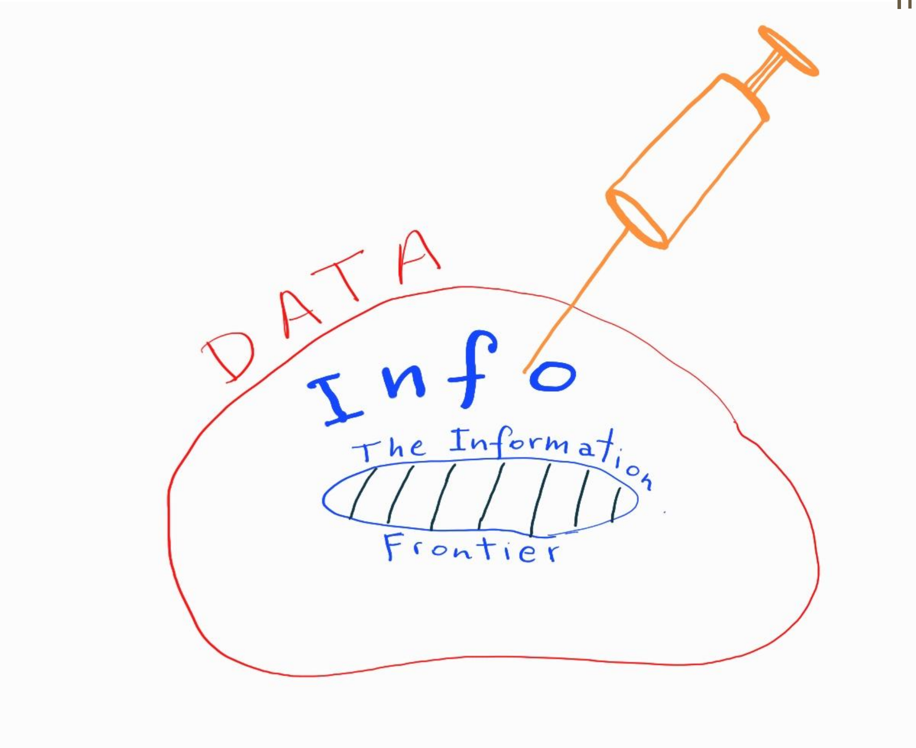
The Information Frontier



Inject *catalysts*:

- Modeling
- Tools & techniques
- Prior info

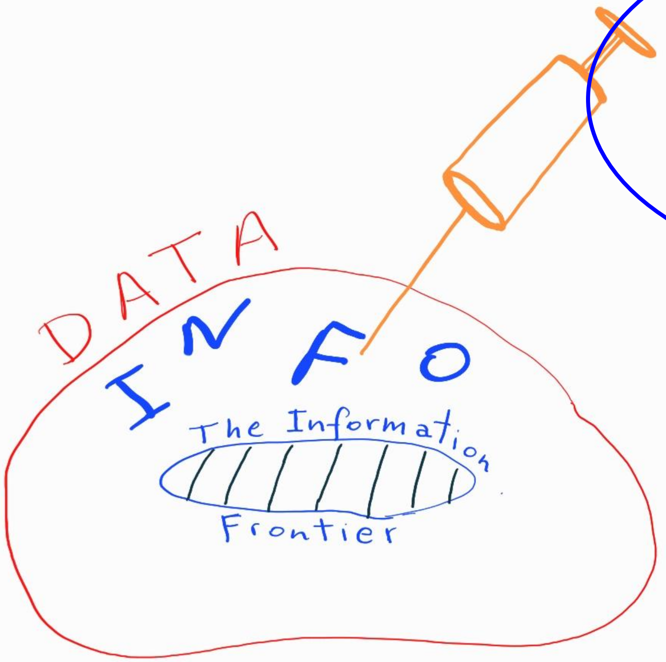
The Information Frontier



Inject *catalysts*:

- Modeling
- Tools & techniques
- Prior info

The Information Frontier

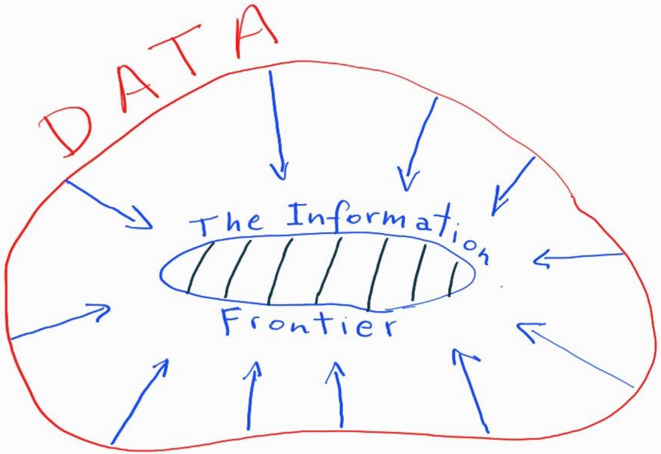


Inject *catalysts*:

- Modeling
- Tools & techniques
- Prior info

The Information Frontier

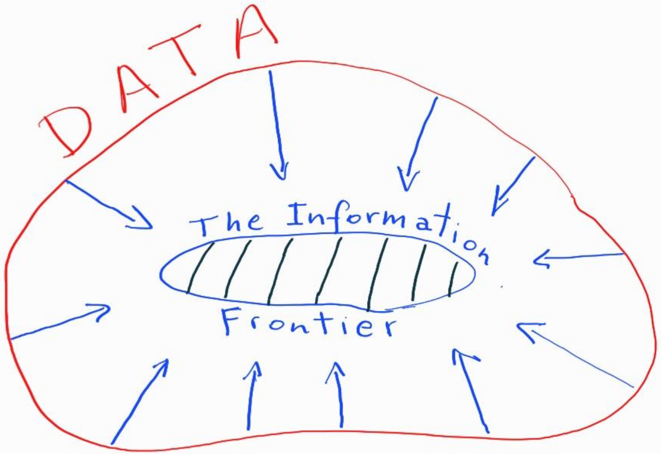
LHC data is very complex and sophisticated



Different tools can explore differently this frontier

The Information Frontier

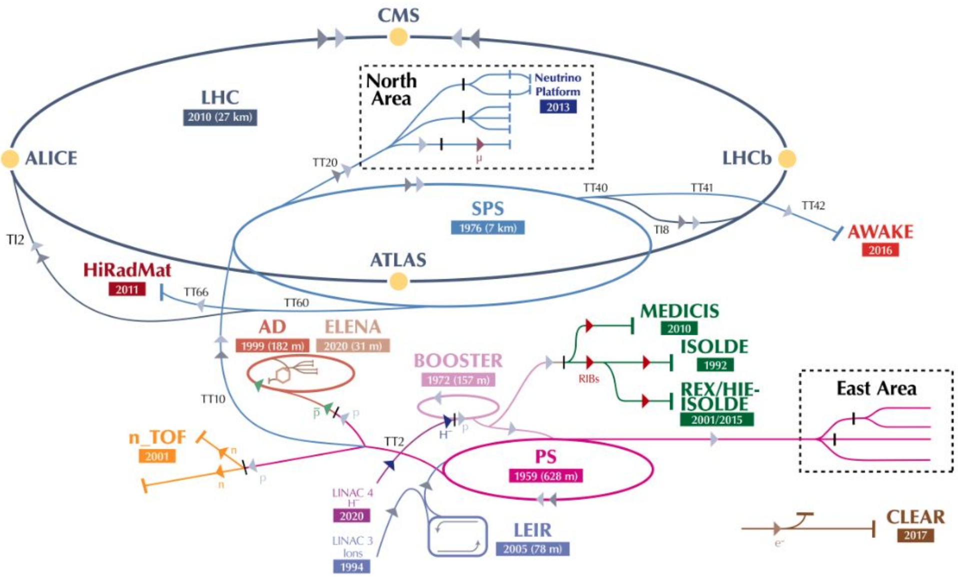
LHC data is very complex and sophisticated



Different tools can explore differently this frontier

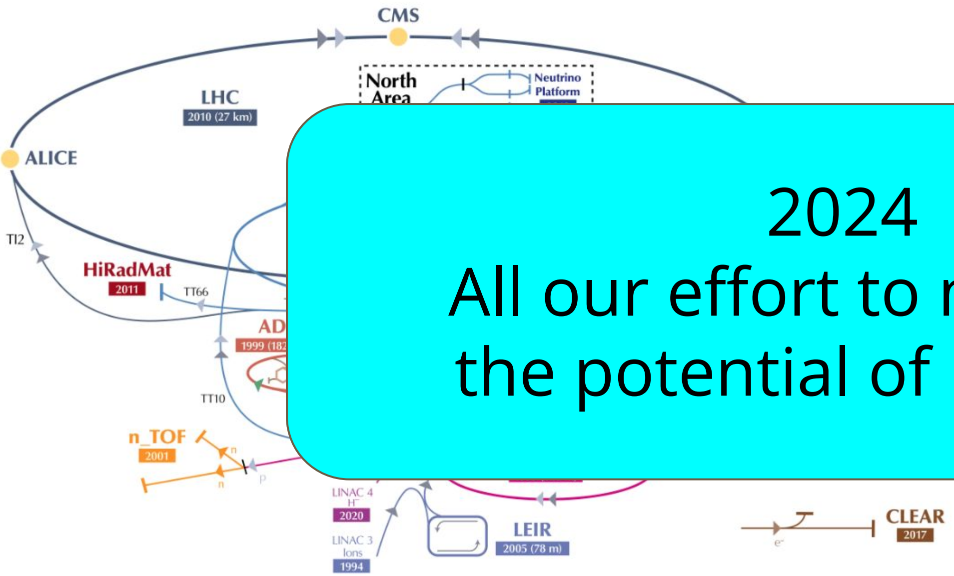
How much effort should we devote to moving the frontier?

LHC History



- 1980's: proposal
- 1995: Approved
- 2008: Started
- Huge effort in coordinating technology achievements
- Outstanding effort in all fields to reach one of the the most outstanding machines ever built by mankind

LHC History

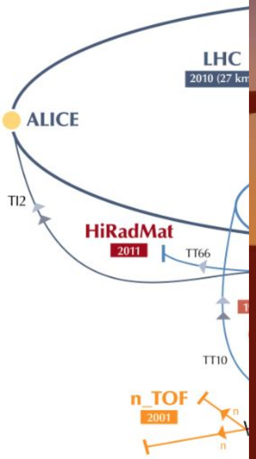


- 1980's: proposal
- 1995: Approved
- 2008: Started

n coordinating
chievements
effort in all fields
of the the most
machines ever built

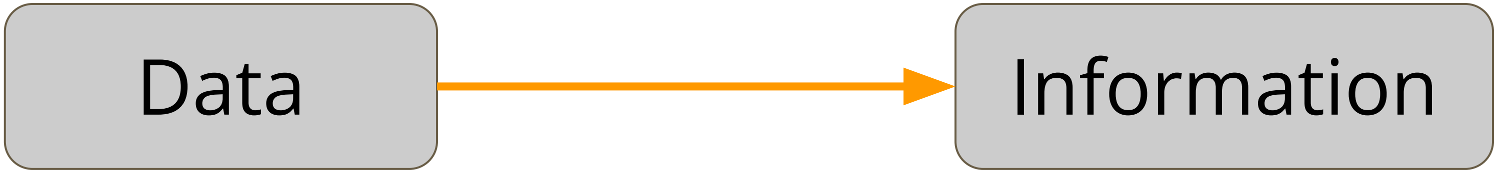
2024
All our effort to maximize
the potential of LHC data

The carrot of a Discovery!

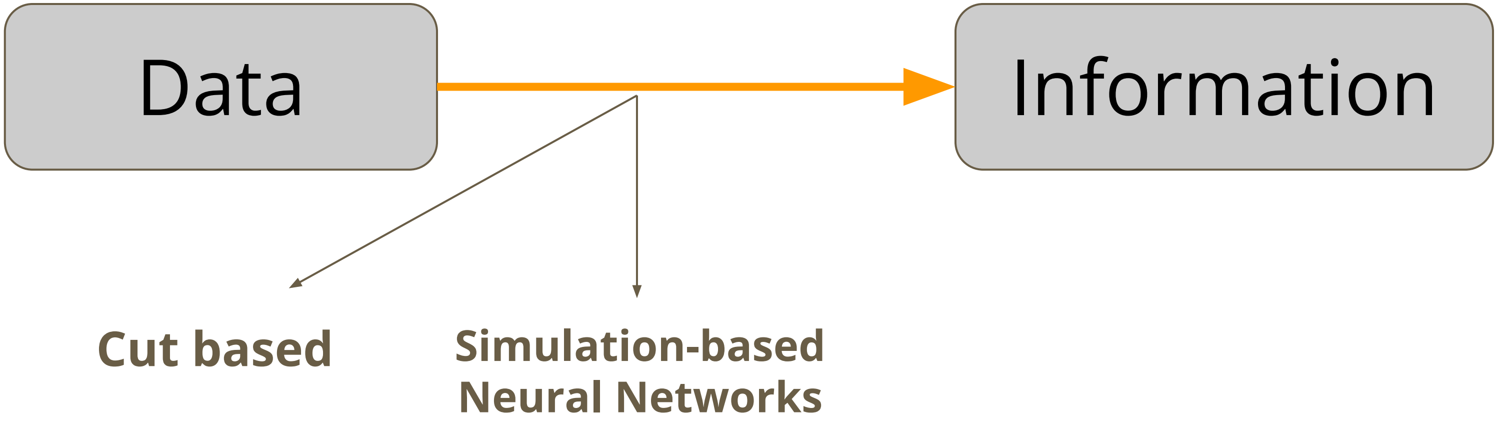


proposal
pproved
arted
n coordinating
chievements
effort in all fields
of the the most
machines ever built

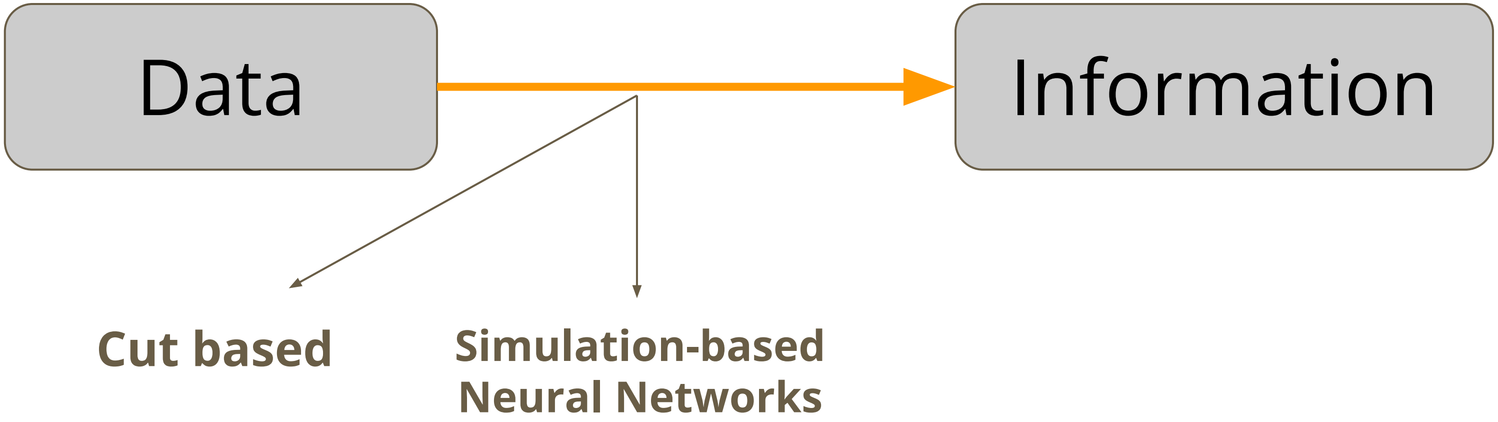
Typical problem at the LHC



Typical problem at the LHC

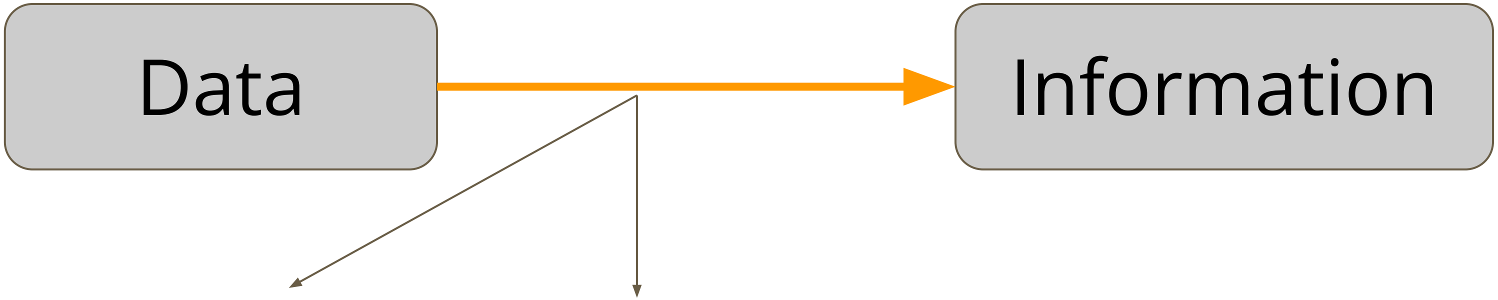


Typical problem at the LHC



Misses to take benefit of correlation at the event-by-event level

Typical problem at the LHC



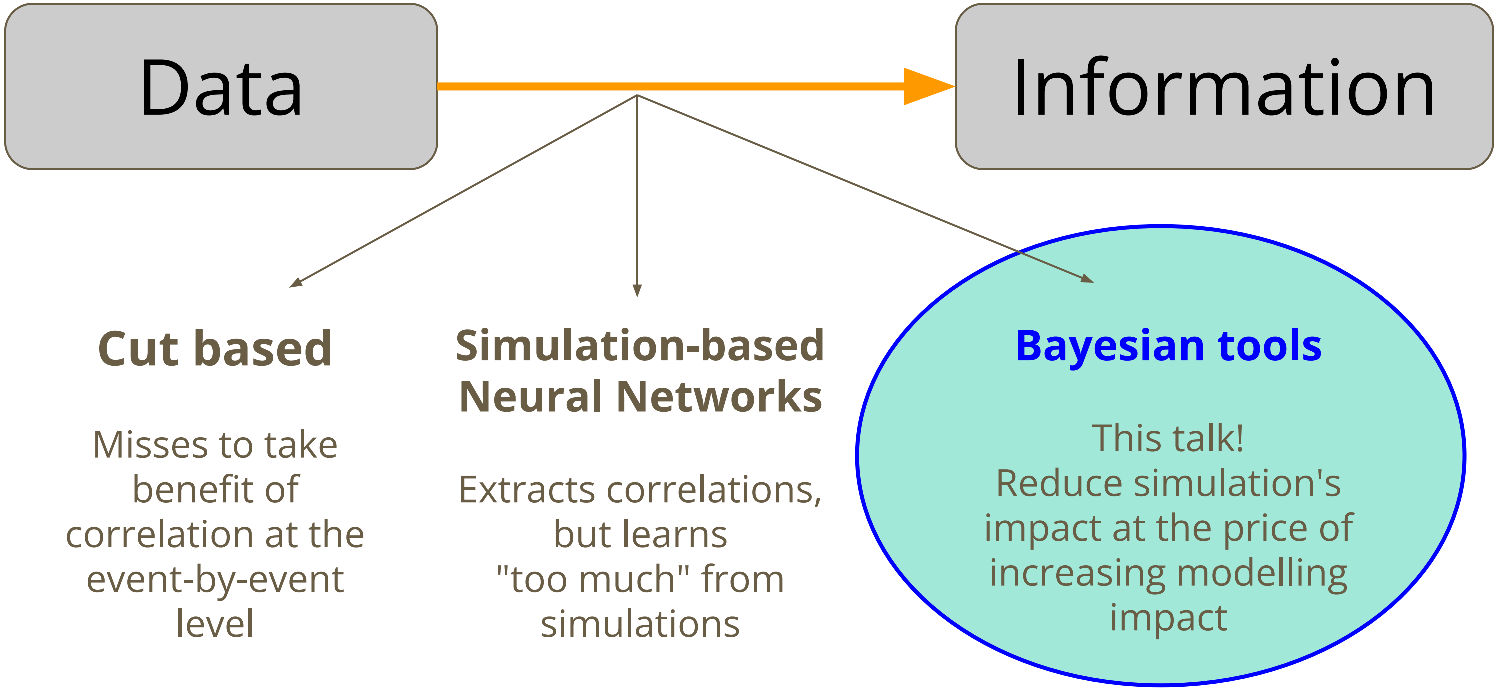
Cut based

Misses to take benefit of correlation at the event-by-event level

Simulation-based Neural Networks

Extracts correlations, but learns "too much" from simulations

Typical problem at the LHC



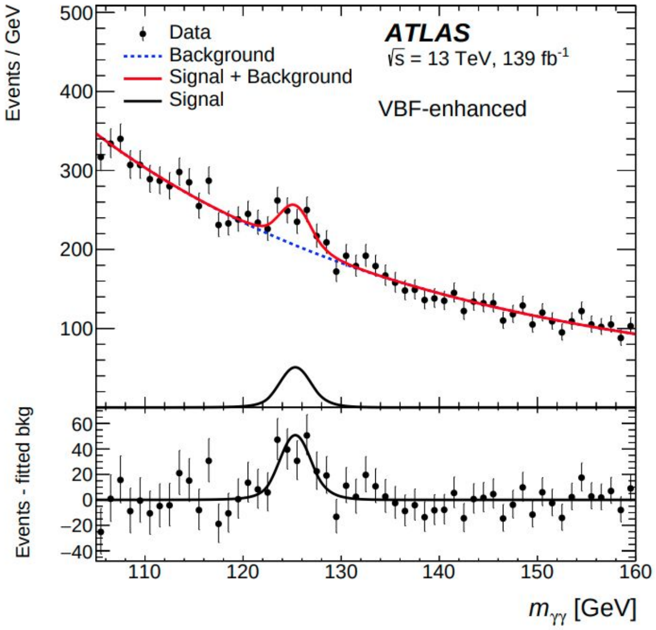
Summary

- Intro
- ABCD method
- Bayesian techniques
  - No Signal & Background region
  - No hard assignment
  - Probability, correlation and prior-knowledge
- $pp \longrightarrow hh \longrightarrow bbbb$ (inspiration & chimera)
  - Compare ABCD Vs Bayesian: multidimensionality!
  - Exploiting prior knowledge
    - Continuity
    - Unimodality
- Outlook
- Conclusions

ABCD Method

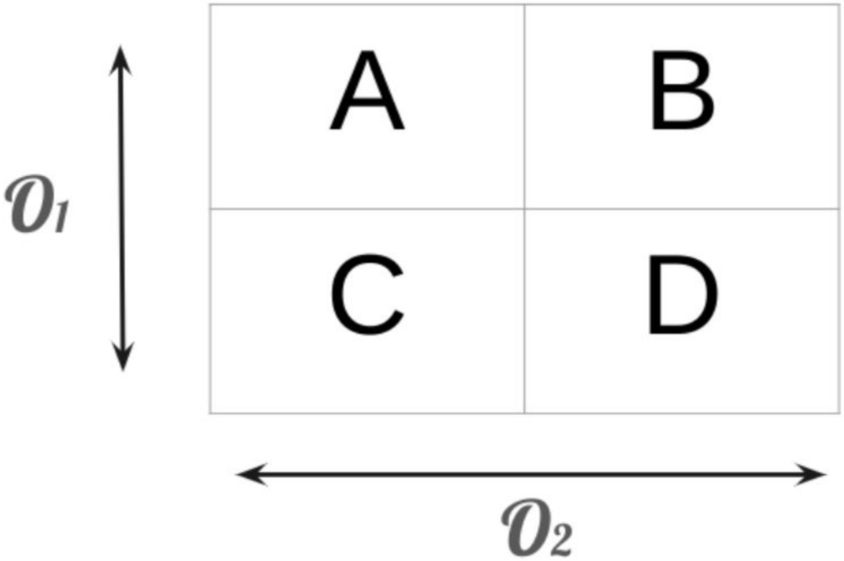
Data-driven methods

If simulations are not reliable



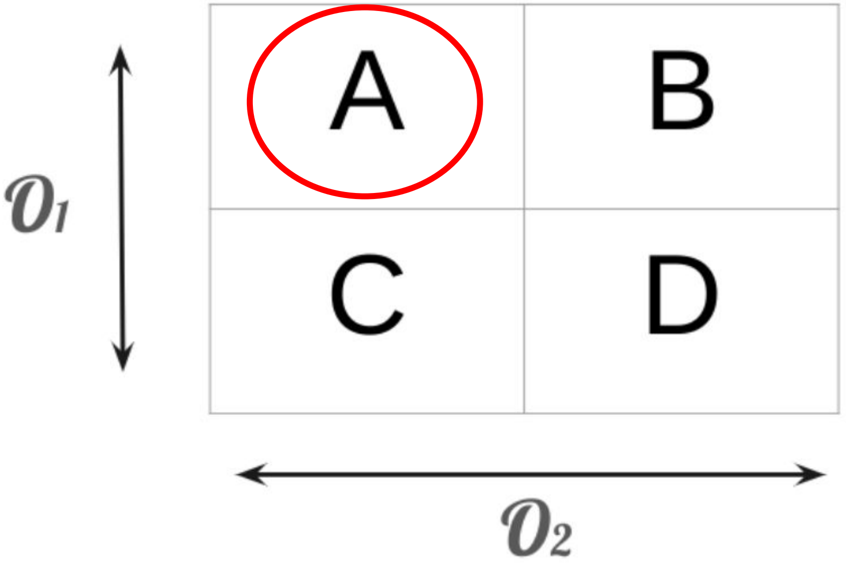
Data-driven methods

If simulations are not reliable



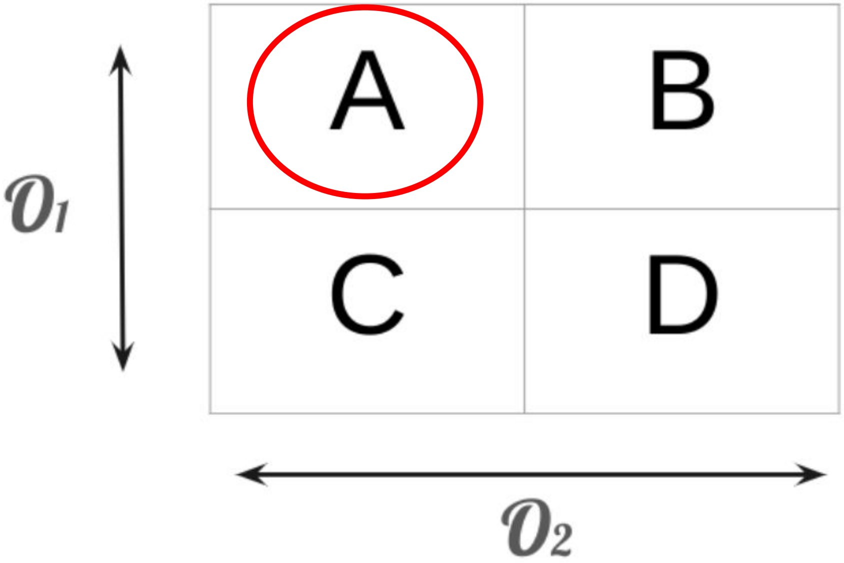
Data-driven methods

If simulations are not reliable



Signal is only in A and its background is easily estimated from the "*control regions*"

ABCD data-driven method



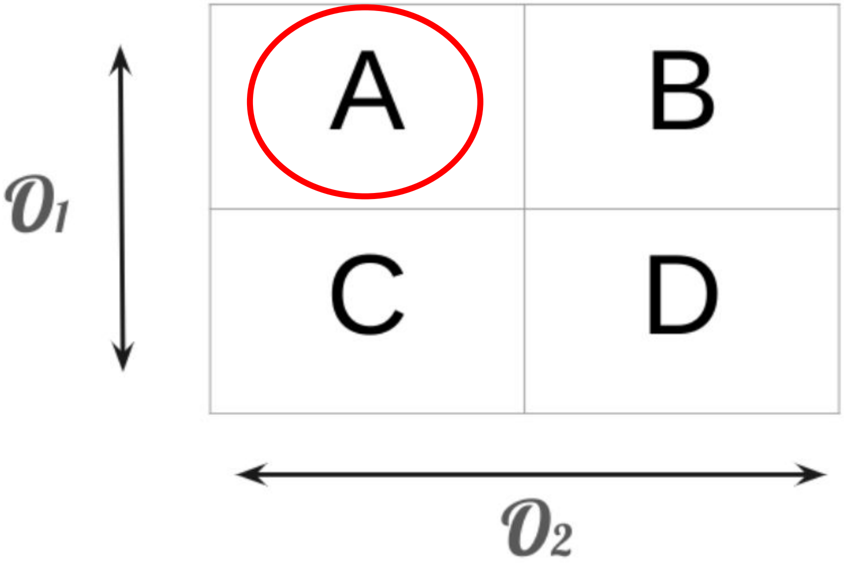
Quite simple

- 2 independent observables
- Signal restricted to A
- Immediately:

$N_A(\text{background}) = N_B * N_C / N_D$

$\text{Signal} = N_A - N_A(\text{background})$

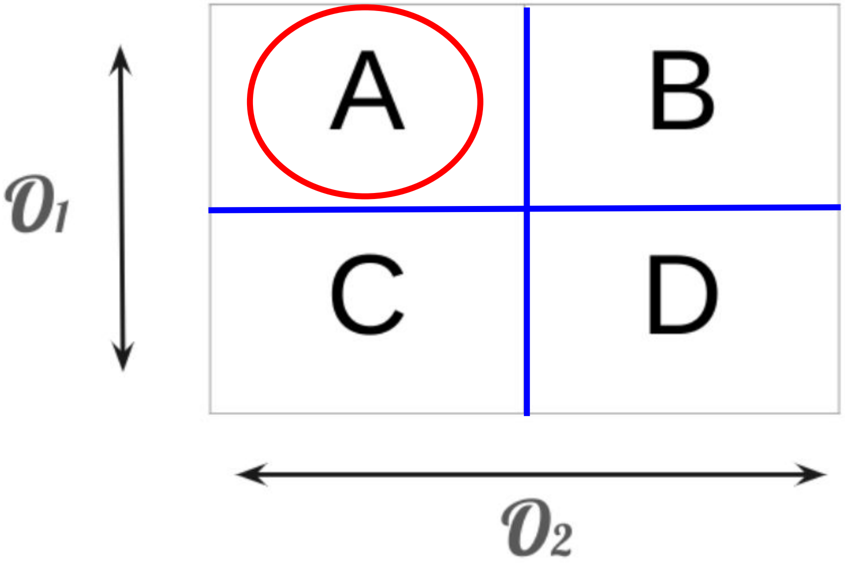
ABCD data-driven method



To notice:

- Prior-knowledge to define A, B, C & D

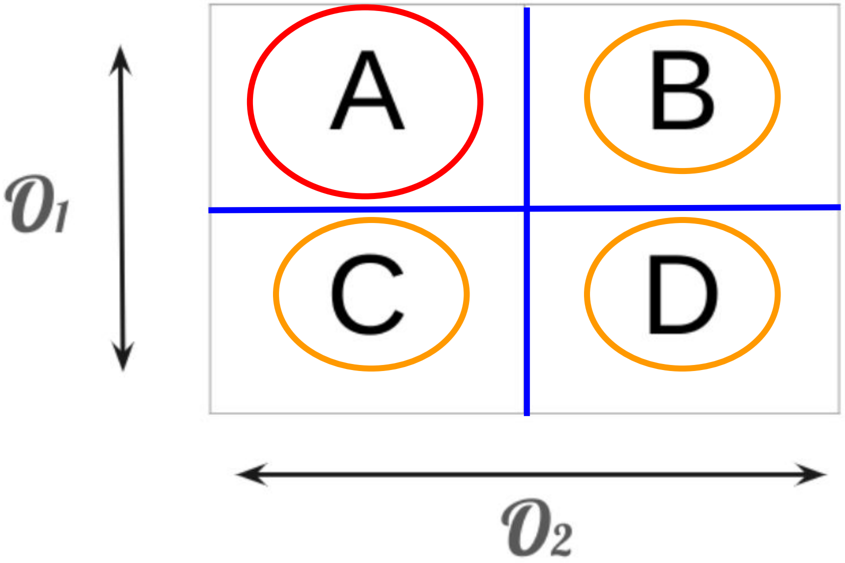
ABCD data-driven method



To notice:

- Prior-knowledge to define A, B, C & D
- Hard cuts (hard assignments)

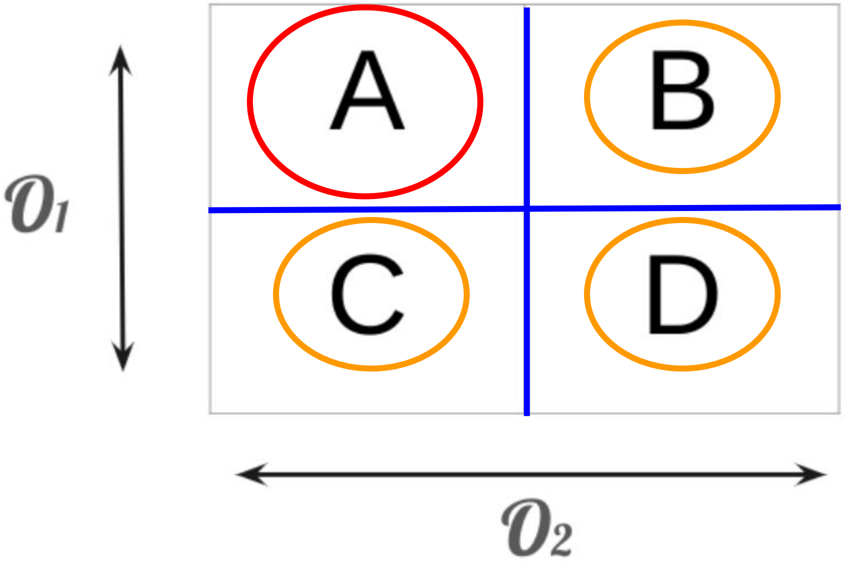
ABCD data-driven method



To notice:

- Prior-knowledge to define A, B, C & D
- Hard cuts (hard assignments)
- Signal and Background regions

ABCD data-driven method



To notice:

- Prior-knowledge to define A, B, C & D
- Hard cuts (hard assignments)
- Signal and Background regions
- Naturally conflictive hypotheses:
  - Regions close-by to have same distributions
  - Regions far away to avoid signal contamination

Bayesian Inference

2402.08001

E.A., L.Da Rold, S. Tanco, M, Szewc, A. Szynkman, S. Tanco, T. Tarutina

Bayesian Inference

Bayes Theorem:

$$p(\theta | X) = \frac{p(X | \theta) * p(\theta)}{p(x)}$$

Bayesian Inference

Bayes Theorem:

$$p(\theta|X) = \frac{p(X|\theta) * p(\theta)}{p(x)}$$

Real utility: (X = data, θ=parameters)

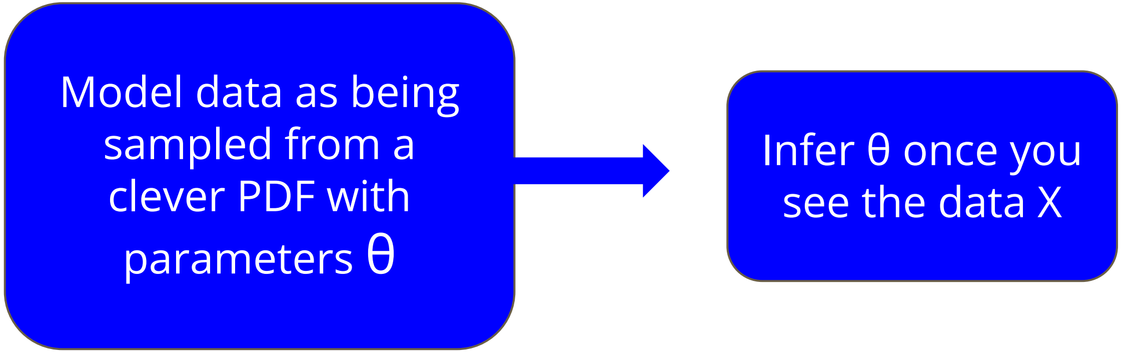
Model data as being sampled from a clever PDF with parameters θ

Bayesian Inference

Bayes Theorem:

$$p(\theta|X) = \frac{p(X|\theta) * p(\theta)}{p(x)}$$

Real utility: (X = data, θ=parameters)

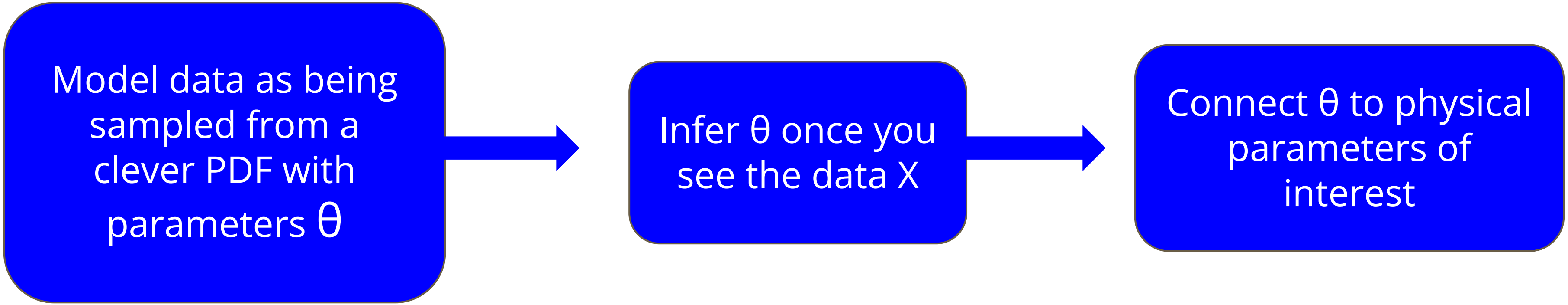


Bayesian Inference

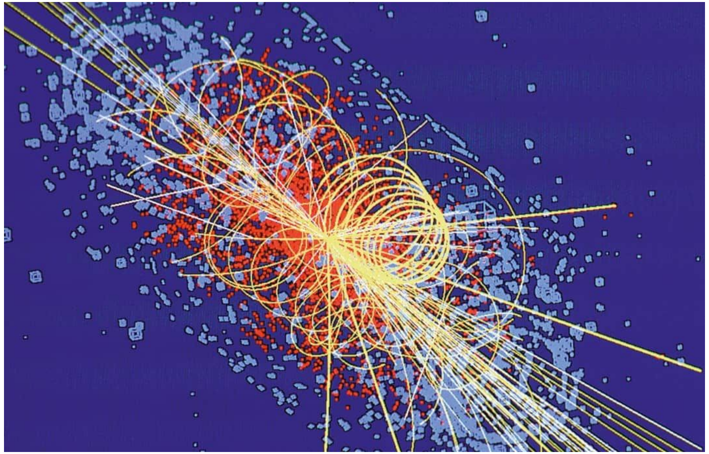
Bayes Theorem:

$$p(\theta|X) = \frac{p(X|\theta) * p(\theta)}{p(x)}$$

Real utility: (X = data, θ=parameters)



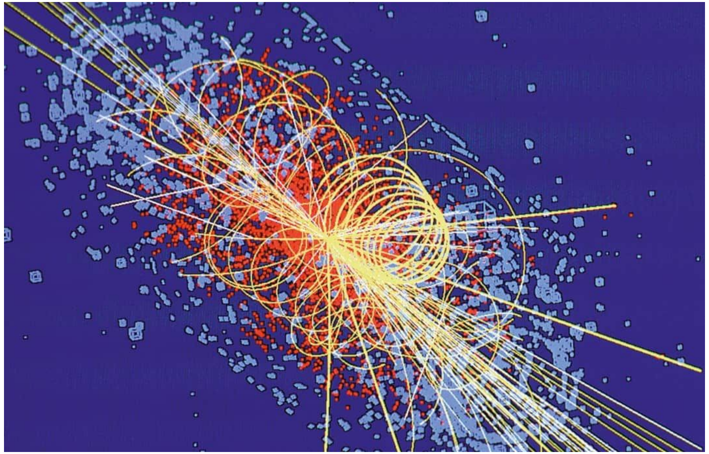
Bayesian Inference: Mixture models



Dataset X:

- Signal
- Few backgrounds

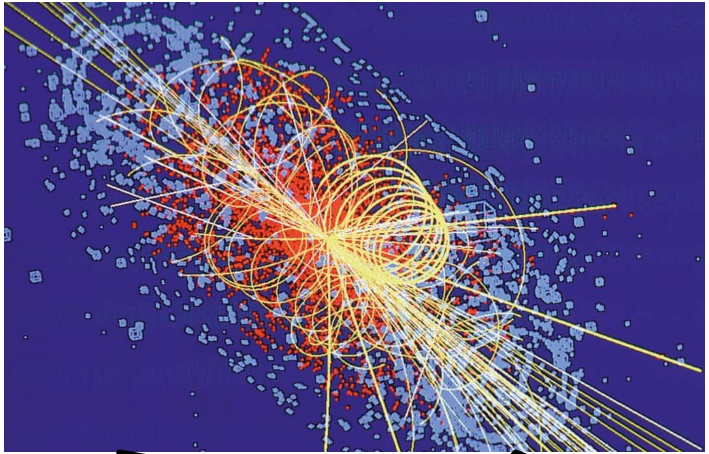
Bayesian Inference: Mixture models



Dataset X:

- Signal
- Few backgrounds
- Each event is either
  signal or one of the backgrounds

Bayesian Inference: Mixture models



Dataset X:

- Signal
- Few backgrounds
- Each event is either

  signal or one of the backgrounds

How to create such a PDF !?

Bayesian Inference: Graphical models

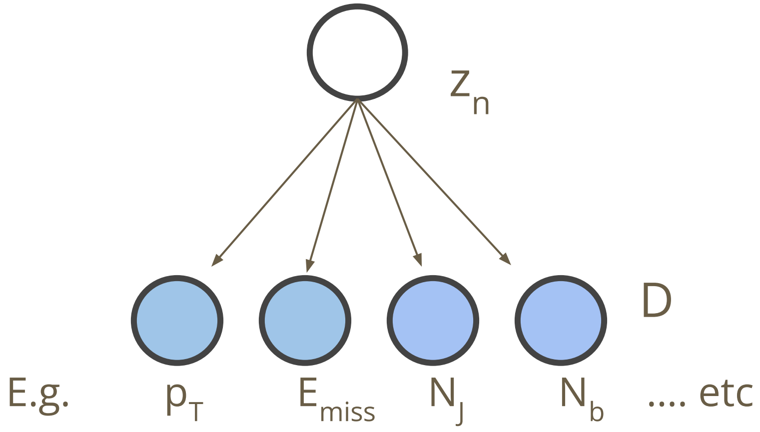
Graphical representation of a PDF to easily visualize the internal structure of the random variables

Bayesian Inference: Graphical models

At each event, sample a multinomial random variable that decides whether is signal or some of the backgrounds

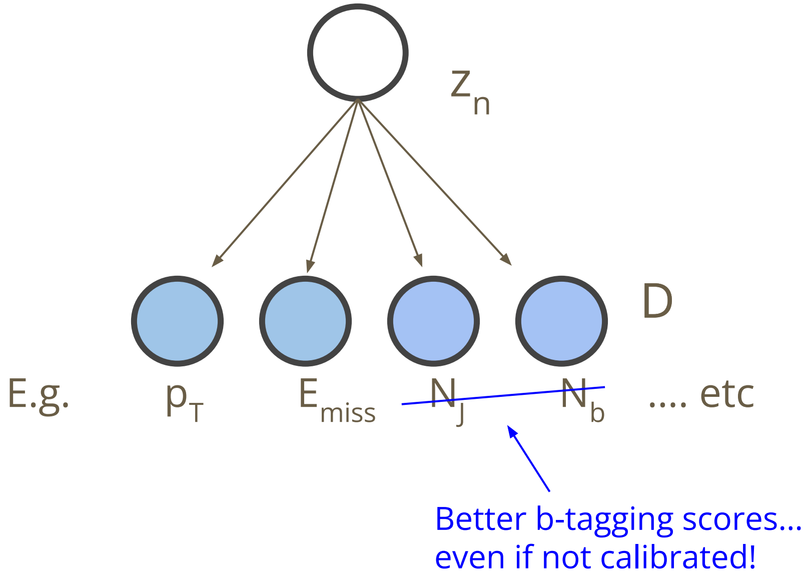
(K classes)

Bayesian Inference: Graphical models



Depending on the class of the event, we sample D random independent variables of what *we measure*

Bayesian Inference: Graphical models



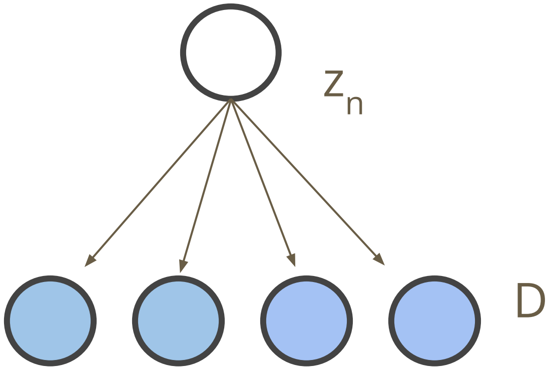
Depending on the class of the event, we sample D random independent variables of what *we measure*

Bayesian Inference: Graphical models

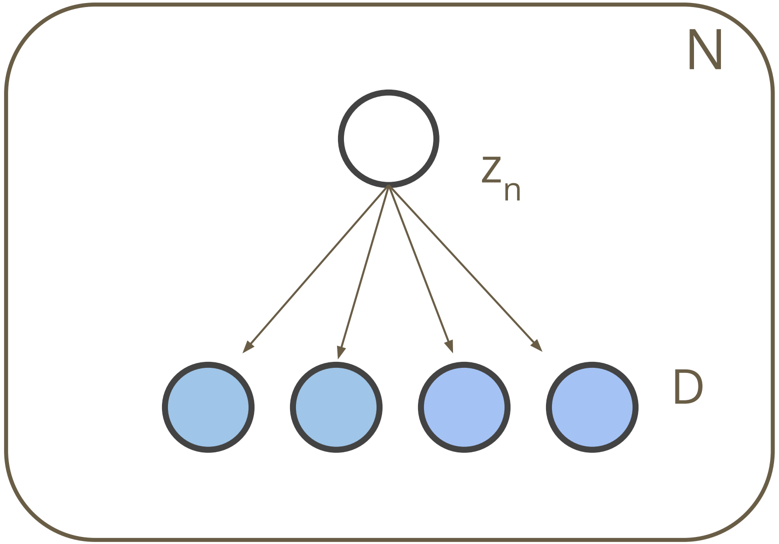
Convention:

Empty circles:
Sampled and unobserved RV

Filled circles:
Sampled and observed RV

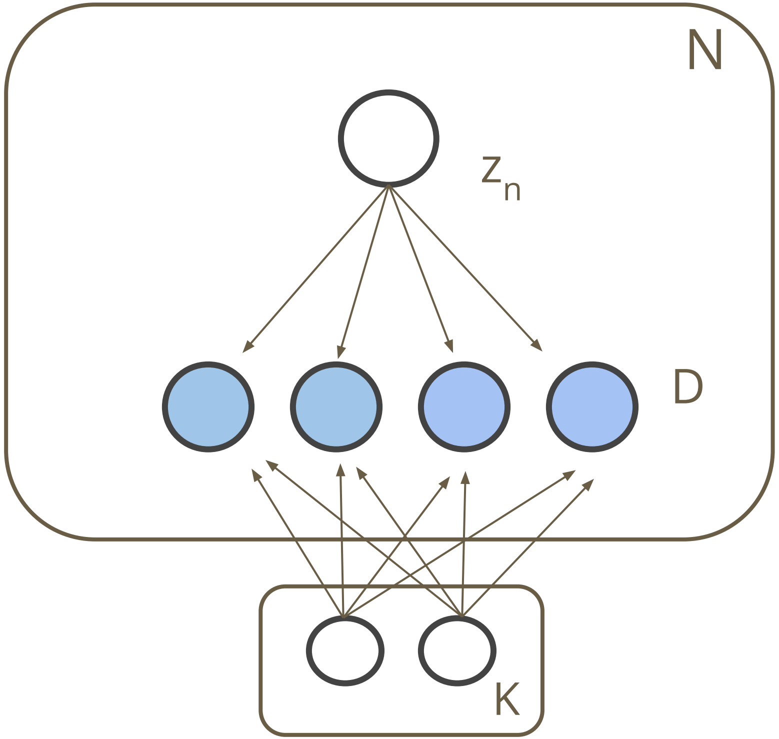


Bayesian Inference: Graphical models



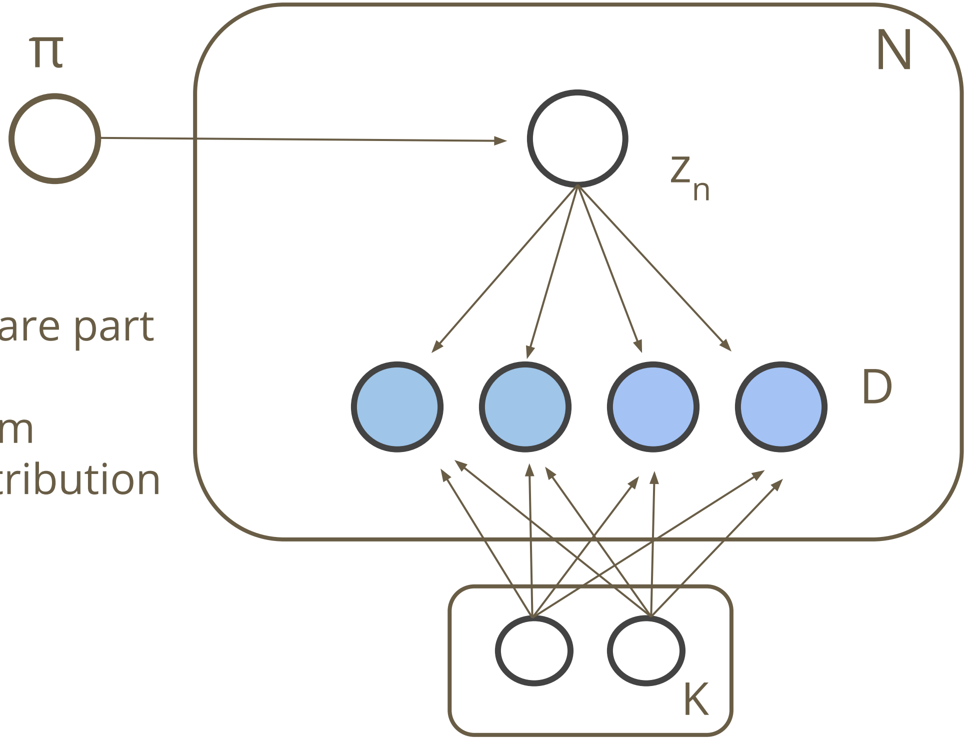
Procedure that is repeated N times

Bayesian Inference: Graphical models



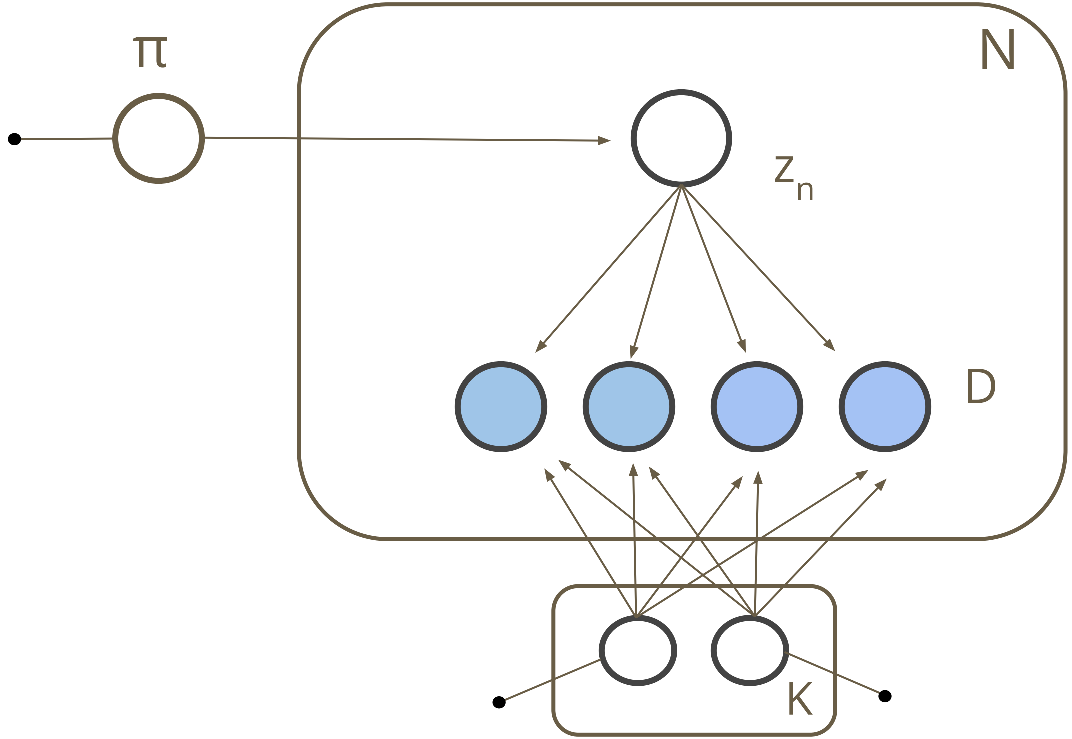
Each one of the K classes has an *expected distribution* over the measured quantities

Bayesian Inference: Graphical models



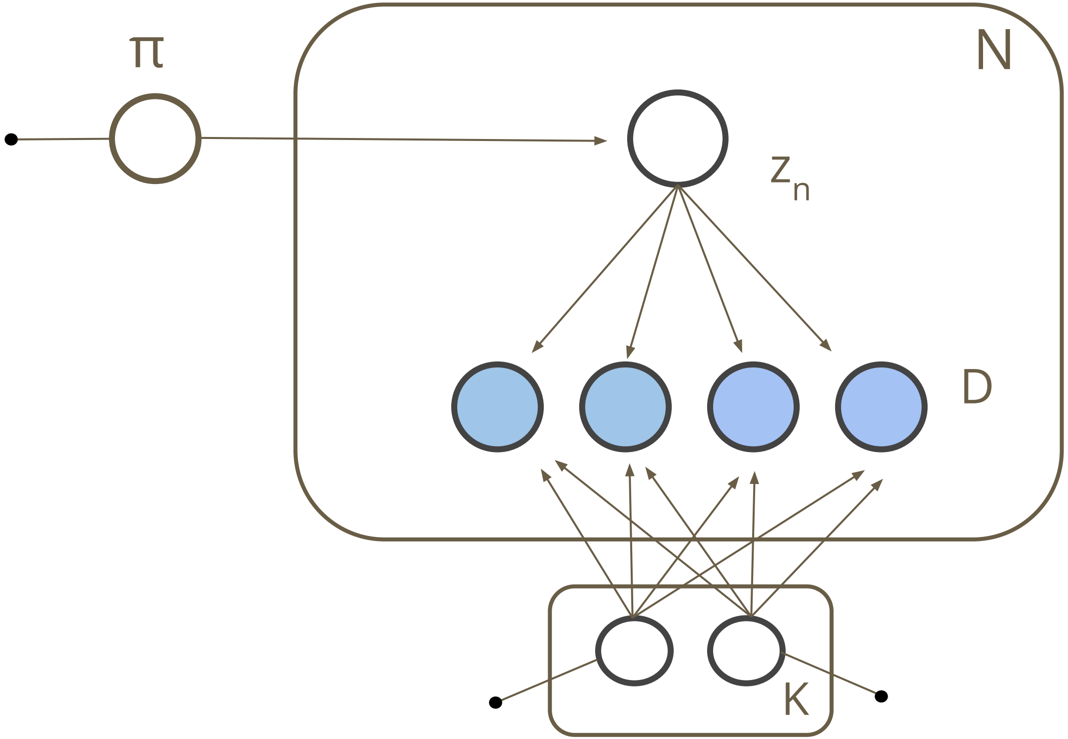
Multinomial parameters are part of the PDF. Sampled from Dirichlet distribution

Bayesian Inference: Graphical models



Mixture Model

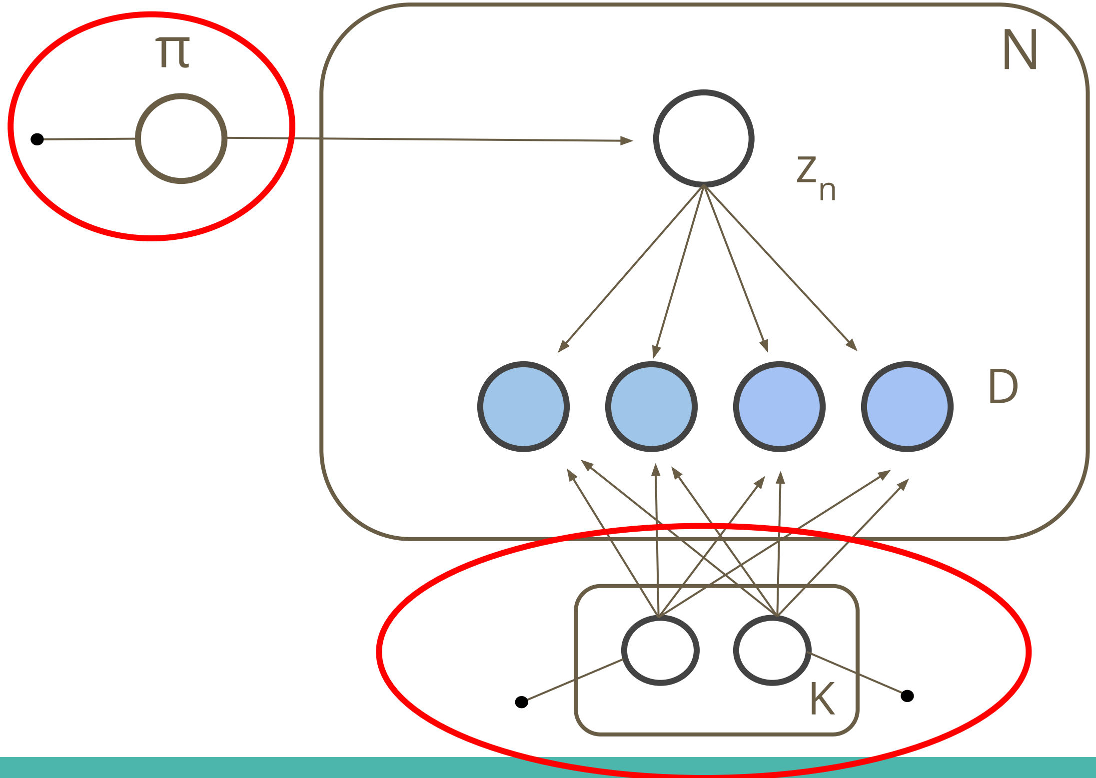
Bayesian Inference: Graphical models



Mixture Model

- Model data as being sampled from a PDF

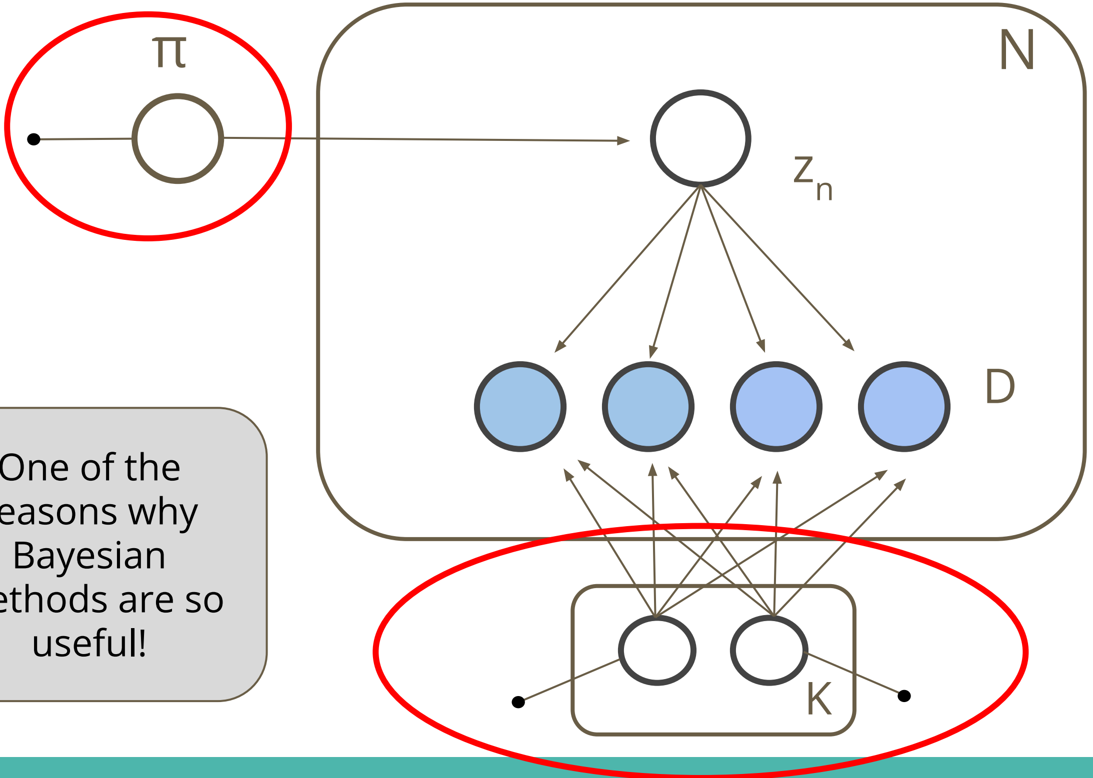
Bayesian Inference: Graphical models



Mixture Model

- Model data as being sampled from a PDF
- Plug our prior knowledge

Bayesian Inference: Graphical models

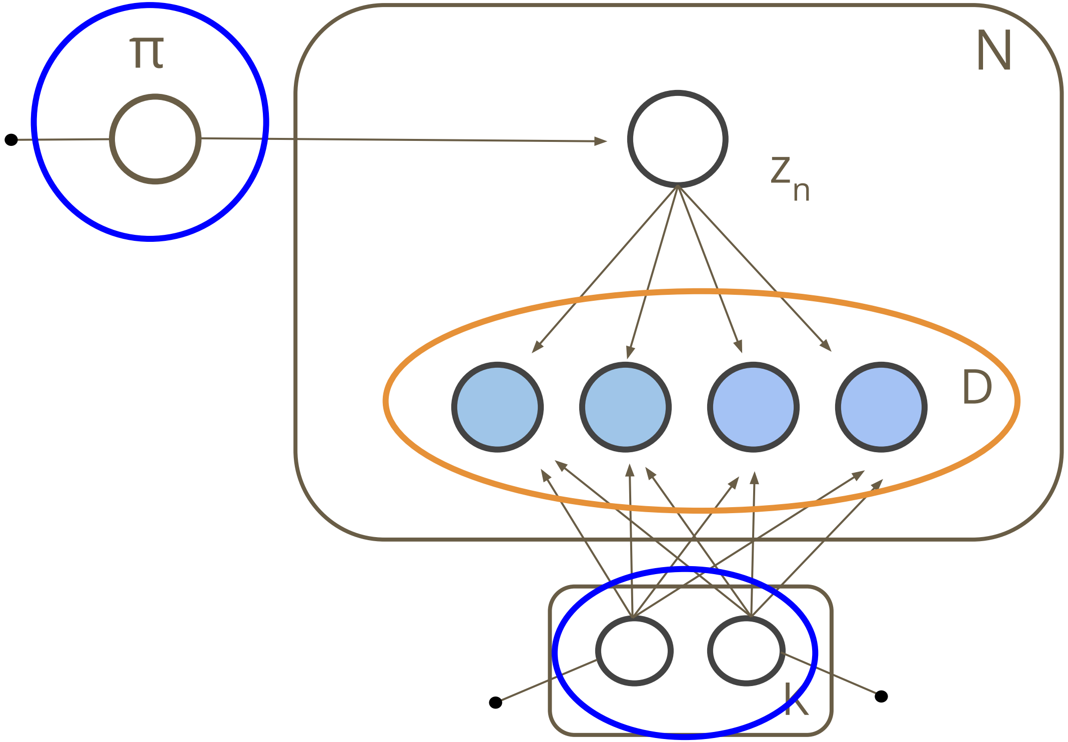


One of the reasons why Bayesian methods are so useful!

Mixture Model

- Model data as being sampled from a PDF
- Plug our prior knowledge

Bayesian Inference: Graphical models

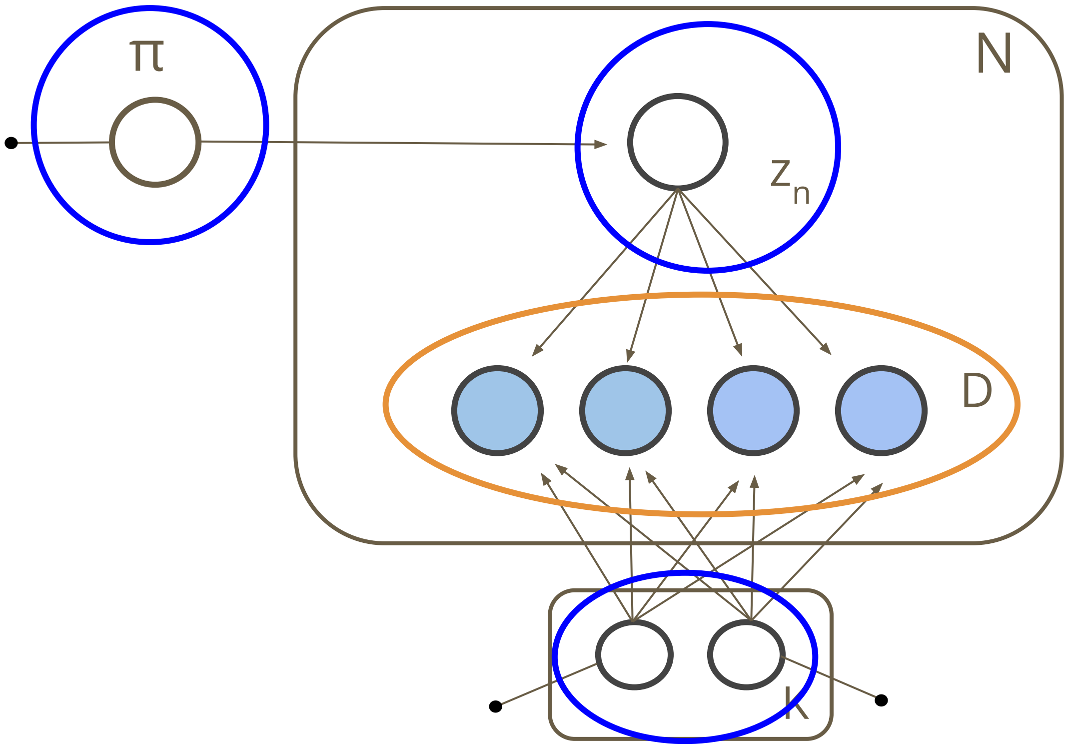


Mixture Model

- Model data as being sampled from a PDF
- Plug our prior knowledge
- Infer the parameters conditioned in the observed data

p(θ|X) = p(X|θ) p(θ) / p(x)

Bayesian Inference: Graphical models



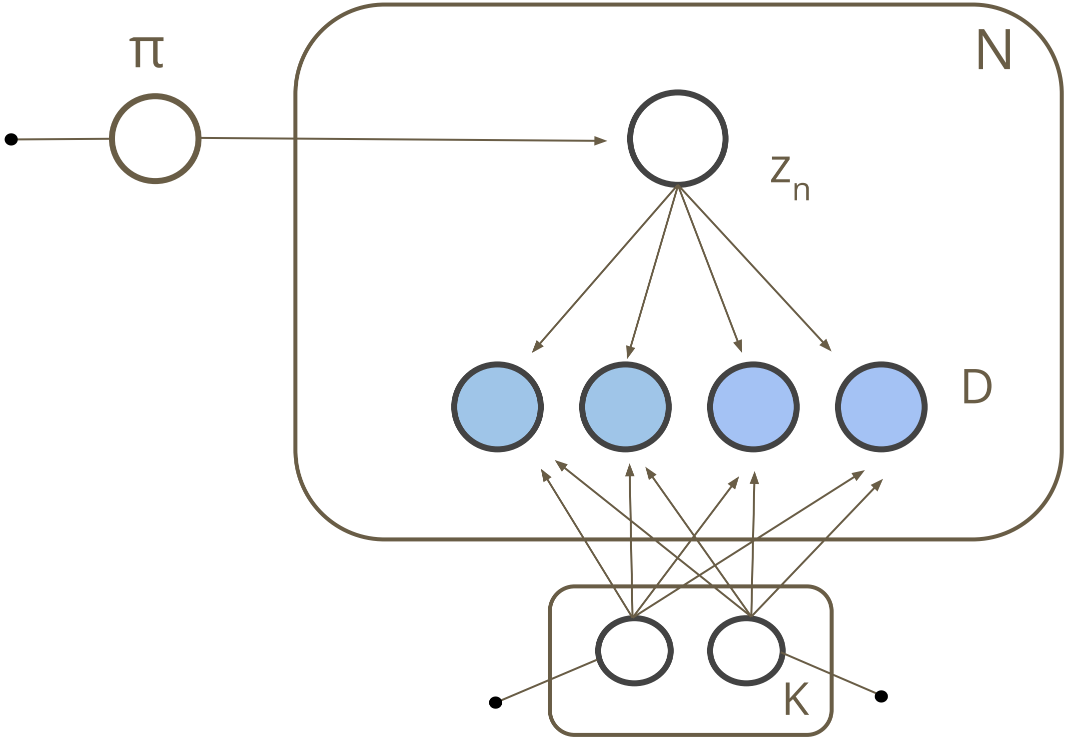
Mixture Model

- Model data as being sampled from a PDF
- Plug our prior knowledge
- Infer the parameters conditioned in the observed data

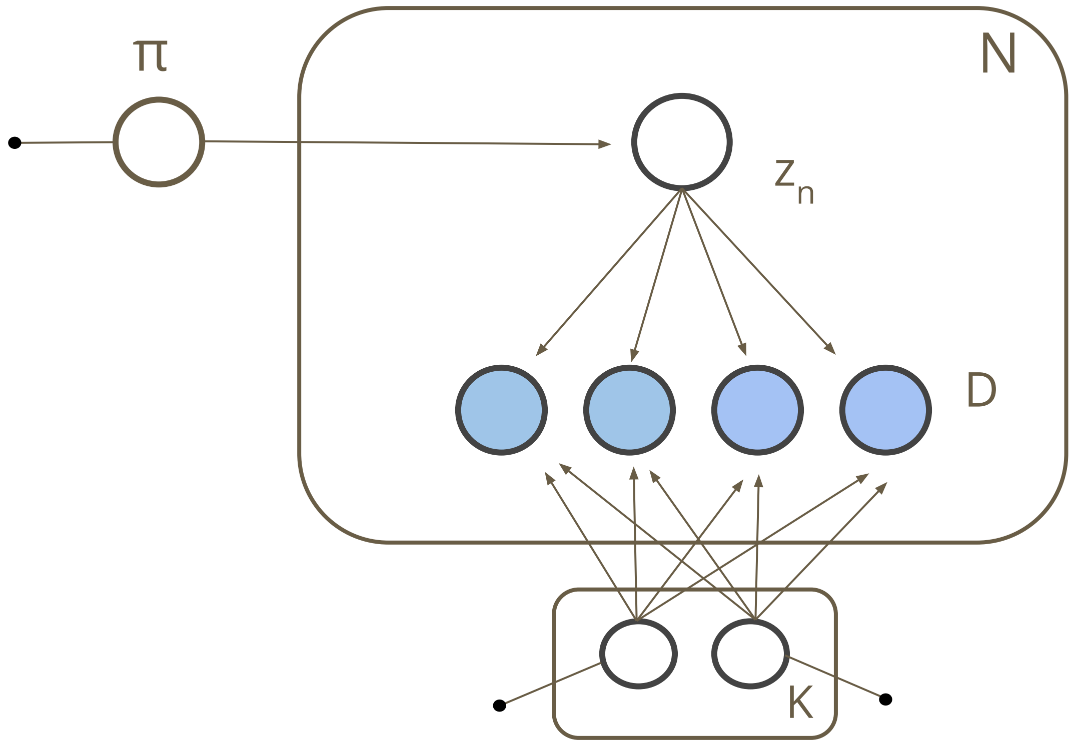
$$p(\theta|X) = p(X|\theta)\, p(\theta) / p(x)$$

- Infer the latent variables

Bayesian Inference

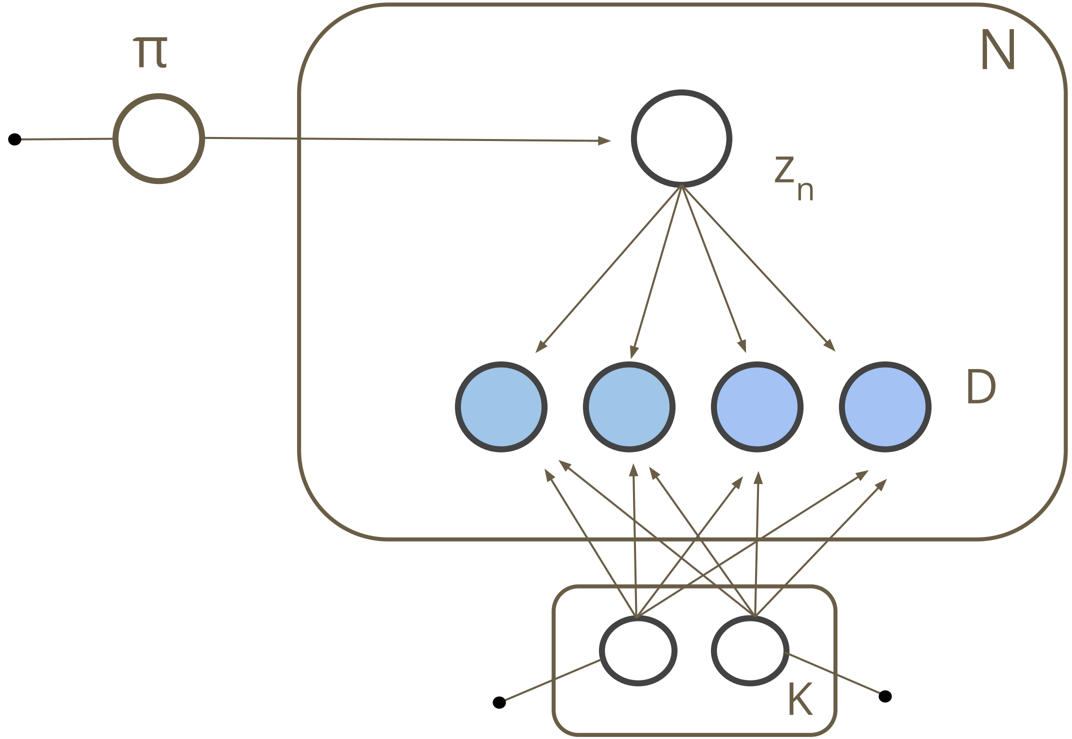


Bayesian Inference



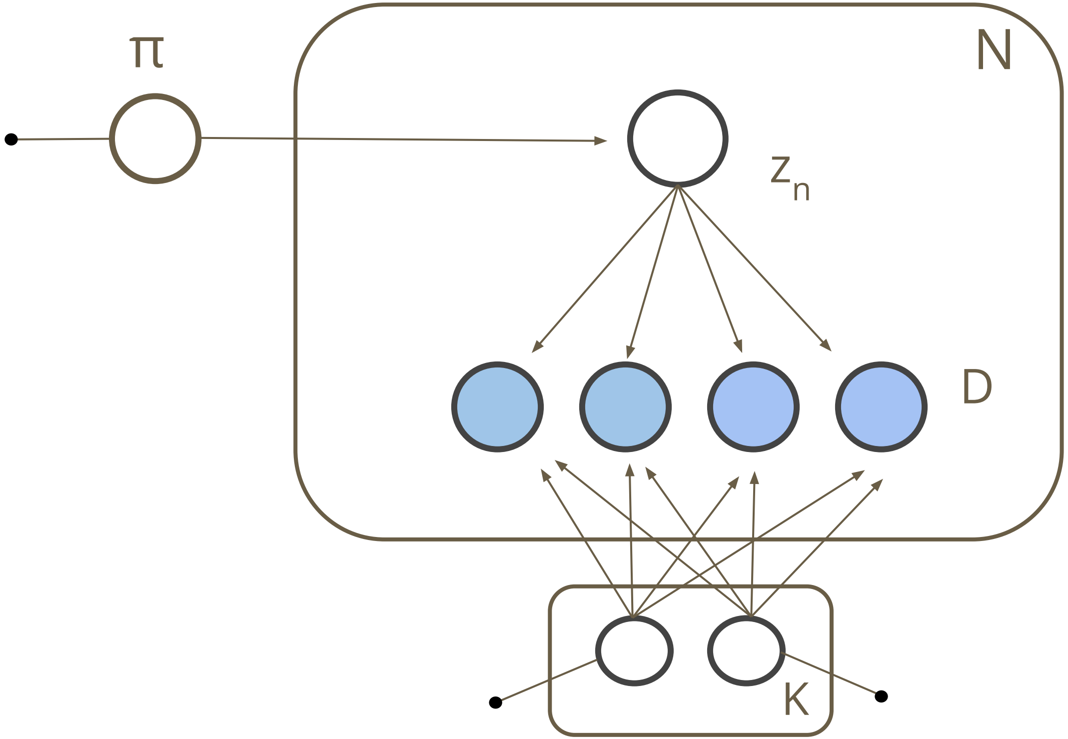
- No hard cuts

Bayesian Inference



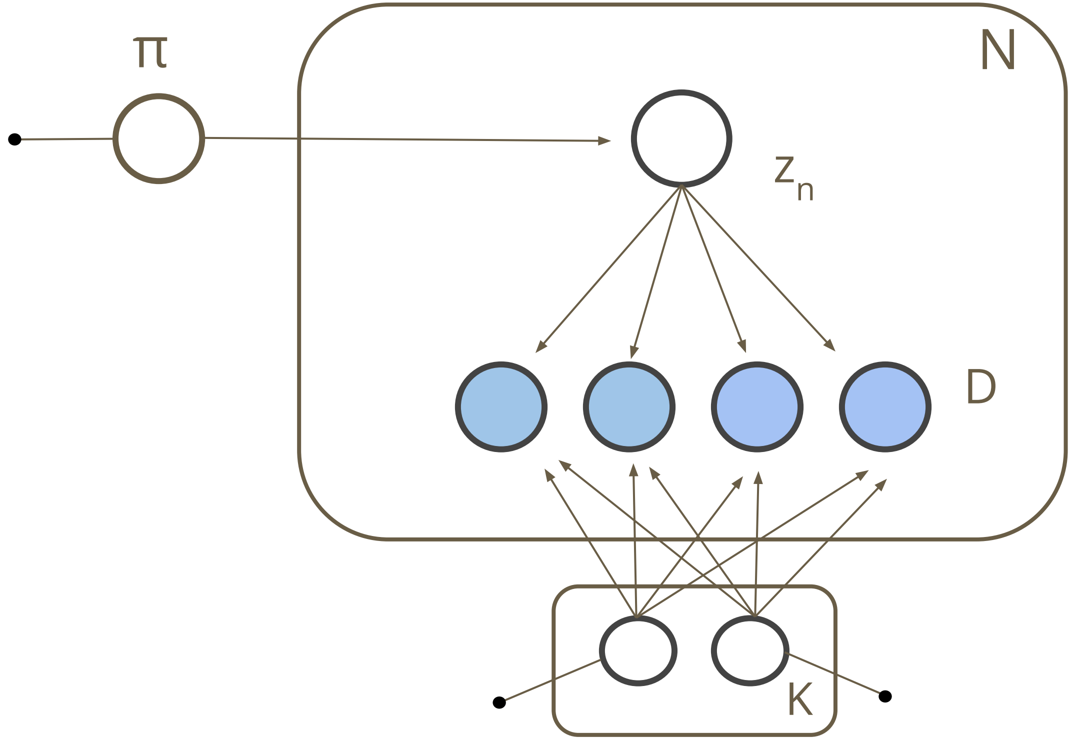
- No hard cuts
- Soft assignments

Bayesian Inference



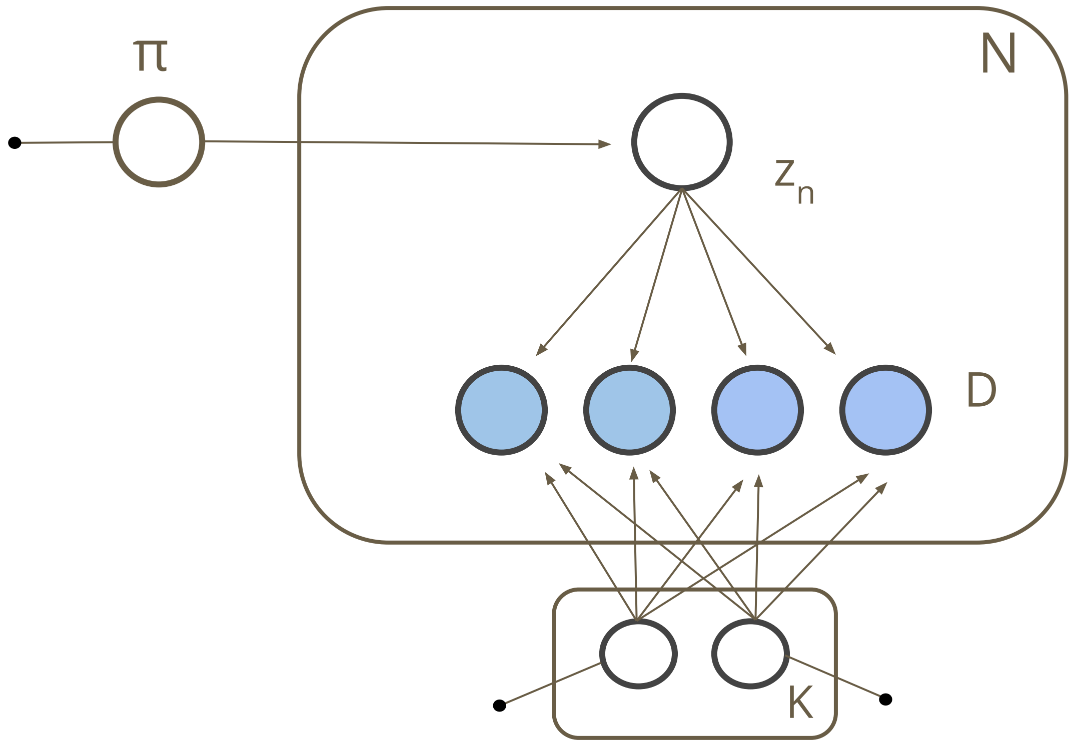
- No hard cuts
- Soft assignments
- No signal/control regions

Bayesian Inference



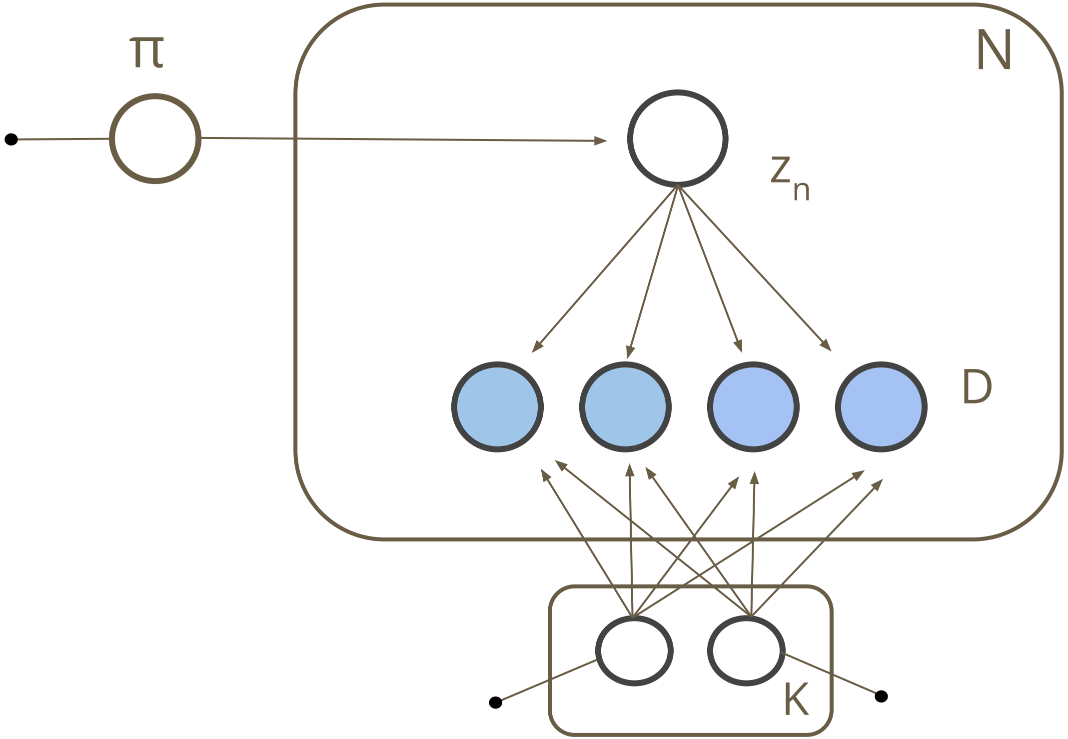
- No hard cuts
- Soft assignments
- No signal/control regions
- K classes & D observables

Bayesian Inference



- No hard cuts
- Soft assignments
- No signal/control regions
- K classes & D observables
- Deployment of data internal structure

Bayesian Inference



- No hard cuts
- Soft assignments
- No signal/control regions
- K classes & D observables
- Deployment of data internal structure
- Controlled injection of prior knowledge

ABCD Vs Bayesian

Improvement & Generalization

ABCD	Bayesian framework
2 observables	D observables
Signal & Background	Signal & K-1 backgrounds
Prior knowledge to define A, B, C & D, and get signal events in A	Visualize, understand and exploit *internal structure of the data*. Plug prior knowledge to *simultaneously* infer classes fractions and posterior distributions
Separated: signal & control regions	Signal & backgrounds can be mixed in all phase space.

$pp \to hh \to bbbb$

(inspiration & chimera)

2402.08001

E.A., L.Da Rold, S. Tanco, M, Szewc, A. Szynkman, S. Tanco, T. Tarutina

$pp \longrightarrow hh \longrightarrow bbbb$

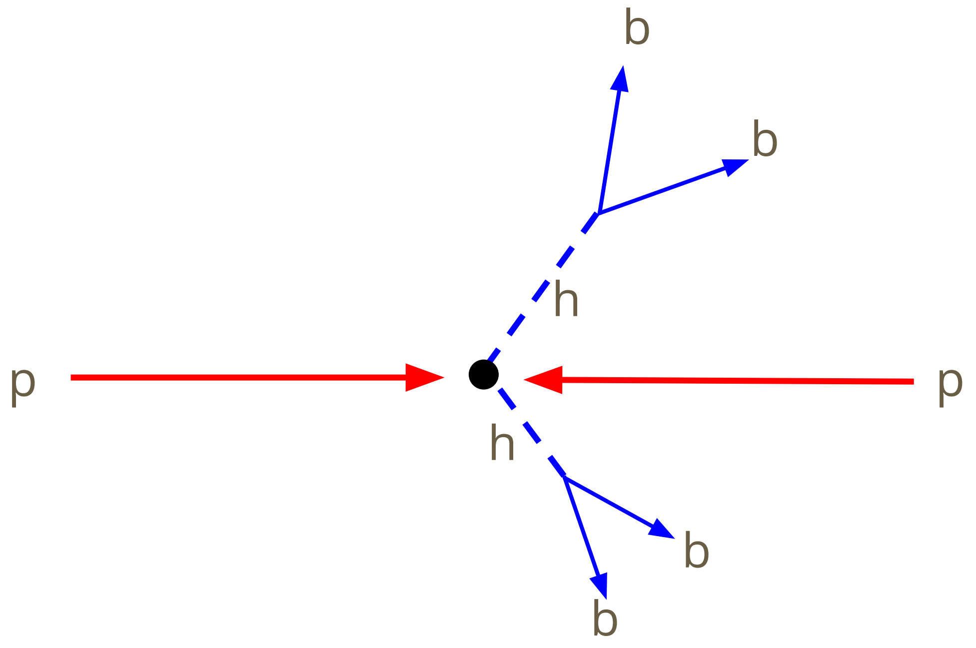
(inspiration & chimera)

Disclaimer:
Toy-model on a toy-problem, just a building block for a chimera enterprise

2402.08001

E.A., L.Da Rold, S. Tanco, M, Szewc, A. Szynkman, S. Tanco, T. Tarutina

$pp \longrightarrow hh \longrightarrow bbbb$

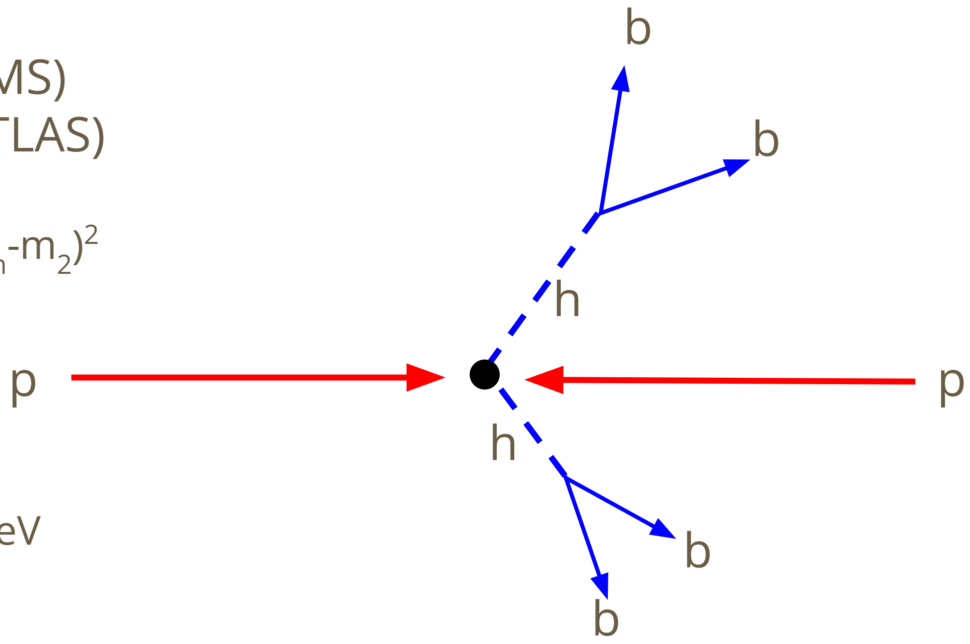


(pictorical)

$pp \longrightarrow hh \longrightarrow bbbb$

2202.09617 (CMS)
2301.03212 (ATLAS)

$\chi^2 = (m_h - m_1)^2 + (m_h - m_2)^2$



O_1: χ > or < 25 GeV

O_2: 3b or 4b

Plus *improvements*

$pp \longrightarrow hh \longrightarrow bbbb$

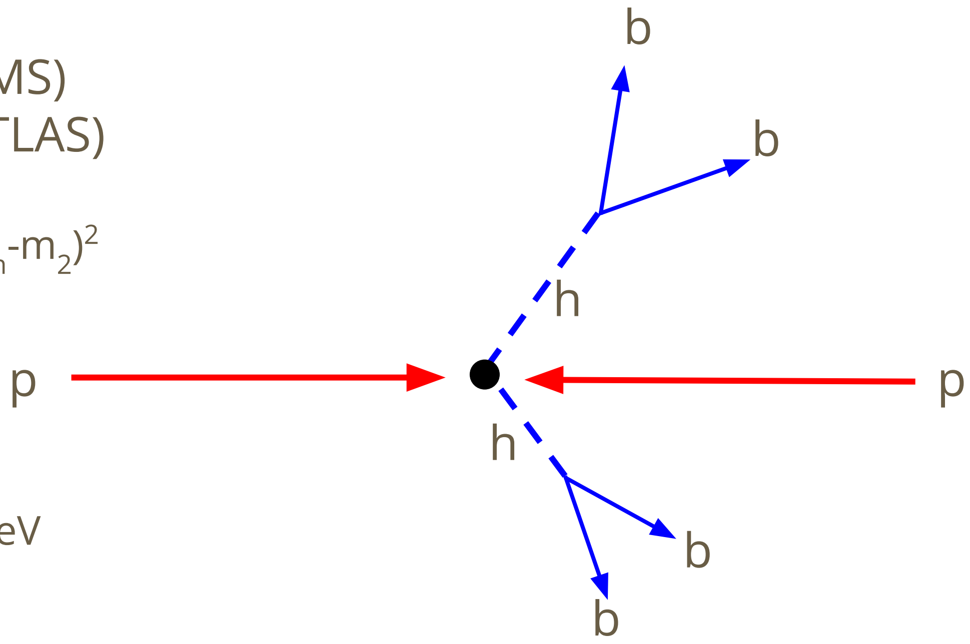
2202.09617 (CMS)
2301.03212 (ATLAS)

$\chi^2 = (m_h - m_1)^2 + (m_h - m_2)^2$

O_1: χ > or < 25 GeV

O_2: 3b or 4b

Plus *improvements*



Bayesian framework

O_1: b-tagging score jet 1

O_2: b-tagging score jet 2

O_3: b-tagging score jet 3

O_4: b-tagging score jet 4

O_5: m_1

O_6: m_2

$pp \longrightarrow hh \longrightarrow bbbb$

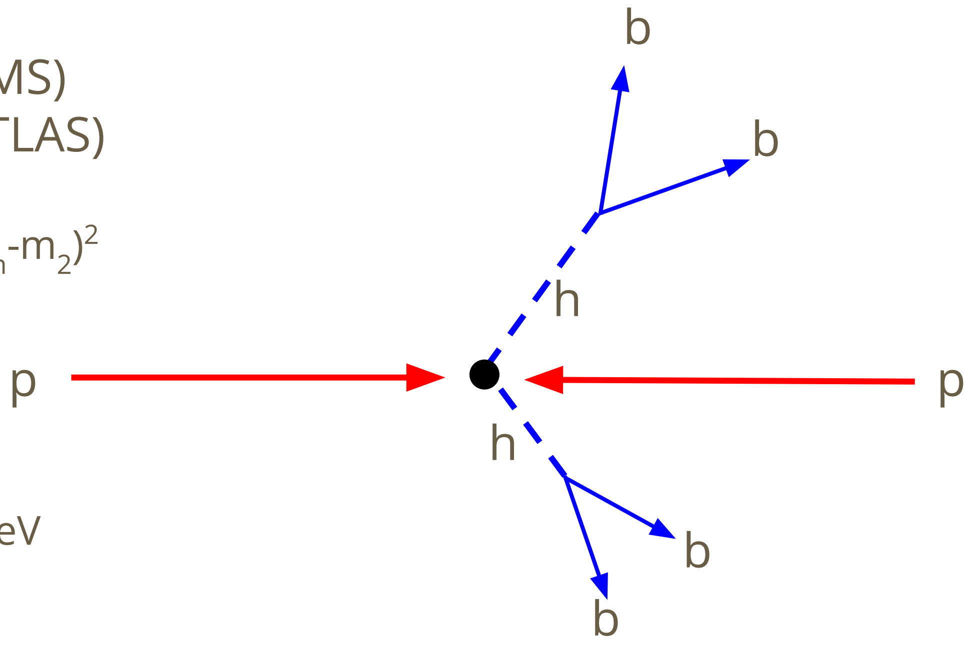
2202.09617 (CMS)
2301.03212 (ATLAS)

$\chi^2 = (m_h - m_1)^2 + (m_h - m_2)^2$

O_1: χ > or < 25 GeV

O_2: 3b or 4b

Plus *improvements*



Bayesian framework

O_1: b-tagging score jet 1

O_2: b-tagging score jet 2

O_3: b-tagging score jet 3

O_4: b-tagging score jet 4

O_5: m_1

O_6: m_2

2D → 6D

Toy-model and Toy-problem

Setup:

Synthetic signal (bbbb) and backgrounds (bbcc, cccc) and play to distinguish signal using ABCD and Bayesian frameworks

Toy-model and Toy-problem

Setup:

Synthetic signal (bbbb) and backgrounds (bbcc, cccc) and play to distinguish signal using ABCD and Bayesian frameworks

Data:

20k events, signal is 1%, 0.5% or 0%

Toy-model and Toy-problem

Setup:

Synthetic signal (bbbb) and backgrounds (bbcc, cccc) and play to distinguish signal using ABCD and Bayesian frameworks

Data:

20k events, signal is 1%, 0.5% or 0%

Toy problem:

m_{bb} ~ N(125 GeV,10 GeV) or ~ Exp(0.003/GeV)

B-tag: 4 x b-scores ~ beta(), sampled from either *bbbb*, *bbcc, cccc*

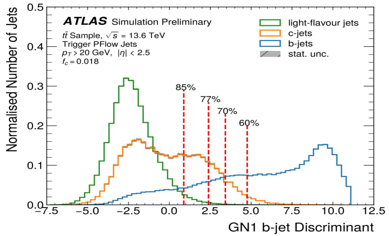
Toy-model and Toy-problem

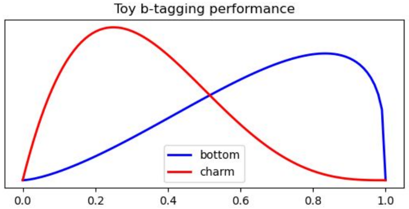
Setup:

Synthetic signal (bbbb) and backgrounds (bbcc, cccc) and play to distinguish signal using ABCD and Bayesian frameworks

Data:

20k events, signal is 1%, 0.5% or 0%





Toy problem:

m_{bb} ~ N(125 GeV,10 GeV) or ~ Exp(0.003/GeV)

B-tag: 4 x b-scores ~ beta(), sampled from either *bbbb, bbcc, cccc*

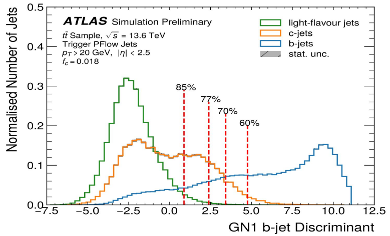
Toy-model and Toy-problem

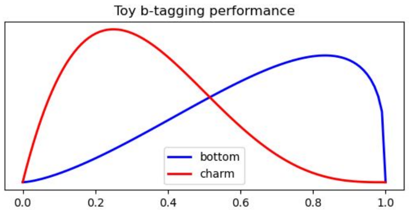
Setup:

Synthetic signal (bbbb) and backgrounds (bbcc, cccc) and play to distinguish signal using ABCD and Bayesian frameworks

Data:

20k events, signal is 1%, 0.5% or 0%





We'll start from biased priors

Toy problem:

m_{bb} ~ N(125 GeV,10 GeV)   or   ~ Exp(0.003/GeV)

B-tag: 4 x b-scores ~ beta(), sampled from either *bbbb*, *bbcc*, *cccc*

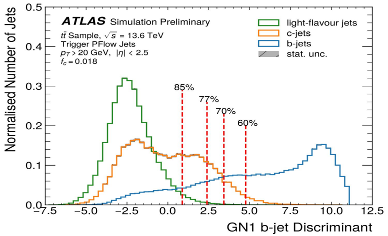
Toy-model and Toy-problem

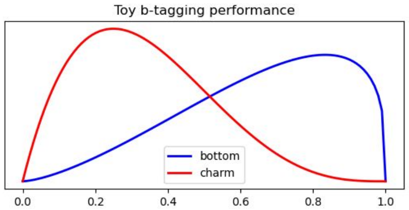
Setup:

Synthetic signal (bbbb) and backgrounds (bbcc, cccc) and play to distinguish signal using ABCD and Bayesian frameworks

Data:

20k events, signal is 1%, 0.5% or 0%





We'll start from biased priors

Toy problem:

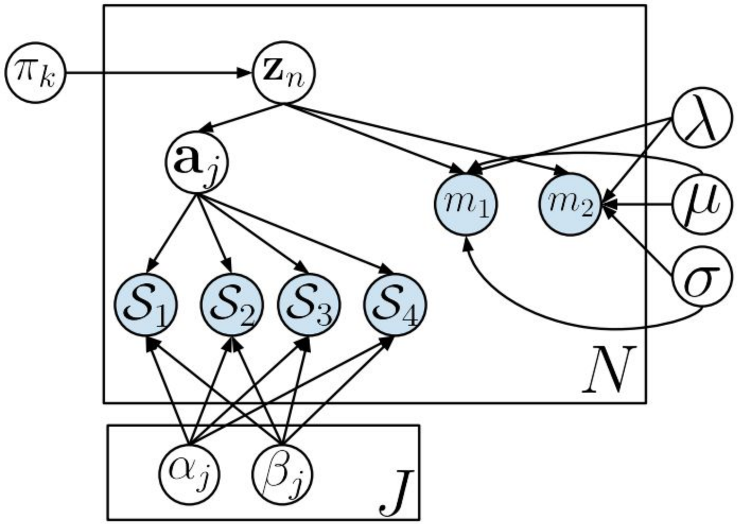
m_{bb} ~ N(125 GeV,10 GeV)   or   ~ Exp(0.003/GeV)

B-tag: 4 x b-scores ~ beta(), sampled from either *bbbb, bbcc, cccc*

Important prior-knowledge

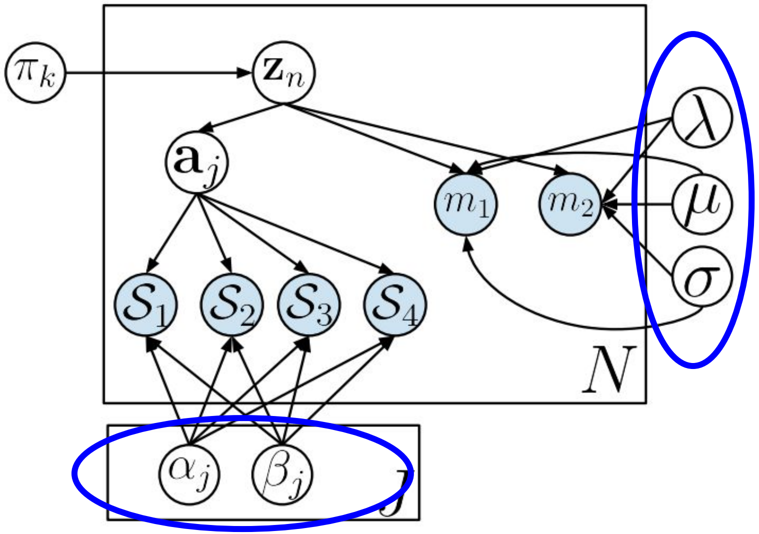
Bayes @ Toy-problem

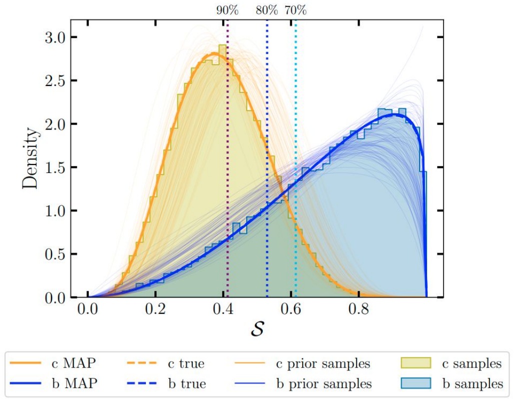
6 Observables



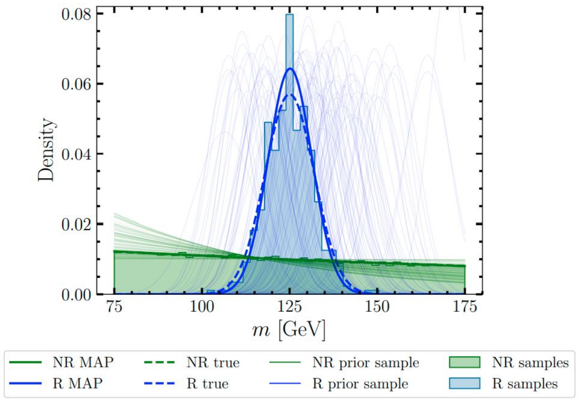
Bayes @ Toy-problem

6 Observables



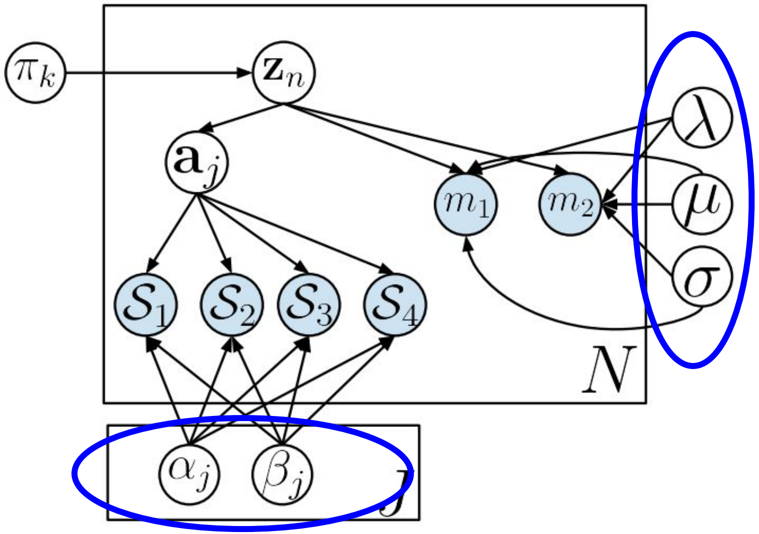


Inference results

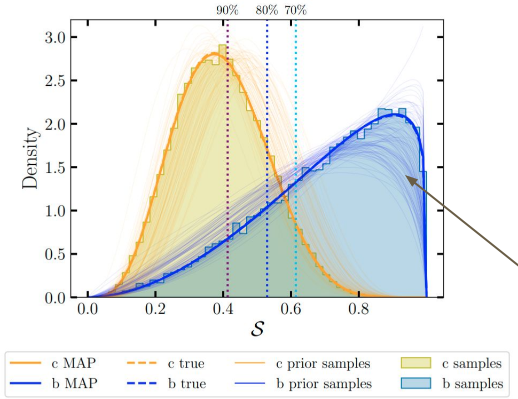


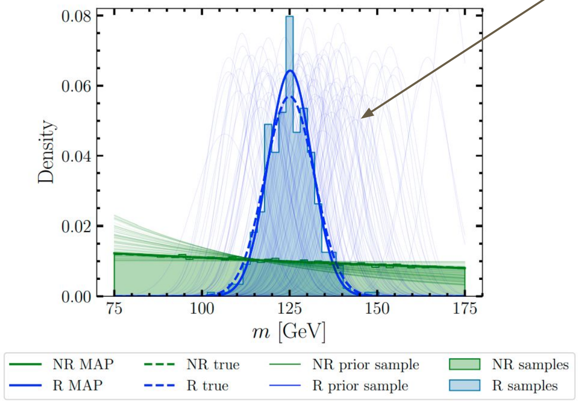
Bayes @ Toy-problem

6 Observables



Inference results

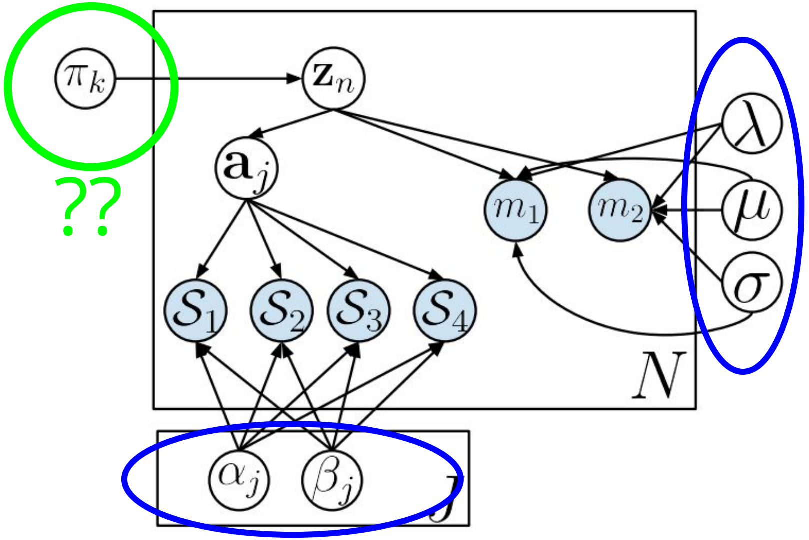




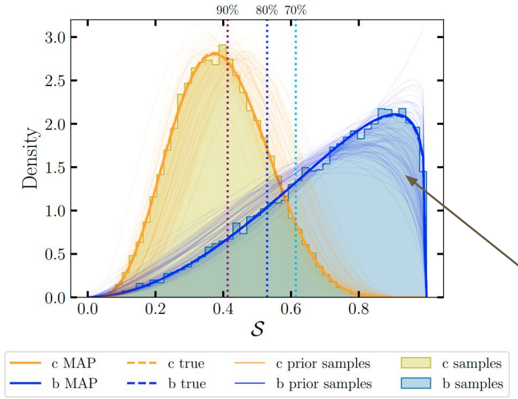
Biased priors, but inference gets correct curves!

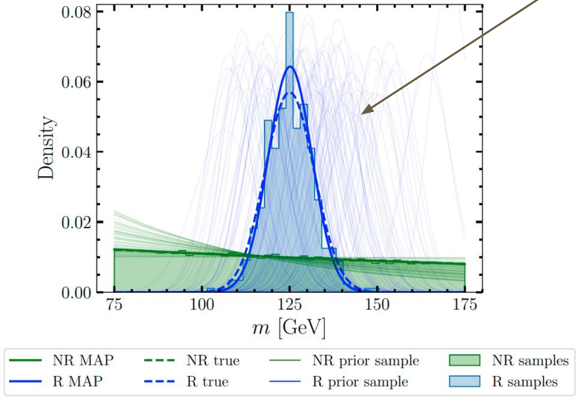
Bayes @ Toy-problem

6 Observables



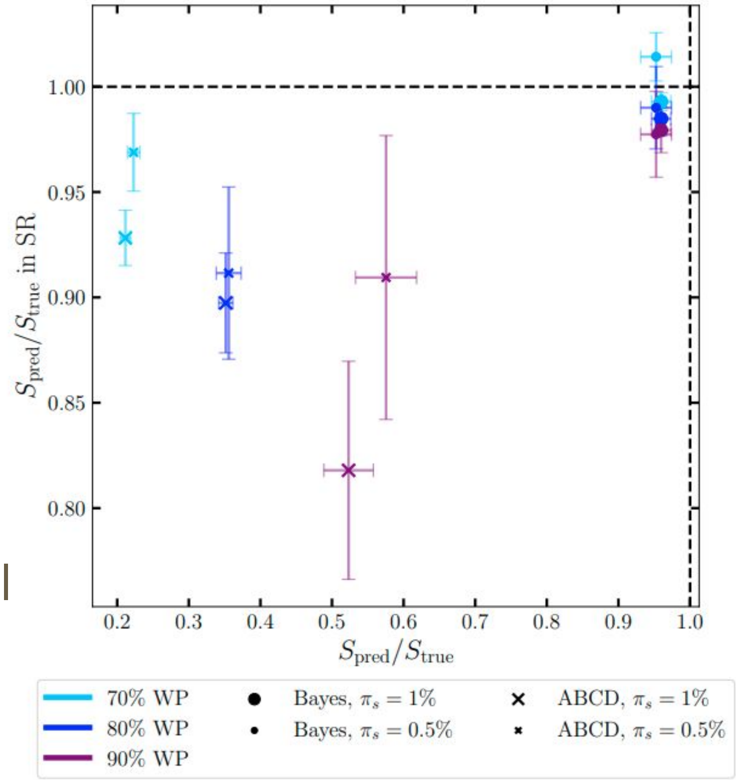
Inference results





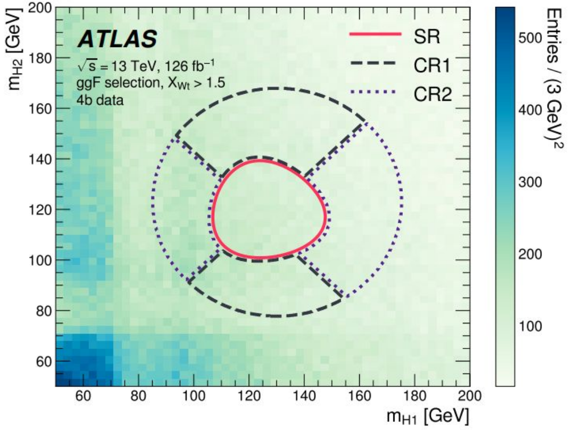
Biased priors, but inference gets correct curves!

ABCD Vs Bayesian framework

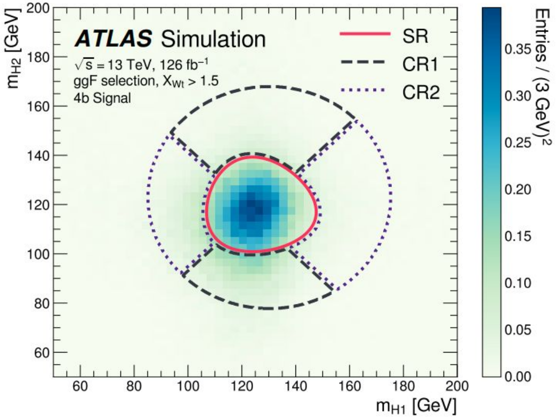


1% & 0.5% signal

Experimental current status

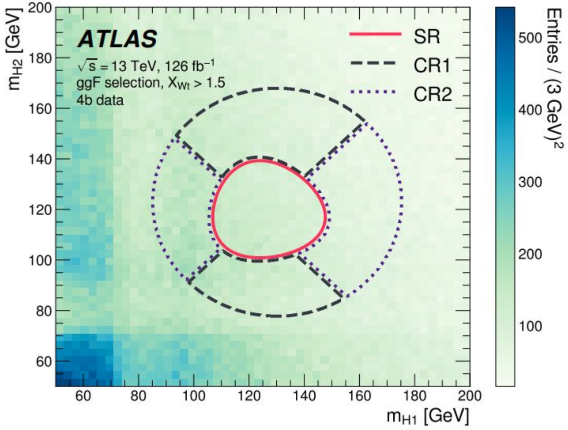


Data

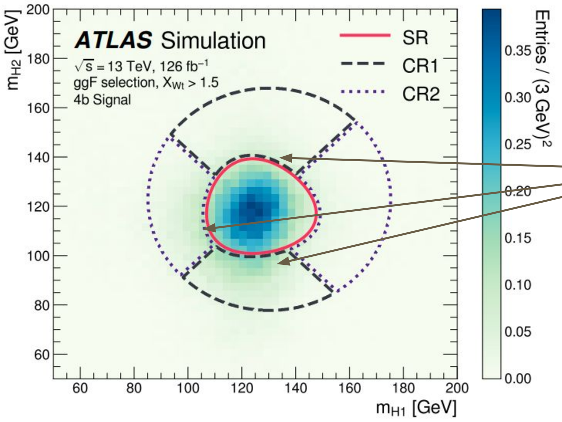


Signal @simulations

Experimental current status



Data

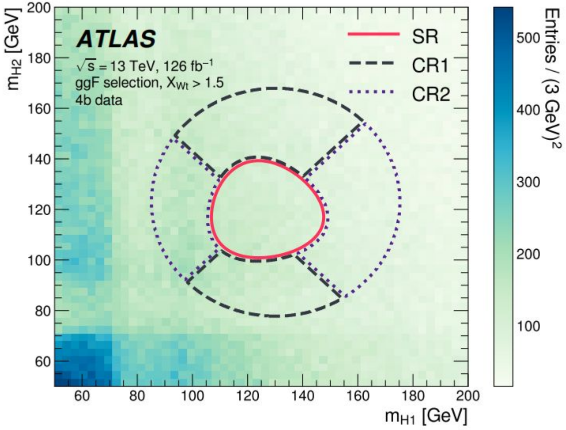


Signal @simulations

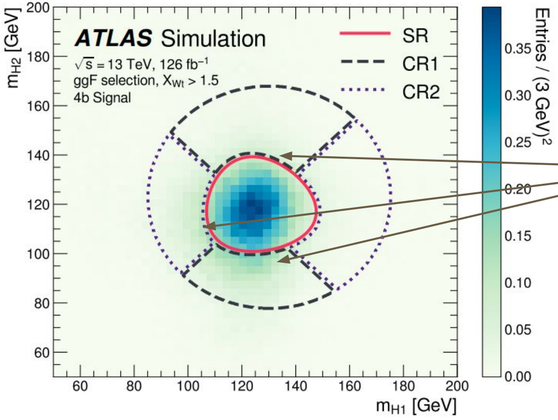
Missing 10-15% of signal !!

(many Eur!)

Experimental current status



Data



Signal @simulations

Missing 10-15% of signal !!

(many Eur!)

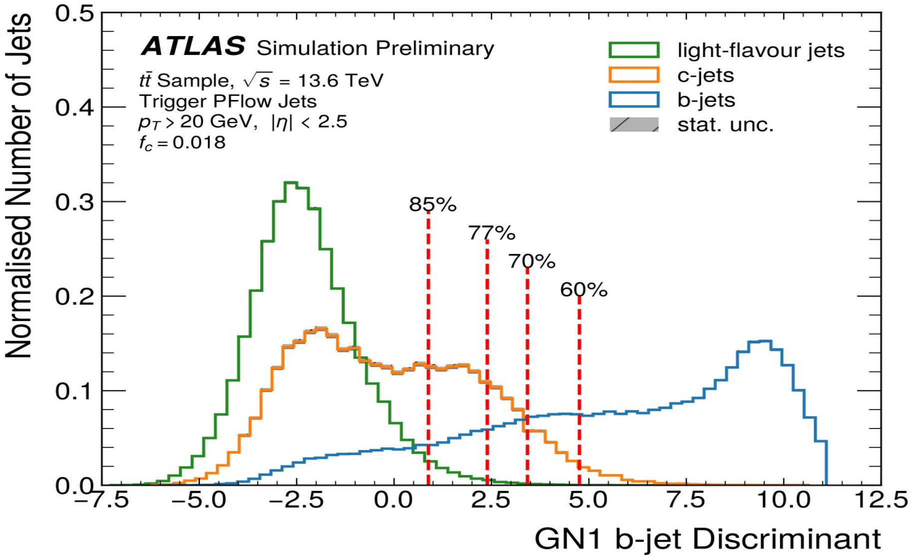
Much more is lost by requiring 4 b-tags at eff=0.77!!

(0.77^4=0.35)

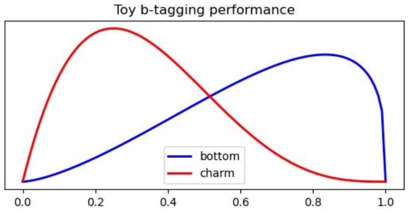
Bayesian exploitation of continuity and unimodality

Yuling Yao + EA
2404.01387

Exploit Continuity and Unimodality

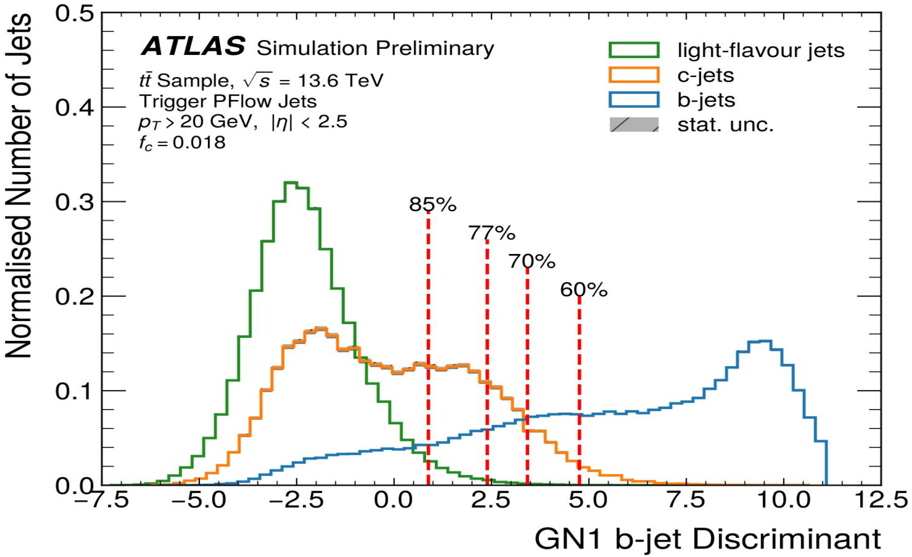


≈



Previous work

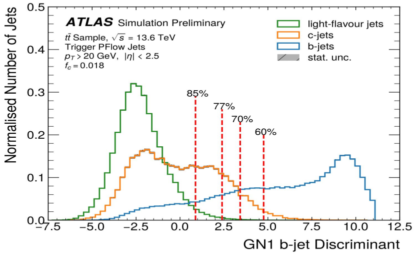
Exploit Continuity and Unimodality



We'll infer these continuous arbitrary distributions!

The leverage: continuity, unimodality & multidimensionality!

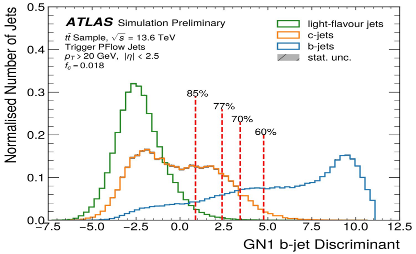
Gaussian Processes



$$f(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^k * det(\boldsymbol{\Sigma})}} * e^{-\frac{1}{2} * ((\boldsymbol{x}-\boldsymbol{\mu})^T \cdot inv(\boldsymbol{\Sigma}) \cdot (\boldsymbol{x}-\boldsymbol{\mu}))}$$

We bin the score and ***x*** contains the distribution values in each bin

Gaussian Processes



We bin the score and ***x*** contains the distribution values in each bin

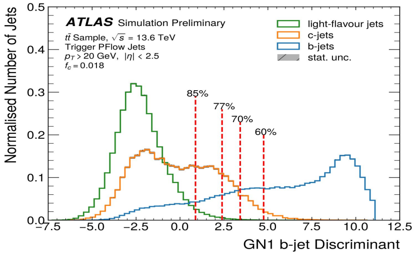
$$f(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^k * det(\boldsymbol{\Sigma})}} * e^{-\frac{1}{2} * ((\boldsymbol{x}-\boldsymbol{\mu})^T \cdot inv(\boldsymbol{\Sigma}) \cdot (\boldsymbol{x}-\boldsymbol{\mu}))}$$

Each bin is sampled around some expected μ

Define uncertainty and how related are neighbouring bins: Continuity!

$$\Sigma^{-1} = \begin{pmatrix} 2 & 1 & 0.5 & 0 & \text{....} \\ 1 & 2 & 1 & 0.5 & 0 & \text{...} \\ 0.5 & 1 & 2 & 1 & 0.5 & 0 & \text{...} \\ 0 & 0.5 & 1 & 2 & 1 & 0.5 & 0 \text{....} \\ \text{....} & \text{.......} & & \text{.....} & & \text{.....} \end{pmatrix}$$

Gaussian Processes



We bin the score and $\boldsymbol{x}$ contains the distribution values in each bin

$$f(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^k * det(\boldsymbol{\Sigma})}} * e^{-\frac{1}{2} * ((\boldsymbol{x}-\boldsymbol{\mu})^T \cdot inv(\boldsymbol{\Sigma}) \cdot (\boldsymbol{x}-\boldsymbol{\mu}))}$$

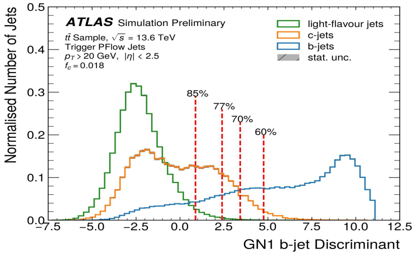
Each bin is sampled around some expected μ

Define uncertainty and how related are neighbouring bins: Continuity!

We can sample continuous curves around a central curve with very few hyperparameters

$$\Sigma^{-1} = \begin{pmatrix} 2 & 1 & 0.5 & 0 & .... & & & \\ 1 & 2 & 1 & 0.5 & 0 & ... & & \\ 0.5 & 1 & 2 & 1 & 0.5 & 0 & ... & \\ 0 & 0.5 & 1 & 2 & 1 & 0.5 & 0 & .... \\ .... & .... & & ..... & & ..... & & \end{pmatrix}$$

Gaussian Processes



$$f(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^k * det(\boldsymbol{\Sigma})}} * e^{-\frac{1}{2} * ((\boldsymbol{x}-\boldsymbol{\mu})^T \cdot inv(\boldsymbol{\Sigma}) \cdot (\boldsymbol{x}-\boldsymbol{\mu}))}$$

We bin the score and ***x*** contains the distribution values in each bin

Each bin is sampled around some expected μ

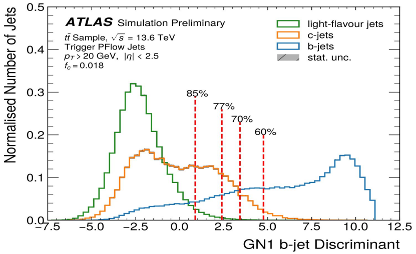
Define uncertainty and how related are neighbouring bins: Continuity!

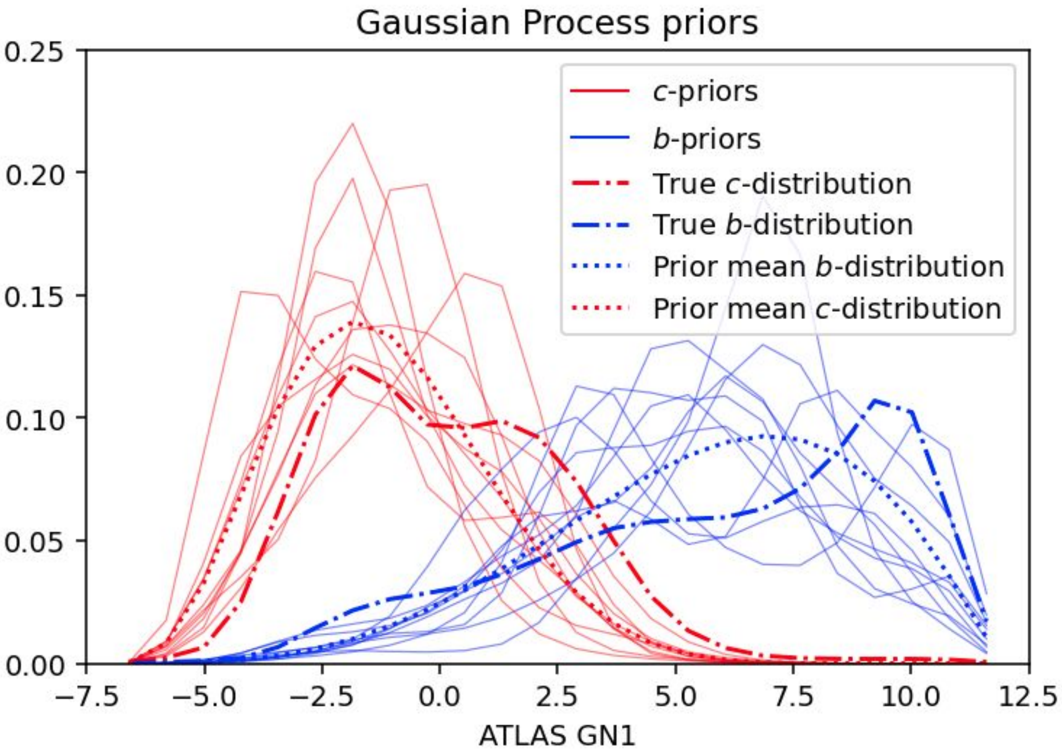
Prior information

We can sample continuous curves around a central curve with very few hyperparameters

$$\Sigma^{-1} = \begin{pmatrix} 2 & 1 & 0.5 & 0 & .... & & & \\ 1 & 2 & 1 & 0.5 & 0 & ... & & \\ 0.5 & 1 & 2 & 1 & 0.5 & 0 & ... & \\ 0 & 0.5 & 1 & 2 & 1 & 0.5 & 0 & .... \\ .... & .... & & ..... & & ..... & & \end{pmatrix}$$

Gaussian Processes

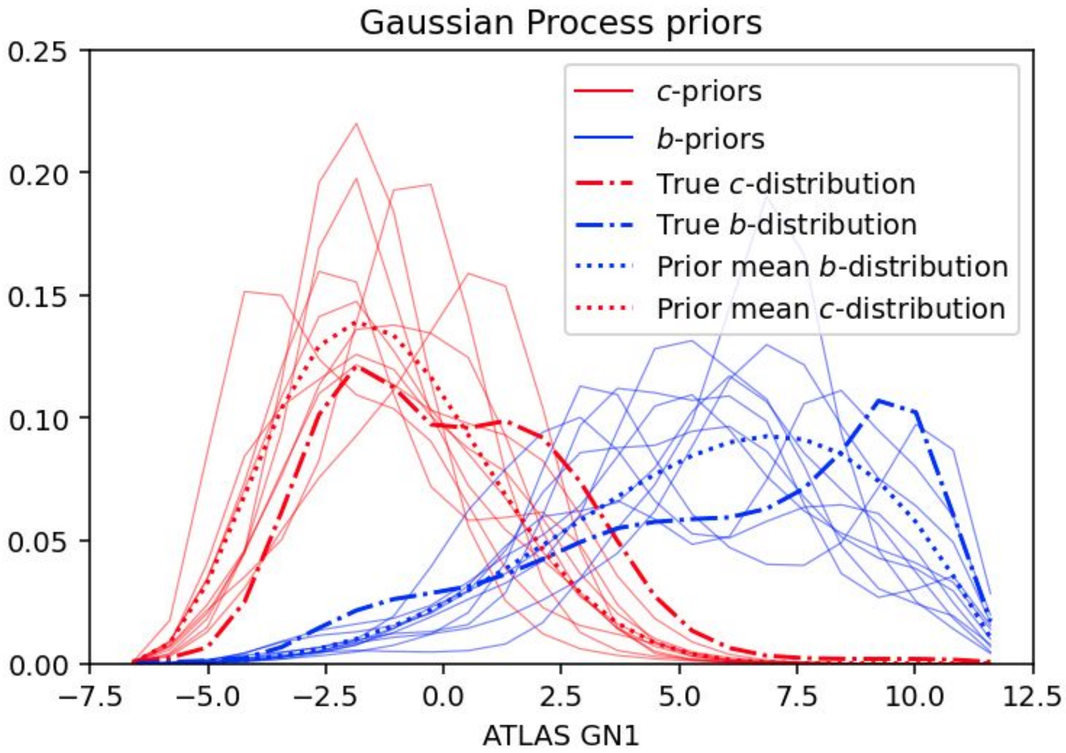




Gaussian Processes

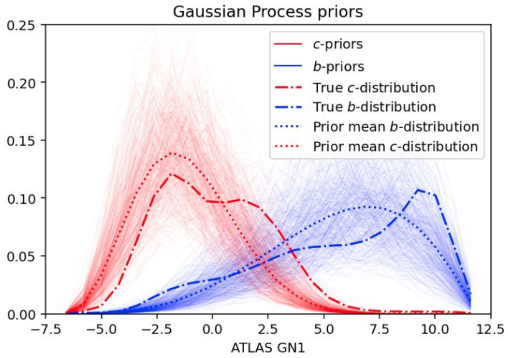
The game:

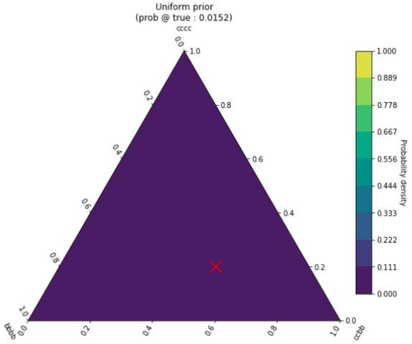
- Starts with biased prior
- The data will shift the posterior to the most likely distribution, which should be the true
- Leverage:
  - Multidimensionality
  - Continuity
  - bbbb, ccbb, cccc



Gaussian Processes: Results

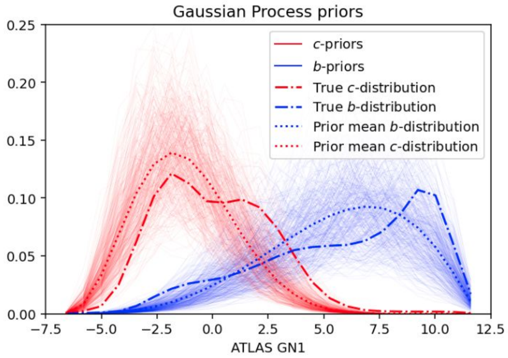
This is how we start

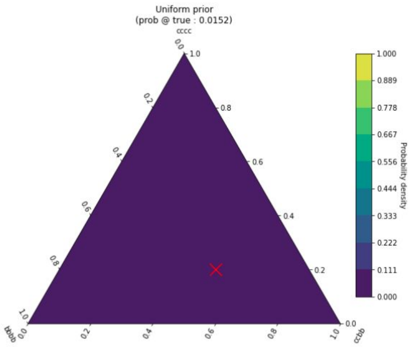




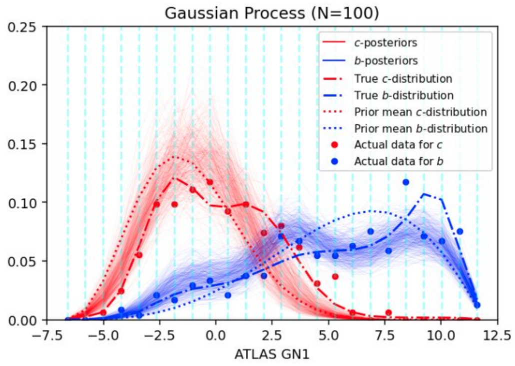
Gaussian Processes: Results

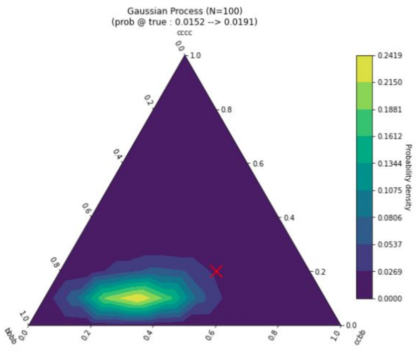
This is how we start





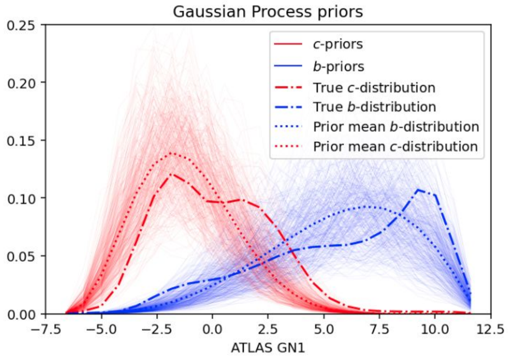
After seeing 100 events

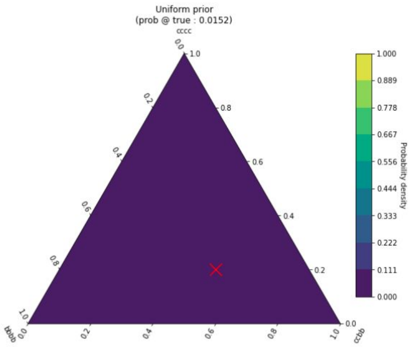




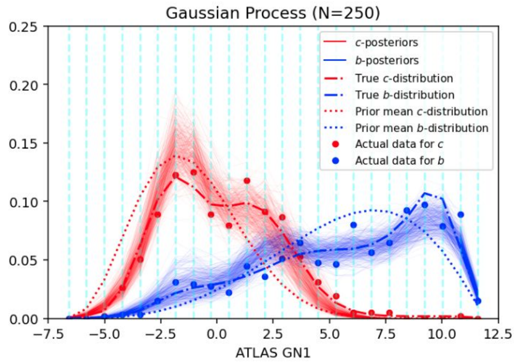
Gaussian Processes: Results

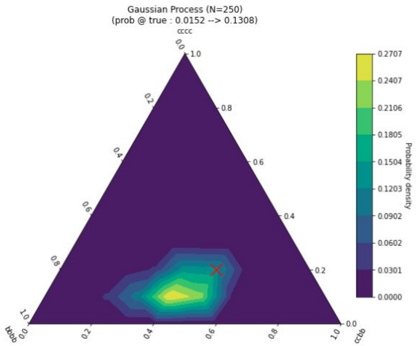
This is how we start





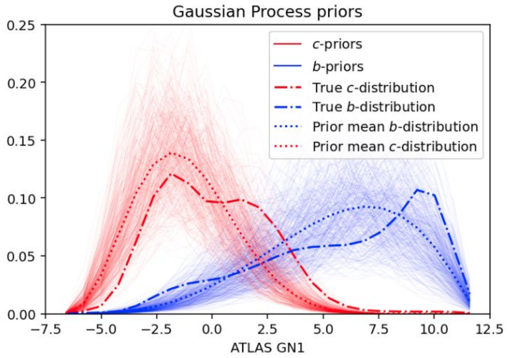
After seeing 250 events

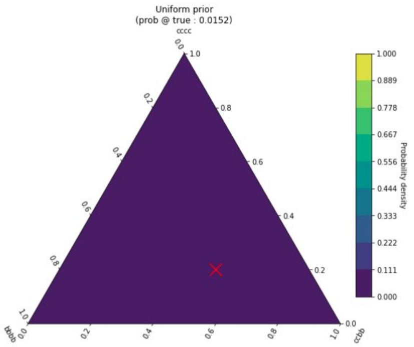




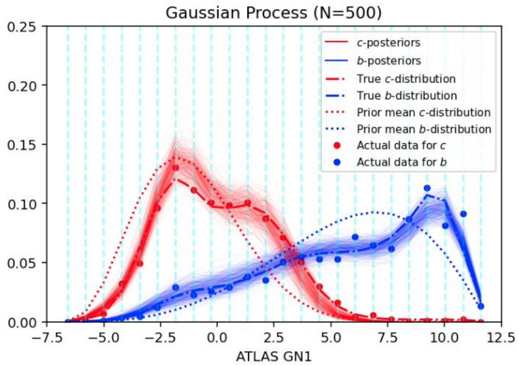
Gaussian Processes: Results

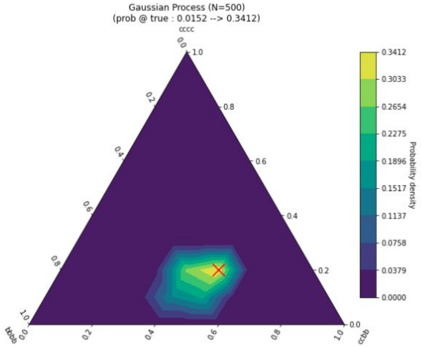
This is how we start





After seeing 500 events

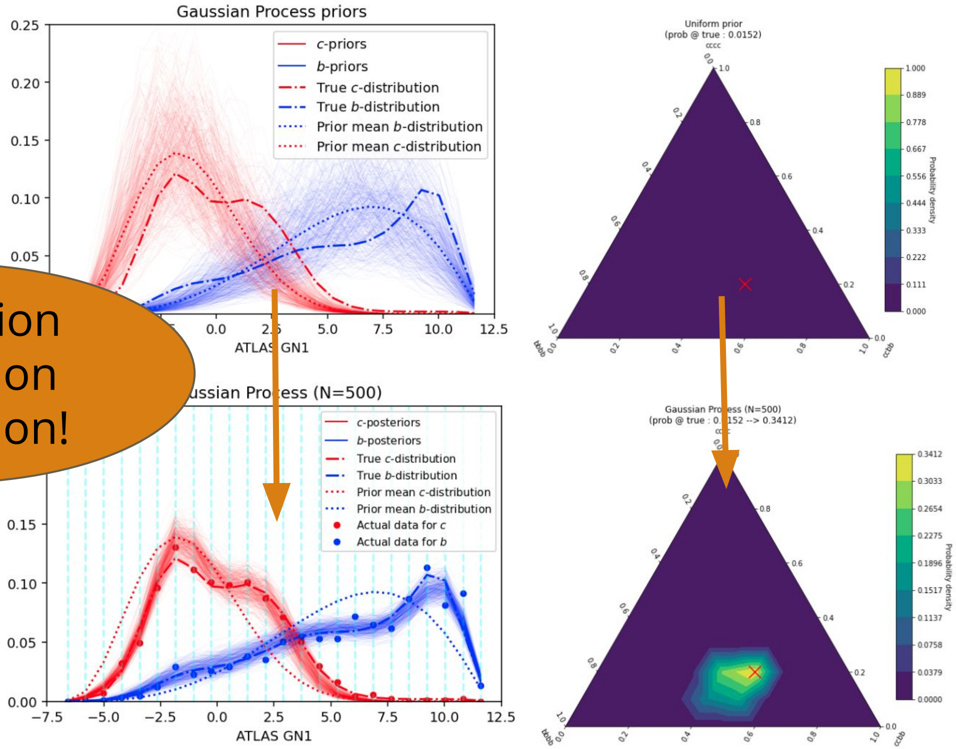




Gaussian Processes: Results

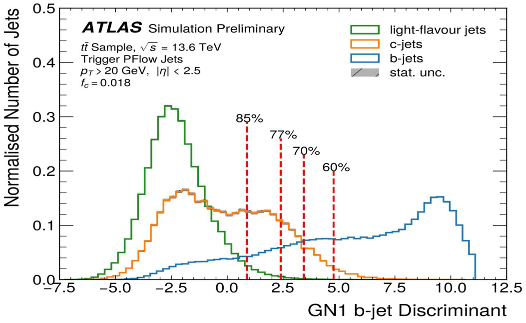
This is how we start

After seeing 500 events



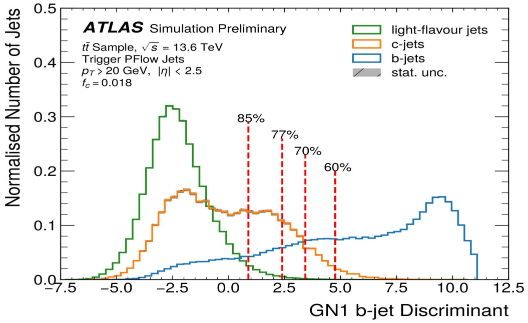
Correlation correlation correlation!

Unimodal model



How to sample unimodal arbitrary continuous curves ?

Unimodal model

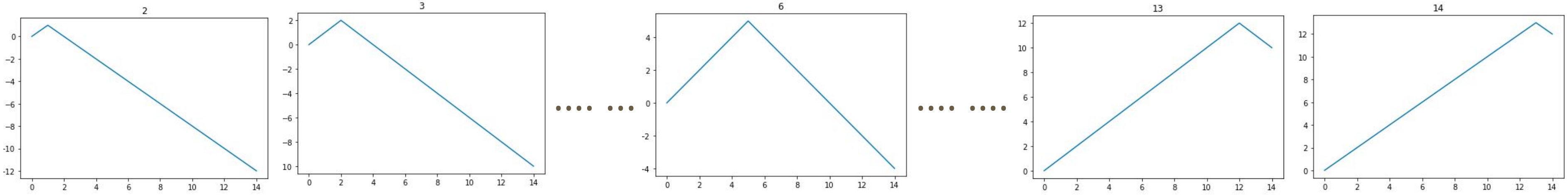


Prior information!

How to sample unimodal arbitrary continuous curves ?

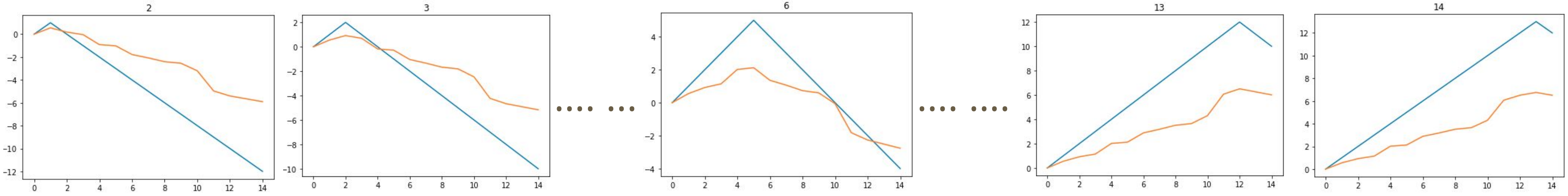
Unimodal model

Construct strict linear unimodal, one for each bin



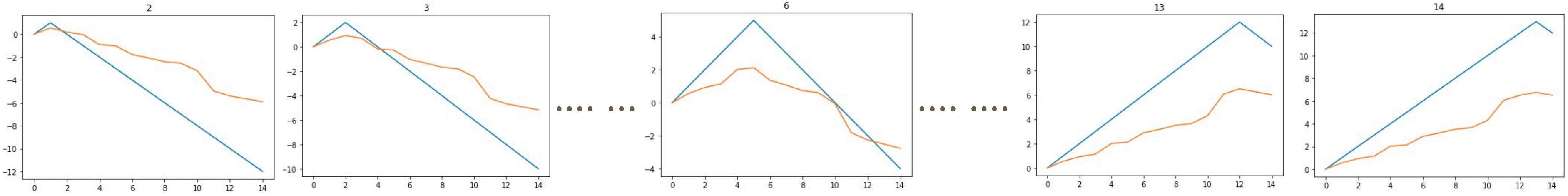
Unimodal model

Allow for randomness with a half normal $|N(0,0.5)|$ at each step

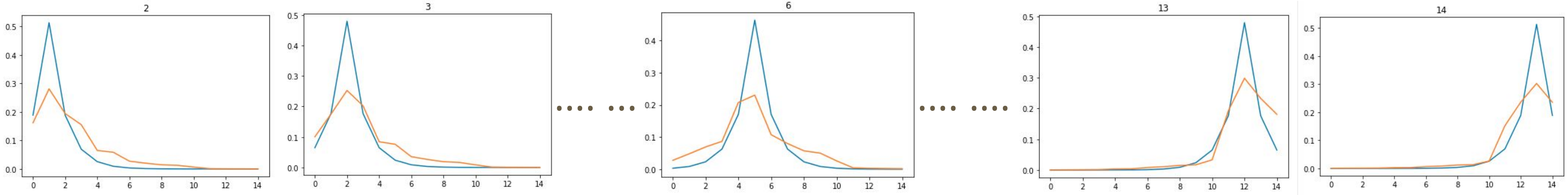


Unimodal model

Allow for randomness with a half normal |*N*(0,0.5)| at each step

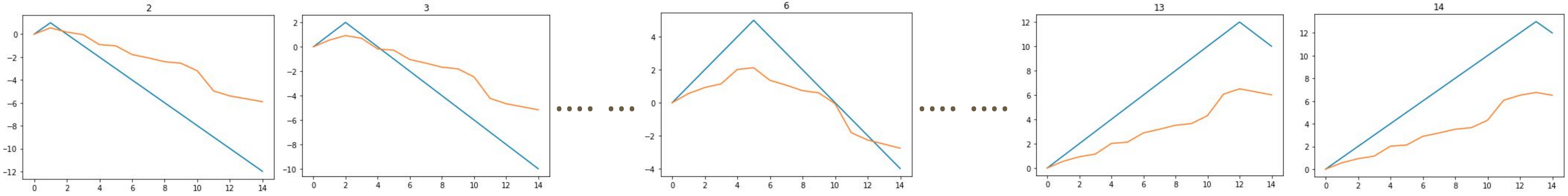


Apply *softmax()* to make them integrate to unity

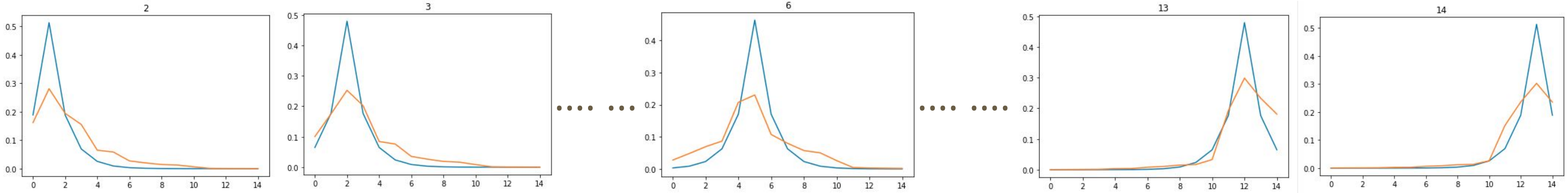


Unimodal model

Allow for randomness with a half normal $|N(0,0.5)|$ at each step



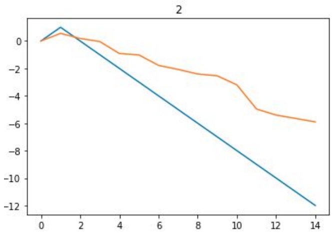
Apply *softmax()* to make them integrate to unity

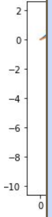


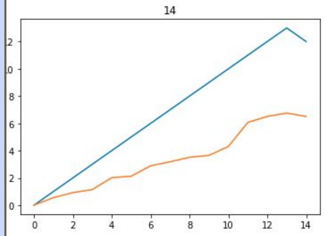
How to have unimodal at any bin and with some freedom of shape in other bins?

Unimodal model

Allow for random



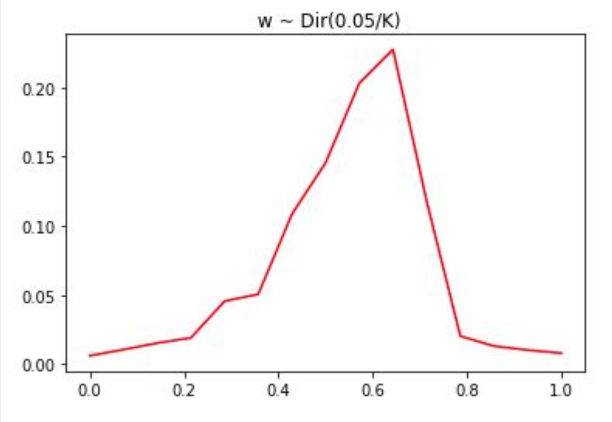




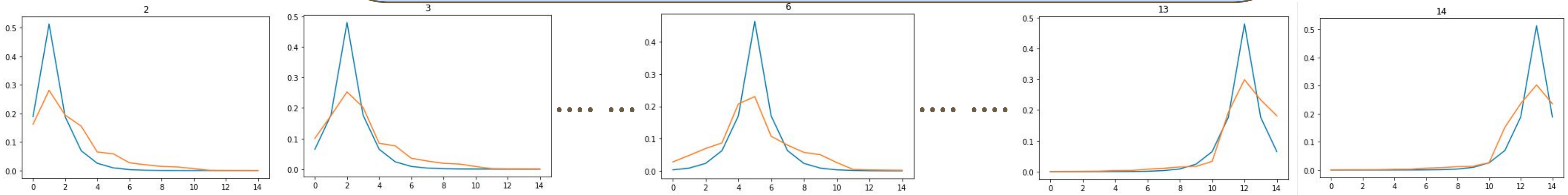
Sum orange curves weighted by

w ~Dirichlet(α)

With α small



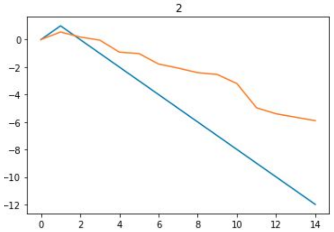
Apply *softmax()* to

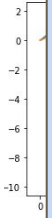


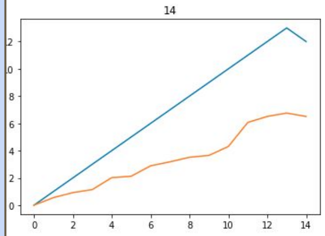
How to have unimodal at any bin and with some freedom of shape in other bins?

Unimodal model

Allow for random



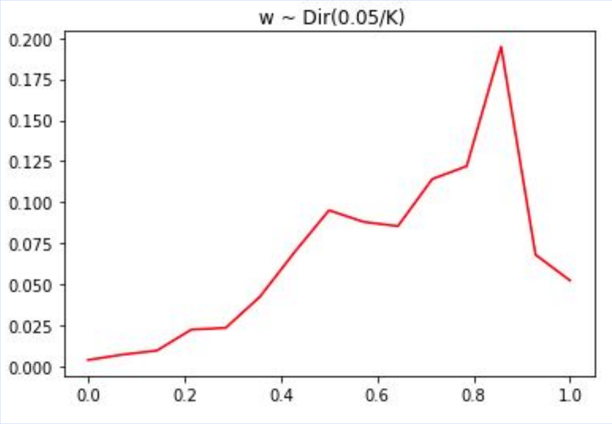




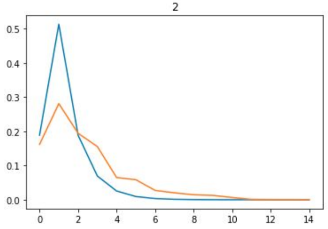
Sum orange curves weighted by

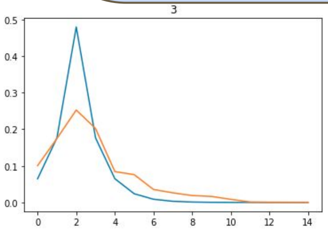
w ~Dirichlet(α)

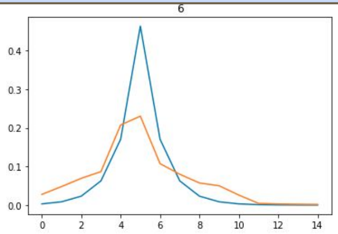
With α small

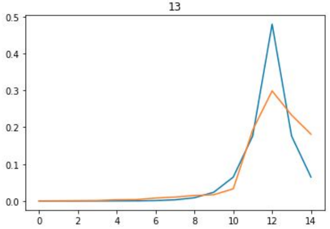


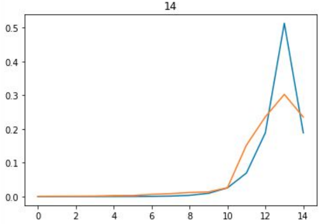
Apply *softmax()* to







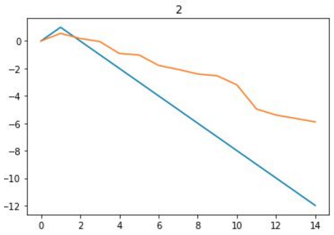


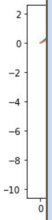


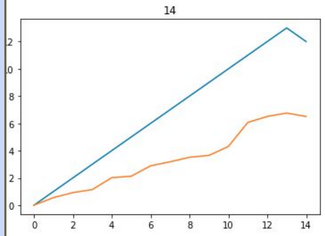
How to have unimodal at any bin and with some freedom of shape in other bins?

Unimodal model

Allow for random



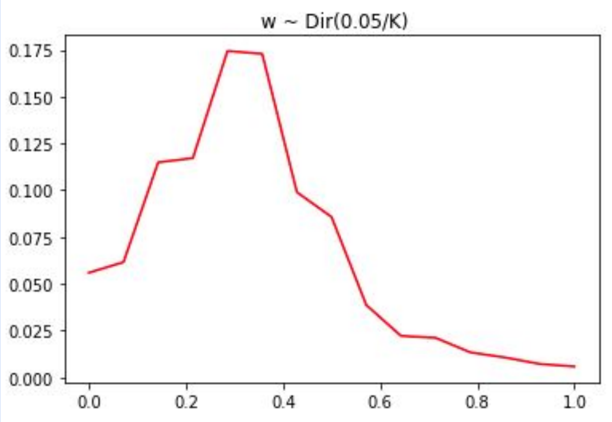




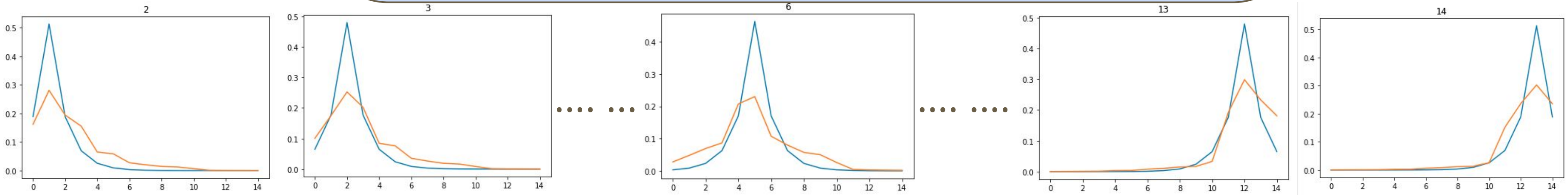
Sum orange curves weighted by

w ~Dirichlet(α)

With α small



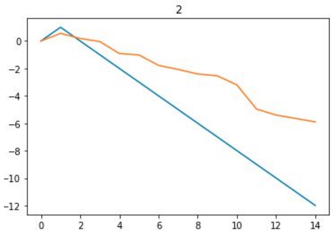
Apply *softmax()* to

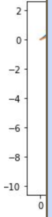


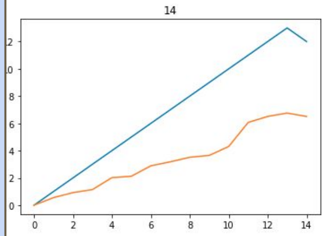
How to have unimodal at any bin and with some freedom of shape in other bins?

Unimodal model

Allow for random



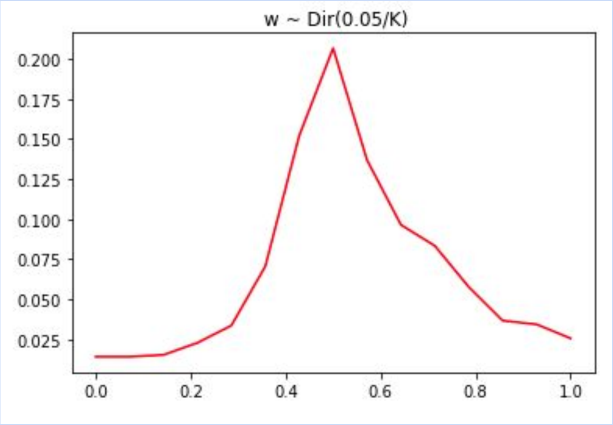




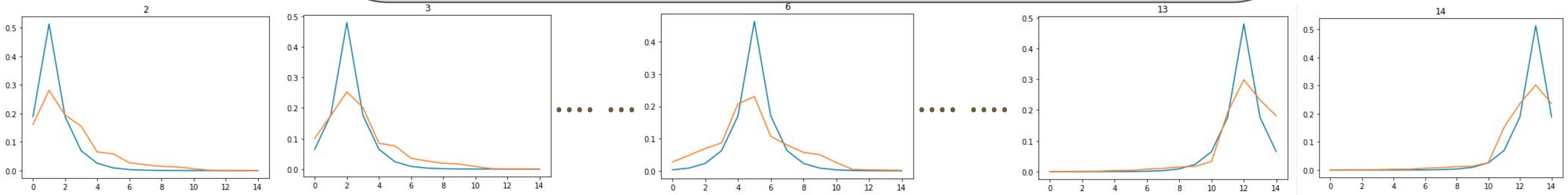
Sum orange curves weighted by

w ~Dirichlet(α)

With α small



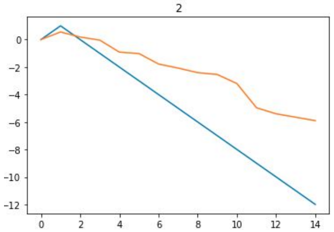
Apply *softmax()* to

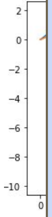


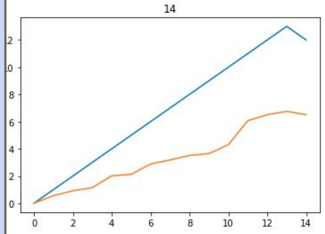
How to have unimodal at any bin and with some freedom of shape in other bins?

Unimodal model

Allow for random



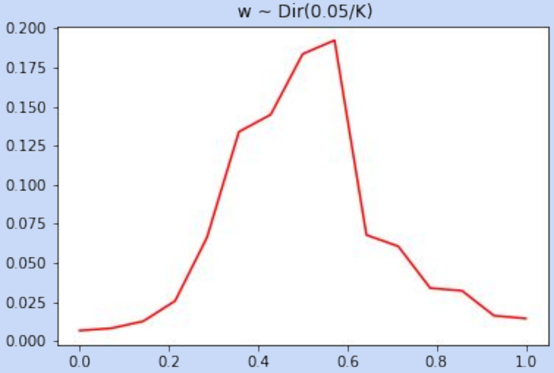




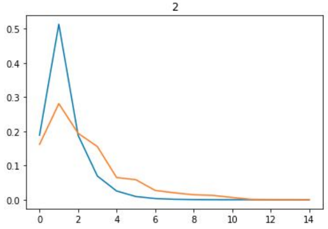
Sum orange curves weighted by

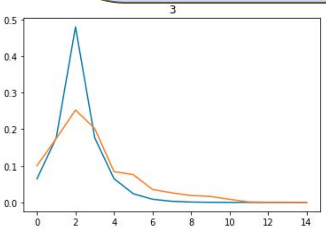
w ~Dirichlet(α)

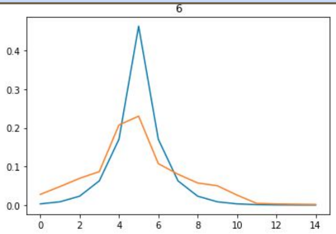
With α small

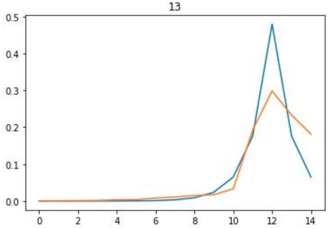


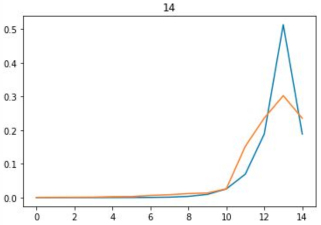
Apply *softmax()* to







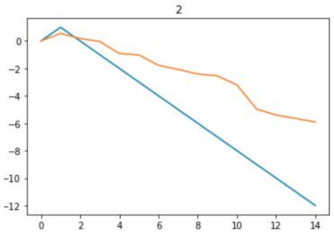
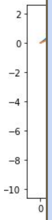




How to have unimodal at any bin and with some freedom of shape in other bins?

Unimodal model

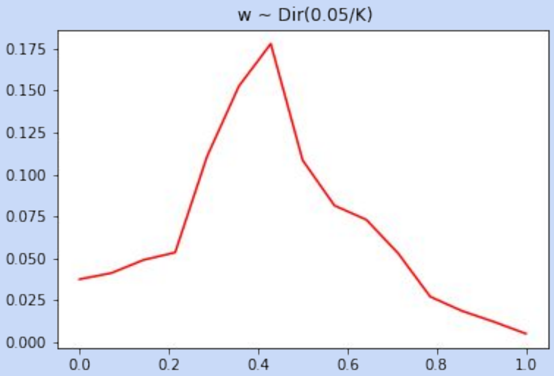
Allow for random

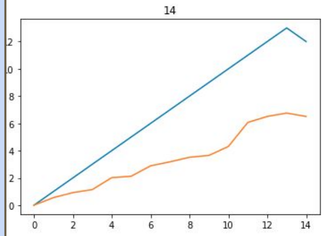



Sum orange curves weighted by

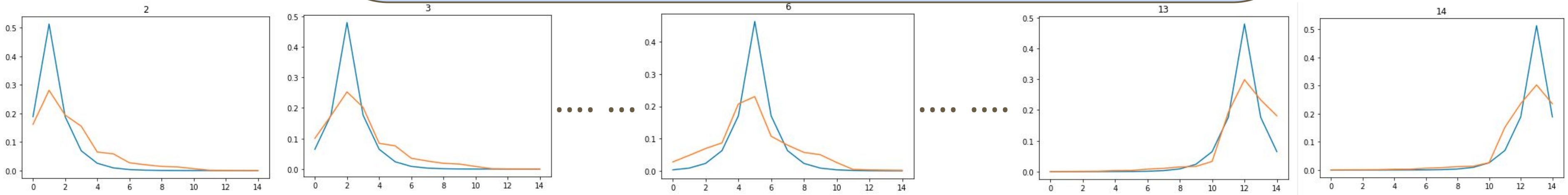
w ~Dirichlet(α)

With α small





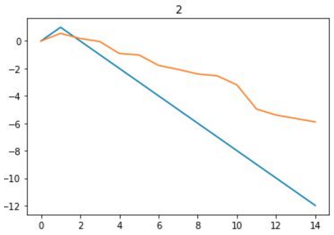
Apply *softmax()* to

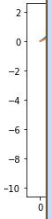


How to have unimodal at any bin and with some freedom of shape in other bins?

Unimodal model

Allow for random

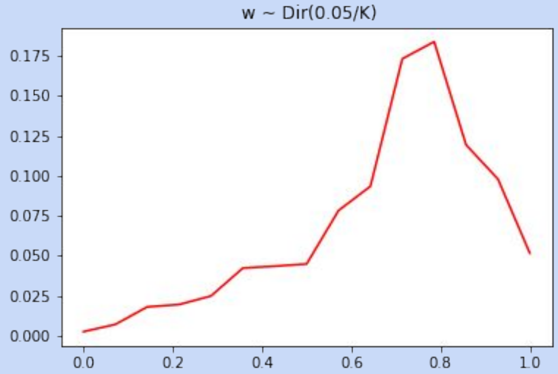


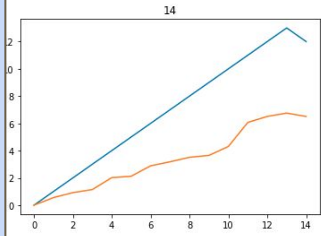


Sum orange curves weighted by

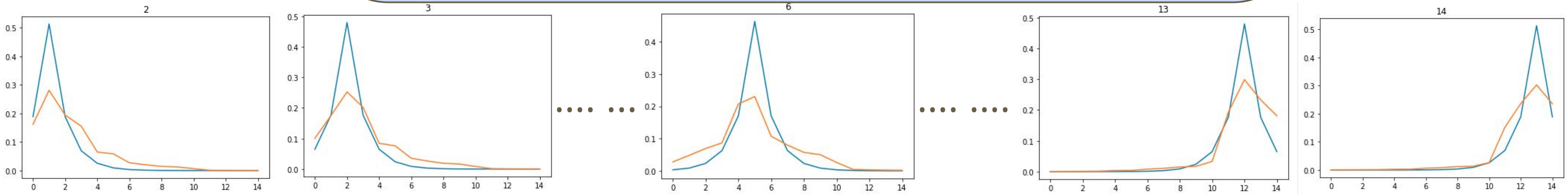
w ~Dirichlet(α)

With α small





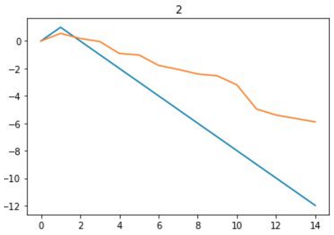
Apply *softmax()* to

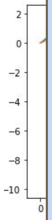


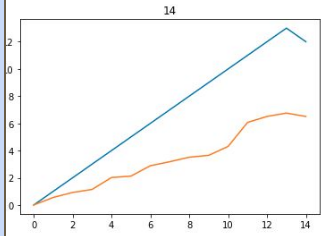
How to have unimodal at any bin and with some freedom of shape in other bins?

Unimodal model

Allow for random



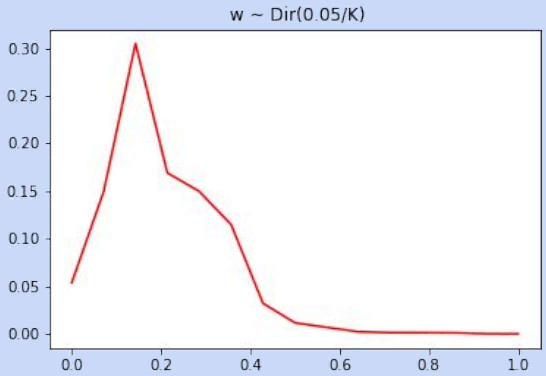




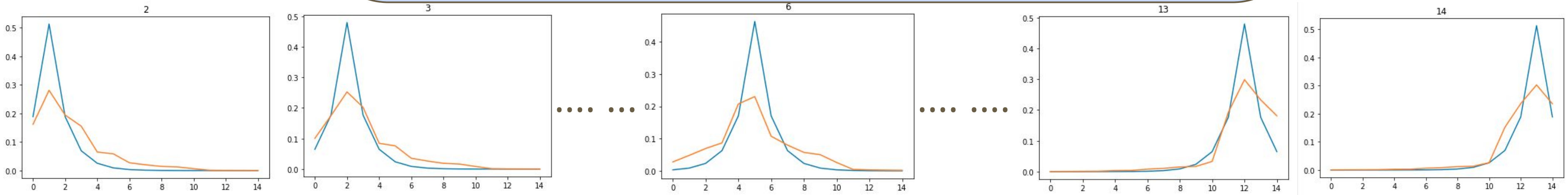
Sum orange curves weighted by

w ~Dirichlet(α)

With α small

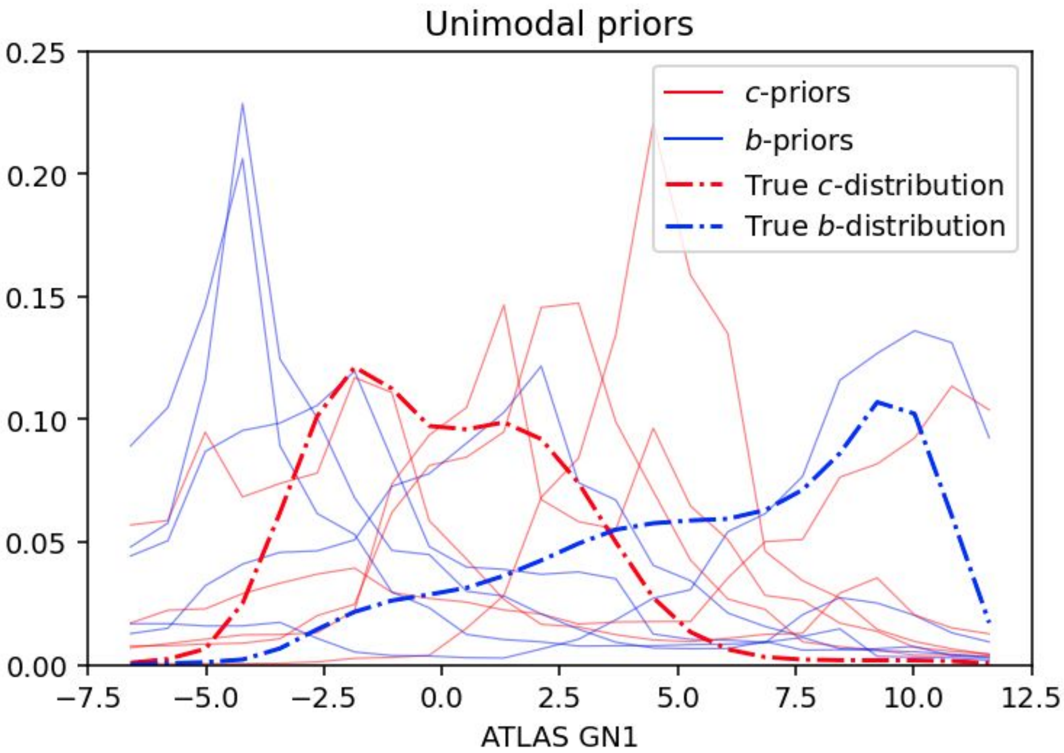


Apply *softmax()* to



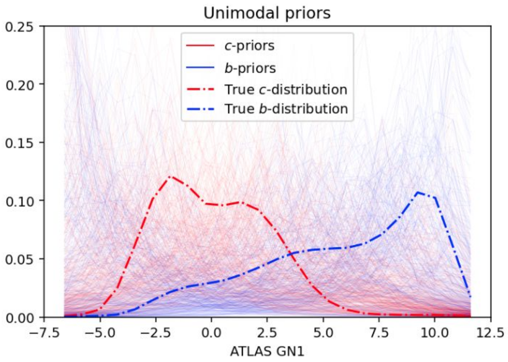
How to have unimodal at any bin and with some freedom of shape in other bins?

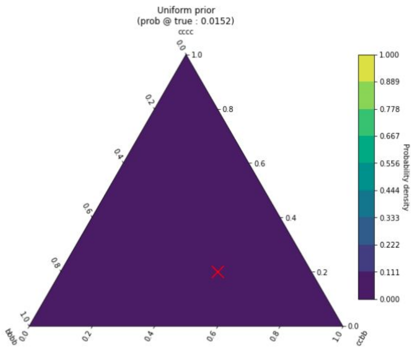
Unimodal model



Unimodal model: Results

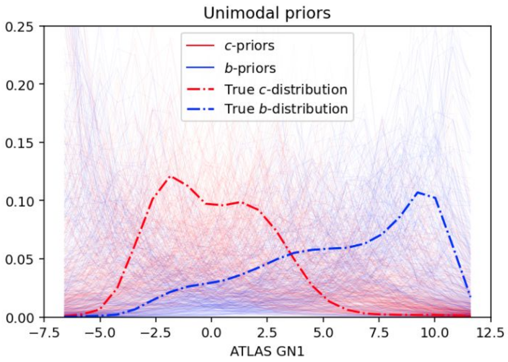
This is how we start

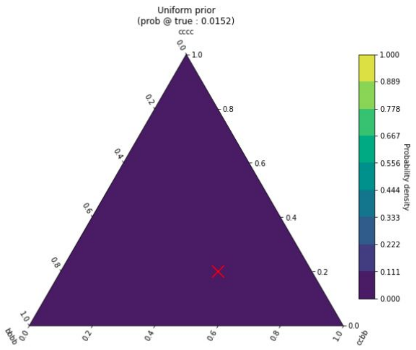




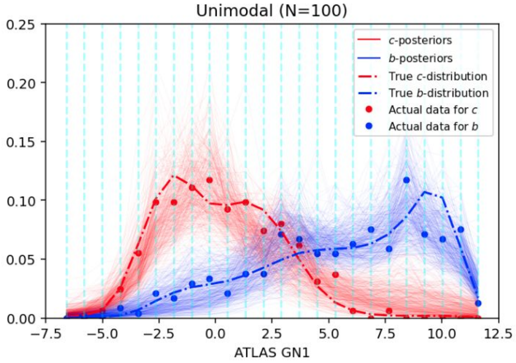
Unimodal model: Results

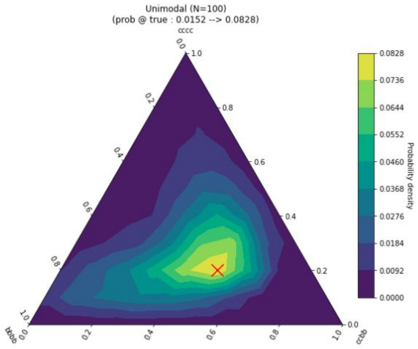
This is how we start





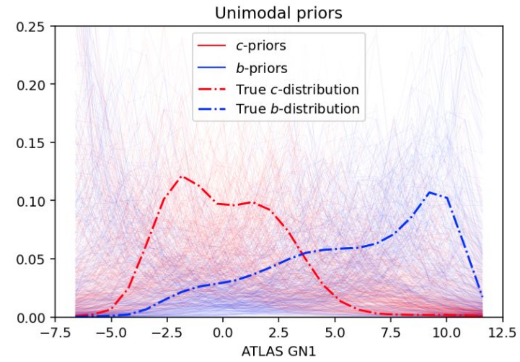
After seeing 100 events

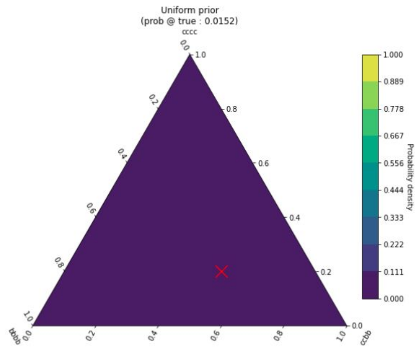




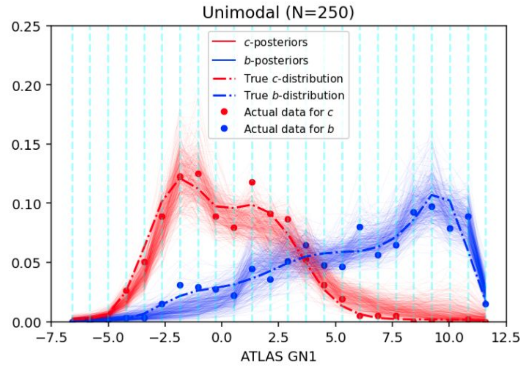
Unimodal model: Results

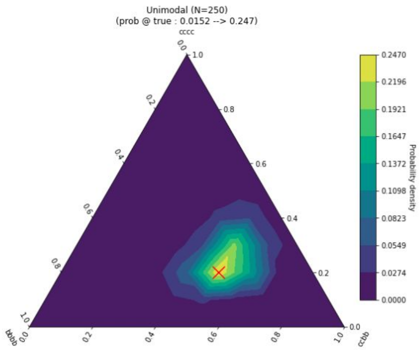
This is how we start





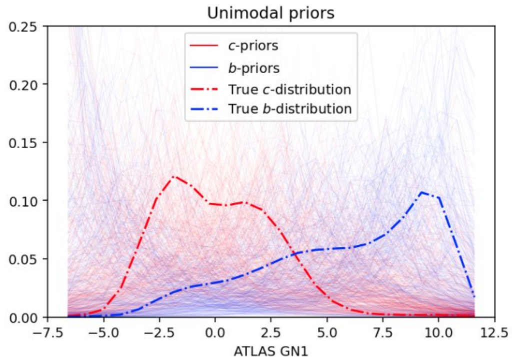
After seeing 250 events

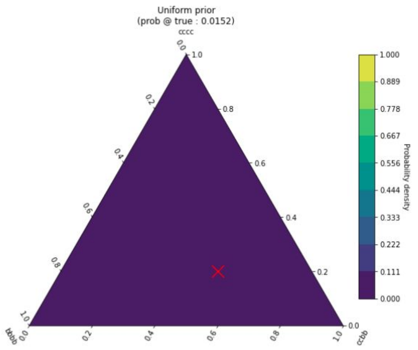




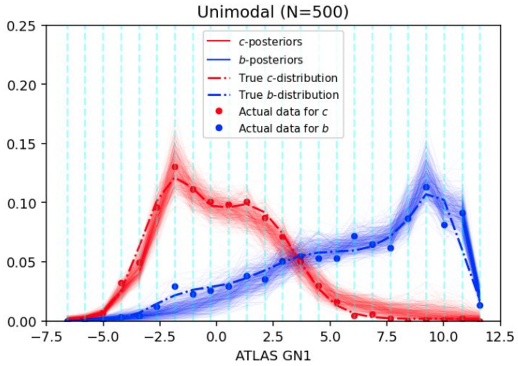
Unimodal model: Results

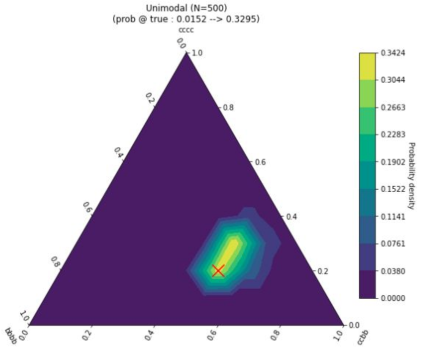
This is how we start





After seeing 500 events



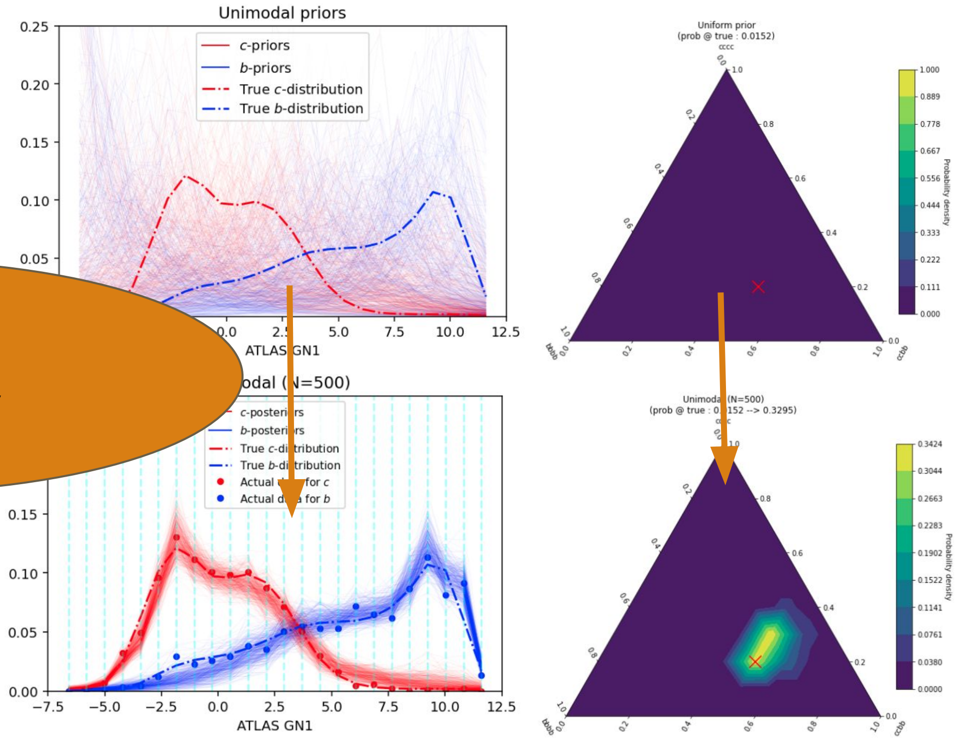


Unimodal model: Results

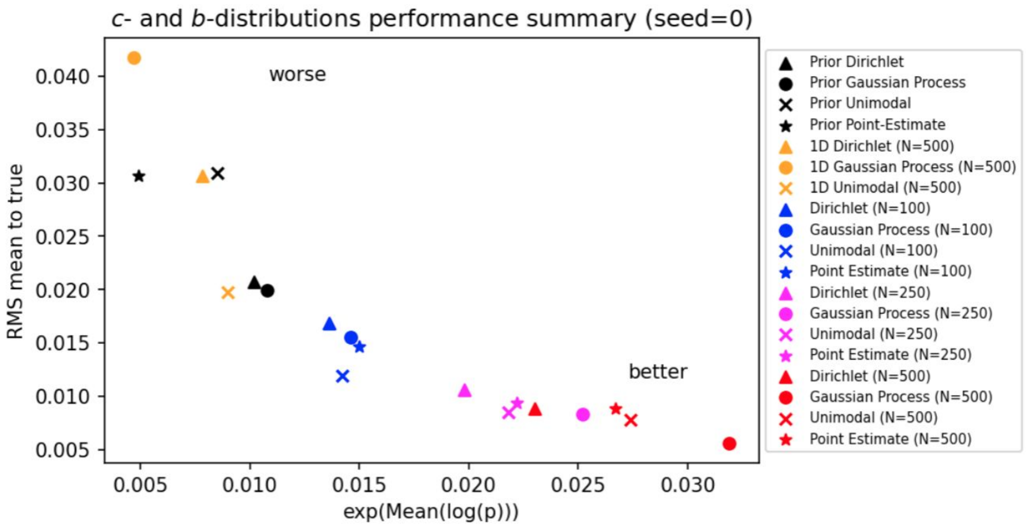
This is how we start

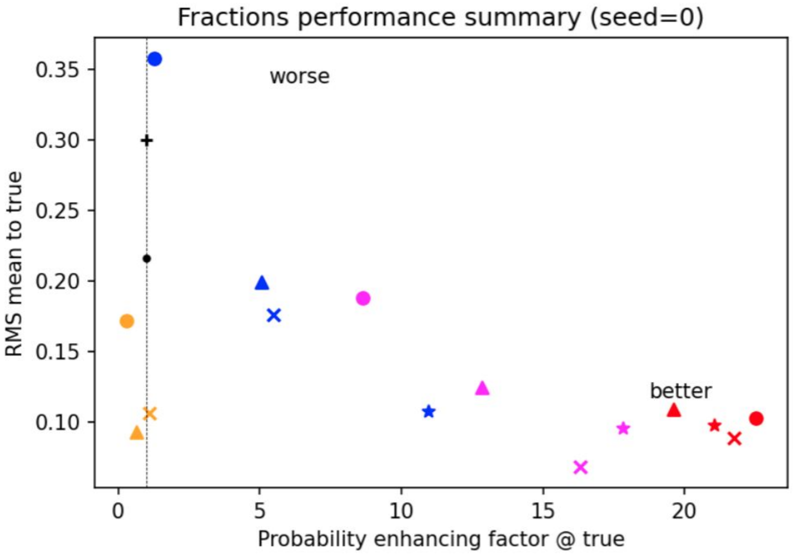
Correlation
+
unimodality
knowledge

After seeing 500 events



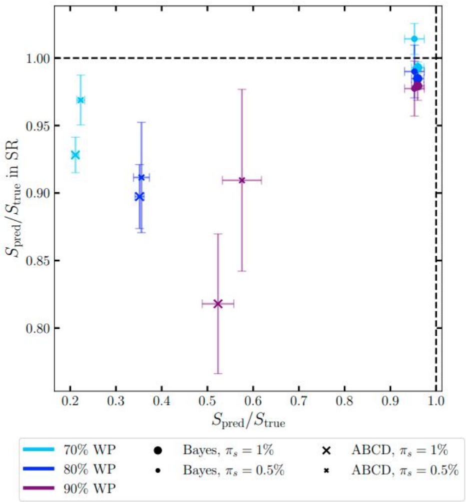
Exploiting prior info: summary results





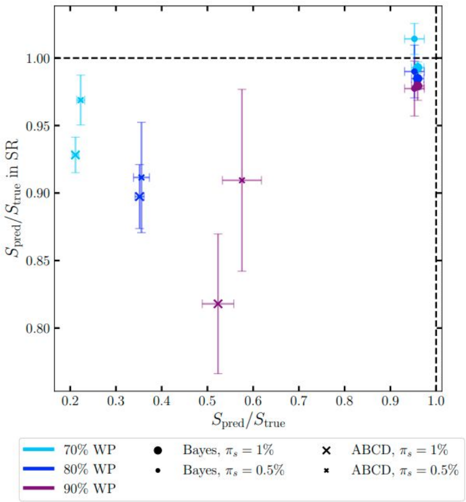
Outlook & Conclusions

Outlook & Conclusions



- Bayesian framework generalizes and improves ABCD method
- Bayesian is a sophisticated data-driven method that exploits multidimensionality +
- Toy-model on a toy-problem inspired in pp > hh > bbbb

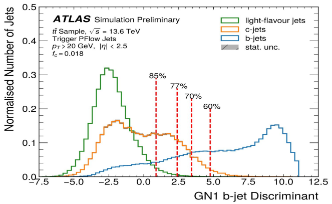
Outlook & Conclusions



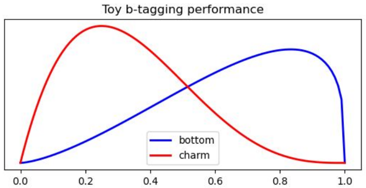
- Bayesian framework generalizes and improves ABCD method

- Bayesian is a sophisticated data-driven method that exploits multidimensionality +

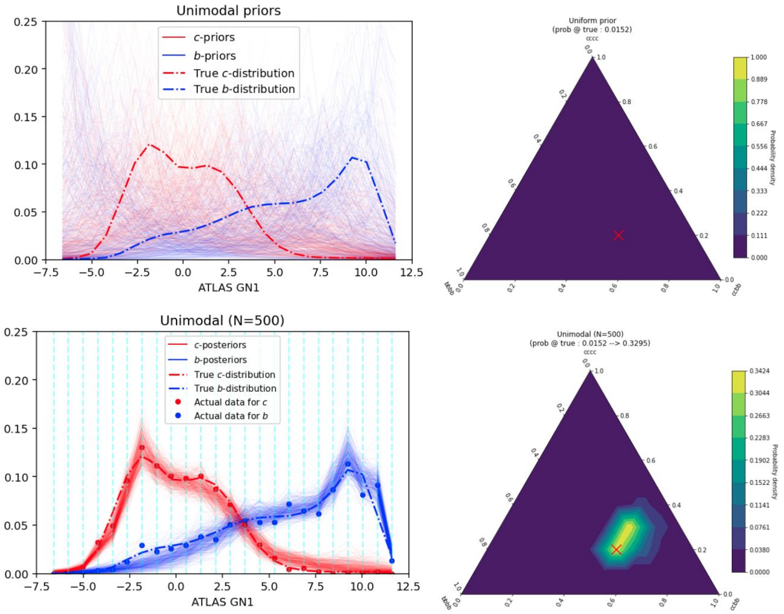
- Toy-model on a toy-problem inspired in pp > hh > bbbb



≈

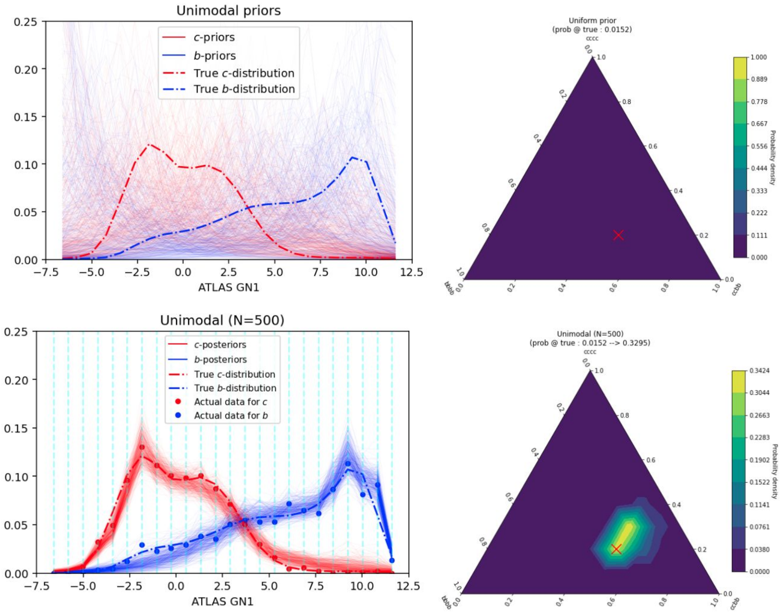


Outlook & Conclusions



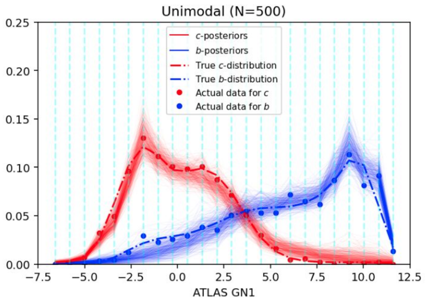
- Fully exploit continuity & unimodality
- Leveraged by
  - Multidimensionality
  - bbbb, bbcc, cccc

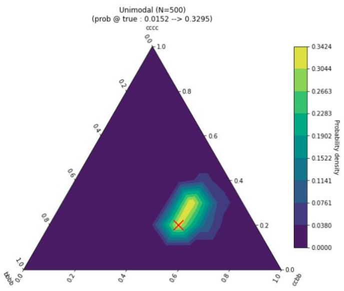
Outlook & Conclusions

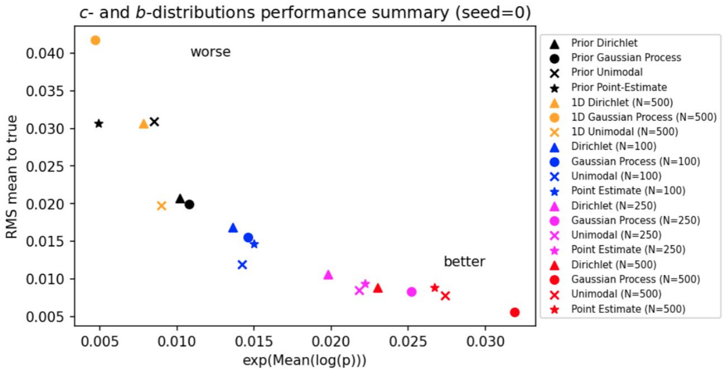


- Fully exploit continuity & unimodality
- Leveraged by
  - Multidimensionality
  - bbbb, bbcc, cccc
- We are extracting correct distributions and fractions
- Using very few assumptions
- Reducing simulations impact

Outlook & Conclusions





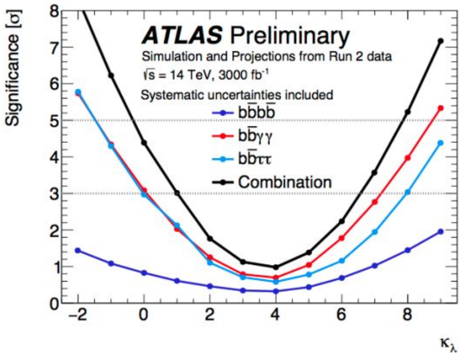


- Fully exploit continuity & unimodality
- Leveraged by
  - Multidimensionality
  - bbbb, bbcc, cccc
- We are extracting correct distributions and fractions
- Using very few assumptions
- Reducing simulations impact

Outlook & Conclusions

Expected significance for SM HH production

	Statistical-only		Statistical + Systematic	
	ATLAS	CMS	ATLAS	CMS
$HH \to b\bar{b}b\bar{b}$	1.4	1.2	0.61	0.95
$HH \to b\bar{b}\tau\tau$	2.5 → 4.0	1.6	2.1 → 2.8	1.4
$HH \to b\bar{b}\gamma\gamma$	2.1 → 2.3	1.8	2.0 → 2.2	1.8 → 2.2
$HH \to b\bar{b}VV(ll\nu\nu)$	-	0.59	-	0.56
$HH \to b\bar{b}ZZ(4l)$	-	0.37	-	0.37
combined	3.5	2.8	3.0 → 3.2	2.6
mmm..	Combined 4.5		Combined 4.0	



pp > hh > bbbb

- Challenging, beautiful, attractive
- Huge enterprise to ride
  - Systematics (pile-up, etc)
  - More backgrounds
  - Integrate steps in unique Bayes
  - etc
- This could also improve bbyy & bbττ!
- Caution with "*I'm a rabbit, i'm a rabbit*" effect (see back-up slides)

Outlook & Conclusions

- Bayesian tools look promising
- No hard cuts. No signal & background regions. No hard-assignments
- Go *analytic* and *probabilistic!*
- Multidimensionality: correlation, correlation, correlation!
- There is more info in the data that what is currently being used ?
- A different approach that may increase pp > hh sensitivity

(We can talk much more about hh!)

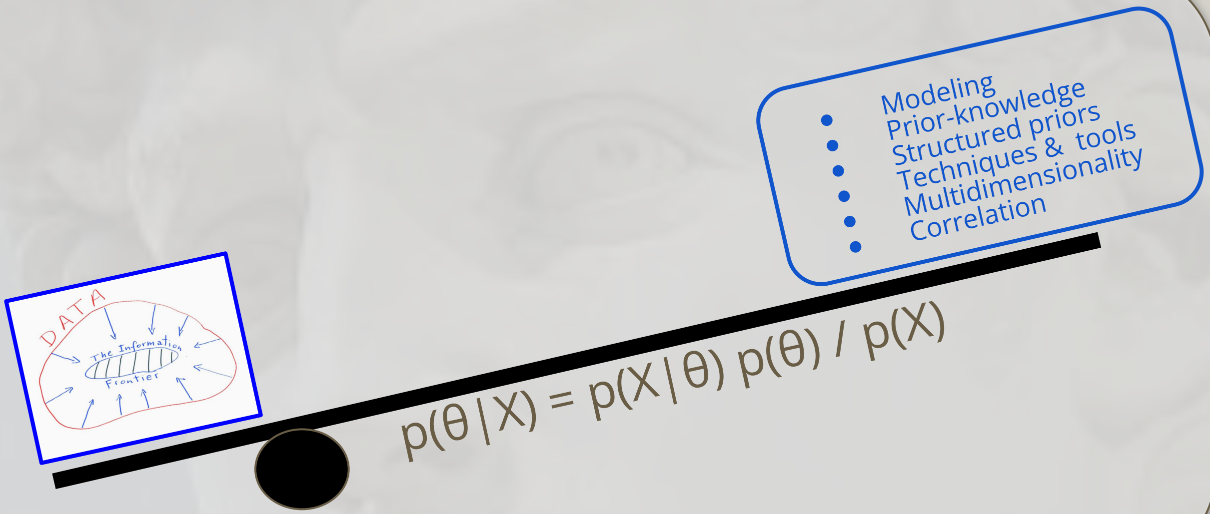
Outlook & Conclusions

change S/√B thinking

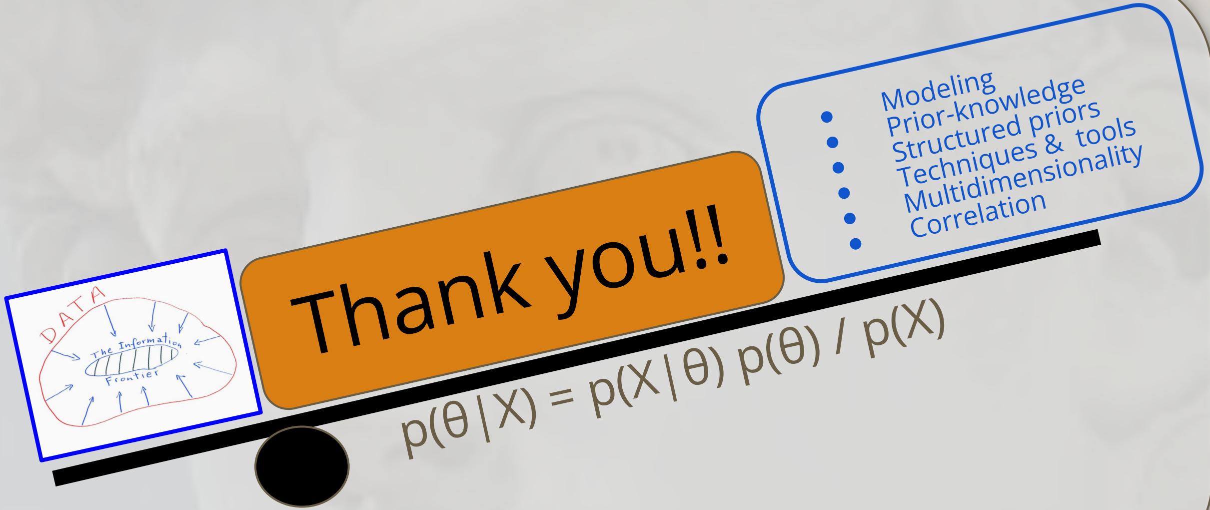
- Bayesian tools look promising
- No hard cuts. No signal & background regions. No hard-assignments
- Go *analytic* and *probabilistic!*
- Multidimensionality: correlation, correlation, correlation!
- There is more info in the data that what is currently being used ?
- A different approach that may increase pp > hh sensitivity

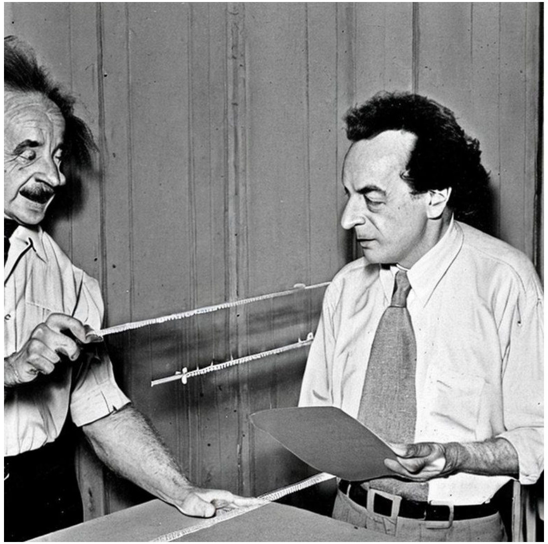
    (We can talk much more about hh!)

The Information Frontier



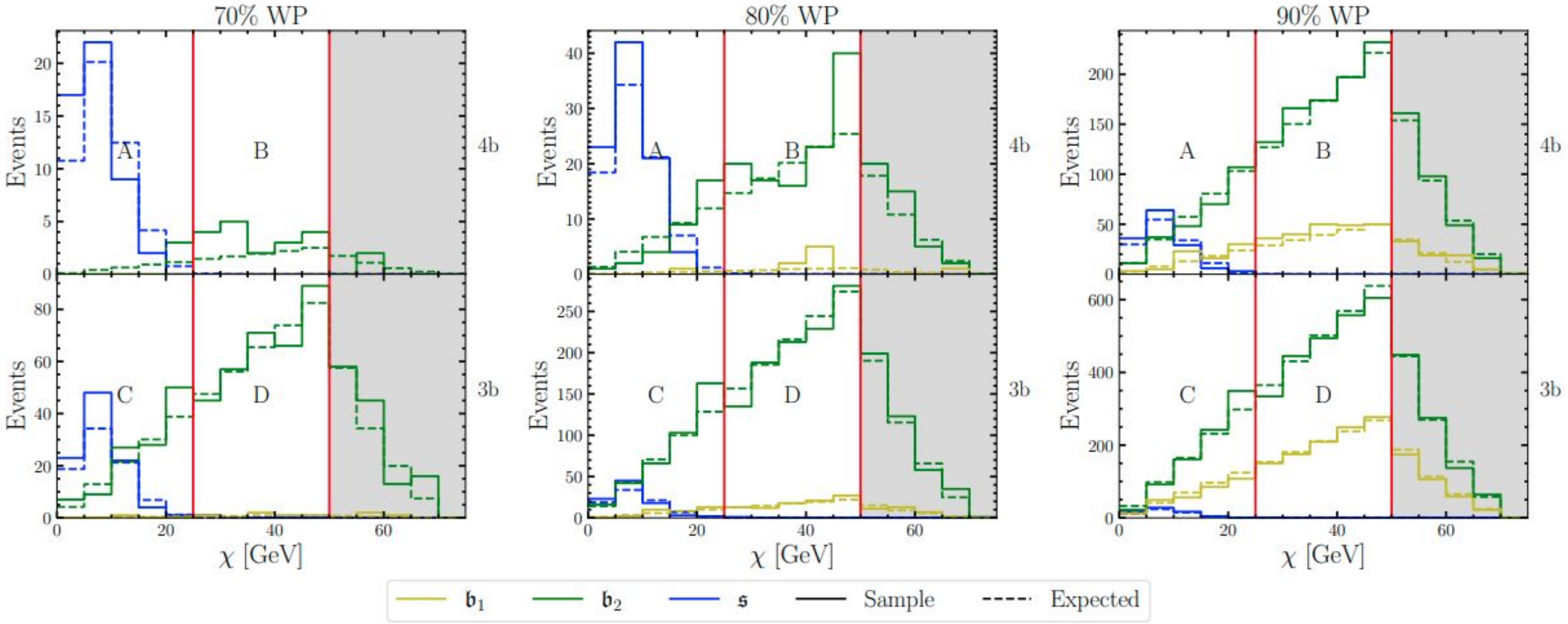
The Information Frontier



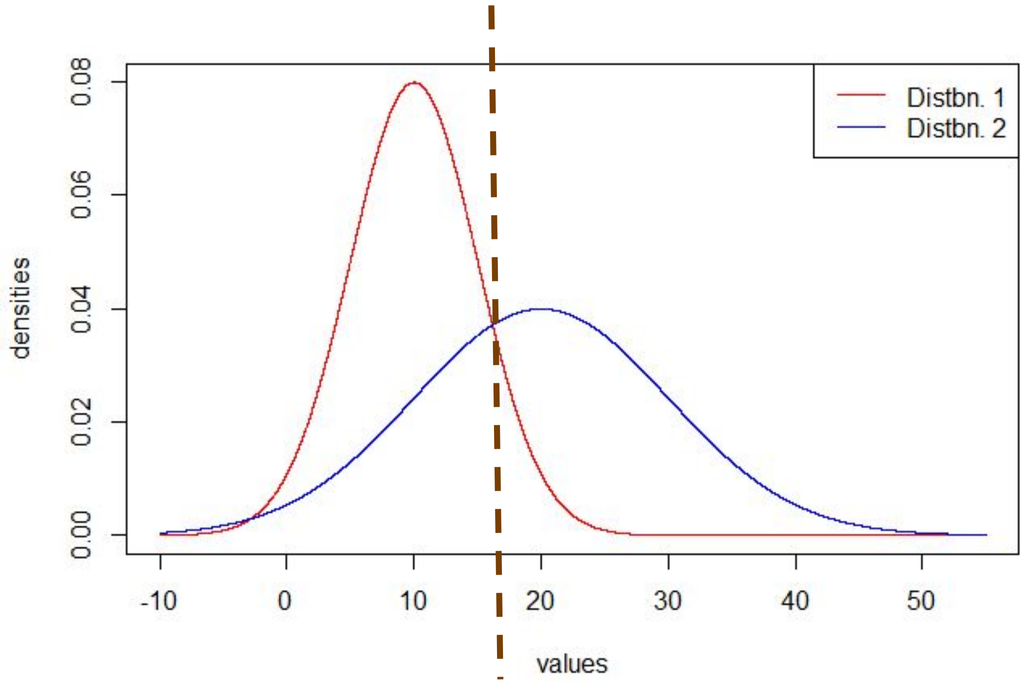


Backup slides

ABCD for WP = 70%, 80% and 90%



Hard- Vs Soft-assignment

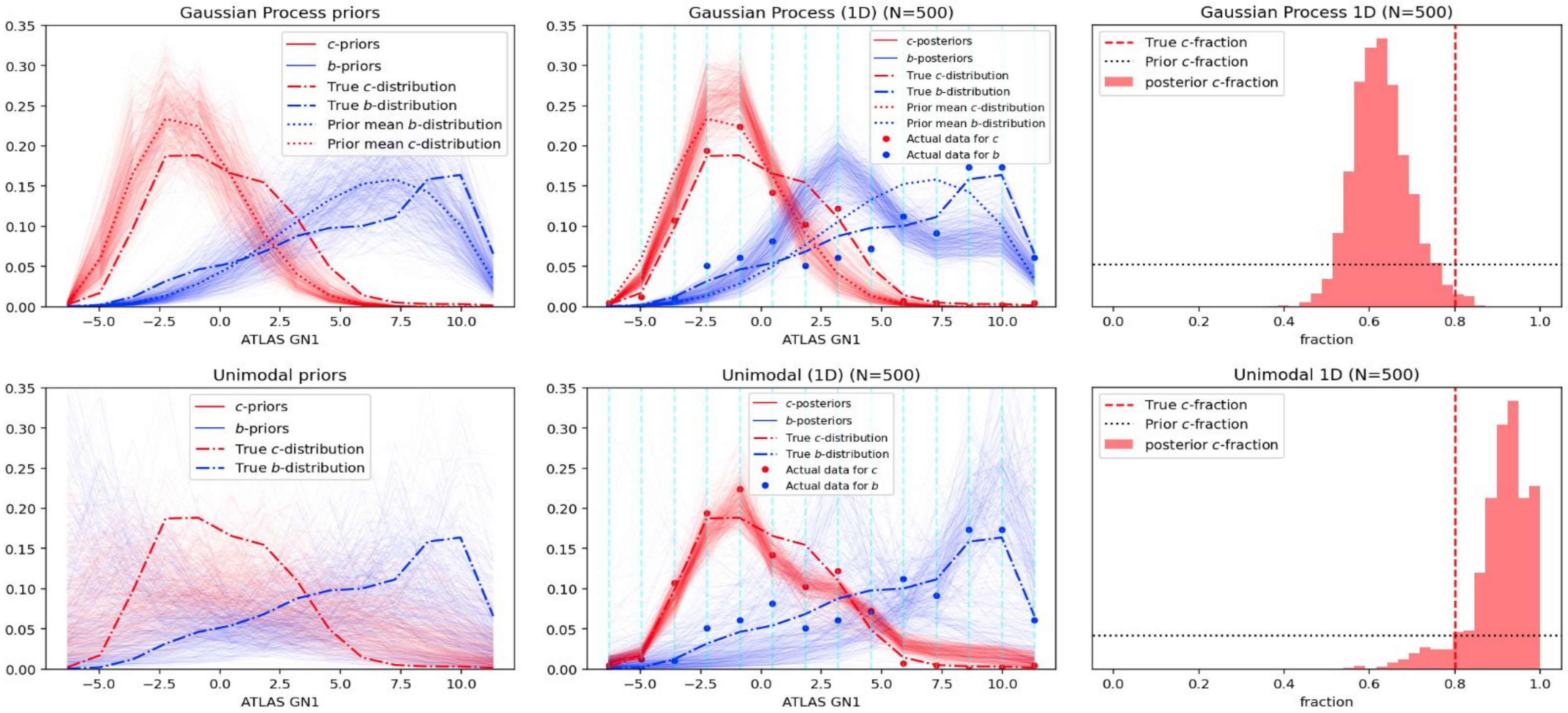


The only way of having same number of events is if blue area on the left equals red area on the right. Very unlikely.

And even in that case the behaviors are different

1D inference problem

The problem has non-identification



Joke

Scotland Yard, FBI and Argentine Federal police are in the world's final Police-detective Contest, in which a rabbit is set free and it has to be found.

First day, FBI takes 2hs and finds the rabbit. How did you do it? Well.. we computed the wind, the trees distribution and the genetic pattern, and we knew where it was going to be. Clap clap clap....👏👏👏

Second day, Scotland Yard takes 30m! How did you do it?! Well, quite easy, we knew its shape, its skills, the forest distribution, we plugged everything to our AI, and we knew exactly where to find it. Wooow...👏👏👏👏👏👏!!!!

And then the third day came the Argentine Federal Police... 30m.. Nothing... 2hs..nothing....10hs... nothing....1 day...nothing... 2days... nothing!!! And after a week they arrived.... [page down]

Joke

Scotland Yard, FBI and Argentine Federal police are in the world's final Police-detective Contest, in which a rabbit is set free and it has to be found.

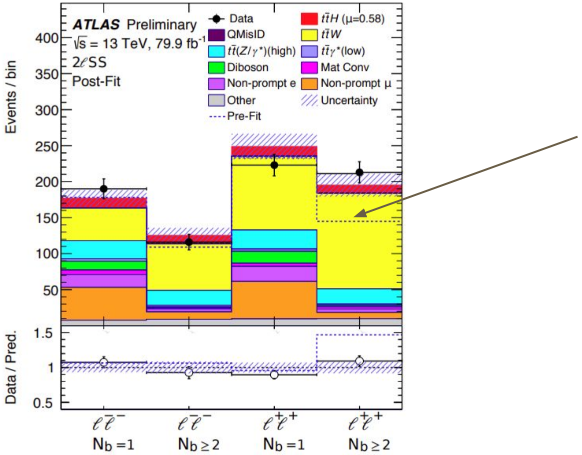
First day, FBI takes 2hs and finds the rabbit. How did you do it? Well.. we computed the wind, the trees distribution and the genetic pattern, and we knew where it was going to be. Clap clap clap....👏👏👏

Second day, Scotland Yard takes 30m! How did you do it?! Well, quite easy, we knew its shape, its skills, the forest distribution, we plugged everything to our AI, and we knew exactly where to find it. Wooow...👏👏👏👏👏👏!!!!

And then the third day came the Argentine Federal Police... 30m.. Nothing... 2hs..nothing....10hs... nothing....1 day...nothing... 2days... nothing!!! And after a week they arrived....

Typical problem at LHC

- **Theory**: SM or BSM
- **Data**: events with p_T, E_{miss}, N_b, N_j, etc.
- **Simulations**: a guide of what to expect of Signal and few Backgrounds



Typical problem at LHC

- **Theory**: SM or BSM
- **Data**: events with p_T, E_{miss}, N_b, N_j, etc.
- **Simulations**: a guide of what to expect
- Expected Signal and few Backgrounds

Typical problem at LHC

- **Theory**: SM or BSM
- **Data**: events with p_T, E_{miss}, N_b, N_j, etc.
- **Simulations**: a guide of what to expect
- Expected Signal and few Backgrounds

Plug theory in simulations and compare to data

Signal region: cuts, selections

Typical problem at LHC

- **Theory**: SM or BSM
- **Data**: events with p_T, E_{miss}, N_b, N_j, etc.
- **Simulations**: a guide of what to expect
- Expected Signal and few Backgrounds

Plug theory in simulations and compare to data

Signal region: cuts, selections

Use shapes from simulations and fit yields

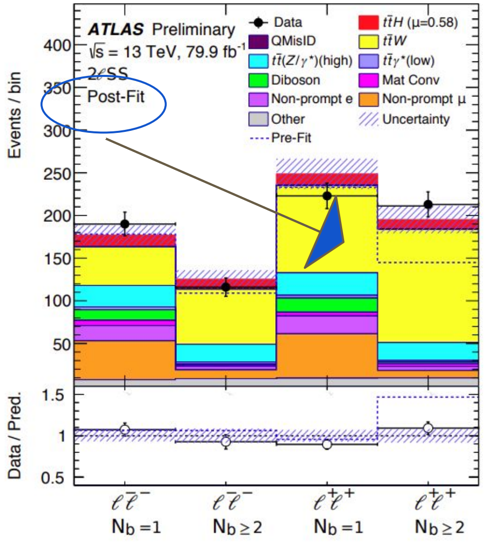
Typical problem at LHC

- **Theory**: SM or BSM
- **Data**: events with p_T, E_{miss}, N_b, N_j, etc.
- **Simulations**: a guide of what to expect
- Expected Signal and few Backgrounds

Plug theory in simulations and compare to data

Signal region: cuts, selections

Use shapes from simulations and fit yields

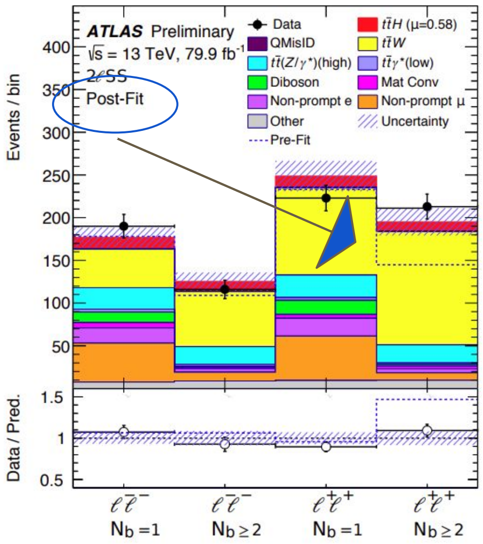


Typical problem at LHC

- **Theory**: SM or BSM
- **Data**: events with p_T, E_{miss}, N_b, N_j, etc.
- **Simulations**: a guide of what to expect
- Expected Signal and few Backgrounds

Plug theory in simulations and compare to data

Signal region: cuts, selections



Train NN, E.g. classifier, and define signal region

Use shapes from simulations and fit yields

Typical problem at LHC

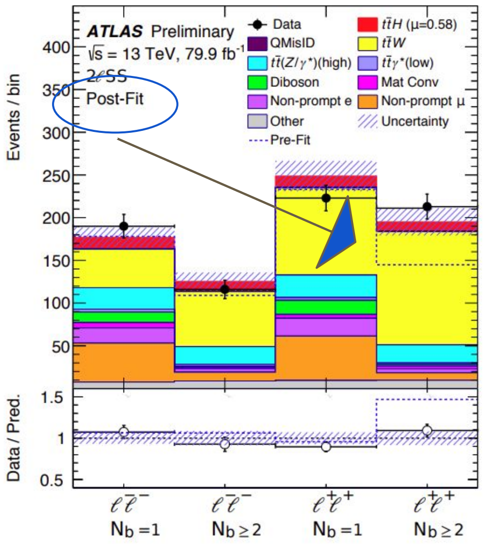
- **Theory**: SM or BSM
- **Data**: events with p_T, E_{miss}, N_b, N_j, etc.
- **Simulations**: a guide of what to expect
- Expected Signal and few Backgrounds

Plug theory in simulations and compare to data

Signal region: cuts, selections

Train NN, E.g. classifier, and define signal region

Use shapes from simulations and fit yields



[Submitted on 28 May 2024]

Is machine learning good or bad for the natural sciences?

David W. Hogg (NYU, MPIA, Flatiron), Soledad Villar (JHU, Flatiron)

Machine learning (ML) methods are having a huge impact across all of the sciences. However, ML has a strong ontology - in which only the data exist - and a strong epistemology - in which a model is considered good if it performs well on held-out training data. These philosophies are in strong conflict with both standard practices and key philosophies in the natural sciences. Here, we identify some locations for ML in the natural sciences at which the ontology and epistemology are valuable. For example, when an expressive machine learning model is used in a causal inference to represent the effects of confounders, such as foregrounds, backgrounds, or instrument calibration parameters, the model capacity and loose philosophy of ML can make the results more trustworthy. We also show that there are contexts in which the introduction of ML introduces strong, unwanted statistical biases. For one, when ML models are used to emulate physical (or first-principles) simulations, they introduce strong confirmation biases. For another, when expressive regressions are used to label datasets, those labels cannot be used in downstream joint or ensemble analyses without taking on uncontrolled biases. The question in the title is being asked of all of the natural sciences; that is, we are calling on the scientific communities to take a step back and consider the role and value of ML in their fields; the (partial) answers we give here come from the particular perspective of physics.