Unsupervised Quark/Gluon Jet Tagging With Poissonian Mixture Models

Machine Learning at GGI

Manuel Szewc, Jozef Stefan Institute

Jožef Stefan Institute

In this talk

I'll introduce a little bit of collider physics and motivate the need for better modelling in a few areas

Detail **one** possible avenue: probabilistic graphical models or Bayesian networks

I'll show one very detailed example from arxiv:2112.11352, you can ask me also about arxiv:2107.00668 where we apply graphical models as a proof-of-principle for unsupervised four-tops extraction (also in last week's talk)

Motivation

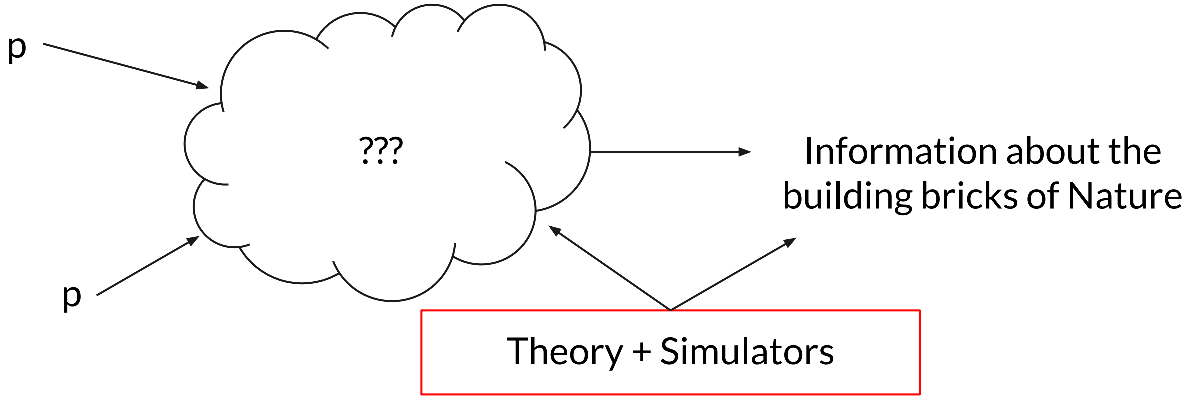
The Standard Model of particle physics represents our understanding of (some of) the most fundamental aspects of Nature.

However, experimental observations (dark matter, neutrino masses and gravity,...) and theoretical issues (θ_{QCD}problem, the Higgs hierarchy problem,...) make clear the limitations of the SM.

How do we explore the SM and beyond?

Motivation

Colliders are among the most powerful tools we have to test and expand our knowledge



Motivation

With colliders, we have been able to test our predictions to astounding precision. The discovery of the Higgs for instance, and the incredible precision in measuring its properties.

However, one may fear we’re getting to the end of the line due to the energies/luminosities achievable and the systematic limitations of our current analysis methodology.

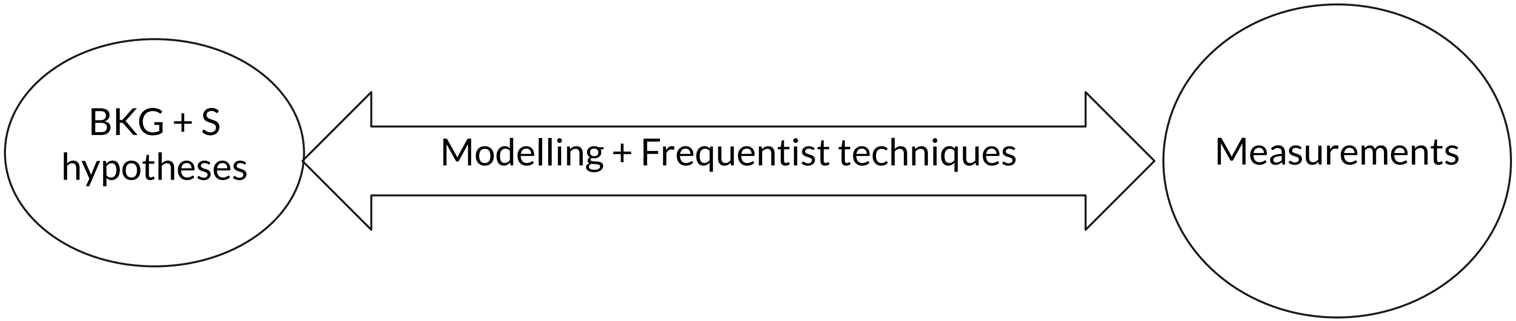
Namely, the need for precision in a high statistics environment can drive the computational costs. And there is no guarantee that the modelling can be precise enough!

MC based strategies for detecting a given signal

BKG-only hypothesis discarded → Discovery (similar to Higgs discovery)

BKG + S hypothesis discarded → Exclusion regions on the specific model parameter space (with possibility of recasting)

BKG + S measured → Parameter estimation → Couplings, **tagger efficiencies**



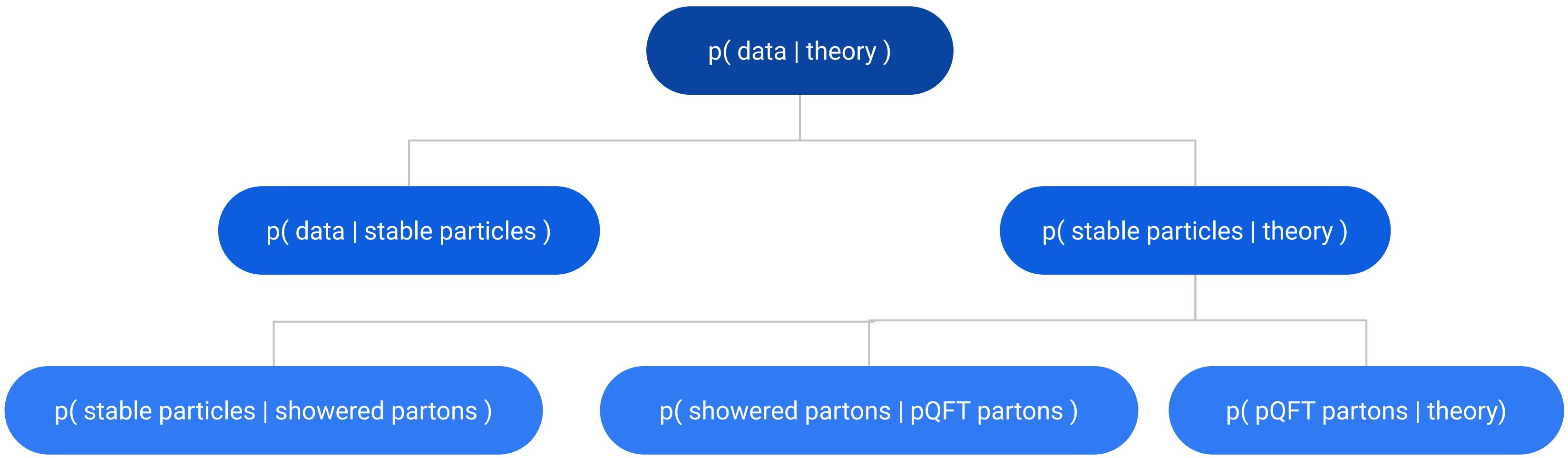
Machine Learning based

Semi-supervised and unsupervised algorithms could ease the dependency on Monte Carlo. Although there is no such thing as a free lunch.

Surrogate models

Surrogate models emulate a given process, and can be preferable to the true process probability density by a variety of reasons: accessibility, speed, storage size.

We already use empirical models for non-perturbative MC simulations (mainly, hadronization) and there are known pitfalls due to additional systematic uncertainties and unphysical behavior.


p(data | theory)
p(data | stable particles)
p(stable particles | theory)
p(stable particles | showered partons)
p(showered partons | pQFT partons)
p(pQFT partons | theory)

Bayesian techniques

They already permeate event generators and their surrogate models extensions: NNPDFs, Bayesian Neural Networks, cINN surrogate models.

Bayesian inference allows for a better evaluation of uncertainties and for improved unfolding (as long as one can make probabilistic statements).

To me, Bayesian always sounds good… Now it is becoming more possible. Differentiable programming is a key development. So are different approaches to inference like Black Box Variational Inference and simulation-based inference where the intractabilities are somewhat sidestepped.

Bayesian techniques for modelling

There is a vast array of literature about Bayesian model building we can take advantage of. Already a lot of examples in HEP (e.g. GANs, VAEs, Shower Deconstruction, LDA, Topic models in general ...)

Always need to keep in mind the same requirements as for event generators:

- **Physically meaningful**: harder to achieve than with event generators. Requires physical bias to be baked into the model.
- **Speed + size**: training should not be too hard nor require so much data as to render simulations preferable

Probabilistic Graphical Models

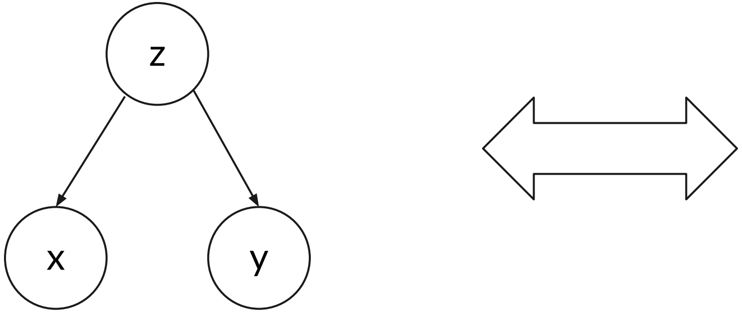
Clear, interpretable, testable way to state our modelling assumptions and incorporate physical bias.

Vast literature on **implementations** and **inference** available.

Has already found many applications in collider physics (see Topic models which has been applied to quark/gluon tagging at high p_T arxiv: 2205.04459, 4-top searches arxiv:1911.09699 and quark/jet modifications Heavy Ions collisions arxiv:2008.08596).

Probabilistic Graphical Models

Association between a **Graph** (we'll deal with Directed Acyclical Graphs or DAGs) and a **joint distribution** or **model**. We call a DAG with an associated distribution a **Bayesian Network**.



$$p(x, y, z) = p(x|z)p(y|z)p(z)$$

Probabilistic Graphical Models

Powerful visualization technique to express assumptions. Useful for designing and motivating models.

Economical representation of the joint distribution that also provides insights into properties of the model, like **conditional independence** from graphical criteria like **d-separation.**

Inference can be expressed efficiently in terms of graphical manipulations.

What I'm interested in...

We can think of **Bayesian Networks for unsupervised learning**.

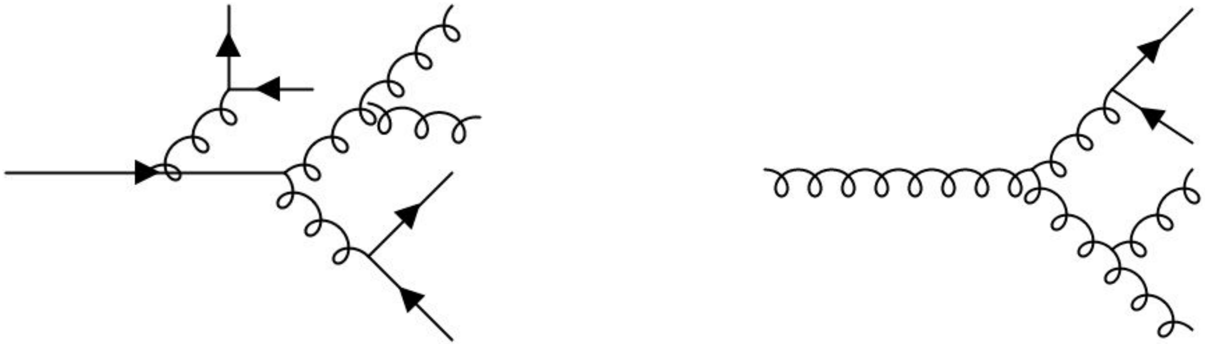
We do not learn with labels. Instead, we want to model the generative process of the data and infer the appropriate underlying classes and hopefully match them with physical processes (non-guaranteed!).

We can then run statistical tests using the learned probability densities (frequentists tests with the Likelihood, but specially Bayesian tests to qualify the learned data distributions).

An example: A Quark/Gluon classifier

Unsupervised quark/gluon jet tagging with Poissonian Mixture Models

Based on E. Alvarez, M. Spannowsky and MS, arxiv:2112.11352



An example: A Quark/Gluon classifier

Unsupervised quark/gluon jet tagging with Poissonian Mixture Models

Based on E. Alvarez, M. Spannowsky and MS, arxiv:2112.11352

- Highly important task for a lot of analyses
- Usually quark/gluon classifiers rely on supervised datasets with approximately well-known observables
- But Monte Carlo bring uncertainties (and also the definition of quark and gluon jets can be problematic!)

Quark vs Gluon jets

Let’s focus on a well-known observable, the iterative SoftDrop multiplicity (C. Frye, A. J. Larkoski, J. Thaler, and K. Zhou, arxiv:1704.06266)

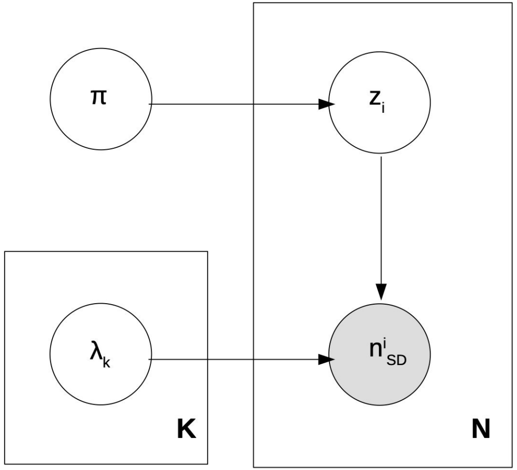
At Leading Logarithmic order, it follows a Poissonean distribution for each type of jet

Let’s consider a mixture model that can be trained on unlabeled data! Ideally, we would match the classes to quark and gluon jets.

As a benchmark, we consider the Quark/Gluon dataset provided by P. T. Komiske, E. M. Metodiev and J. Thaler, arxiv:1810.05165

The mixture model

$$p(X) = \prod_{i=1}^{N} \sum_{k=\{q,g\}} \pi_k \text{Poisson}(n^i_{SD}; \lambda_k)$$

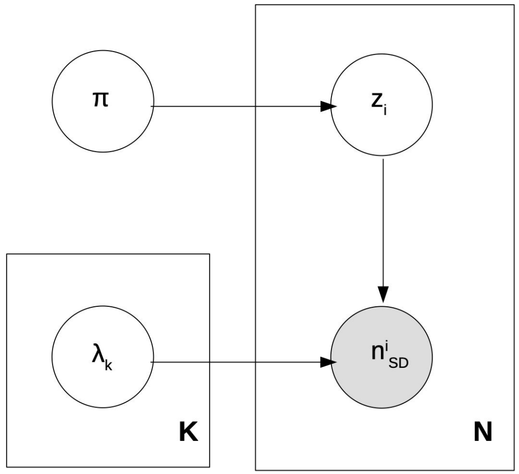


The mixture model

Here we don't include any priors (but we consider uniform priors on a limited range)

Mode degeneracy is avoided by identifying gluon jets with the largest rate found

Mode collapse is (mostly) avoided because there are only two classes and the problem is rich enough



Supervised and Unsupervised Metrics

We can define a **quark/gluon tagger** from the learned model

$$p(z = \text{quark}|n_{SD}, \pi^{\text{MLE}}, \lambda^{\text{MLE}}) = \frac{\pi_q^{\text{MLE}} \text{Poisson}(n_{SD}, \lambda_q^{\text{MLE}})}{\sum_{k=\{q,g\}} \pi_k^{\text{MLE}} \text{Poisson}(n_{SD}, \lambda_k^{\text{MLE}})}$$

We can also obtain the **learned data probability density**

$$p(n_{\text{SD}}|\text{data}) = \pi_g^{\text{MLE}} \, \text{Poisson}(\lambda_g^{\text{MLE}}) + (1.0 - \pi_g^{\text{MLE}}) \, \text{Poisson}(\lambda_q^{\text{MLE}})$$

Supervised and Unsupervised Metrics

With a tagger we can compute the usual supervised metrics (accuracy, AUC, etc)

But more interestingly, we can compare the learned data density with the measured density to obtain unsupervised metrics

$$d_H(p,q) = \frac{1}{\sqrt{2}}\sqrt{\sum_{n_{SD}=0}^{\infty}\left(\sqrt{p(n_{SD})} - \sqrt{q(n_{SD})}\right)^2}$$

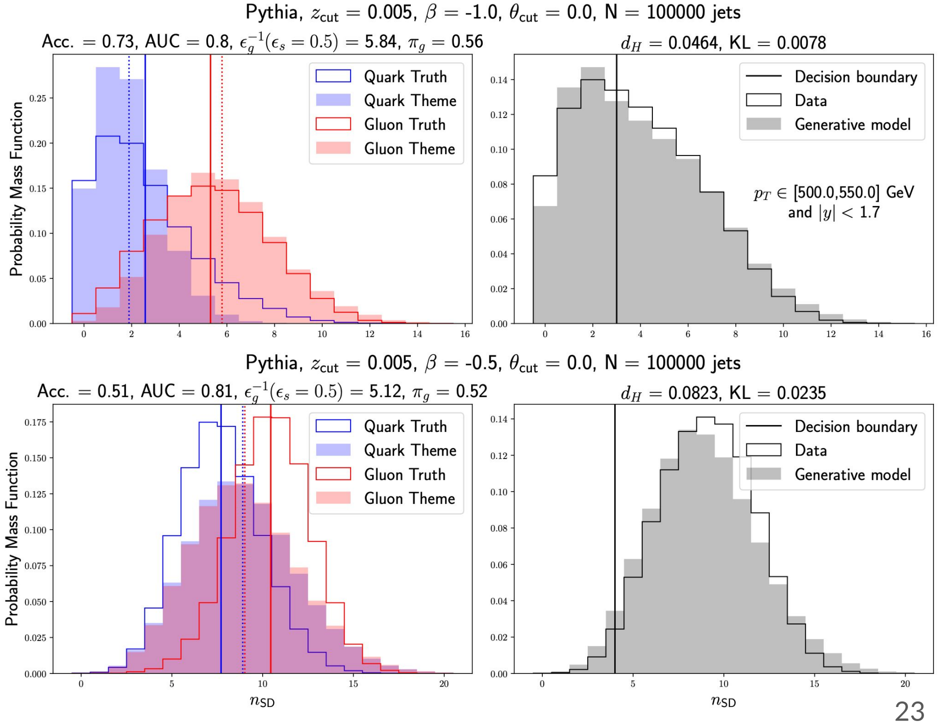
$$\mathrm{KL}(p||q) = -\sum_{n_{SD}=0}^{\infty} p(n_{SD})\mathrm{Ln}\left(\frac{q(n_{SD})}{p(n_{SD})}\right)$$

Good and bad examples

We show learned distributions for the underlying processes and for the data

Supervised metrics are for validation, but unsupervised metrics are the ones we care about.

AUC does not permit good hyperparameter selection

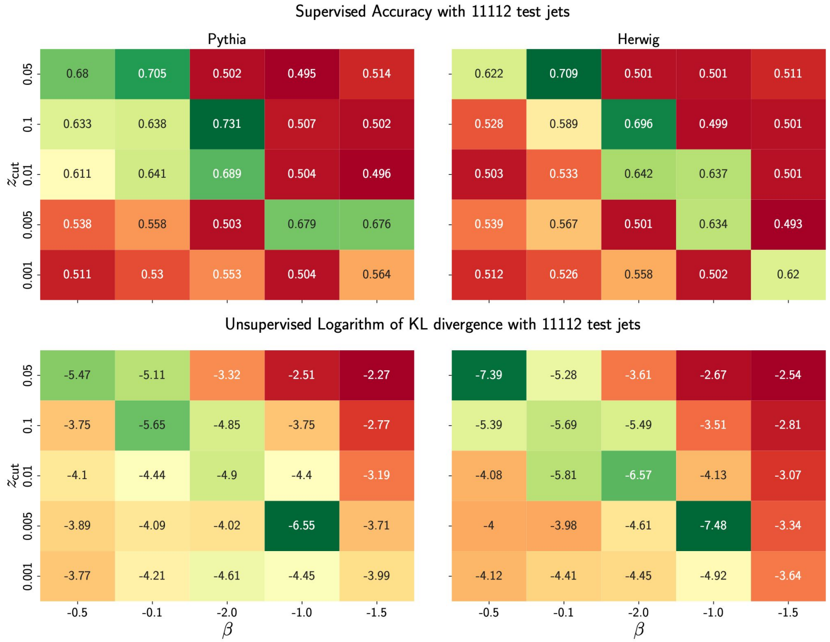


Supervised vs Unsupervised metrics

There is a good correlation between supervised and unsupervised metrics

They allow us to select hyperparameters where the model is good at explaining the data **and** at distinguishing quarks from gluons

Model performance is to a large extent **dataset independent**

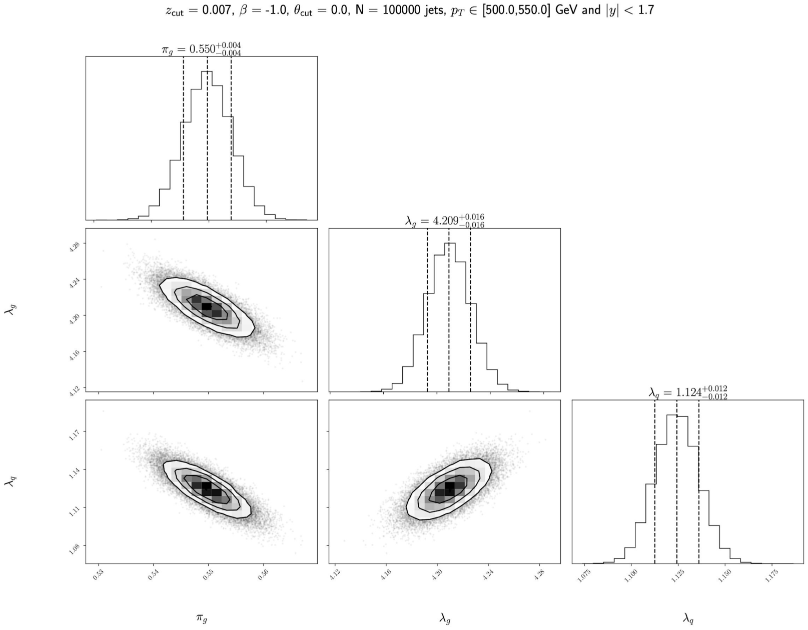


Full Bayesian analysis

We can go from MLE to full Bayesian analysis (with uniform priors)

We observe the correlations between random variables

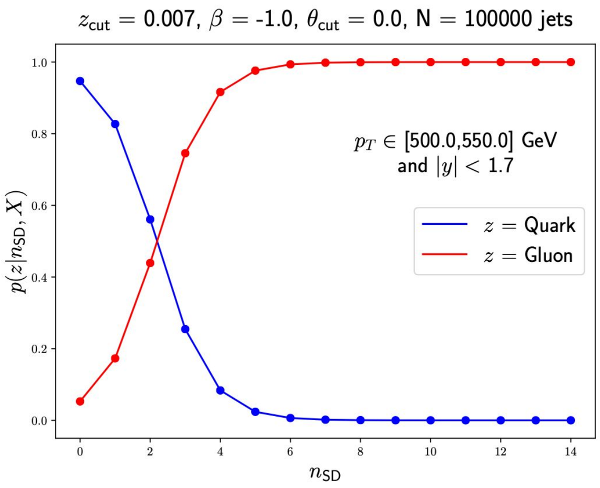
But very narrow distributions! Likelihood-driven case



We can also marginalize to get a better tagger

This is a tagger that considers the full posterior distribution

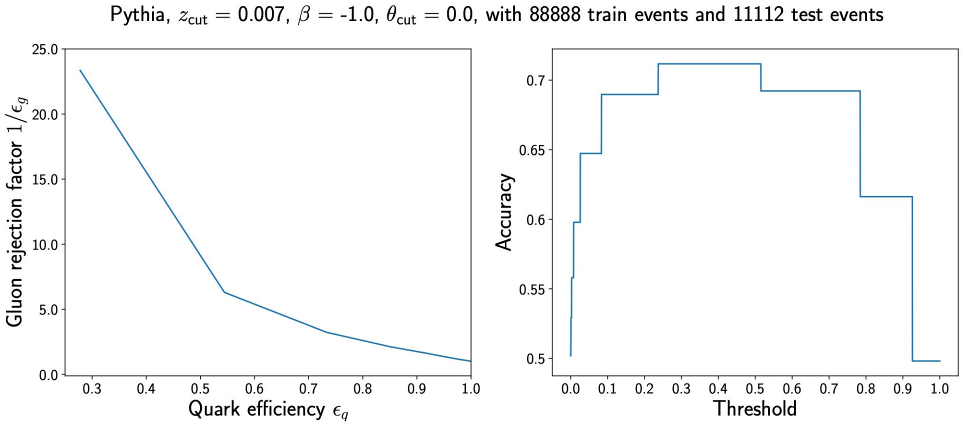
$$p(z|n_{\mathrm{SD}}, X) \approx \frac{1}{T}\sum_{t=1}^{T} p(z|n_{\mathrm{SD}}, \pi_g^t, \lambda_g^t, \lambda_q^t)$$



The classifier performance

We observe how the performance is best around the meaningful probabilistic threshold

Remember, this is unsupervised learning.



What did we learn?

We can obtain a quark/gluon classifier directly from data assuming a Poisson mixture model.

This classifier can be optimized with data-driven metrics, resulting in accuracy in the 0.65-0.70 range.

This classifier is robust against detector effects.

We could incorporate these unsupervised methods to traditional analyses (either at the Likelihood level or by computing Bayesian tests)

Main issues

The dataset provides a narrow bin. For a more realistic implementation, more kinematic information should be included either as another set of variables or by binning the dataset into different subsets.

The Poisson hypothesis is only approximately true… How do we deal with deviations? It is not that important for tagging but it is for Monte Carlo tuning.

All in all

A good, solid, unsupervised classifier which is really easy and cheap to implement

It can be extended with other observables provided we have some good understanding on how to model them

Can be part of a functional definition of quarks and gluons

We have not implemented the learned distributions in statistical tests as we are dealing with a classification problem, but dealing with BSM searches would be a different matter

All in all

Two states of the art collide! It's very interesting to see what can happen.

ML can enhance statistical analyses either by taking full advantage of Monte Carlo distributions or reducing dependence on them.

I tend to value simpler models where we can incorporate domain knowledge and where unknowns are easier to catch.

However, it's hard to generalise to not so nice distributions… Deep Exponential Families and Black Box Variational Inference could provide a trade-off between power and interpretability.

Conclusions

Collider experiments are providing us with incredible amounts of measurements.

To take advantage of it, simulations and analyses techniques need to step-up.

Bayesian graphical models provide clear modelling and fast inference.

A possible tool among many, whenever it is convenient to apply them.

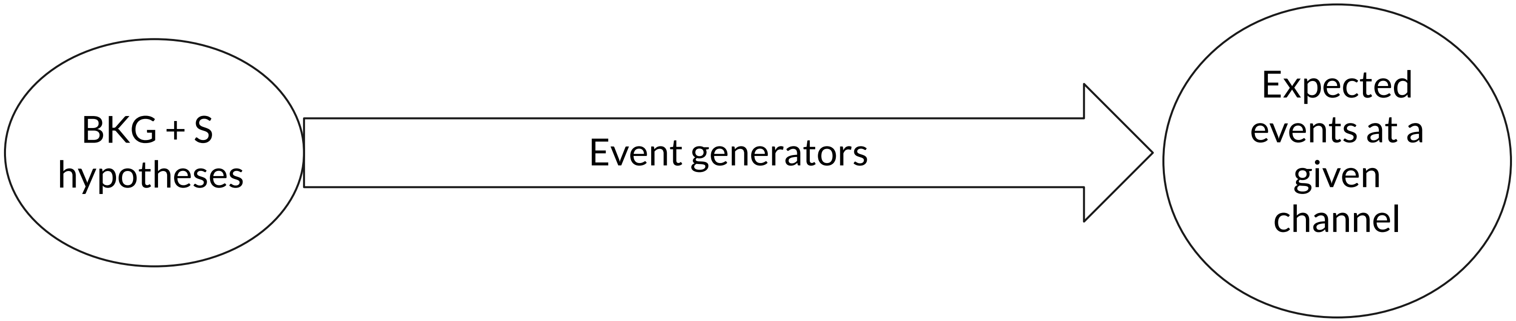

Thank you!
©QUINO

Backup slides

Monte Carlo based strategies

We perform dedicated analyses with specific topologies in mind.

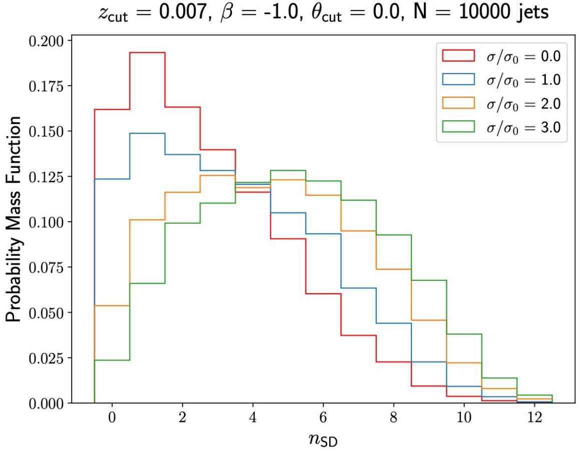
Dedicated searches rely on good modelling: the simulation pipeline is fundamental.

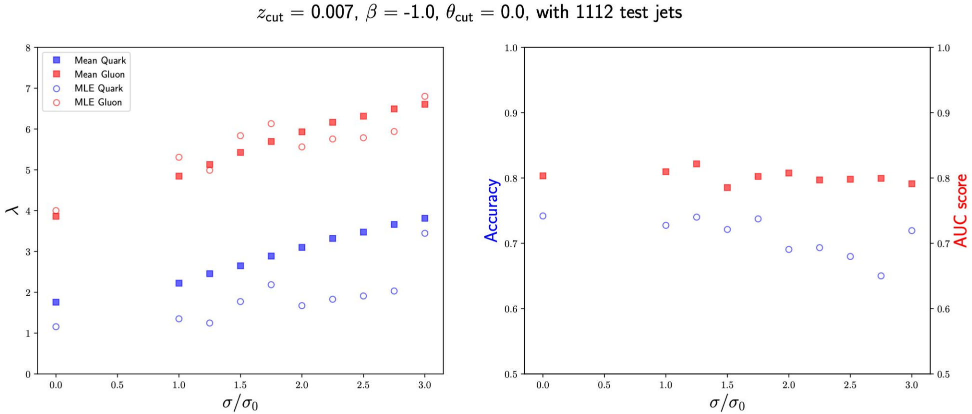


Monte Carlo based strategies

With the expected events, we can turn from estimation to inference. In collider studies, there is a primacy of frequentist methods to estimate or set limits to the parameters of interest or even for model selection.

Detector Effects





Full Bayesian Analysis: Posterior predictive