Lecture 25

Reminder: Please fill out the course survey!!

HCP applications of Normalizing Flows

- Anomaly Detection → see next lecture
- Unfolding (Heidelby group): learn invertible transformation between partons
- Phase space integration/sampling (part of event generation)
\[ \int d^d x \, f(x) \]

Monte Carlo integration in high dimensions is expensive/inefficient. Have to sample a lot if \( f(x) \) is sharply peaked and inefficiently localized on lower dimensional subspace. Multimodal sparsity.

Idea: if \( \int p(x) \, f(x) \, dx \)

Learn \( p(x) \) to hit true \( p(x) \)

Importance sampling: \( \frac{1}{N} \sum \frac{f(x)}{p(x)} \)
New (final) Topic: Anomaly Detection @ UHC

"Model-Independent Searches for NP @ UHC"

LHC -- huge dataset

How do we find anomalies (new physics) if we don't know what we're looking for?

If we do know what we're looking for

Suppose X is heavy, charged, stable

maybe know decay signature

characteristic feature of X event, e.g. hits in tracker
Use features to define a "search region":

\[ x_1 > x_1^{cut} \]
\[ x_2 > x_2^{cut} \]
\[ \vdots \]
\[ x_n > x_n^{cut} \]

Expect reduced SM (B)
Enhanced signal model (S)

Now look at data: \( N_{obs} \)

Need to know precisely

"SM background prediction"

\[ \frac{N_{obs} - B}{\sqrt{B}} > 1 \]

Use Poisson probability distribution

\[ \text{prob}(N_{obs} | B) \]

"p-value"

To compute probability of \( N_{obs} \)

Under null hypothesis:

\[ \sigma \sim \sqrt{B} \]

Suppose \( Y \) is invisible.

\( N_{jets} \)

\( H_T \) (hadronic energy)

\( \Sigma T \)
How to predict $B$?

- Data driven (don't rely on simulations)

Sideband by estimation

```
E.g., "control region"
```

If $m_{d_0}$ is smooth for $b$

- Add localized for signal

Fit $m_{d_0}$ in CR

Interpolate to get $B$ in SR.

Rely on simulations

- Calculating SM does it
- Generically events
- SIDIS detector

- Not precise enough...

Most analyses combine data-driven & simulation.

$\rightarrow 99.99\%$ all searches for NP @ LHC are like this.
Example of a signature that has no explicit search for it yet:

\[ m_Z > m_{X,Y} \]

So \(X, Y\) decay products are collimated and highly boosted.

Could be setting in date \( \epsilon > 50\) significance.

Q: Could we design a model-agnostic search that would discover this model?
Existing model indep searches EHC - 2 kinds

1. Bump hunt
   - Assume signal located in m_j
   - Assume b_j is smooth.

Almost fully model-independent
   - Assume signal located in m_j

2. "general search", "MUSIC"
   (ATLAS), (CMS)
   bin data into many categories (thousands)
   compare against SM simulation in each bin
   compute p-values...

pos - signal model independent
cons - huge look elsewhere effect if many bins
   - heavily simulation dependent, optimistic
   - challenging...

Idea: train classifier on data vs. simulation

optimal classifier is

\[ \frac{p(x | H)}{p(x | \text{data})} \rightarrow \frac{p(x | \text{SM})}{p(x | \text{data})} \]

optimal test for deviation from SM