Mass or Model:
Signal Diversity in Dark Matter Searches

Thomas Edwards, Bradley Kavanagh, and Christoph Weniger

Edwards, Kavanagh & Weniger 1805.04117
Edwards & Weniger 1712.05401
github.com/cweniger/swordfish
github.com/bradkav/WIMpy_NREFT
github.com/tedwards2412/benchmark_free_forecasting
Experimentally constraining Dark Matter (DM) can be difficult

- No non-gravitational signal of DM present
- Many experiments producing data
- Huge landscape of acceptable models
- Not easy to currently accept or rule out particular models (maybe some MOND models with GWs)
Benchmark-free forecasting provides a global view of the parameter space

What do we want to know?

1. What are the optimal set of experiments to maximise the possibility of a DM discovery?

2. Once a discovery is made, what is the most constraining set of experiments i.e. how can we best test its particle nature?
Traditional Methods: We are limited to testing a number of “benchmark” scenarios

- Limitation is in computational time to compare points in the parameter space
- Distance between two points described by some form of Test Statistic (TS)
- Typically choose set of well motivated points

<table>
<thead>
<tr>
<th>$m_X$</th>
<th>100 GeV</th>
<th>1 TeV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$10^{-46}$</td>
<td>$10^{-47}$</td>
</tr>
<tr>
<td>$</td>
<td>F^-</td>
<td>^2$</td>
</tr>
<tr>
<td>$q^2 / 4m_X^2</td>
<td>F^+</td>
<td>^2$</td>
</tr>
<tr>
<td>$q^2 / 4m_X^2</td>
<td>F^-</td>
<td>^2$</td>
</tr>
<tr>
<td>$q^4 / m_X^4</td>
<td>F^+</td>
<td>^2$</td>
</tr>
<tr>
<td>$q^4 / m_X^4</td>
<td>F^-</td>
<td>^2$</td>
</tr>
<tr>
<td>$</td>
<td>F^+</td>
<td>^2$</td>
</tr>
<tr>
<td>$q^4 / 4m_X^4</td>
<td>F^+</td>
<td>^2$</td>
</tr>
<tr>
<td>$q^4 / 4m_X^4</td>
<td>F^-</td>
<td>^2$</td>
</tr>
</tbody>
</table>
Difficulty comes from computing the likelihood ratio via Monte-Carlo

$$\text{TS}(\theta')_{D(\theta)} \equiv -2 \ln \frac{\mathcal{L}(D(\theta)|\theta')}{\max_{\theta''} \mathcal{L}(D(\theta)|\theta'')} \approx \left( \vec{\theta}_2 - \vec{\theta}_1 \right)^T \mathcal{I} \left( \vec{\theta}_2 - \vec{\theta}_1 \right) \approx \| x(\theta) - x(\theta') \|^2 $$

- The **Fisher Information** then acts a metric on the space of model parameters

- Only **locally true**, higher order corrections become important for larger distances

- Euclideanized signal translates distance metric in model parameter space to the signal space
Fast computation can be achieved through Euclideanized Signals

- **$D_{i,j}$** - Signal and background covariance matrix (plus poisson noise)
- **$S_i$** - Signal in the $i$th bin
- **$E_i$** - Exposure in the $i$th bin
- **$R = 0.1$** - Fudge factor to deal with both signal dominated + signal limited regimes

\[
 x_i \equiv \left( \sum_j (D^{-1/2})_{i,j} S_j E_j \right) \left( 1 + \frac{R \cdot S_i}{R \cdot S_i + B_i + K_{ii} E_i} \right)
\]
Euclideanized signals approximately match the standard log-likelihood ratio test statistic

- Approximation tested by considering a large number of random models (illustrated with 3 bins)

- Background kept constant whilst covariance function and signal randomly generated

- Flat exposure used:
  - **Signal limited**: No covariance and high signal to noise
  - **Systematics limited**: Low signal to noise with high covariance
  - **Poisson limited**: Low signal to noise with no covariance
It is now extremely fast to calculate any confidence contour if the space is sampled enough.

- Once the BallTree is constructed, simply select a point and call for all other points within the required distance.

- Sampling must be high - approximately 10 points per 1-sigma region.
It is now extremely fast to calculate any confidence contour if the space is sampled enough.

- Once within
- Samp

\[ \log_{10}(m_X) \text{ [GeV]} \]

\[ \log_{10}(\sigma_{SI}) \text{ [cm}^2] \]

90% CL UL XENON1T

90% CL UL XENONnT
We can now perform the pairwise comparison of many points. What questions can we answer?

1. Benchmark Free forecasting statements can be easily made i.e. compare all benchmark models at once.

2. Quantify the signal diversity of experiments in terms of the number of two-sigma discriminable regions.
We can now perform the pairwise comparison of many points. What questions can we answer?

1. Benchmark Free forecasting statements can be easily made i.e. compare all benchmark models at once
Case Study: Non-Relativistic Effective Field Theory for Direct Detection Experiments

Experimental Setup: We use XENON-nT and DarkSide20k like detector setup.

Considered interactions: We consider operators 1 (spin-independent), 4 (spin-dependent), and 11 as well as magnetic dipole and millicharge DM models.
Benchmark-free statements can be made performing a series of simple look ups

1. Sample a large number of points

2. Euclideanize the signals according to the detector specifications

3. Construct BallTree to perform fast distance calculations

4. For each point of model M, check if there is a point in model S within the specified significance

\[
-2 \ln \max_{\theta' \in \Omega_S} \frac{\mathcal{L}(D(\theta)|\theta')}{\max_{\bar{\theta}} \mathcal{L}(D(\bar{\theta})|\theta')} \sim \min_{\theta' \in \Omega_S} \| x(\theta) - \bar{x}(\theta') \|^2
\]
Simultaneously constrain the mass and DM-nucleon interaction in only a small region of parameter space

\[ m_\chi = 5 \text{ GeV} \]

- Mass discrimination is not possible roughly above 100-200 GeV
- Characterisation of the DM-nucleon interaction is only possible below 100-200 GeV for regions just below the current limit

\[ m_\chi = 50 \text{ GeV} \]
Simultaneously constrain the mass and DM-nucleon interaction in only a small region of parameter space

\[ m_\chi = 5 \text{ GeV} \]

- Mass discrimination is not possible roughly above 100-200 GeV
- Characterisation of the DM-nucleon interaction is only possible below 100-200 GeV for regions just below the current limit

\[ m_\chi = 50 \text{ GeV} \]
We can now perform the pairwise comparison of many points. What questions can we answer?

2. Quantify the signal diversity of experiments in terms of the number of **two-sigma discriminable regions**
We can approximate the number of discriminable regions using the sampled points in the space.

- In one dimension it is simple to estimate the number of one sigma regions.

- Simply count the number of points within each one sigma of each point and sum over the inverse of this number, $w_i$. 

\[ N = \sum_i w_i \]
Quantifying Signal Diversity using approximate “tiling” of 2-sigma discriminable regions

1. Sample the model parameter space. The sampling must be fine enough such that there are more than 10 points within every confidence contour.

2. Euclideanize the signals using experimental parameters such that each parameter point has an associated new vector $x$.

3. The radius of any confidence contour is defined by $r$. Calculate the number of points within $r$ to give $w_i$

4. The volume is then approximated by $v$. $c_{ff}$ is the filling factor of hypersphere.

$$r_{\alpha}(\mathcal{M}) = \sqrt{\chi^2_{k=d, ISF}(1 - \alpha)}$$

$$\nu_{\mathcal{M},X}(\Omega_{\mathcal{M}}) = c_{ff} \sum_i w_i$$
Infomertic Venn Diagrams: Visualisation of the degeneracy breaking abilities of future experiments

- An argon detector adds **significant** degeneracy breaking abilities
- Venn diagrams present an easy way to see the number of 2-sigma discriminable areas
Conclusion

- The **pairwise comparison** of points in can be made highly efficient using modern clustering algorithms.

- **Euclideanized signals** are a good approximation to the TS allowing for **benchmark-free forecasting** statements.

- Both **mass reconstruction** and **operator discrimination** is only possible is a small part of the parameter space for near future Xenon and Argon Direct Detection Experiments.

- **Venn diagrams** provide an extremely useful and general visualisation of the degeneracy breaking abilities of experiments.