Model-Independent Searches with Background Matrix Elements

James “Jamie” Gainer, University of Florida
Work by Dipsikha Debnath, JG, Konstantin Matchev
arXiv:1405.5879

LHCP 2014: June 3, 2014
Goal: Discover new physics without a specific signal hypothesis
Why?
• Many ideas for physics beyond the Standard Model

(Murayama)
• Many ideas for physics beyond the Standard Model

• So far the LHC has yet to find this new physics.

(Murayama)
• Many ideas for physics beyond the Standard Model

• So far the LHC has yet to find this new physics.

• Most analyses are optimized for a particular model (e.g. mSUGRA).

(Murayama)
• Many ideas for physics beyond the Standard Model

• So far the LHC has yet to find this new physics.

• Most analyses are optimized for a particular model (e.g. mSUGRA).

• Can we discover “Not Yet Thought Of” theories?
• We would like to be as sensitive to new physics as possible

• Without optimizing for a particular signal model

• We take our inspiration from the Matrix Element Method.
In a mathematically well-defined sense, the best choice of test statistic for distinguishing between two hypotheses (like “signal” and “background”) is the likelihood ratio/ discriminant

\[ \Lambda(E) = \frac{L(H_1 \mid E)}{L(H_0 \mid E)} \]

where \( H_0 \) and \( H_1 \) are two alternative hypotheses and \( E \) is the data in an experiment.
In a mathematically well-defined sense, the best choice of test statistic for distinguishing between two hypotheses (like “signal” and “background”) is the likelihood ratio/ discriminant

\[ \Lambda(E) = \frac{L(H_1 \mid E)}{L(H_0 \mid E)} \]

where \( H_0 \) and \( H_1 \) are two alternative hypotheses and \( E \) is the data in an experiment.

Suggests that the likelihood will be an optimal variable.
Matrix Element Method

• In particle physics, the likelihood we use the expression for probability on the right:

\[
P(x_i | A(B)) = \frac{1}{\sigma(A(B))} \sum_{k,l} \int dx_1 dx_2 \frac{f_k(x_1)f_l(x_2)}{2sx_1x_2} \\
\times \left[ \prod_{\text{all } j} \int \frac{d^3q_j}{(2\pi)^32E_j} \right] \times \left[ \prod_{\text{visible } j} T(\{q_j\}, \{p_j\}) \right] \\
\times |\mathcal{M}_{A(B),kl}(\{q_j\})|^2,
\]

• Normalized to the total cross section

• With integrals over transfer functions, invisible momenta, etc.

• Use of this likelihood = “Matrix Element Method”
• In the Neyman-Pearson Lemma we needed a likelihood ratio.

• Need to know both signal and background likelihood to compute this ratio

• For signal independence, use background likelihood as a test statistic.

• Matrix element variables still “know a lot” about the background so should be optimal at rejecting background
As an example, we consider 20 event pseudoexperiments
Background Likelihood

Sum of log MEKD values in a pseudoexperiment

Processes: gluon fusion Higgs $\rightarrow 4\ell$
Processes: $q\bar{q} \rightarrow 4\ell$ background
Background Likelihood

Processes: $q\bar{q} \rightarrow 4\ell$ background, $\Gamma_Z \rightarrow \Gamma_Z/5$
For each 20 event pseudoexperiment, we calculate the sum of the

Background Likelihood

logarithms of background squared matrix elements evaluated for backgrounds

125 GeV Higgs

Small Z Width

Sum of log MEKD values in a pseudoexperiment
Background Likelihood

the particular kinematics of the event.
We perform this calculation using the MEKD package.
Background Likelihood

Avery, Bourilkov, Chen, Cheng, Drozdetskiy, JG, Korytov, Matchev, Milenovic, Mitselmakher, Park, Rinkevicius, and Snowball


Chen, Cheng, JG, Korytov, Matchev, Milenovic, Mitselmakher, Park, Rinkevicius, and Snowball

Phys. Rev. D89 (2014) 034002

http://mekd.ihepa.ufl.edu
We take our 20 “data” events and evaluate the sum of the logarithms of the sum of log MEKD values in a pseudoexperiment.
of background squared matrix elements, obtaining “Data Value.”
Background Likelihood

p-value is shaded region: fraction of pseudoexperiments
Background Likelihood

with more extreme values of the test statistic than “Data Value”
• p-value from likelihood distribution calculated from Monte Carlo pseudoexperiments

• so reducible backgrounds, detector effects, NLO (if your MC has it) etc. are included in the test statistic distribution, though not in the calculation of the test statistic
Flattening the Background

• I’ve argued for using the background matrix element as a “test statistic” for discovering signals in a model-independent way

• Now I’m going to present some related ideas in which we use similar tools to obtain flat background distributions
Why?
When your background is flat...
It’s easy to discover an unexpected signal.
Example: Football Scores


- Excess in scores that end in “7”, “3”, or “4” evidence that scores are quantized in units of “7” or “3”.
Example: Football Scores


• Broad peak at (1,1). Suggests that scores quantized in units of “1”; not as many scores per game.
• We learned about the structure of football(s) from deviations from flatness in distributions.

• Can we do the same in particle physics?

• How do we make background distributions flat?
Background Ranking

- Want to flatten the background distribution of ME-based variable
Background Ranking

\[ r(\mathcal{E}) = \int_0^{M_E(\mathcal{E})} dM_E \frac{dN}{dM_e} \]
Background Ranking

\[ r(\mathcal{E}) = \int_0^{M_E(\mathcal{E})} dM_E \frac{dN}{dM_e} \]
Background Ranking

\[ r(\mathcal{E}) = \int_{0}^{M_{E}(\mathcal{E})} dM_{E} \frac{dN}{dM_{e}} \]
If we take our normalization from data, deficits will also indicate signals.
Simpler Approach: Quantile Bins

- A mini-version of ranking. Make quantile bins in $\text{ME}(\varepsilon)$ and other variables (here four-lepton invariant mass).

Like the NFL scores example above.
Simpler Approach: Quantile Bins

- A mini-version of ranking. Make quantile bins in $ME(\xi)$ and other variables (here four-lepton invariant mass).

Here we consider 150 BG events.

Background Only: One Pseudoexperiment

Here we consider 150 BG events.
Simpler Approach: Quantile Bins

• A mini-version of ranking. Make quantile bins in ME(ξ) and other variables (here four-lepton invariant mass).

Quantile bins made with large MC set— not the data.
Simpler Approach: Quantile Bins

- A mini-version of ranking. Make quantile bins in ME(ε) and other variables (here four-lepton invariant mass).

Relatively flat with some fluctuations.
Simpler Approach: Quantile Bins

- A mini-version of ranking. Make quantile bins in ME(ε) and other variables (here four-lepton invariant mass)

On average, BG is flat (that was our goal!)
Simpler Approach: Quantile Bins

- A mini-version of ranking. Make quantile bins in ME(\(\epsilon\)) and other variables (here four-lepton invariant mass).

Now we consider 150 events in total, but 60 are signal.
Simpler Approach: Quantile Bins

- A mini-version of ranking. Make quantile bins in ME(ε) and other variables (here four-lepton invariant mass).

Normalization from data: signal also causes deficits.
Simpler Approach: Quantile Bins

- A mini-version of ranking. Make quantile bins in $\text{ME}(\varepsilon)$ and other variables (here four-lepton invariant mass).

Signal + Background pseudoexperiments averaged.
Simpler Approach: Quantile Bins

- To see more concretely how quantile bins help, we consider
- 50 gluon fusion signal four-lepton events
- 150 background (q\bar{q}) four-lepton events
Simpler Approach: Quantile Bins

- “By eye” it seems like there is
- an excess in m4l below ~130 GeV
- a deficit in m4l above ~140 for high $|M|^2$ values
Simpler Approach: Quantile Bins

- After quantile binning (according to the background distribution)

- No real deficit in upper right hand corner

- Moderate excess in 2\textsuperscript{nd} m4l column, especially the lower ranked bins.
Of course we can also obtain flat distributions by weighing contributions to (potentially multivariable) distributions by the inverse of the background PDF.

To explain this, I am going to invoke Leonardo da Vinci and aliens…
Conclusions

• Important to develop model independent ways to search for signal

• This can be done in a straightforward, sensitive way using the background matrix element.

• Related techniques allow backgrounds to be flattened—giving an intuitive and general understanding of the significance of possible signals.
  
  • Ranking
  
  • Quantile Binning
  
  • Reweighting

• Looking forward to the discovery of unexpected signals at the LHC!