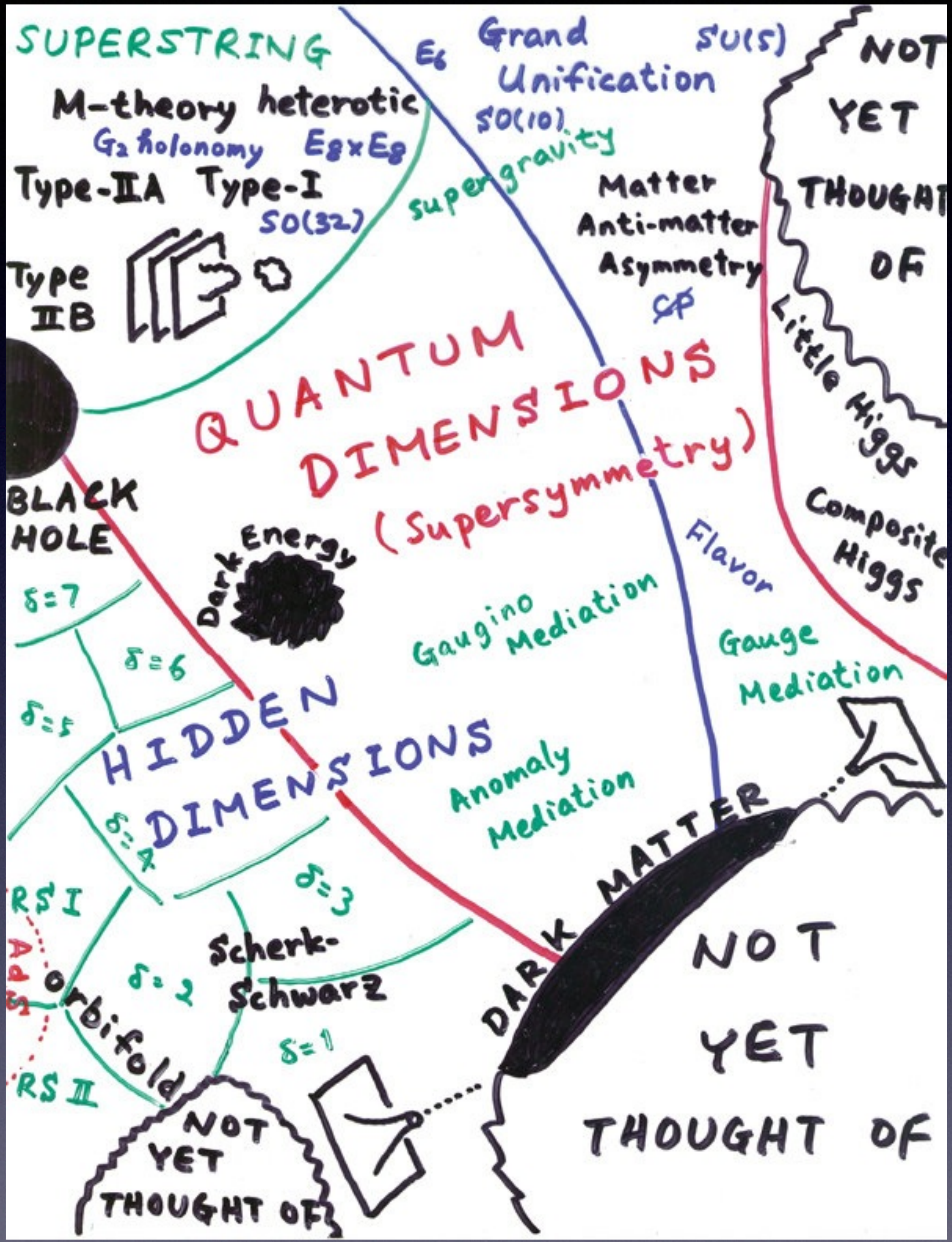
Model-Independent Searches with Background Matrix Elements

James “Jamie” Gainer, University of Florida

Work by Dipsikha Debnath, JG, Konstantin Matchev

Pheno 2014: May 6, 2014

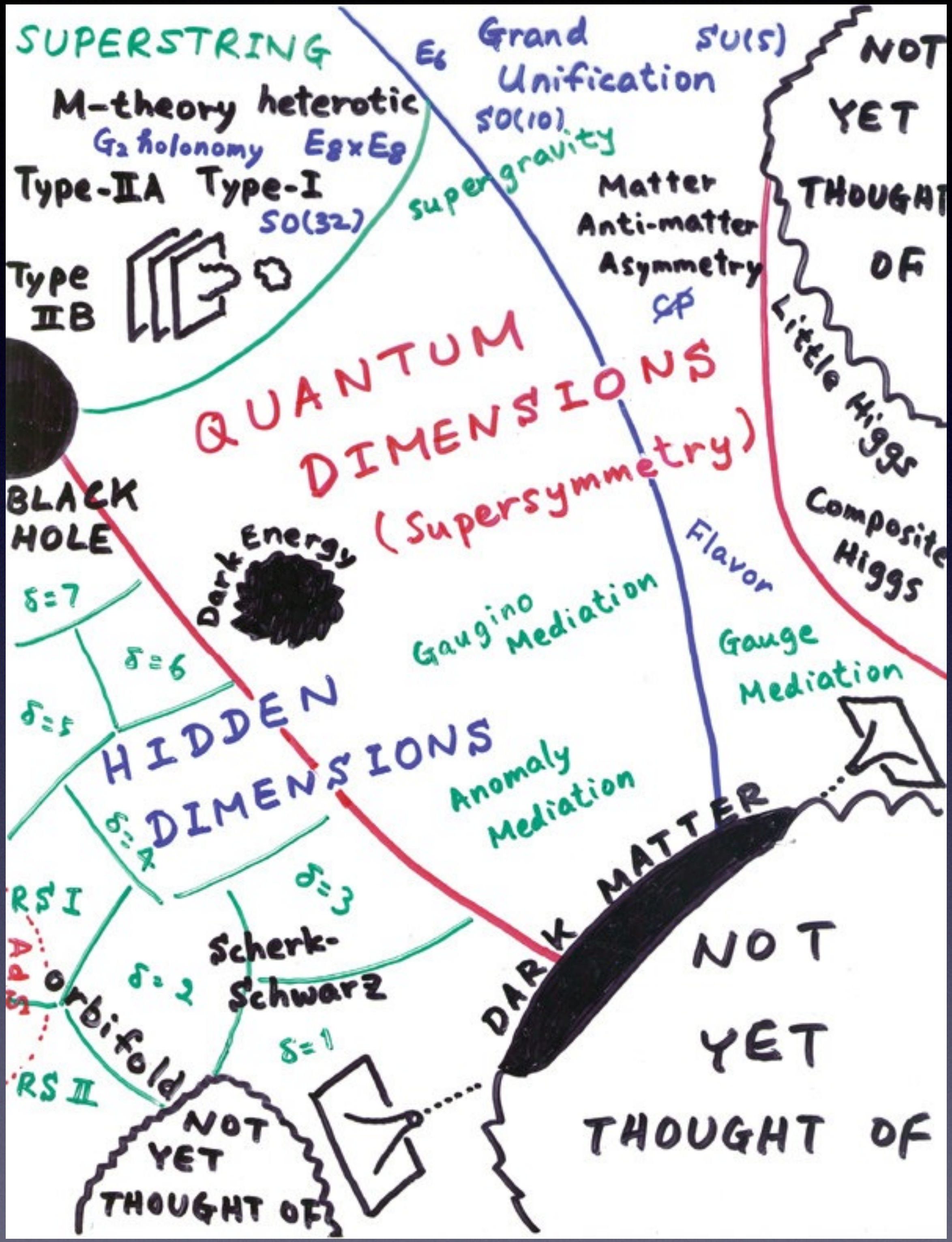
- Many ideas for physics beyond the Standard Model



(Murayama)

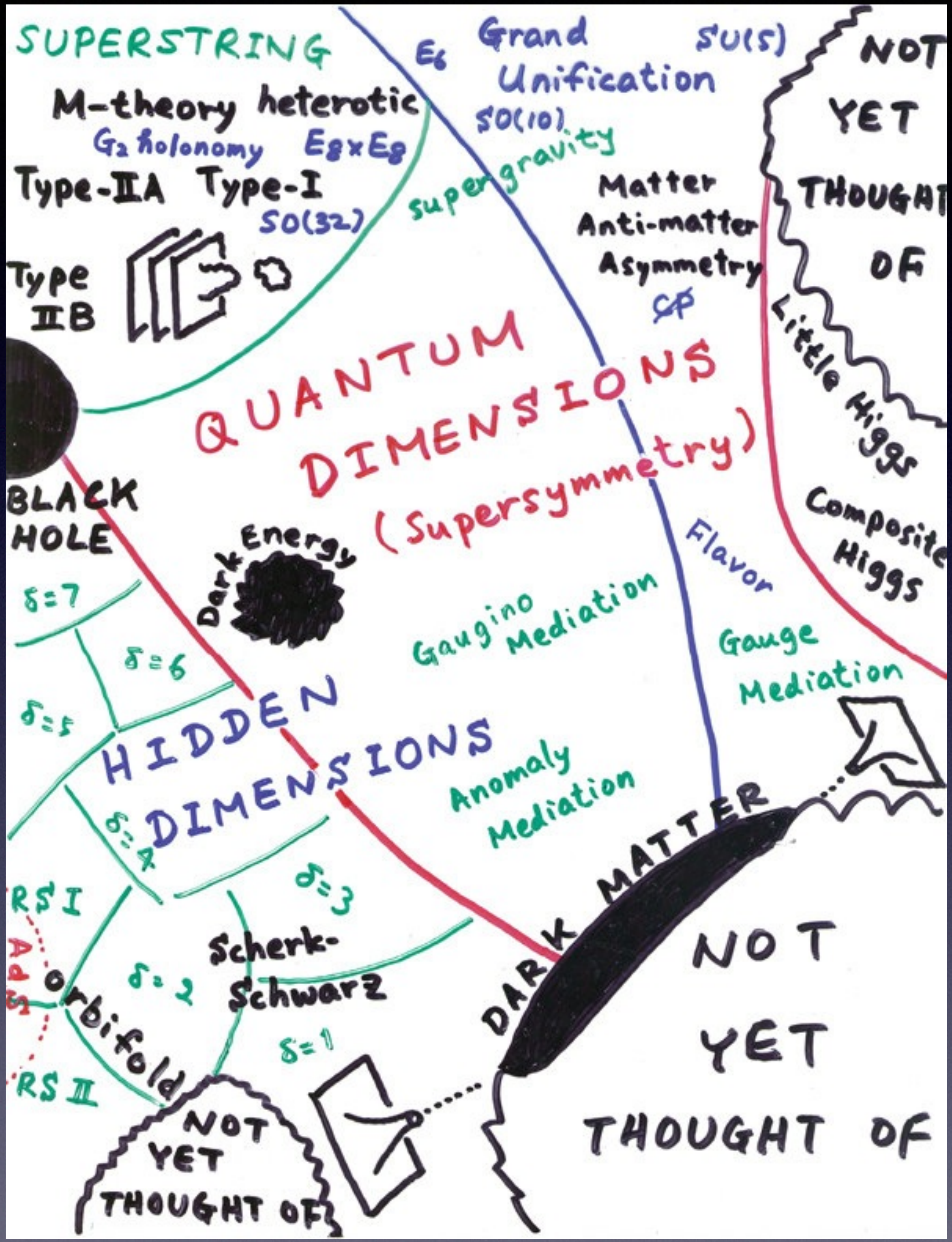
- Many ideas for physics beyond the Standard Model

- 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 signature (e.g. mSUGRA).

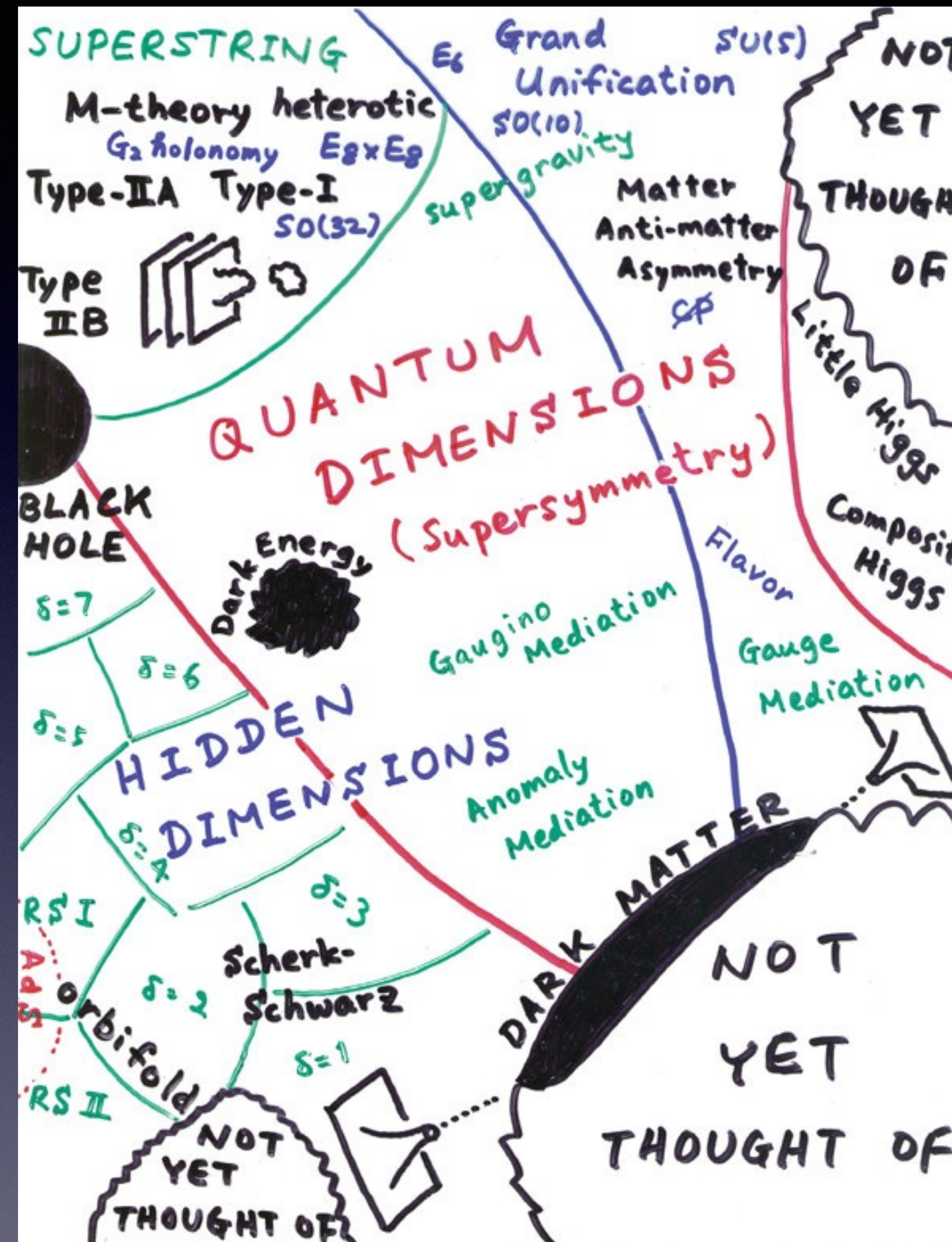


(Murayama)

- 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 signature (e.g. mSUGRA).



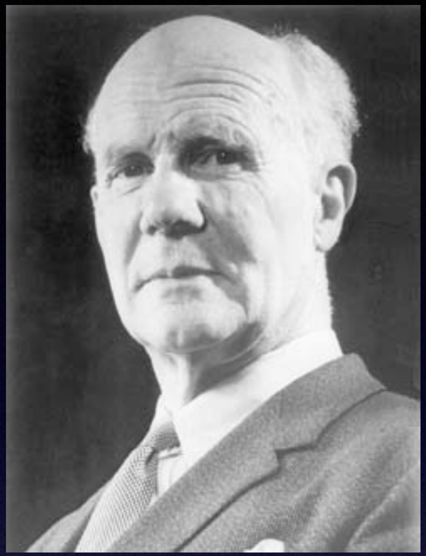
- Can we discover "Not Yet Thought Of" theories?

- 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, which you probably just heard about

Neyman-Pearson Lemma



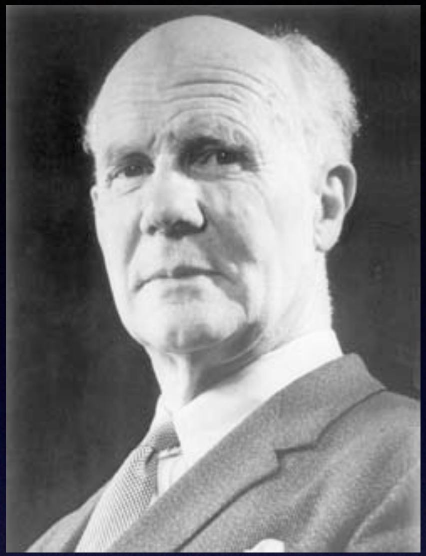
Actually Neyman and Pearson were roughly the same age. Google works in mysterious ways...

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

Neyman-Pearson Lemma



Actually Neyman and Pearson were roughly the same age. Google works in mysterious ways...

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

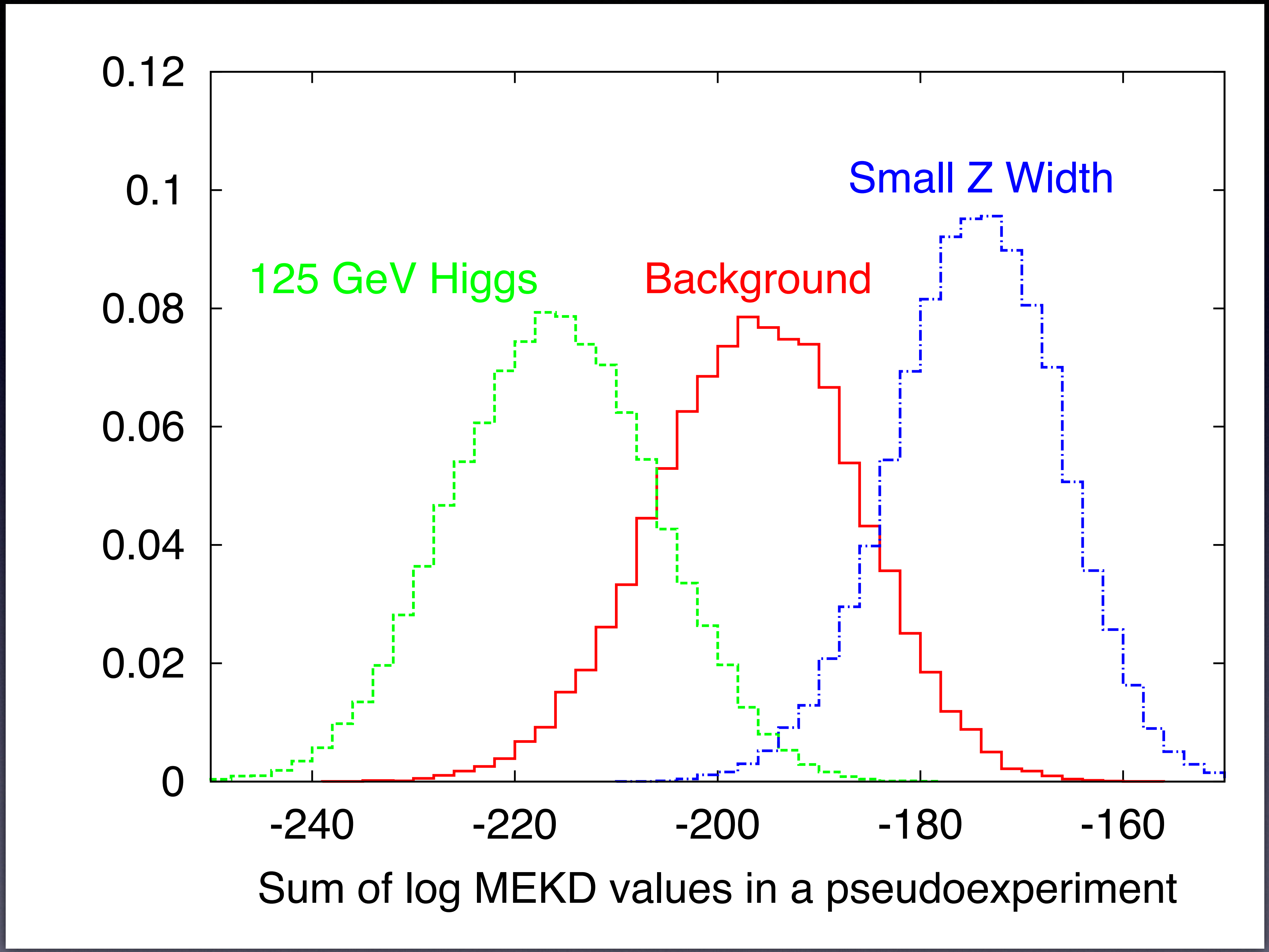
- In particle physics, the likelihood/ we use the expression for probability on the right:

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

- Normalized to the total cross section
- With integrals over transfer functions, invisible momenta, etc.
- Use of this likelihood = “Matrix Element Method”

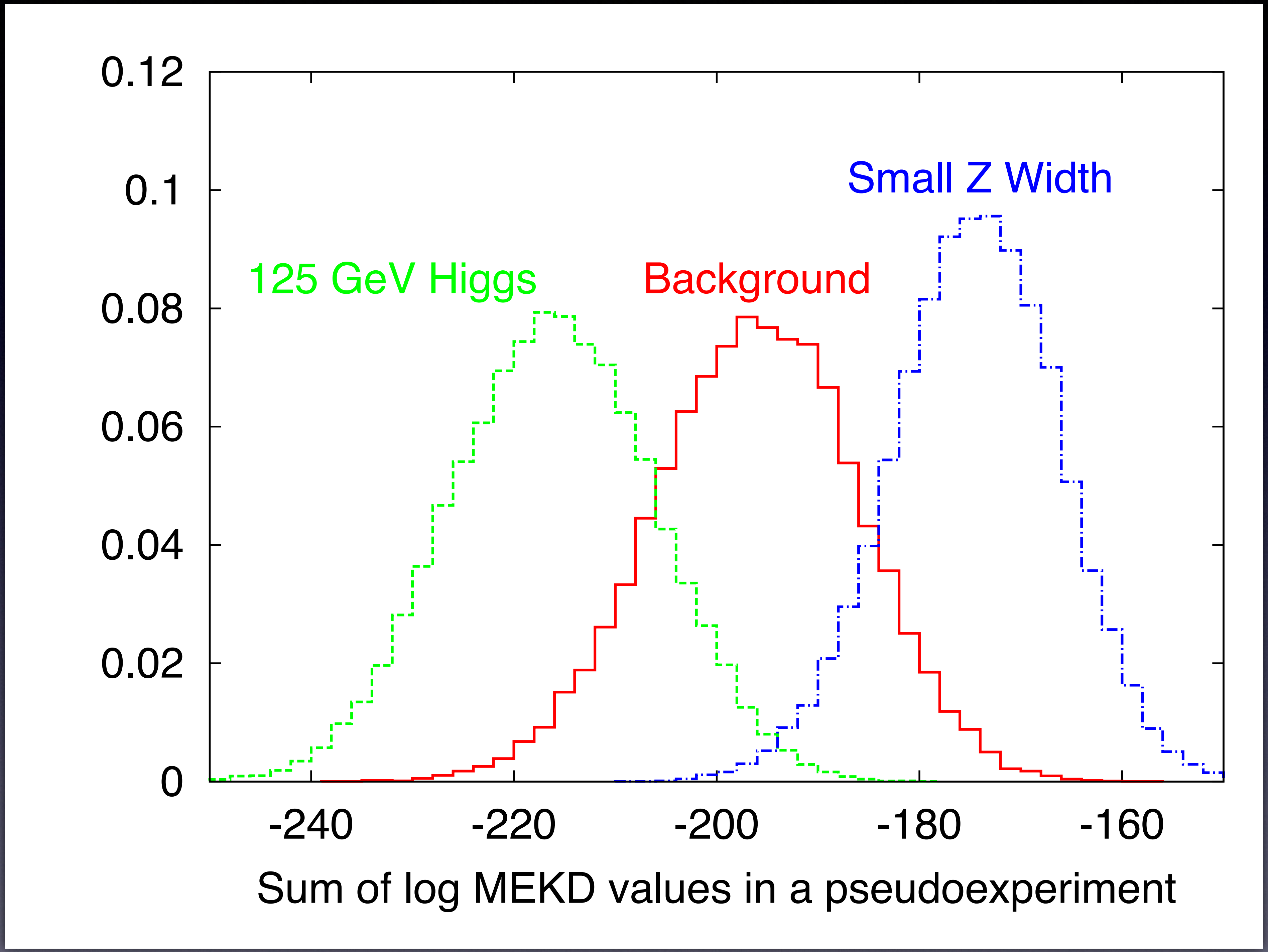
- In the Neyman-Pearson Lemma we needed a likelihood **ratio.**
- Need to know both signal and background likelihood to compute this ratio
- For signal independence, use **background likelihood** as a test statistic.
- Matrix element variables still “know a lot” about the background so should be optimal at rejecting background

Background Likelihood



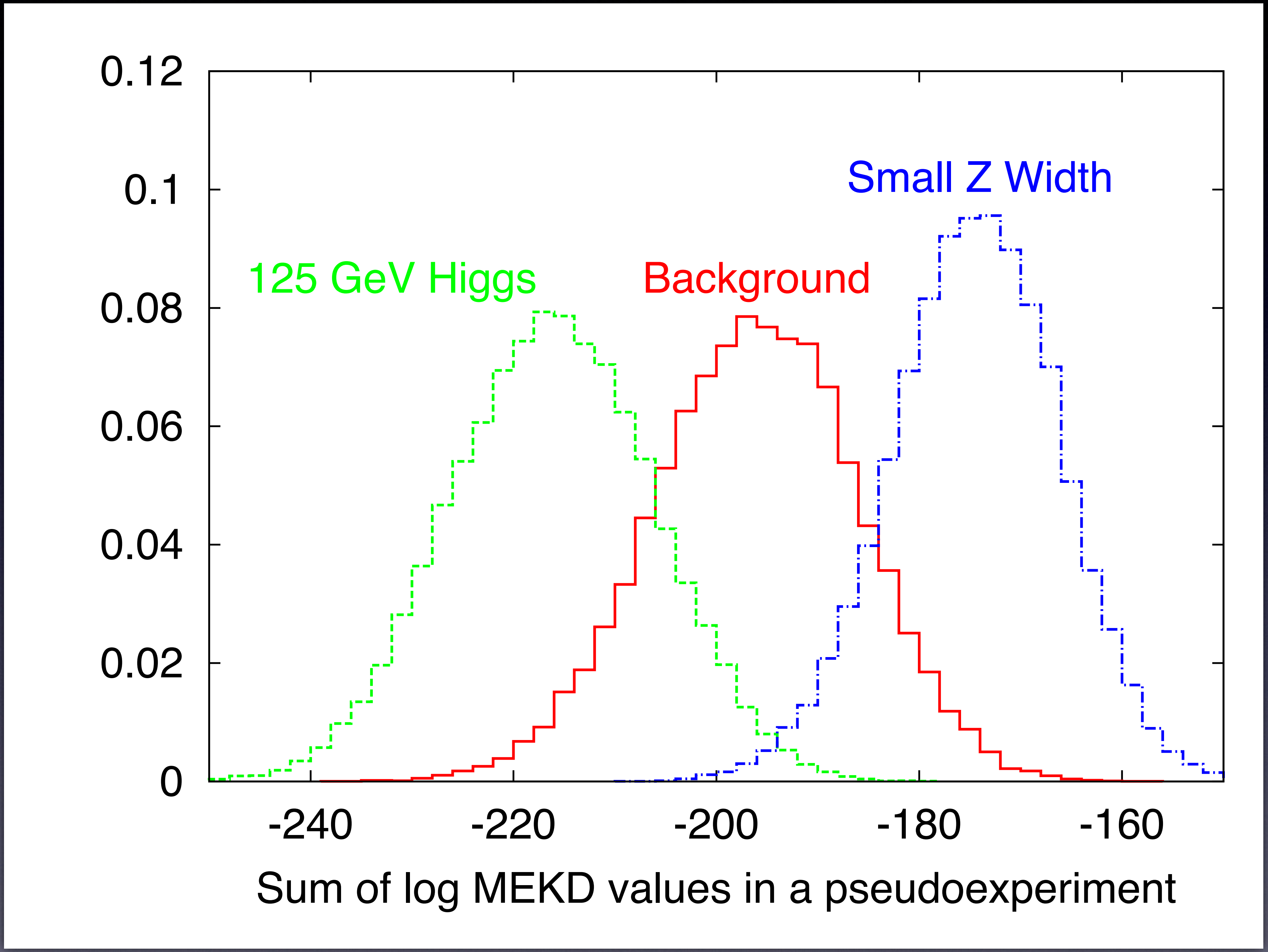
As an example, we consider 20 event pseudoexperiments

Background Likelihood



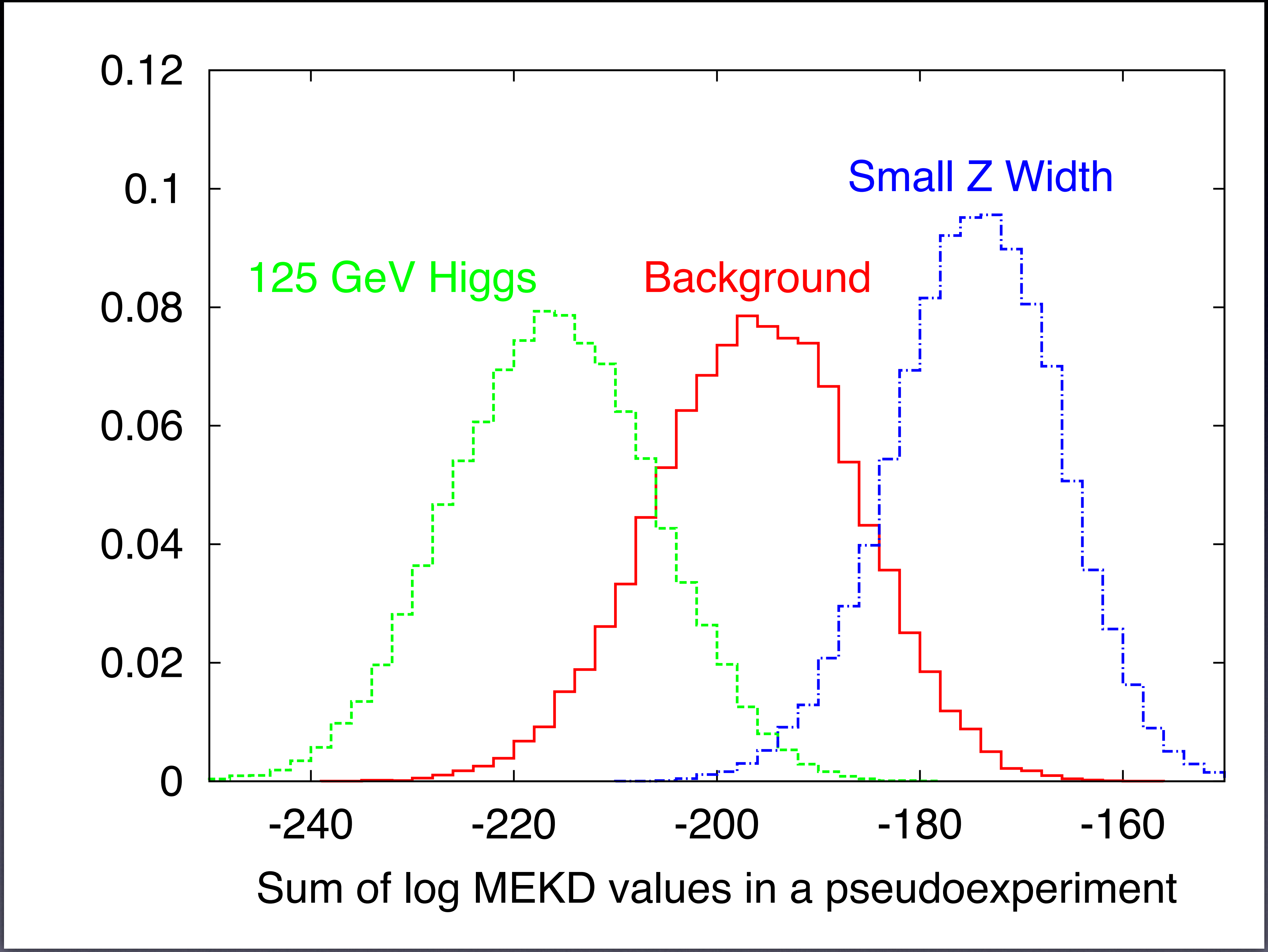
Processes: gluon fusion Higgs → 4ℓ

Background Likelihood



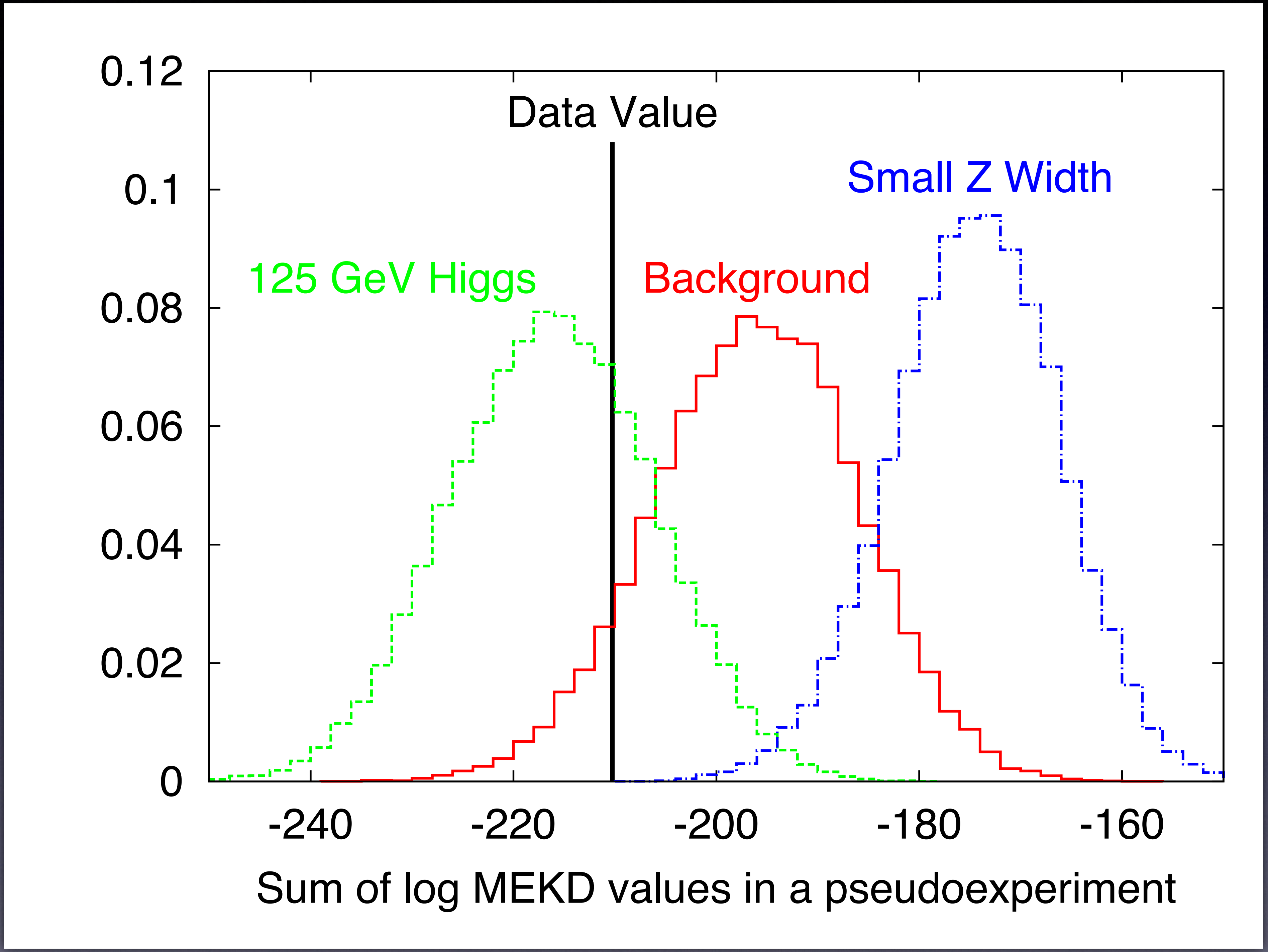
Processes: $q\bar{q} \rightarrow 4\ell$ background

Background Likelihood



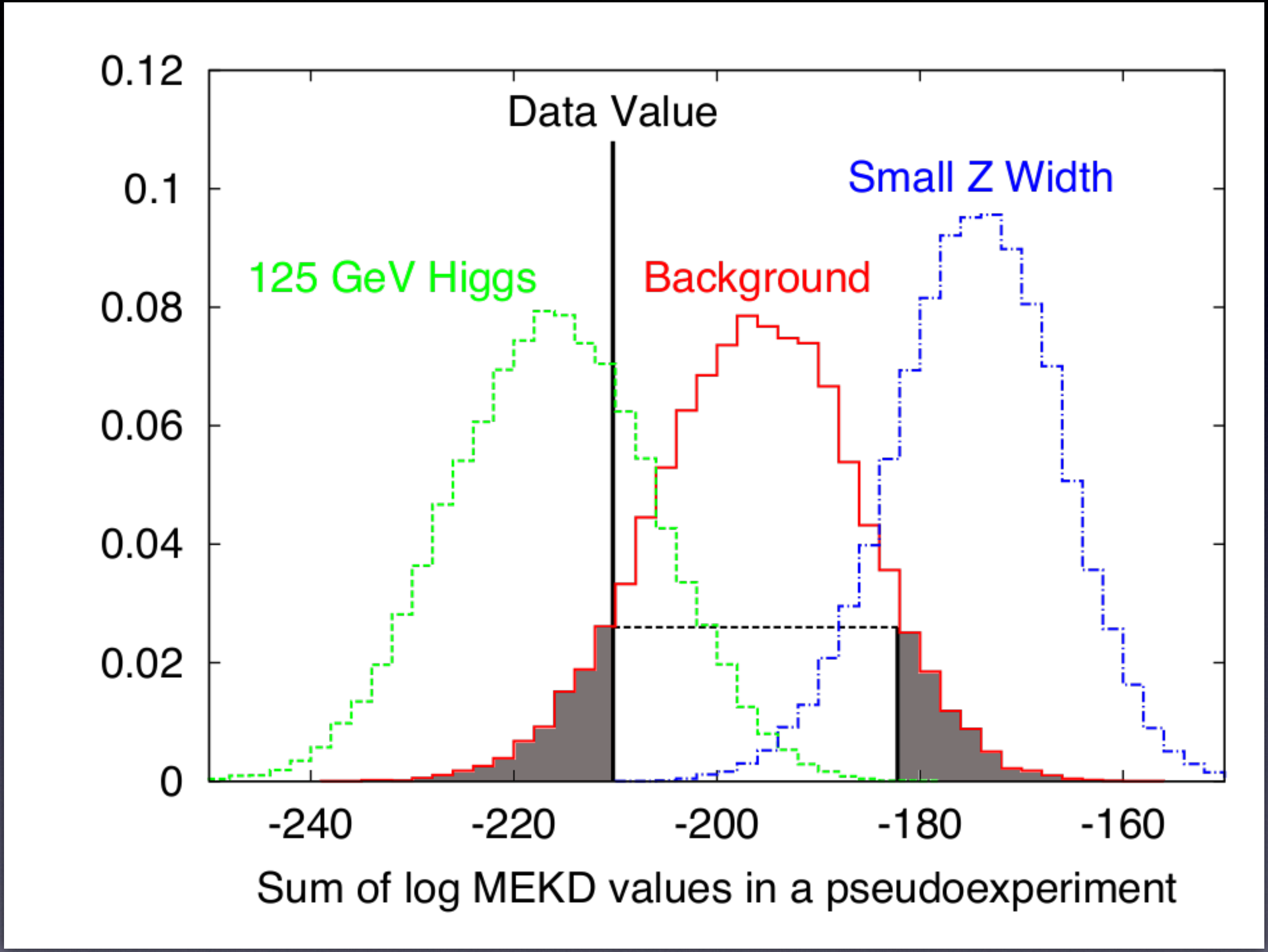
Processes: $q\bar{q} \to 4\ell$ background, $\Gamma_Z \to \Gamma_Z/5$

Background Likelihood



We take our 20 “data” events and evaluate the sum of MEKD values.

Background Likelihood



p-value is shaded region

- p-value from likelihood distribution calculated from Monte Carlos

- so reducible backgrounds, detector effects, NLO (if your MC has it) etc. are included automatically

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

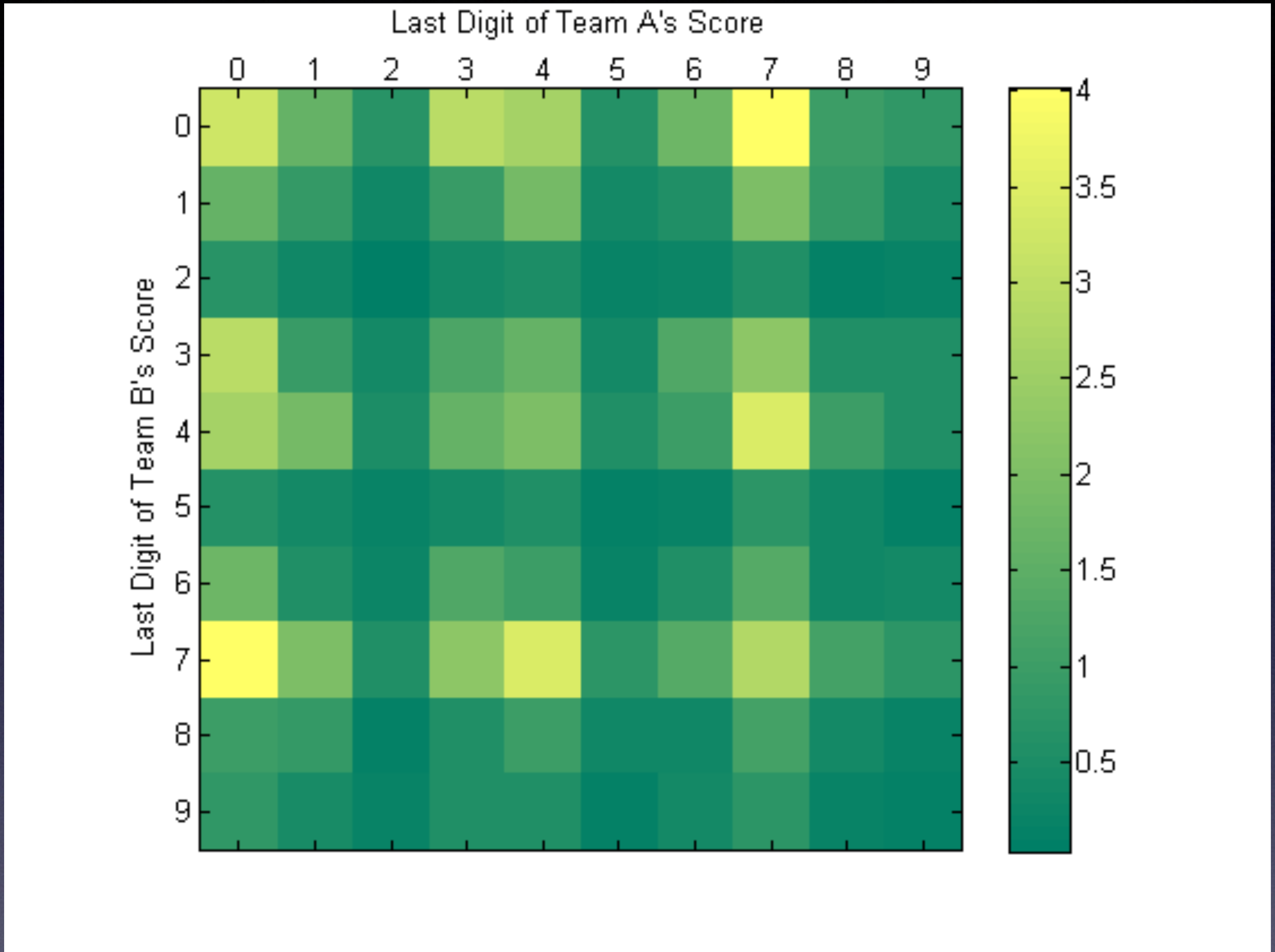
- Now I’m going to present some related ideas in which we use similar tools to obtain flat background distributions

Why?

When your background is flat…

It’s easy to discover an unexpected signal.

Example: NFL Scores



http://blogs.mathworks.com/community/2013/01/07/football-squares-with-matlab/

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

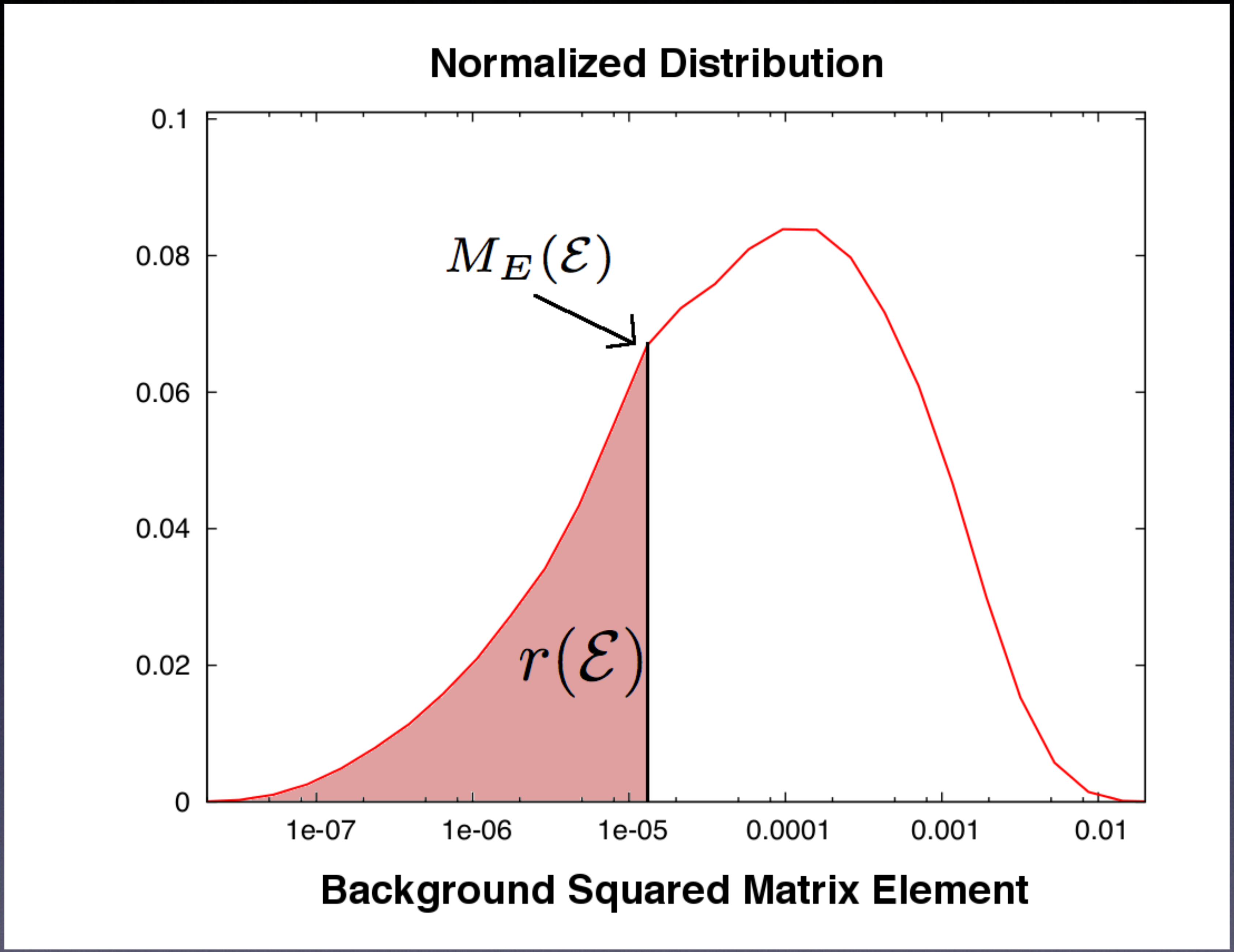
December 23, 1972
13 - 7 Steelers!!!

- We learned about the structure of football from deviations from flatness in distributions.

- Can we do the same in particle physics?

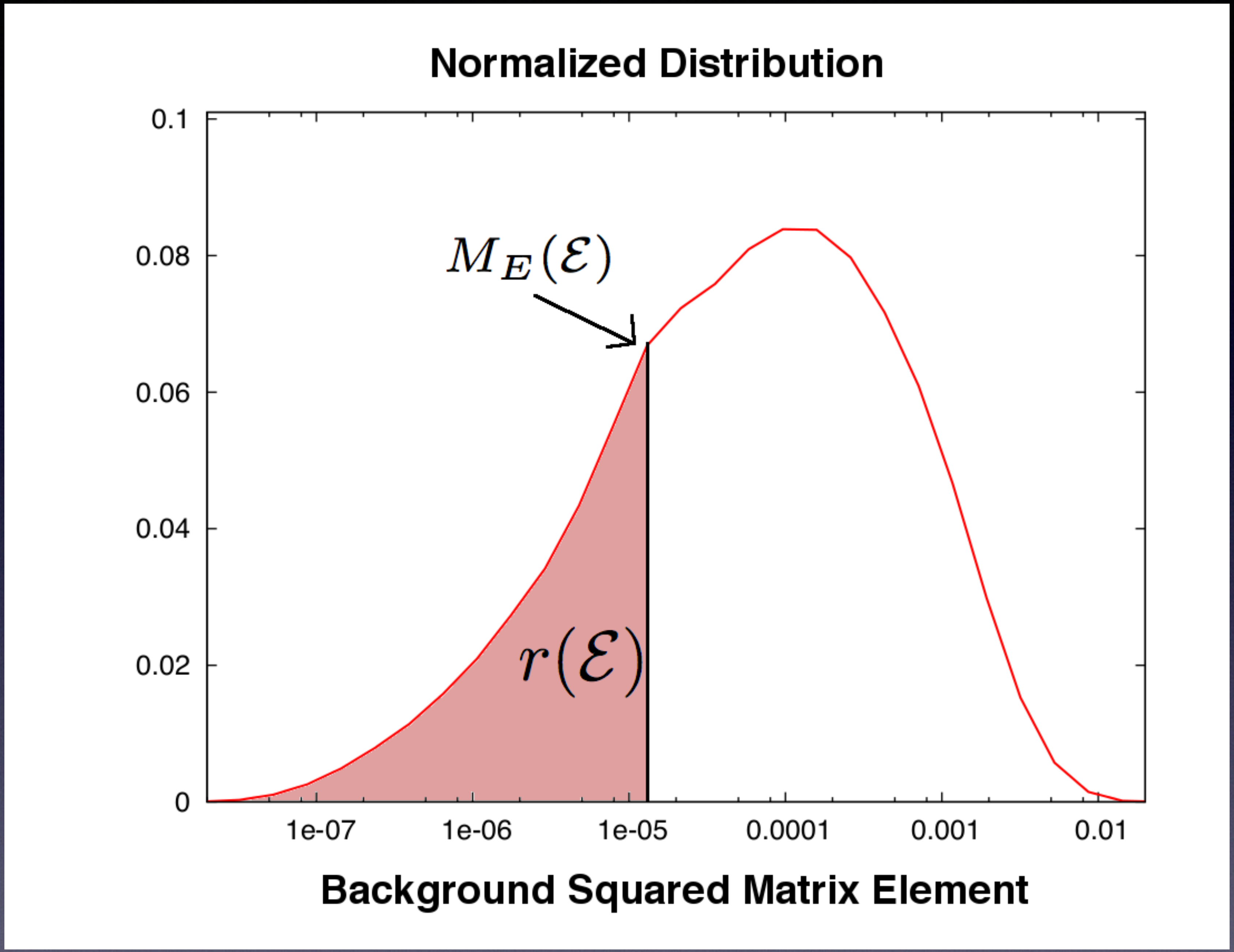
- How do we make background distributions flat?

Background Ranking



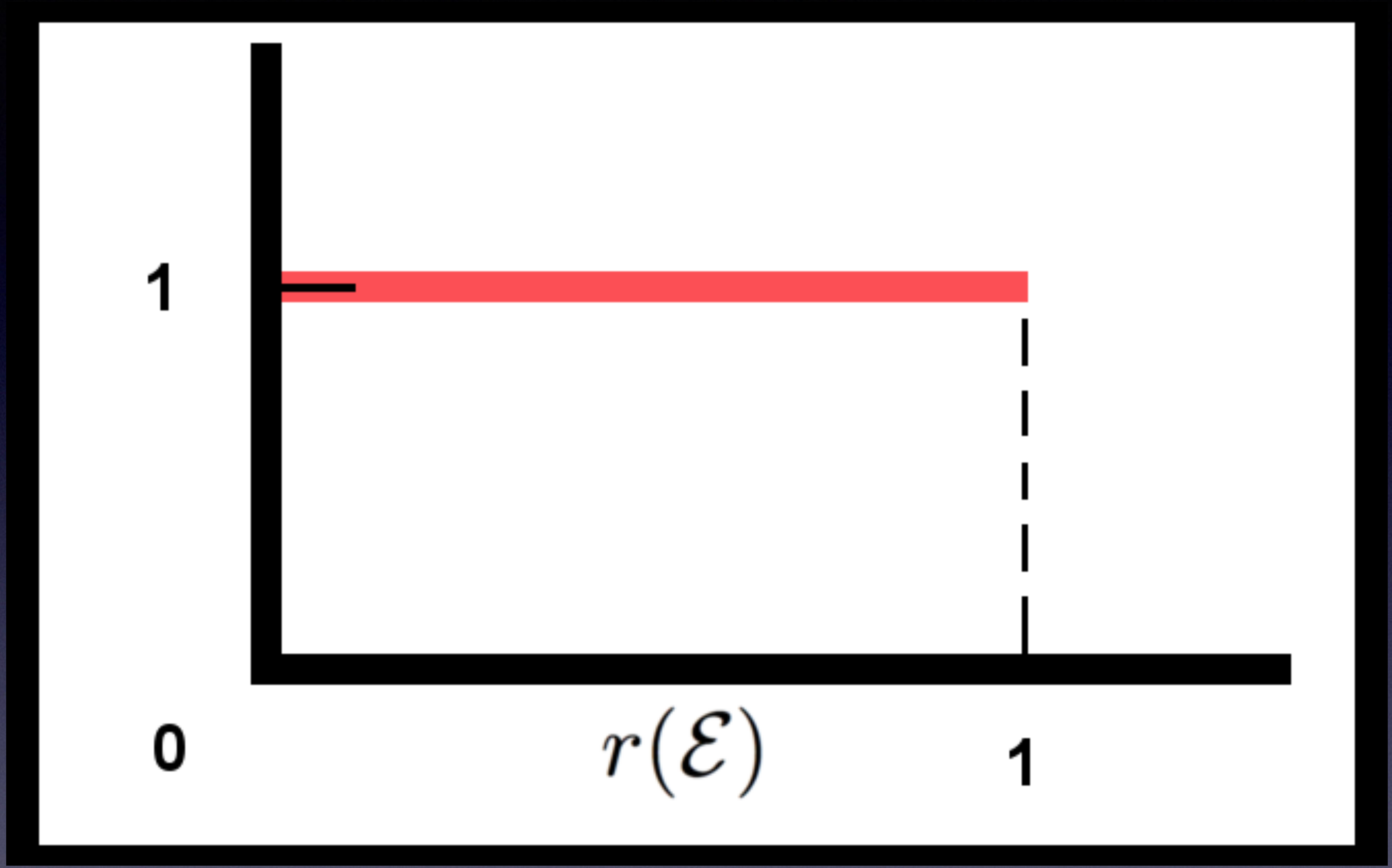
- Want to flatten the background distribution of ME-based variable

Background Ranking



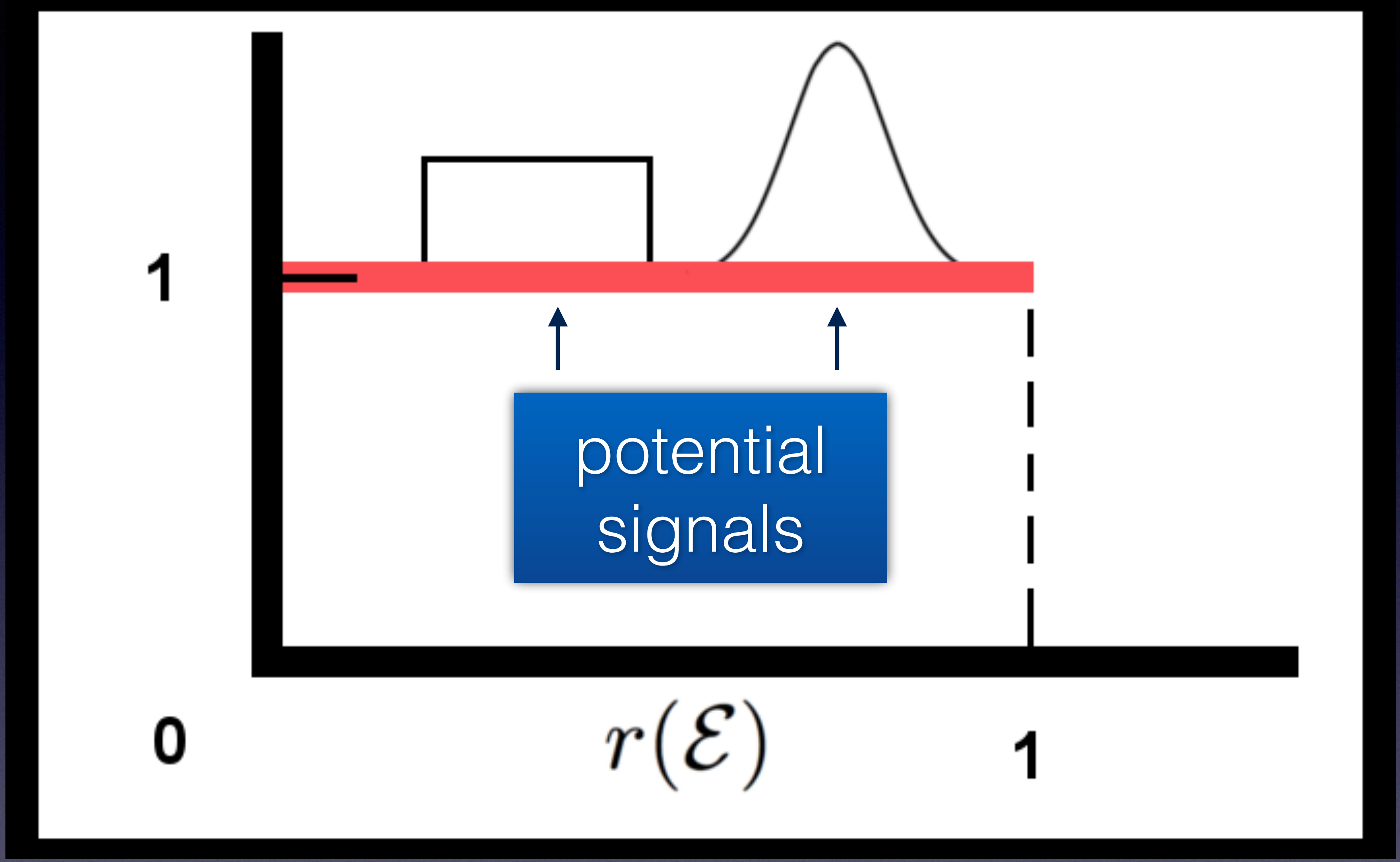
$$r(\mathcal{E}) = \int_0^{M_E(\mathcal{E})} dM_E \frac{dN}{dM_e}$$

Background Ranking



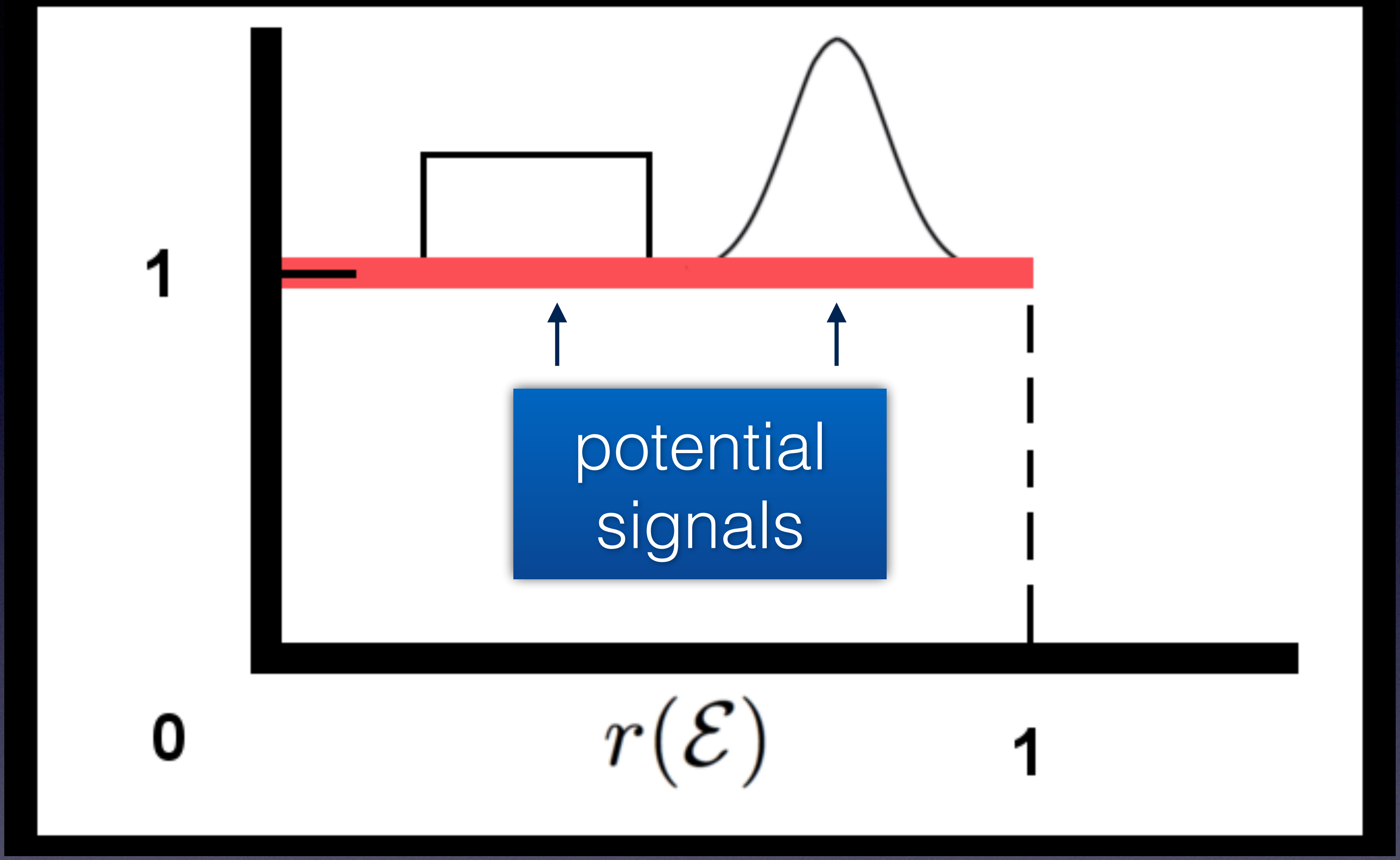
$$r(\mathcal{E}) = \int_0^{M_E(\mathcal{E})} dM_E \frac{dN}{dM_e}$$

Background Ranking



$$r(\mathcal{E}) = \int_0^{M_E(\mathcal{E})} dM_E \frac{dN}{dM_e}$$

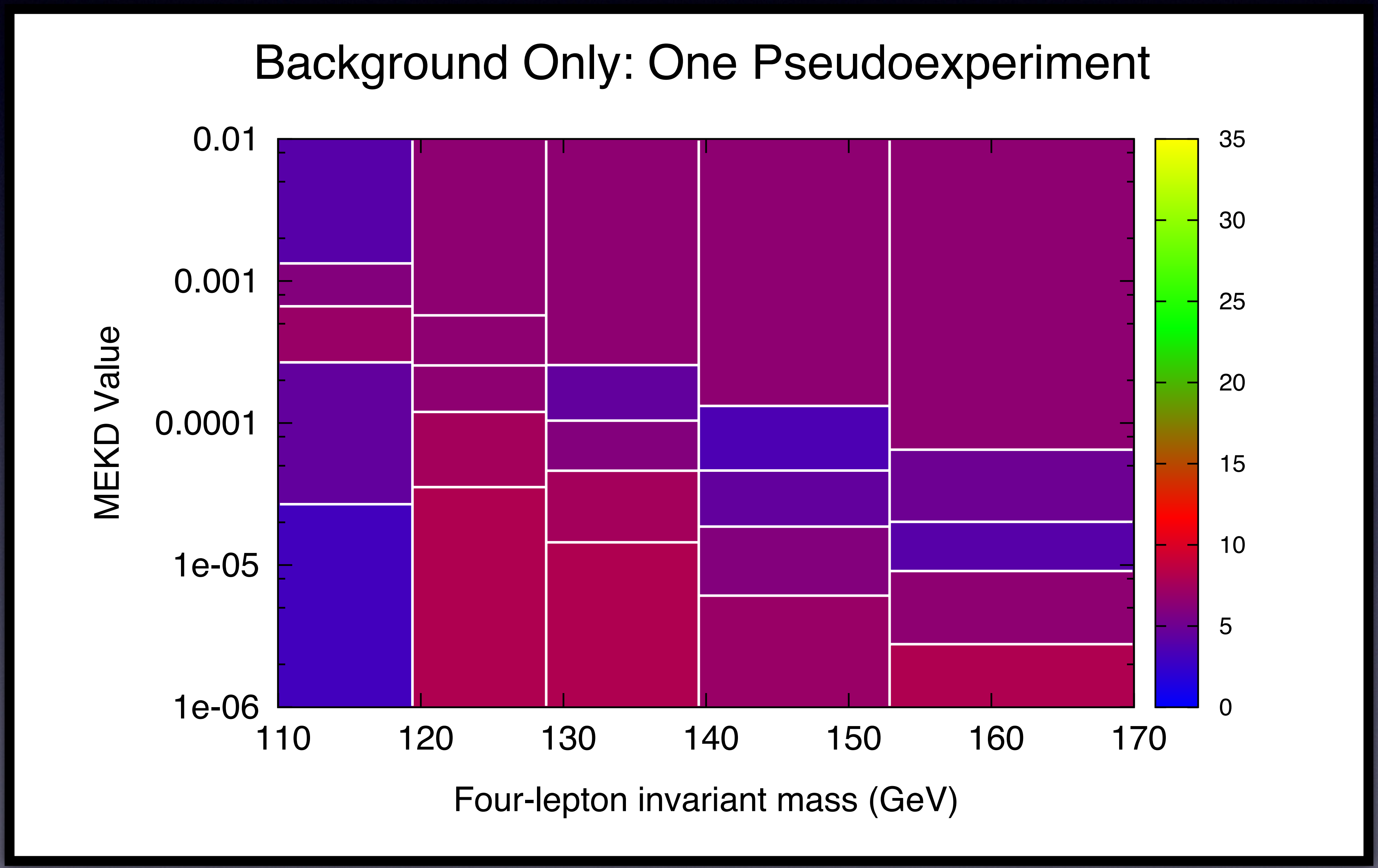
Background Ranking



If we take our normalization from data, deficits will also indicate signals.

Simpler Approach: Quantile Bins

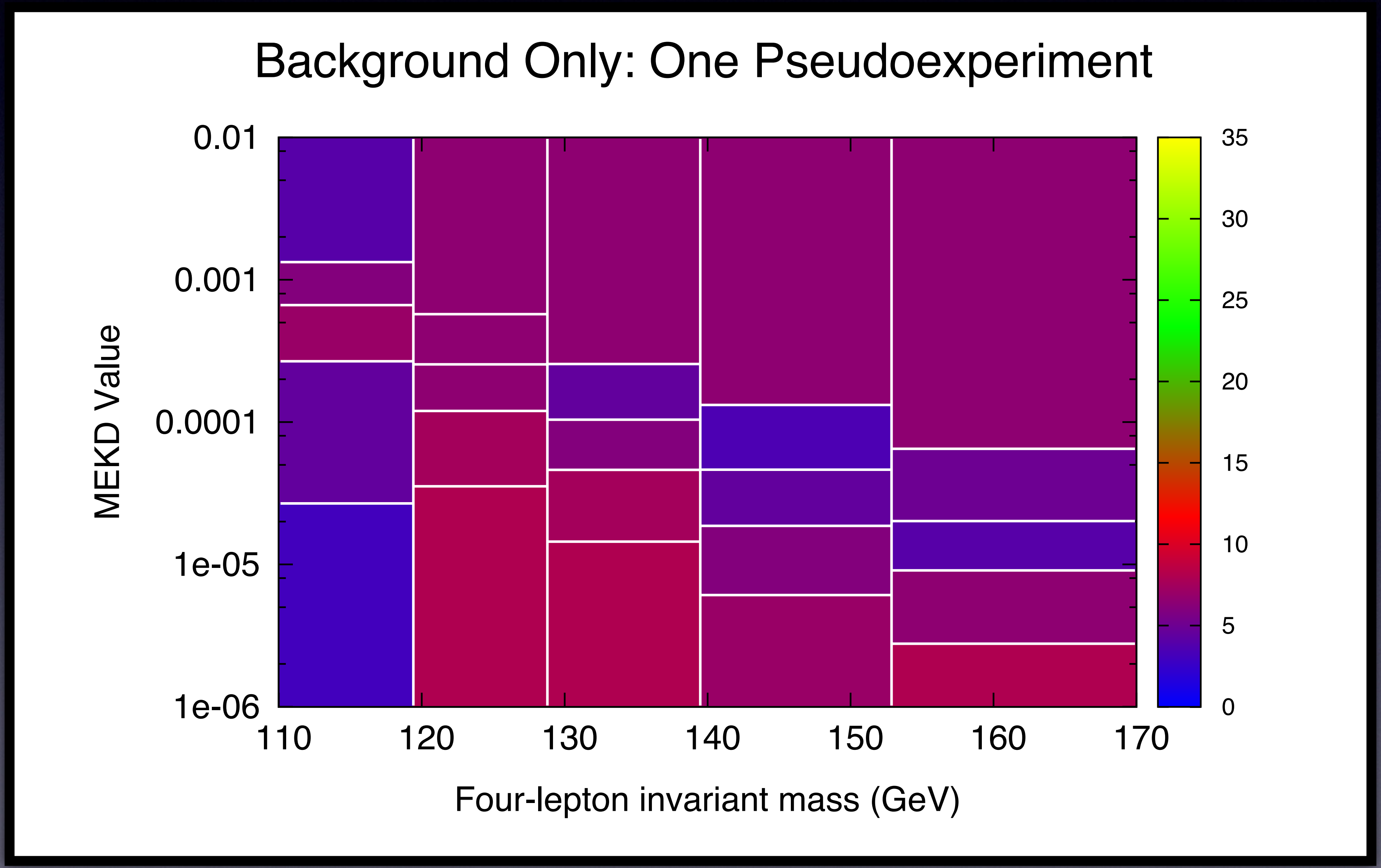
- A mini-version of ranking. Make quantile bins in ME(ℰ) and other variables (here four-lepton invariant mass)



Like the NFL scores example above.

Simpler Approach: Quantile Bins

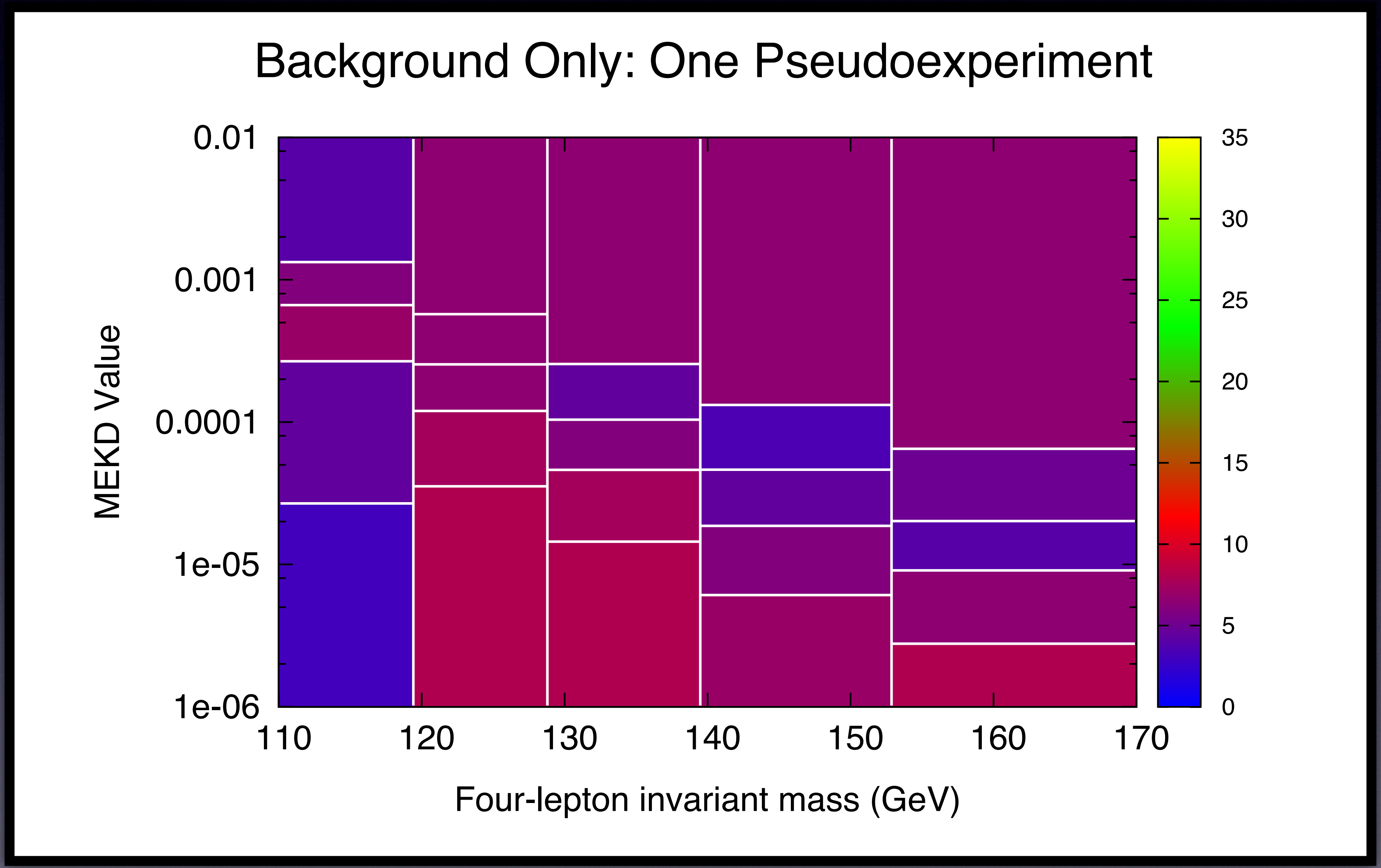
- A mini-version of ranking. Make quantile bins in ME(ℇ) and other variables (here four-lepton invariant mass)



Here we consider 150 BG events.

Simpler Approach: Quantile Bins

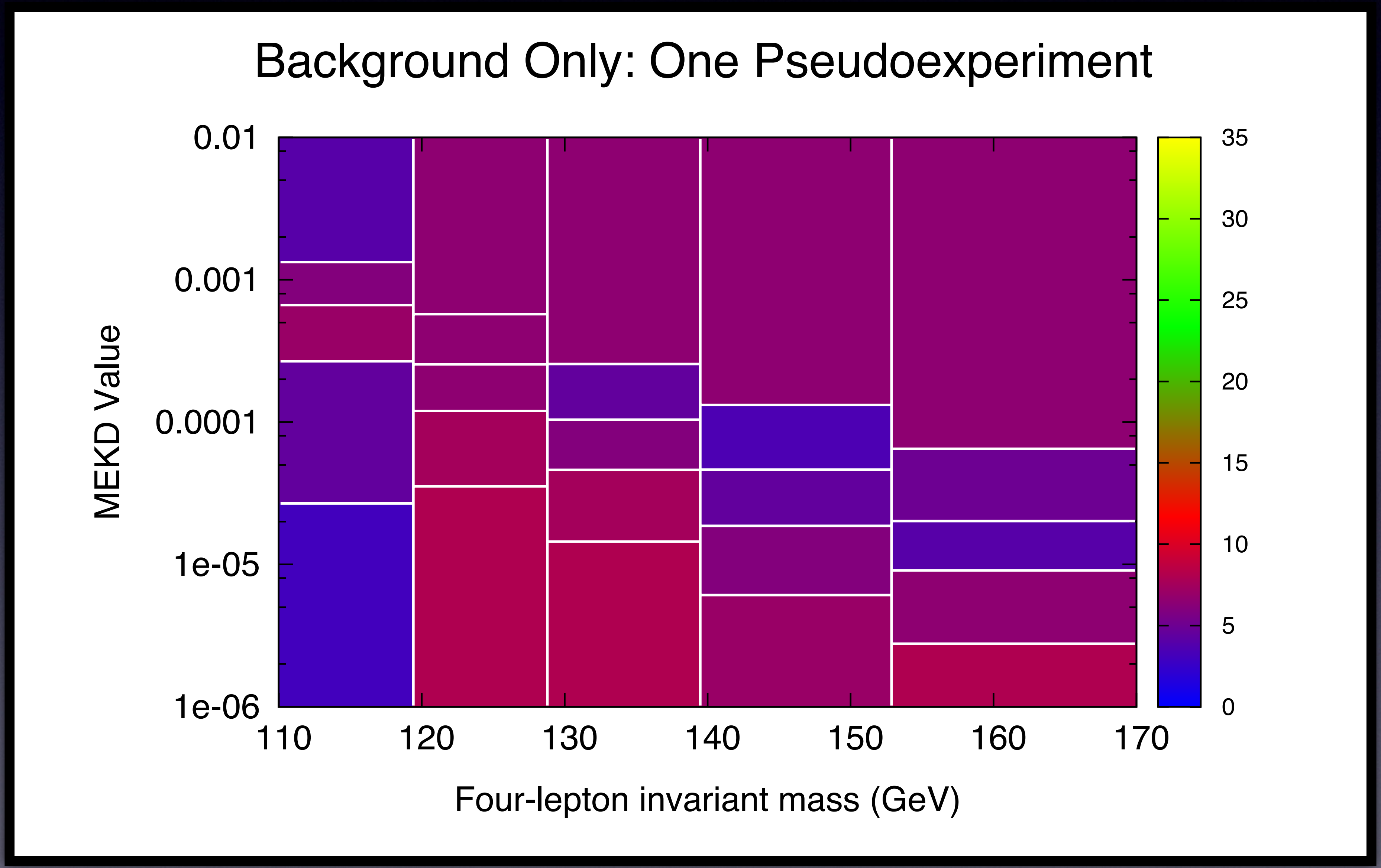
- A mini-version of ranking. Make quantile bins in ME(ℇ) and other variables (here four-lepton invariant mass)



Quantile bins made with large MC set— not the data.

Simpler Approach: Quantile Bins

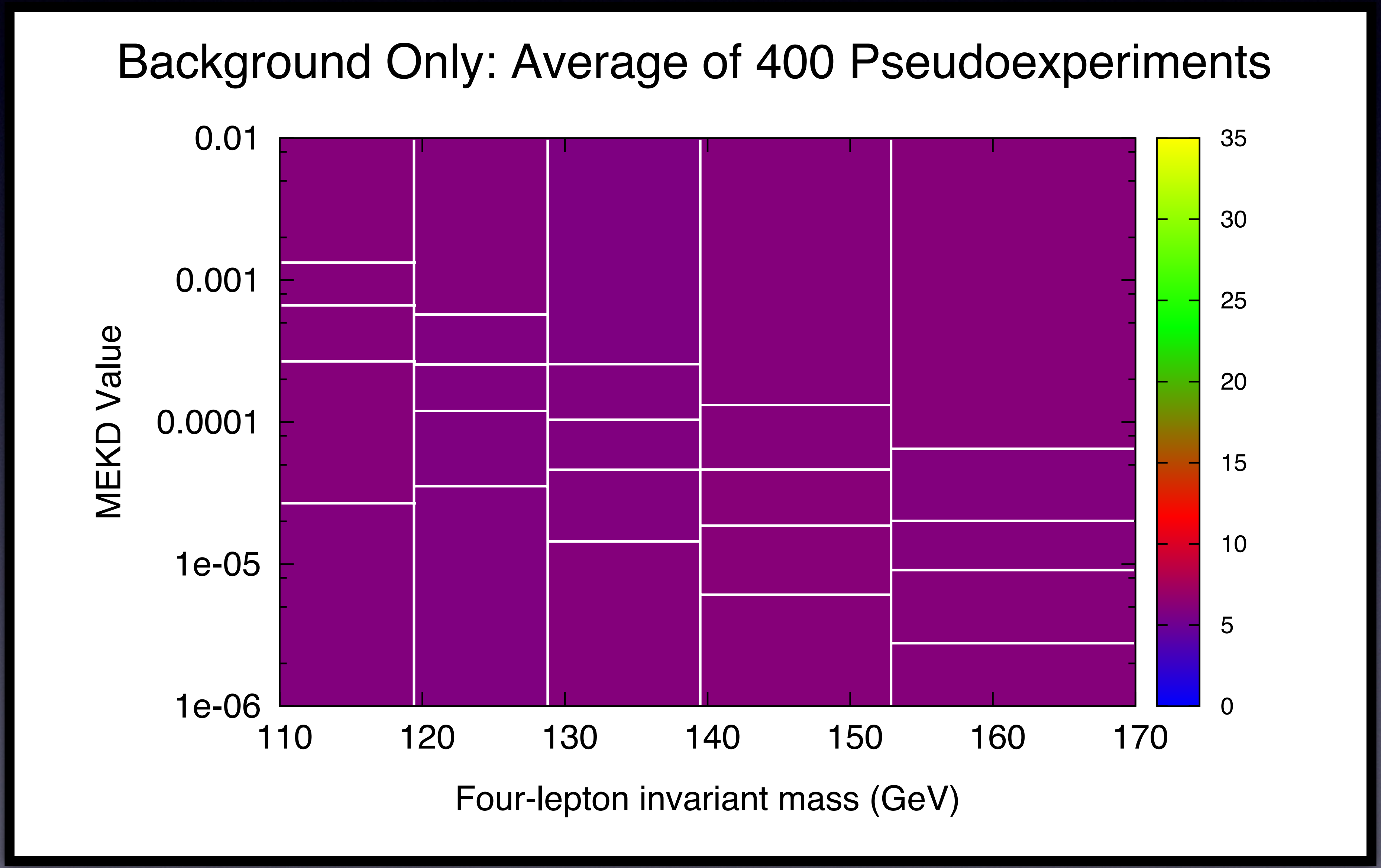
- A mini-version of ranking. Make quantile bins in ME(Ɛ) and other variables (here four-lepton invariant mass)



Relatively flat with some fluctuations.

Simpler Approach: Quantile Bins

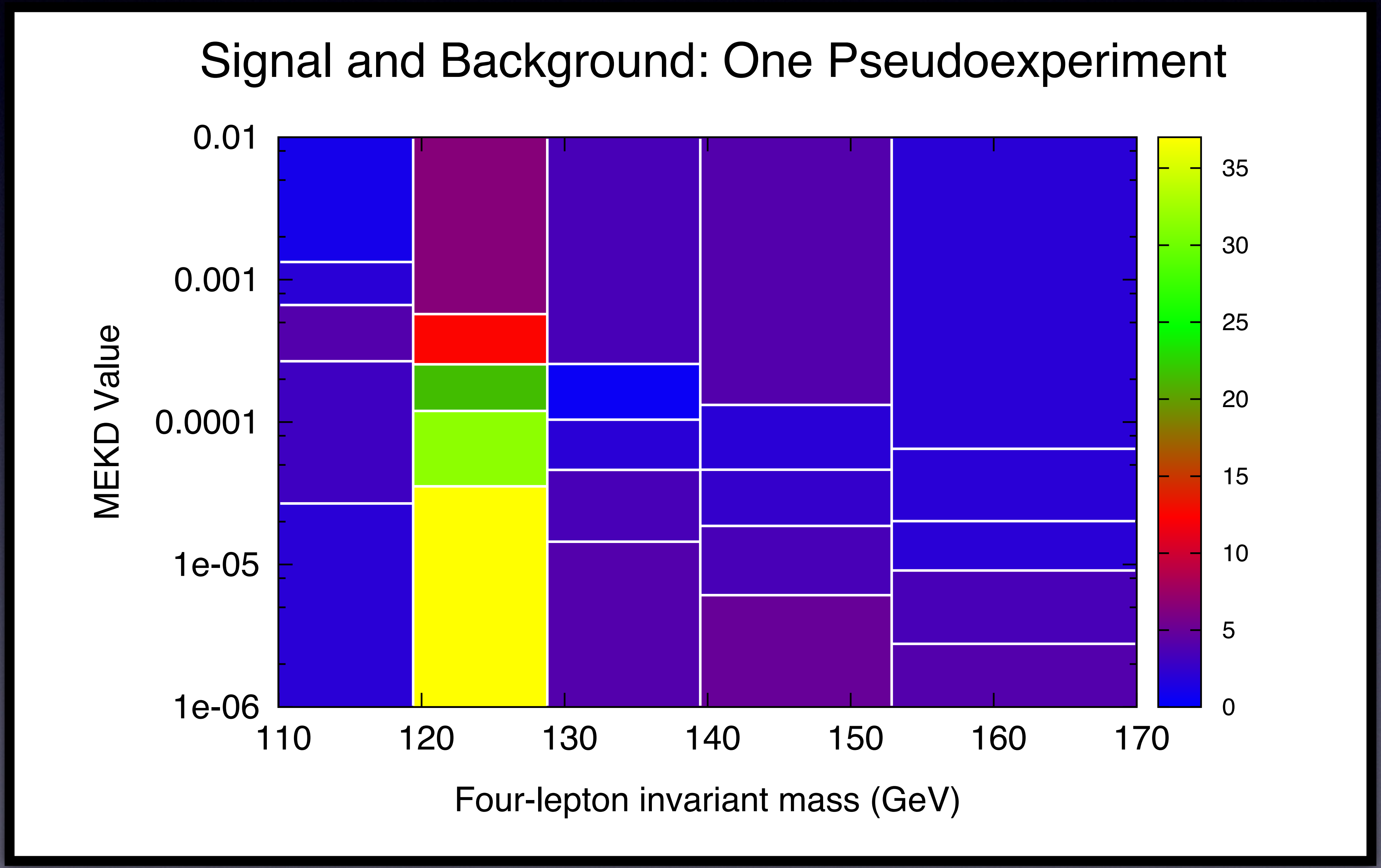
- A mini-version of ranking. Make quantile bins in ME(Ɛ) and other variables (here four-lepton invariant mass)



On average, BG is flat (that was our goal!)

Simpler Approach: Quantile Bins

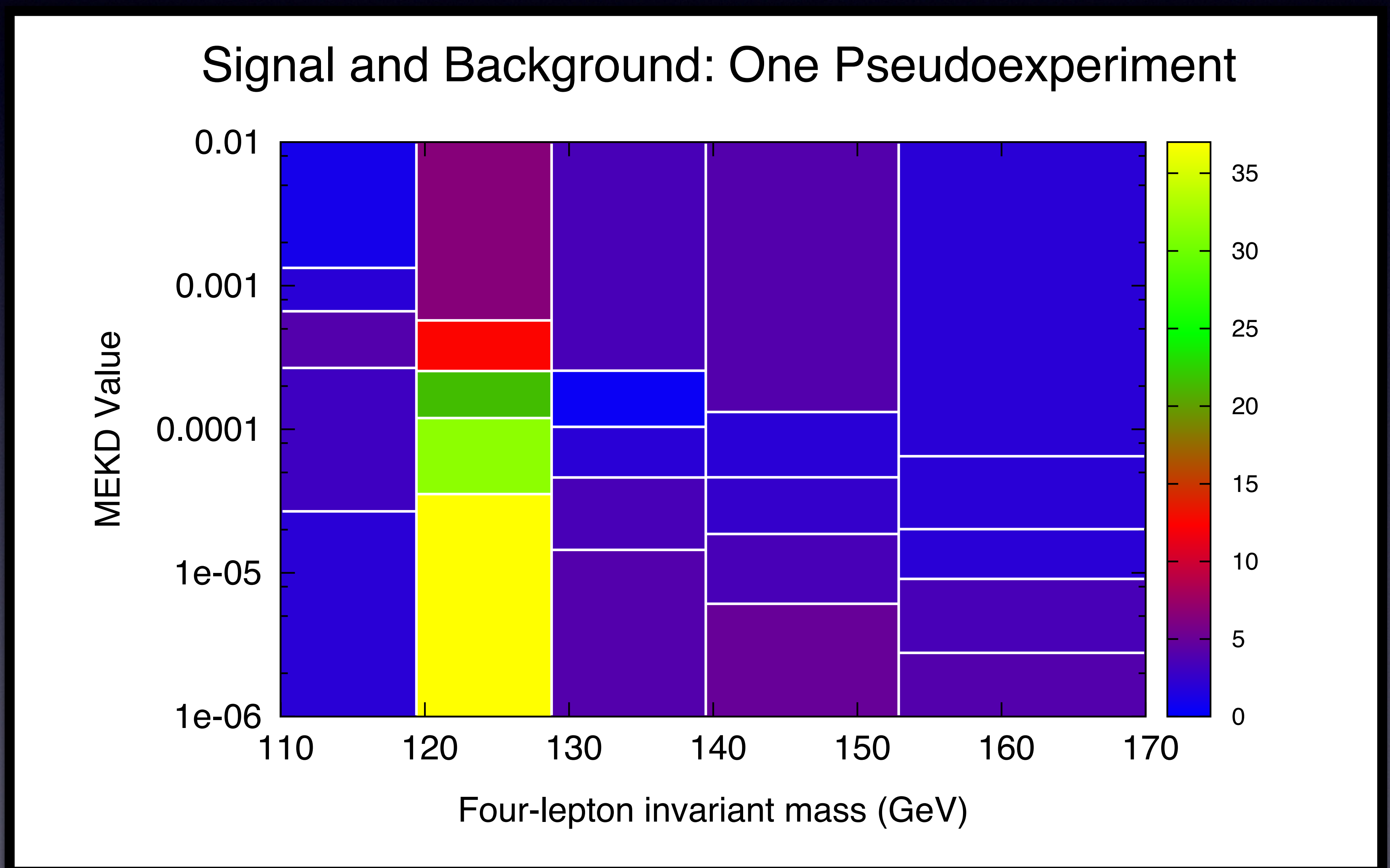
- A mini-version of ranking. Make quantile bins in ME(ℇ) and other variables (here four-lepton invariant mass)



Now we consider 150 events in total, but 60 are signal.

Simpler Approach: Quantile Bins

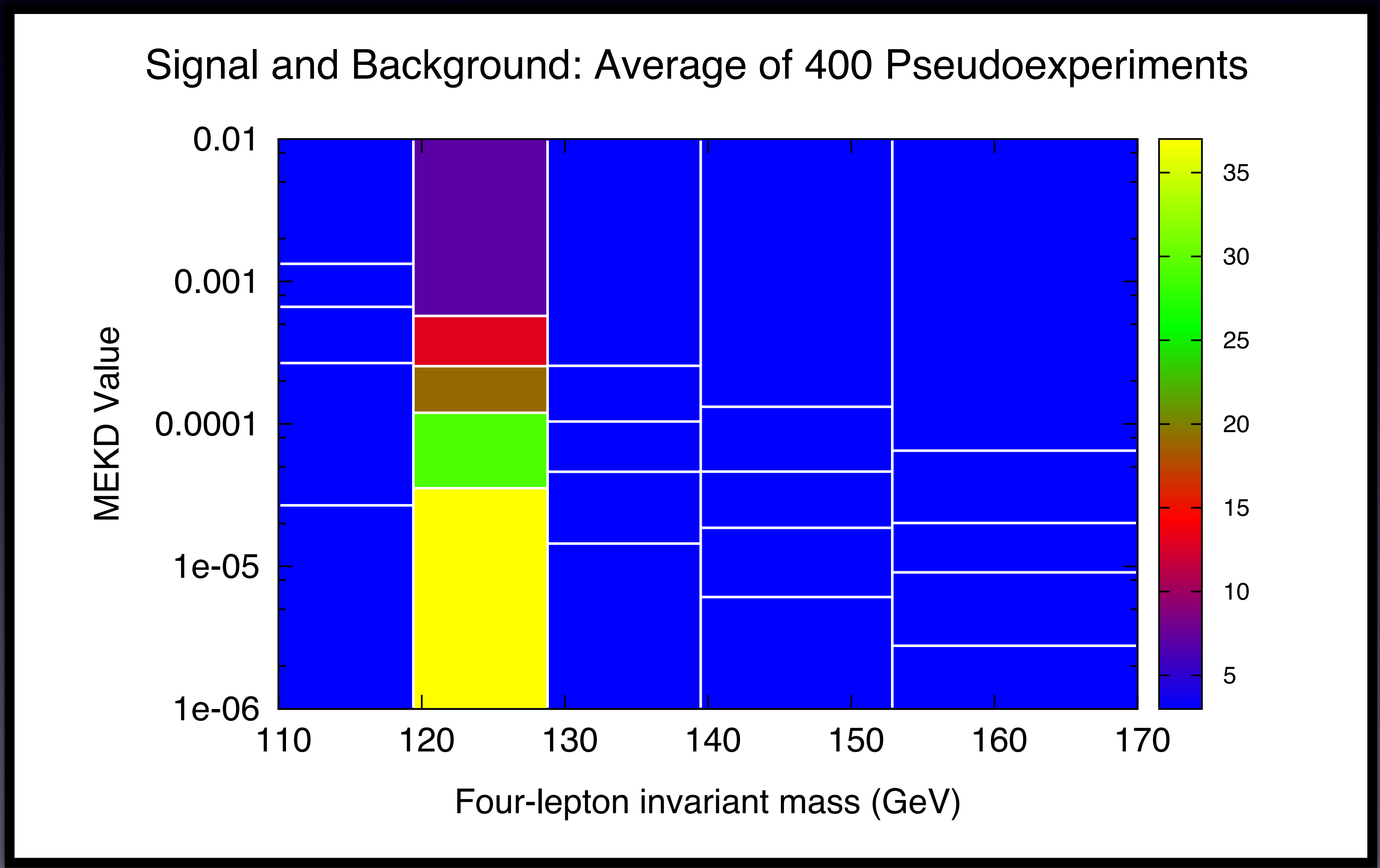
- A mini-version of ranking. Make quantile bins in ME(ℇ) and other variables (here four-lepton invariant mass)



Normalization from data: signal also causes deficits.

Simpler Approach: Quantile Bins

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

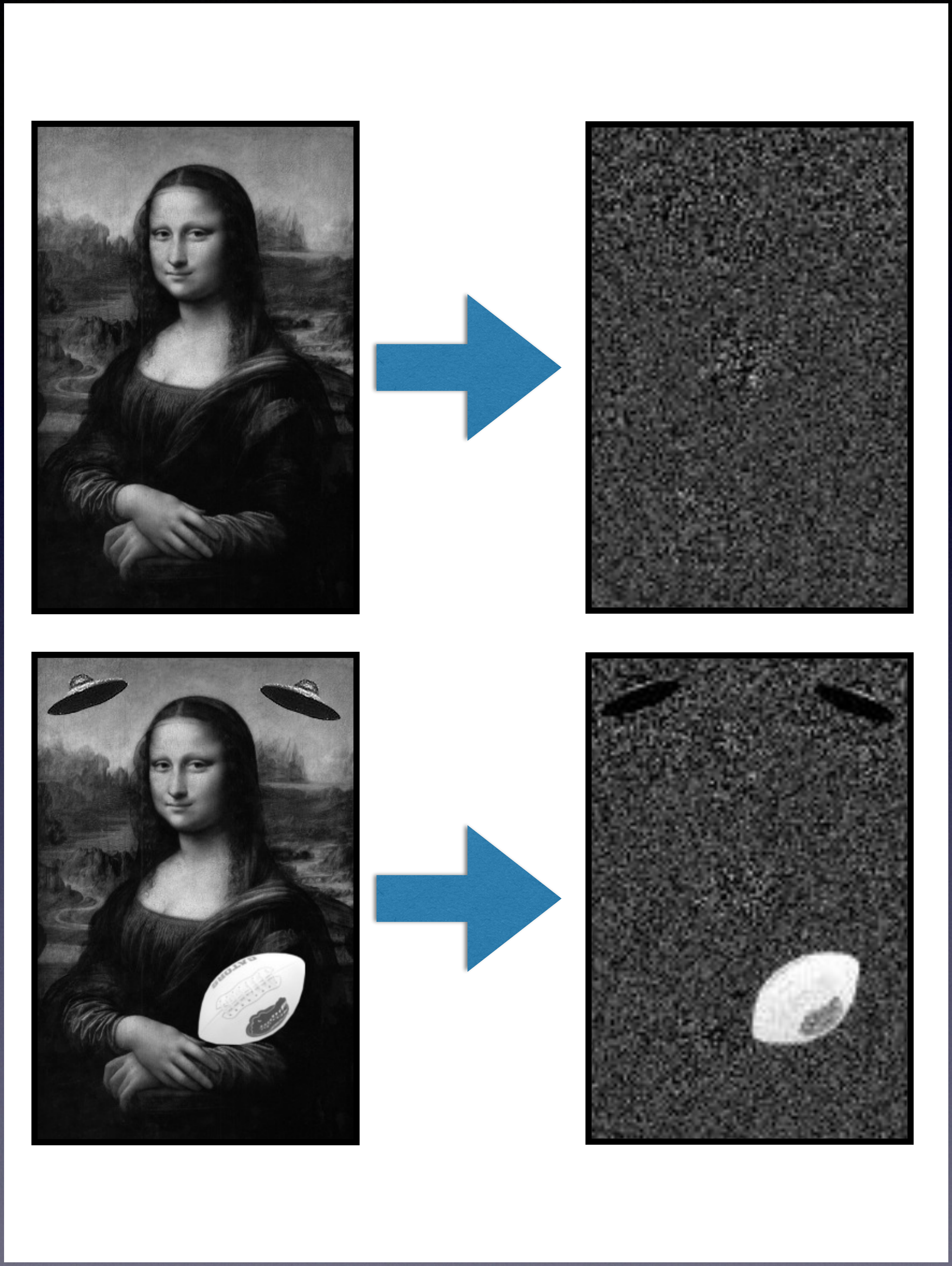


Signal + Background pseudoexperiments averaged.

Analytic Flattening

- 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!