Optimal Observables for Events with Missing Energy

Bob McElrath
University of California, Davis

Work(s) in progress with Alexandre Alvaes, Hsin-Chia Cheng, Oscar Eboli, Jack Gunion, Guido Marandella, and Tilman Plehn.

Vancouver Linear Collider Workshop 2006

An Imaginary Discovery

Let us imagine that we have an excess in a many-body final state. (For instance, $b\bar{b}l^+l^-\not{E}_T$). What is it?

- Is it SUSY?

- Is it UED?

- Is it Little Higgs?

- Is it a leptoquark?

- Is it squark+leptoquark production?

In this talk I ask: how do we tell what it is and how do we quantify how well we know?

Model Independence

While SUSY/UED/Little Higgs are interesting, they are the wrong questions to ask at the level where one has observed an excess.

Instead we wish to ask the following

1 . What is the phase space structure?

2 . What is the spin structure?

To answer these we first must answer

3 . What is the spectrum of the missing energy?

and to answer *that* we must answer

4 . What are the masses appearing in the phase space?

How do we do Data Analysis?

Experimentalists and theorists alike generally ask: "given this pile of 4-vectors, how do I combine them to solve for an interesting quantity?" – This path leads to plots, histograms, likelihoods, neural networks, linear approximations, gaussian errors, monte carlo, and hopefully, a measurement.

Most analyses come down to: Find a transformation $f(p_1^\mu, \ldots, p_N^\mu)$ that when histogrammed, gives a simple function.

For instance in spin studies,

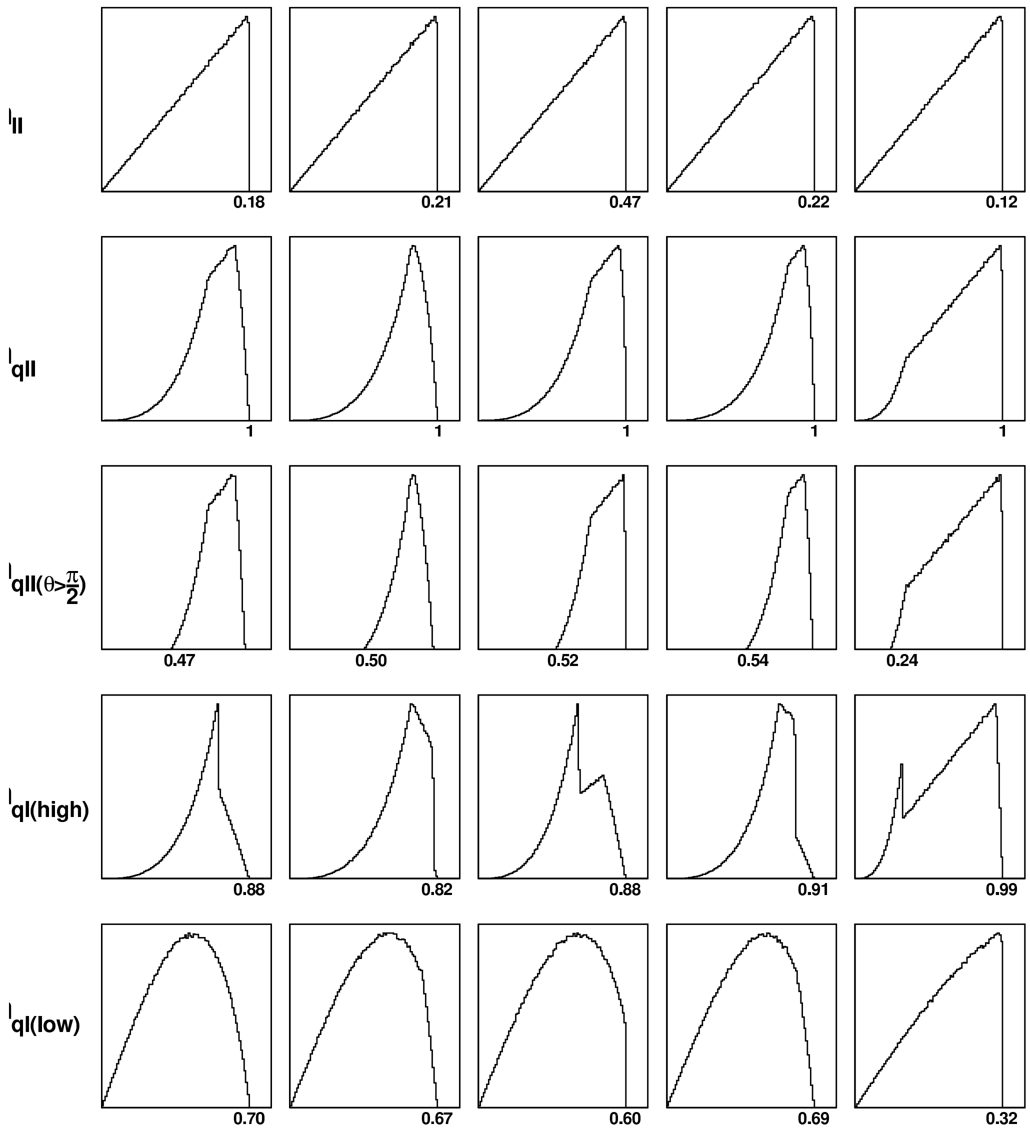
$$f(p_1^\mu, \ldots, p_N^\mu) = \cos\theta = \frac{E_1 E_2 - p_1^\mu p_{2\mu}}{|\vec{p}_1||\vec{p}_2|}, \qquad \frac{d\sigma}{df} = 1 \pm a\cos\theta \pm b\cos^2\theta$$

or resonances

$$f(p_1^\mu, \ldots, p_N^\mu) = p_{ij}^2 = (p_i^\mu + p_j^\mu)^2, \qquad \frac{d\sigma}{df} = \frac{1}{(p_{ij}^2 - M^2)^2 + M^2\Gamma^2}.$$

By now it is clear that there *may be no simple functions*. These methods are also *guaranteed to lose information*. Furthermore, some things worth measuring *are not simple*.

A SUSY case (Barr Method)



There is a menu of possible things to plot and curves to fit.

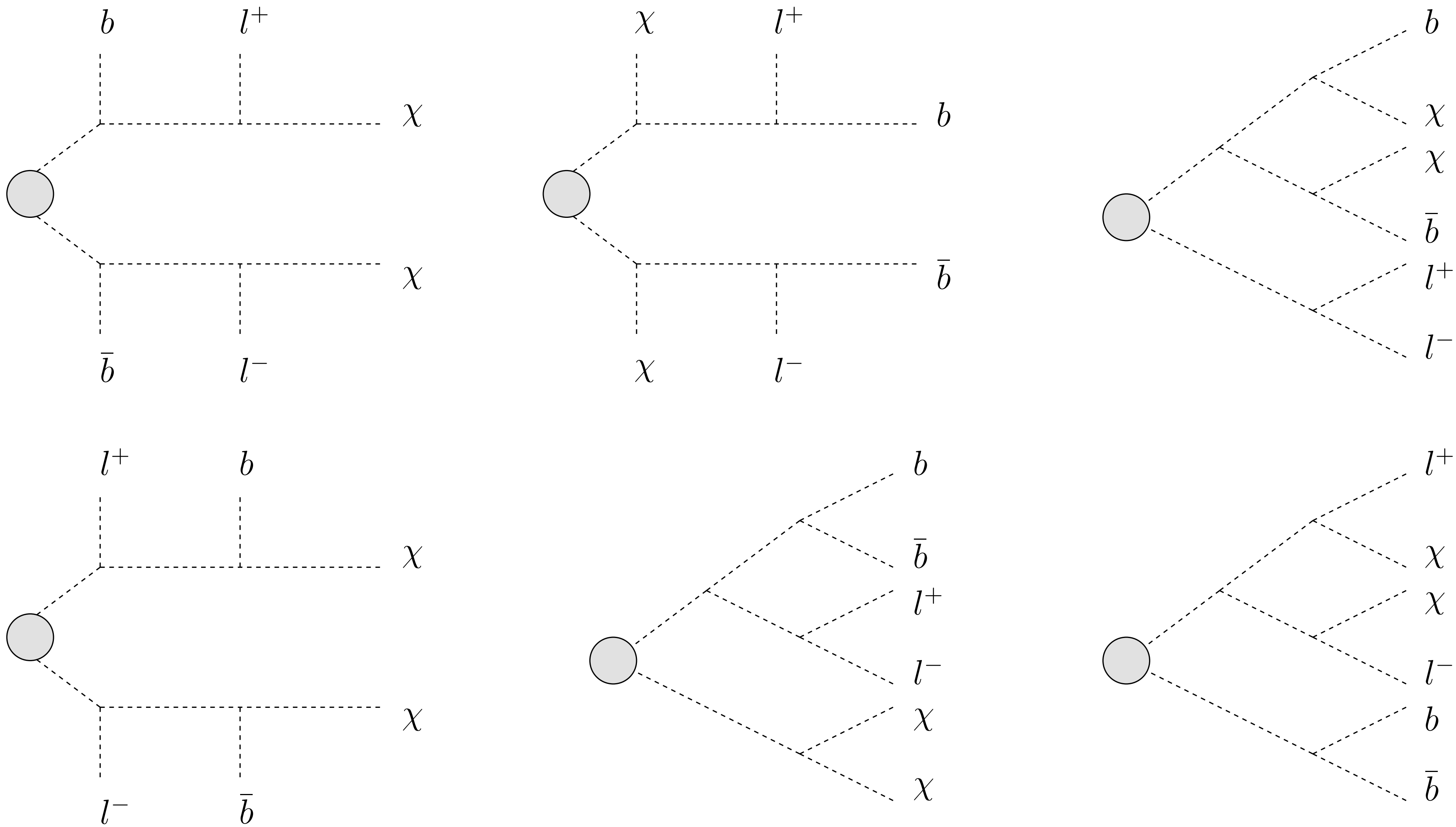
Strong assumptions about the phase space structure, and where a final state particle appears in the decay chain.

Can we trust the statistics when the analysis chooses from this menu? What is the trial factor?

The number of things to plot/fit is *exponential* in the number of external fields.

[Gjelsten, Miller, Osland `hep-ph/0410303`]

Some of the Topologies for $b\bar{b}l^+l^-\not{E}_T$



A Theorist's Perspective on Data Analysis

Every particle discovery comes down to measuring a multi-dimensional probability distribution for observables.

Therefore I ask:

Given this *theory*, what are the statistics underlying it, and what is the optimal way to extract information?

Cross Sections as Probability Densities

A cross section generally is given by

$$\sigma = \frac{1}{F}\int \left|\mathcal{M}(p_0^\mu, p_i^\mu, \mathbf{Y})\right|^2 \left(\prod_i \frac{d^3\vec{p}_i}{(2\pi)^3 2E_i}\right)(2\pi)^4\delta^4(p_0^\mu - \textstyle\sum_i p_i^\mu)$$

for some initial state momenta p_0^μ and final state momenta p_i^μ. This is a zero-dimensional projection of a high-dimensional phase space, and as such contains very little information! Buried in here somewhere is all the information that is to be had.

Cross Sections as Probability Densities

A cross section generally is given by

$$\sigma = \frac{1}{F}\int \left|\mathcal{M}(p_0^\mu, p_i^\mu, \mathbf{Y})\right|^2 \left(\prod_i \frac{d^3\vec{p}_i}{(2\pi)^3 2E_i}\right)(2\pi)^4\delta^4(p_0^\mu - \textstyle\sum_i p_i^\mu)$$

for some initial state momenta p_0^μ and final state momenta p_i^μ. This is a zero-dimensional projection of a high-dimensional phase space, and as such contains very little information! Buried in here somewhere is all the information that is to be had.

Let us do a little rearrangement to retain all information in the high-dimensional space.

$$P(\vec{p}_1, \ldots, \vec{p}_N) = \frac{1}{\sigma}\frac{d\sigma}{\prod_i d^3\vec{p}} = \frac{(2\pi)^{4-3N}}{2^N F\sigma \prod_i E_i}\left|\mathcal{M}(p_0^\mu, p_i^\mu, \mathbf{Y})\right|^2 \delta^4\left(p_0^\mu - \textstyle\sum_i p_i^\mu\right).$$

this is a *probability density* expressing the probability of a particular configuration of momenta. For N external particles, it is a $3N-4$ dimensional space.

Probability Densities for Hadron Colliders

The previous equations assumed all initial and final state momenta were known. e.g. a lepton collider. At hadron or photon colliders this is not the case. So we must integrate over the initial state as well.

$$P_{had}(\vec{p}_1,\ldots,\vec{p}_N,x_1,x_2) = \frac{1}{\sigma}\frac{d\sigma}{dx_1 dx_1 \prod_i d^3\vec{p}}$$
$$= \frac{(2\pi)^{4-3N}}{2^N F\sigma \prod_i E_i} f_{i1}(x_1) f_{j2}(x_2) \left|\mathcal{M}_{ij}(p_0^\mu,\ldots,p_N^\mu,\mathbf{Y})\right|^2 \delta^4\left(p_0^\mu - \sum_i p_i^\mu\right).$$

for parton i and j having Parton Density Functions f_{i1} and f_{j2} respectively and $p_0^\mu = \sqrt{s}(x_1+x_2; 0, 0, x_1-x_2)$.

Probability Densiities with Missing Energy

If one expects new physics to explain the Dark Matter component of the universe, one generically expects a dark matter particle, with non-zero mass to escape the detector.

Therefore in events with missing particles, we must *project* the previous probability densities onto the space of L measured particles

$$P_{\text{meas}}(\vec{p}_1, \ldots, \vec{p}_L) \;=\; \frac{(2\pi)^{4-3N}}{2^N F\sigma \prod_l E_l} \int \left|\mathcal{M}_{ij}(p_0^\mu, \ldots, p_N^\mu, \mathbf{Y})\right|^2 \delta^4\left(\textstyle\sum_i p_i^\mu\right) \prod_m \frac{d^3\vec{p}_m}{E_m}$$

for lepton colliders or

$$P_{\text{meas,had}}(\vec{p}_1, \ldots, \vec{p}_L) = \frac{(2\pi)^{4-3N}}{2^N F\sigma \prod_l E_l}$$
$$\times \int f_{1i}(x_1) f_{2j}(x_2) \left|\mathcal{M}_{ij}(p_0^\mu, \ldots, p_N^\mu)\right|^2 \delta^4\left(\textstyle\sum_i p_i^\mu\right) dx_1 dx_2 \prod_m \frac{d^3\vec{p}_m}{E_m}$$

for hadron or photon colliders.

Summary of What We Know

These probability densities are fundamentally what we measure. Let us denote this "complete probability density" by

$$P(\mathbf{x}, \mathbf{x}' | \mathbf{Y}) = P(\mathbf{x}' | \mathbf{x}, \mathbf{Y}) \, P(\mathbf{x} | \mathbf{Y})$$

where $\mathbf{x} = \{\vec{p}_1, \ldots, \vec{p}_L\}$ are our observables for L visible particles, $\mathbf{x}' = \{\vec{p}_{L+1}, \ldots, \vec{p}_N\}$ are "missing" observables*, and $\mathbf{Y} = \{M_1, \Gamma_1, M_2, \Gamma_2, etc\}$ is the set of Lagrangian parameters.

We measure

$$P(\mathbf{x} | \mathbf{Y}) = \frac{1}{\sigma(\mathbf{Y})} \frac{d\sigma}{d\mathbf{x}} = \int P(\mathbf{x}, \mathbf{x}' | \mathbf{Y}) d\mathbf{x}'.$$

If we have observed some events $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ we can write the distribution of missing observables

$$P(\mathbf{x}' | \mathbf{X}, \mathbf{Y}) = \int P(\mathbf{x}, \mathbf{x}' | \mathbf{Y}) \simeq \frac{1}{N} \sum_{i=1}^{N} \frac{P(\mathbf{x}_i, \mathbf{x}' | \mathbf{Y})}{P(\mathbf{x}_i | \mathbf{Y})}$$

But this is only meaningful if we know $\mathbf{Y}$!

*subject to constraints from the $\delta^4(\sum_i p_i^\mu)$

Connections to Traditional Methods

Traditional, 1-dimensional analyses rely on approximate factorization of $P(\mathbf{x}|\mathbf{Y})$. For instance if I want to measure a mass, I want to write

$$\begin{aligned} P(\mathbf{x}|\mathbf{Y}) &= P((p_i+p_j)^2|M,\Gamma)P(\{\mathbf{x}\}-\{(p_i+p_j)^2\}|\mathbf{Y}-\{M,\Gamma\}) \\ &\quad +\epsilon(\mathbf{x}|\mathbf{Y}) \end{aligned}$$

If $P(\mathbf{x}|\mathbf{Y})$ factorizes in this manner ($\epsilon(\mathbf{x}|\mathbf{Y})=0$) then $(p_i+p_j)^2$ is a *sufficient statistic* of M and Γ.

$\epsilon(\mathbf{x}|\mathbf{Y})$ is in general non-zero, but might be shown to be small (for instance, $\epsilon(\mathbf{x}|\mathbf{Y})$ will contain finite-width effects, interference, and NLO corrections which are all small).

The only interesting factorization for missing-energy events is

$$P(\mathbf{x}|\mathbf{Y}) = P(\{(p_i+p_j)^2\}|\{M_k,\Gamma_k\})P(everything\ else)$$

because narrow propagators vary quickly with $(p_i+p_j)^2$ while numerators are slowly-varying. *Masses do not factorize from each other.*

Connections to Traditional Methods II

From Meade, Reece, **hep-ph/0601124**

$\langle H_t \rangle$: red

$\langle |\not{E}_T| \rangle$: blue

$\langle M_{\text{eff}} \rangle$: purple

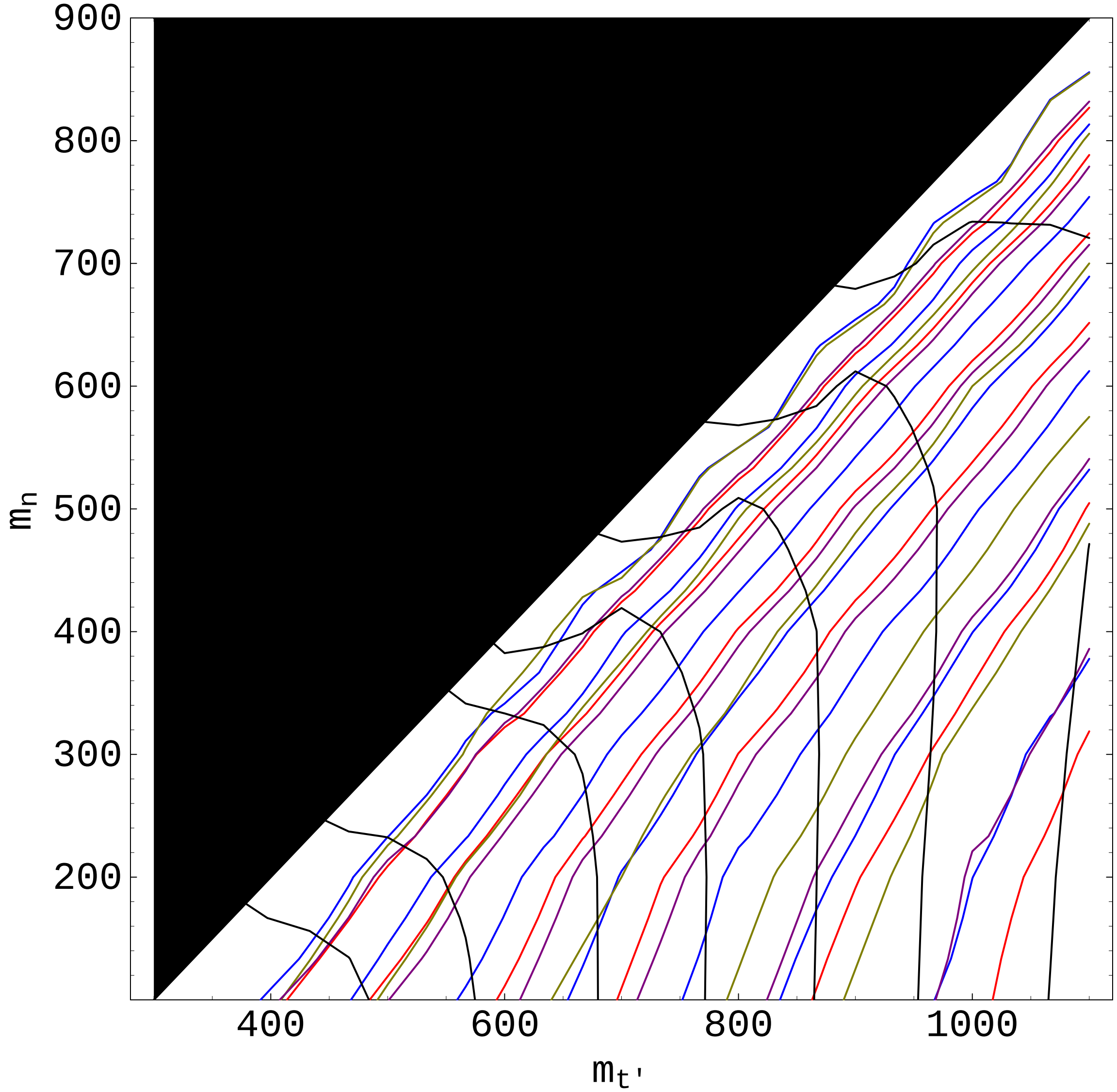
M_{T2}^{max}: gold

σ: black

Mass Degeneracies are not fundamental (Alves, Éboli, Plehn, **hep-ph/0605067**)

This is the statement that

$$\frac{\partial}{\partial M} P(\vec{p}_1, \ldots, \vec{p}_N | \mathbf{Y}) \neq 0$$



Constructing a Likelihood

Our task is to determine $\mathbf{Y}$ for a given set of events $\mathbf{X}$. The Neyman-Pearson lemma tells us that the most powerful statistic for differentiating two hypotheses $\mathbf{Y}$ and $\mathbf{Y}^{(n-1)}$ is the ratio of two Likelihoods. Our "complete data" Likelihood is

$$L(\mathbf{Y}|\mathbf{X},\mathbf{X}') = \prod_{i=1}^{N} P(\mathbf{x}_i, \mathbf{x}_i'|\mathbf{Y}).$$

But this isn't useful yet since $\mathbf{X}'$ is unknown. In statistics literature this is referred to as the "missing data" problem. A solution is the *EM Algorithm*. It consists of two steps

- Compute the *E*xpectation of the Likelihood above, given the observed data $\mathbf{X}$ and some guess for parameters $\mathbf{Y}^{(n-1)}$:

$$Q(\mathbf{Y},\mathbf{Y}^{(n-1)}) = E\left[\log P(\mathbf{X},\mathbf{X}'|\mathbf{Y})|\mathbf{X},\mathbf{Y}^{(n-1)}\right]$$

- *M*aximize this expectation over $\mathbf{Y}$ to determine a new guess $\mathbf{Y}^{(n)}$.

EM Algorithm

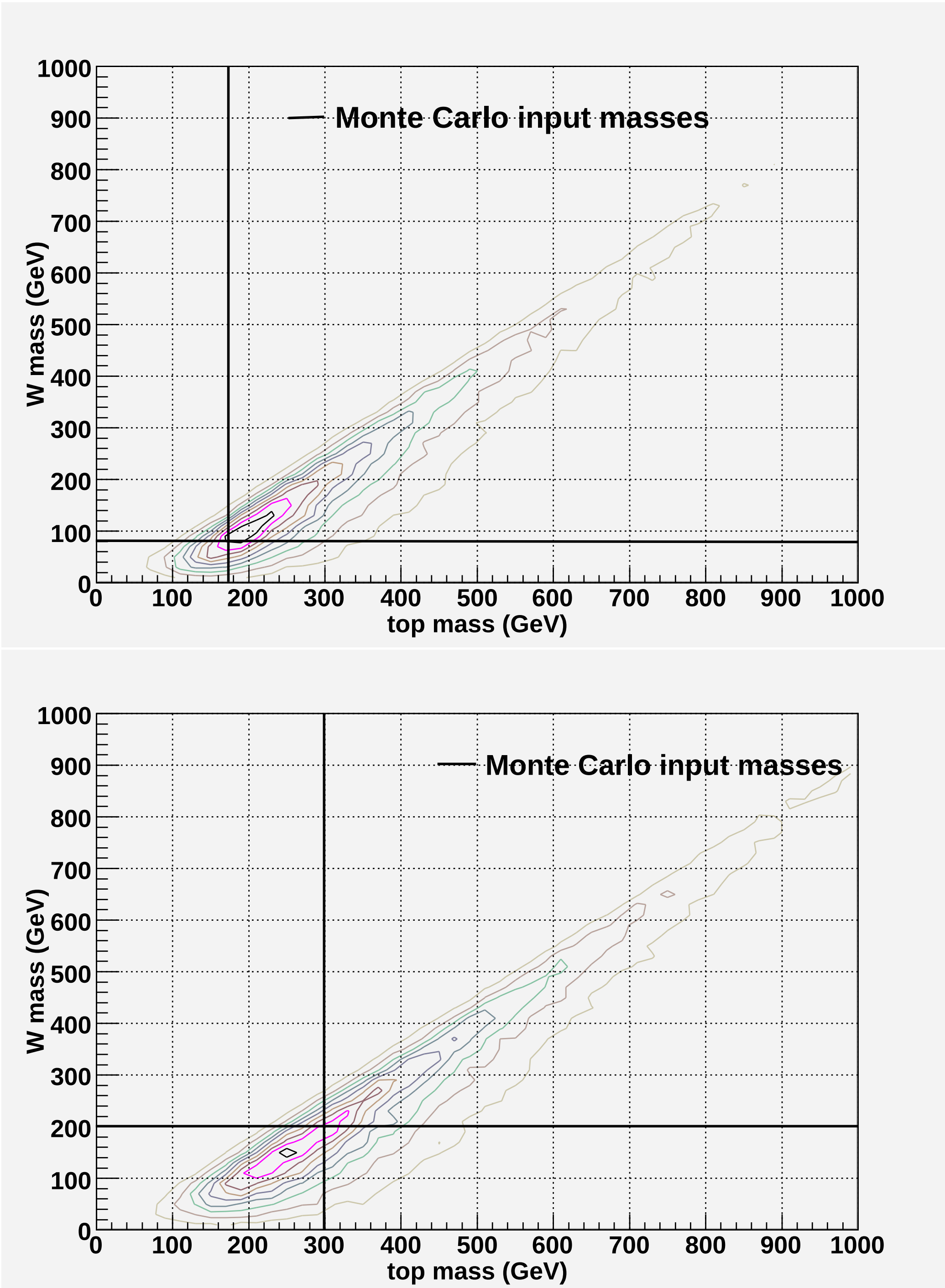
To determine $\mathbf{Y}$ we must maximize

$$
\begin{aligned}
Q(\mathbf{Y}, \mathbf{Y}^{(n-1)}) &= E\left[\log P(\mathbf{X}, \mathbf{X}'|\mathbf{Y})|\mathbf{X}, \mathbf{Y}^{(n-1)}\right] \\
&= \int \log P(\mathbf{X}, x'|\mathbf{Y}) P(x'|\mathbf{X}, \mathbf{Y}^{(n-1)})\, dx' \\
&= \frac{1}{N} \sum_{i,j=1}^{N} \frac{\int \log P(x_i, x'|\mathbf{Y}) P(x_j, x'|\mathbf{Y}^{(n-1)})\, dx'}{\int P(x_j, x'|\mathbf{Y}^{(n-1)})\, dx'}
\end{aligned}
$$

over $\mathbf{Y}$. Note that $Q(\mathbf{Y}, \mathbf{Y}')$ is an analytic function of $\mathbf{Y}$. It is guaranteed to converge to the correct $\mathbf{Y}$.

Note that if I choose the *wrong* $\mathbf{Y}^{(n-1)}$, and I maximize some likelihood, I will converge to some other $\mathbf{Y}^{(n)}$ that is also not correct. (e.g. I cannot generally obtain the correct $\mathbf{Y}$ in one step)

This algorithm is iterative, and we are forced to an iterative solution since if we don't know the masses, then we don't know the missing momentum spectrum. But if we don't know the missing momentum spectrum, then we don't know the masses.



These graphs are obtained by doing a Monte Carlo over missing momentum, assuming a *fixed* distribution for missing momentum. I plot an estimator for M_W vs an estimator M_t.

- When masses are low, peak of distribution occurs *above* the actual masses.
- When masses are high, peak of distribution occurs *below* the actual masses.

This is because the distribution of missing momentum is more consistent with $\simeq$ 200 GeV masses!

Systematic Analysis

The most exciting phrase to hear in science, the one that heralds new discoveries, is not "Eureka!" (I've found it!), but "That's funny..."
– Isaac Asimov

The variables usually considered in SUSY/DM studies are almost useless to look at and develop an intuition about. There are no sidebands, no peaks. Everything looks like a falling p_T spectrum, with a broad bump if you're lucky.

We need to be able to slice and dice this analysis at stages before a final result is presented. A human will notice strange bumps in distributions that a Likelihood analysis will not. A human may recognize when strange bumps are systematic errors...

Strongly coupled theories (Technicolor) have higher resonances (ρ/ρ'). Would we notice?

Systematic Understanding

Once $\mathbf{Y}$ is obtained, one can perform detailed studies that may seem impossible with traditional methods. For instance

- Is there only one missing particle, or more? Plot:

  $$f(m^2) = \int P(\mathbf{x}, \mathbf{x}'|\mathbf{Y})\delta(m^2 - \left|m_b^2 + M_W^2 - 2M_W E_b\right|_{\vec{p}_W=0})d\mathbf{x}d\mathbf{x}'$$

  where the argument of the delta function is evaluated in the $W^\pm$ rest frame. Multiple invisible particles will give multiple bumps.

- Are there multiple overlapping signals? e.g. $\tilde{t} \to b\chi^+ \to bl^+\chi_0$ with $t \to bW^+ \to bl^+\nu$. Plot:

  $$f(m_1, m_2) = \int P(\mathbf{x}, \mathbf{x}'|\mathbf{Y})\delta(m_1 - |p_1 + p_2|)\delta(m_2 - |p_1 + p_2 + p_3|)d\mathbf{x}d\mathbf{x}'$$

  Multiple signals (of the same kinematic topology) will give multiple bumps.

- Can one identify side-bands for background estimation?

Other Reasons to use Likelihoods

Given two hypotheses $P(x|\mathbf{Y})$ and $P'(x|\mathbf{Y}')$, they can be directly compared by constructing their Likelihood ratio, as usual

$$R = \frac{L(x|\mathbf{Y})}{L'(x|\mathbf{Y}')}$$

which tells us if we can reject one hypothesis or the other. This only works *after* we have determined the paramters $\mathbf{Y}$ and $\mathbf{Y}'$, so we already know the missing momentum spectrum in each case.

- Wide classes of Likelihood Ratio (LR) tests

- The LR statistic is invariant under reparameterization.

Resolution Functions

Running Monte Carlo and detector simulations for every possible signal, and every possible choice of parameters $\mathbf{Y}$ for each signal is impractical.

But, this multiply-specifies the information we need. e.g. The detector simulation is insensitive to the details of the hard scattering matrix element. It is not necessary to simulate Z' at 500 GeV and also an off-shell Z at $M_Z = 500$ GeV because the detector response is *identical*.

I propose instead that we think hard about *resolution functions*. Our probability densities of interest then become

$$P_{meas}(x, x'|\mathbf{Y}) = \int P(w, x'|\mathbf{Y}) R(w, x|\mathbf{Z})\, dw.$$

Where the resolution function $R(w, x|\mathbf{Z})$ describes how the final-state partons w get mapped into detector objects x.

Resolution Functions ctd...

With a resolution function our Likelihood becomes

$$Q(\mathbf{Y}, \mathbf{Y}^{(n-1)})$$
$$= \frac{1}{N} \sum_{i,j=1}^{N} \frac{\int \log P(w, x'|\mathbf{Y}) P(v, x'|\mathbf{Y}^{(n-1)}) R(w, x_i|\mathbf{Z}) R(v, x_j|\mathbf{Z})\, dw dv dx'}{\int P(x_j, x'|\mathbf{Y}^{(n-1)}) R(w, x_j|\mathbf{Z})\, dw dx'}$$

This involes $2N$ integrals, which can be efficiently evaluated using many techniques.

Under many circumstances, resolution functions factorize. e.g.

$$R(\vec{p}_{e1}, \vec{p}_{e1}) = R(\vec{p}_{e1}) R(\vec{p}_{e2})$$

$$R(\vec{p}_b, \vec{p}_{\overline{b}}, \vec{p}_{e^+}, \vec{p}_{e^-}) = R(\vec{p}_b, \vec{p}_{\overline{b}}) R(\vec{p}_{e^+}) R(\vec{p}_{e^-})$$

One might need to generate detector simulation for $b\overline{b}$ pairs in any configuration, but then you *don't have to repeat it* when analyzing any signal with 2 b's! (e.g. $LQ_3\, LQ_3 \to b\overline{b}\tau^+\tau^-$)

Computational Complexity and Resolution Functions

For an N-body final state, the computational complexity of running a full detector simulation is $\mathcal{O}(e^N)$.

With resolution functions one only needs to detector simulation for *factorizable* pieces.

- As long as they are well-separated, leptons always factorize.
- Jets do not factorize due to color reconnections.
- Resolution functions may be taken from data.

Summary/Conclusions

- Due to non-factorization of probability densities with missing energy, there generally are no one-parameter sufficient statistics.

- Because we understand the probability densities responsible for the data, HEP analyses are ideally suited to likelihood analyses.

- If we are only capable of finding SUSY, due to our ability to run Monte Carlo or our limited analysis techniques, we may not be capable of finding "something else", and we risk shoe-horning any and all signals into a SUSY bin.

- Detector resolution functions can reduce our model dependence, and may reduce the computational complexity of analyses.