Summary of Columbia REU Research

ATLAS: Summer 2014
Outline

- Brief Analysis Summary
- Overview of Statistical Analysis
- Comparisons between the Asymptotics and the Toys
- Applications for Mass Dependent Cuts
Quest for the Randall-Sundrum Graviton

- Search for a new RSG particle around 1 TeV

\[ RSG \rightarrow HH \rightarrow b\bar{b}b\bar{b} \]

- Decay products will become “Lorentz boosted” causing the standard \( \Delta R = 0.4 \) jets to overlap

- Move into the boosted regime with the b-quarks from each H contained in a large-R jet with \( \Delta R = 1.0 \)
The Standard Cuts

Following are the standard cuts for the analysis now:

- MV1 > 0.70
- Leading jet $p_T > 350$ GeV
- Subleading jet $p_T > 250$ GeV
- Track jet $p_T > 20$ GeV
- Coupling: $c = 2.0$
Hypothesis Testing

Two hypotheses: null hypothesis: $H_0$, and alternative hypothesis: $H_1$

Assuming $H_0$ true, calculate the probability, $p$, that $H_0$ is true given the data

Compare $p$ to the critical value, $\alpha$ (i.e., $\alpha = 0.05$)

Two Conclusions:

1. If $p < \alpha$ => reject the $H_0$ in favor of $H_1$
2. If $p > \alpha$ => fail to reject $H_0$
Choices of the Null Hypothesis

- For Discovery, let $H_0 =$ there is no RSG
- For Limit Setting, let $H_0 =$ there is a RSG with signal strength $\mu$ and we just missed it.

For example: consider a histogram with $N$ bins...

$$\nu_i = b_i + \mu s_i$$

$\nu_i$ : probability of an event landing in $i^{th}$ bin

$b_i$ : probability of background event being in $i^{th}$ bin

$s_i$ : probability of signal event being in $i^{th}$ bin
The number of events in each of the bins follows a Poisson distribution.

Independent probabilities => multiply them together.

Nuisance parameters ($\theta$): values that have to be derived from the data and are not known a priori.

$$L(\mu, \theta) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} e^{-u_k}$$
Motivation for Test Statistic

- Quantify the agreement of the null and alternative hypothesis through their ratio
  \[ \lambda(\mu) = \frac{L(\mu, \theta'')}{L(\mu', \theta')} \]

  \[ \theta'' : \theta \text{ that maximizes } L \text{ for a given } \mu \]

  \[ \mu' \text{ and } \theta' : \mu \text{ and } \theta \text{ values that together maximize } L \]

- Since multiplication is computationally expensive, take the logarithm
  \[ t_\mu = -2 \ln(\lambda(\mu)) \]
Refining the test statistic

- If $\mu' < 0$, this does not support the null hypothesis that there is a signal.

- If $\mu' > \mu$, this should support our null that we have a signal strength of at least $\mu$.

$$
\tilde{q}_\mu = \begin{cases} 
-2 \ln \frac{L(\mu, \theta''(\mu))}{L(0, \theta''(0))} & \mu' < 0 \\
-2 \ln \frac{L(\mu, \theta''(\mu))}{L(\mu', \theta')} & 0 \leq \mu' \leq \mu \\
0 & \mu' > \mu 
\end{cases}
$$
Toys Method

- Scan over a range of $\mu$'s for each mass point and find the pdfs for $H_0$ and $H_1$ and the $\tilde{q}_{\mu}^{\text{obs}}$ value.
- Find the $CL_s$ value for each of these mu scans as

$$CL_s = p_s/(1-p_b)$$

$p_s$ : probability that $\tilde{q}_{\mu}$ is greater than $\tilde{q}_{\mu}^{\text{obs}}$

$p_b$ : probability that $\tilde{q}_{\mu}$ is less than $\tilde{q}_{\mu}^{\text{obs}}$

- Determine which $\tilde{q}_{\mu}$ corresponds to a $CL_s$ of 0.05
Increasing the $\mu$ makes it harder to observe a $\tilde{q}_\mu$ value this extreme and decreases the $\mathrm{CL}_S$ value.
Finding the 95% CL_{s} value

From these CL_{s} values for all of the \( \mu \) scans, found the \( \mu \) corresponding to a CL_{s} = 0.05

Continued adding \( \mu \) scans where higher accuracy was desired until the error was less than 1%.

\[
err = 100\% \times \frac{\mu_{up} - \mu_{low}}{2 \mu_{current}}
\]
Repeat this procedure for every RSG potential mass point that we are considering to find the “Brazil plot” for the toys.

Cuts:
MV1 > 0.75
c = 1.0
Generate Brazil plots more quickly by using an asymptotic approximation allowing us to find the cross section using a formula.

Cuts:
MV1 > 0.75
c = 1.0
Comparing Asymptotics vs. Toys

Test how well the asymptotics approximates the toys by overlaying the plots on previous two slides.

Overlaying the two Brazil Plots
Asymptotics vs. Toys

Compare the toys and asymptotics for two different MV1 cuts of 0.50 and 0.90 to test if the discrepancy is due to the limited background statistics in the signal region.

Overlaying the two Brazil Plots for MV1 cut of 0.50

Overlaying the two Brazil Plots for MV1 cut of 0.90

88 background events

18 background events
Ratio Comparisons

- Appear to follow basically the same trend
- So the discrepancy apparently is not due to the background...
Scaling the signal

Test if discrepancy is due to small number of expected signal events by scaling the signal by 30.
Investigating Mass Dependent Cuts

Test what $p_T$ cut we can make on the leading large R jet to maintain 90% efficiency
Sensitivity for cuts $\text{MV1} = 0.70$, track jet $p_T > 20 \text{ GeV}$
Conclusions:

- The toys appear to do a better job than the asymptotics for the signal sample size that we are dealing with, although it may not be possible to use them with all the systematics for the analysis.

- Using different cuts for the different RSG masses that we are considering can improve our efficiency substantially.

- Many thanks to the Columbia REU program and my supervisor, Dr. Emily Thompson, for providing this summer research opportunity at CERN!
