Statistical Analysis Procedures Used for the Search for the Randall-Sundrum Graviton in the Boosted Regime

Nicole Hartman

Supervisor: Dr. Emily Thompson

August 1, 2014

Abstract

In the analysis looking for the signature $X \rightarrow HH \rightarrow b\bar{b}b\bar{b}$, where $X$ is the Randall-Sundrum Graviton (RSG), this paper reveals the preliminary results and preparation work done to optimize the $p_T$ cut that we can make on the leading jet to optimize the limits that we can set on the cross section for the RSG particle. The limit setting is done using a frequentist approach and an asymptotics approximation, and this paper also examines the accuracy of this approximation by doing some comparisons with the toys to see if we are justified in using the asymptotics when unblinding the analysis.
1 Introduction

The Standard Model incorporates the strong, electromagnetic, and weak forces to provide the framework for particle physics that has been proven to astounding accuracy through a plethora of experiments. However, despite its experimental success, it must be an incomplete description of nature for it fails to incorporate the last fundamental force: gravity. The Standard Model still can make accurate predictions because of gravity’s vanishingly small strength in comparison with the other forces. For example, force strengths can be quantified by their coupling constants, which are approximately $1$, $10^{-2}$, and $10^{-6}$ for the strong, electromagnetic, and weak forces, respectively [10]. Gravity, on the other hand, has a coupling constant of $10^{-40}$, thirty-four orders of magnitude smaller than that of the weak force! These coupling constants have no pattern, and this mystery of gravity’s infinitesimally small strength is known as the hierarchy problem in particle physics.

A solution proposed by Randall and Sundrum suggests that our apparently three-dimensional world actually has five-dimensions with two tiny curled up extra dimensions. The gravitational force might actually be stronger than it appears, but its strength is concealed from us by manifesting itself in the inaccessible extra dimensions. We can infer the veracity of these claims by finding the approximately 1 TeV, spin-2 graviton that Randall and Sundrum’s theory predicts. This mass is within the energy scale of the Large Hadron Collider and has therefore motivated our experimental quest for this aptly named Randall-Sundrum graviton (RSG).

1.1 Large Hadron Collider

The Large Hadron Collider (LHC) allows us to search for new particles, such as the Randall-Sundrum graviton, by accelerating and colliding protons and then using the relation $E = mc^2$ to use the high energies to create new particles of a given mass $m$. The last run of the LHC went up to energies of 8 TeV in the protons’ center of mass frame. However, the protons are not fundamental particles, but composed of quarks and held together by gluons. Therefore, it is the individual colliding quarks that we are interested in studying, but they can only carry a fraction of the total protons energy, so we can never attain the full center of mass energy for a single quark-quark, quark-gluon, or gluon-gluon “event” under consideration. If protons are limited in this manner, why do we use protons instead of a fundamental particle, like the electron? Since the LHC is a circular collider, the charged particles that it accelerates are susceptible to synchrotron radiation according to Eq 1,
\[ P_{\text{loss}} = \frac{E^4 \text{ current}}{m^4 R} \]  

(1)

where \( E \) is the energy, \( m \) is the mass, and \( R \) is the radius of the accelerator [7]. The power lost is proportional to \( m^{-4} \). Therefore, accelerating the proton, which is about 2000 times as heavy as the electron, allows us to decrease our power lost through synchrotron radiation by roughly 10 orders of magnitude. Furthermore, the dependence of the energy on the factor \( \frac{1}{R} \) illustrates the motivation to build larger colliders to decrease power loss. And describing the LHC as “Large” is no understatement, for it has a radius of 5.4 miles which allows it to accelerate protons to nearly the speed of light [6].

1.2 ATLAS

Once the LHC has generated these potentially interesting events, we then need a detector to capture this data so that physicists can then analyze it to compare the reality of the universe with the reigning theories. The ATLAS detector is one of the detectors at the LHC, composed of a gigantic cylinder 25 m tall and 44 m long [5]. A cut away of the detector is shown in Figure 1.

![Figure 1: Cut away of the ATLAS detector [5]](image_url)

Protons travel along the central axis of the cylinder, and then as the particles decay, they are forced to travel through the detector components. The
detector acts like a braking system on the particles, and we can determine the particle type from where it deposits its energy. The ATLAS detector consists of an inner detector which examines the tracks and vertices of charged particles, the electromagnetic calorimeter where photons and electrons deposit most of their energy, the hadronic calorimeter where hadrons leave the majority of their energy, and lastly the muon detector where muons leave their signatures.

The ATLAS detector utilizes the right-handed spherical coordinate system. The z axis points along the beam pipe and then the azimuthal angle, $\phi$, goes from the vertical around the beam pipe from 0 to $2\pi$, while the polar angle, $\theta$, starts from the $+z$ axis, which points in the direction of the protons velocity, and then rotates from 0 to $\pi$ radians to point towards the -$z$ axis. We can also derive other variables which are more useful for describing the particles in our detector.

The rapidity, $y$, is a measure of the angle in our detector, and is given by Eq 2,

$$y = \frac{1}{2} \ln\left(\frac{E + p_z}{E - p_z}\right)$$

where $E$ is the energy of the particle and $p_z$ is the projection of the particle’s momentum along the beam axis [1]. The rapidity measures how far we are from the center of the detector, and we can use it to define the $\Delta R$ according to Eq 3.

$$\Delta R = \sqrt{(\Delta y)^2 + (\Delta \phi)^2}$$

1.3 Jet Reconstruction

Of the particles that ATLAS detects, for our analysis we are particularly interested in hadronic jets. When the constituents of the proton collide, they can interact to produce a quark or a gluon. These objects both interact via the strong force, which unlike the other forces that we are familiar with, actually grows with increasing distance. Therefore, it becomes more energetically favorable for new particles to pop up out of the vacuum so that the particles can bind to each other in this shorter distance range. This interesting property of the strong force explains why we never observe free quarks in nature, a principle called “confinement.” The quarks or gluons can never exist in isolation, but will continue to form hadrons in a cascading reaction known as “hadronization.” Therefore, a hadronic jet is a quark or gluon that has formed this stream of quarks or gluons and will subsequently deposit most of its energy in the hadronic calorimeter. The jet is then reconstructed by
looking at a cluster of energy deposited in the detector, taking the cell with the highest $p_T$ as the jet axis and then incorporating all of the surrounding energy clusters with in a predefined $\Delta R$ radius as members of the same jet.

Since we are interested in looking for heavy new particles, we want to look for those events where the patrons, or parts of the proton, carry a large fraction of the protons energy. This event is called the hard interaction. However, in the caverns of the collider, hard interactions are not the only interactions present, even if they are the ones that we are interested in. The underlying events, or soft interactions, encompass the collisions that occur between patrons that do not carry a significant fraction of the protons energy and the interactions between patrons that take place even without the constituents actually coming in contact. These non-interesting interactions from other parts of the proton are not the only source of contamination that we have to deal with when isolating our jet, for the LHC does not merely collide one proton with another, but rather groups of protons, called bunches, to increase the likelihood of a hard interaction in each bunch crossing that occurs. The pile-up refers to the other proton proton collisions where a hard interaction did not occur. We can eliminate the jets originating from pile-up and soft interactions by only making a $p_T$ cut on the jets that we select since we expect hard interactions to carry a larger portion of the jet’s momentum. However, this will not save us from the soft interaction spilling over partially into our delta-R cone defining our jet of interest and contaminating it. To minimize this possibility, ATLAS also uses a collection of jet grooming algorithms to remove contamination from the reconstructed jets. One of the methods used to remove pile-up and soft interaction contributions from our jet is called jet trimming.

In jet trimming, the original jet with reconstructed momentum $p_T$ of some predefined radius is then further subdivided into smaller calorimeter jets with smaller $\Delta R$ radiiuses. For example, our analysis uses a $\Delta R$ of 0.3 for these smaller calorimeter jets. We do not expect the soft event and pile-up jets to carry a large fraction of the total momentum of the full jet. Therefore, we can define a cut, $f_{cut}$, and if $p_T^i/p_T^{jet} < f_{cut}$, then we remove this calorimeter jet from the reconstructed jet. This procedure is illustrated in Figure 2.

1.4 Analysis Overview

ATLAS and CMS’s announcement of the discovery of the Higgs Boson on July 4th, 2012, completed the experimental verification the particles predicted by the SM and garnered the particle physics community the 2014 Nobel Prize in Physics. Since the Higgs is the newest member on the particle family scene, its properties are the least known. Therefore, any new physics arising
from discrepancies from the SM will likely arise from probing the Higgs field through examining resonant di-Higgs production. In this analysis, we are looking for a Randall-Sundrum graviton (RSG) by probing the signature of an RSG particle decaying to two Higgs Bosons which then decay into four b quarks.

Actually, \( H \to b\bar{b} \) is the predominant decay mode for the Higgs in the SM, occurring with a probability of 58\%, [8]. However, this signature was not one of the channels used for the Higgs discovery because of the overwhelmingly large background of a gluon producing two b quarks, colloquially called QCD background. In looking for the decay \( RSG \to HH \to bb\bar{b} \), QCD still represents our largest background, but in looking for new physics, our signal cross-section will be larger than in SM production, to allow us to extract the signal from background.

We expect the graviton to have a mass of about 1 TeV, and only 250 GeV of its energy will be used for the rest mass of the Higgs bosons. The remainder of the graviton’s energy will then be given to the momentum of the two Higgs bosons. Therefore, each Higgs boson will see the space in the direction that it is traveling as Lorentz contracted, and the two b quarks that it decays to will have a smaller angular separation distance so that we won’t be able to resolve them using the ATLAS standard \( \Delta R = 0.4 \) jets. This decrease in jet resolution for Lorentz boosted events is shown in Figure 3 [9].

Therefore, instead of trying to resolve each of the four b quarks individually, we can instead make a large-\( R \) jet with \( \Delta R = 1.0 \) for each of the two Higgs bosons. Then we can use b-tagging inside of this large-\( R \) jet to identify the track jets of the two b-quarks, but this time using a more selective \( \Delta R \) of 0.3 to define the smaller calorimeter jets. Finally, by examining the number of events in the signal region, we can either discover the graviton, or set a
2 Statistical Analysis

The most obvious way to discover a new particle would be to look at the data compared to simulation of SM expectations and look for a discrepancy. Because of the small number of expected signal events in our data sample, we utilize sophisticated statistical analysis techniques to garner the most information possible from the data to best utilize the time and resources of the engineers and physicists at CERN.

2.1 Hypothesis Testing

In our quest for the Randall Sundrum graviton, we can ask ourselves two basic questions: did we discover this new particle, or, given the data and a predefined confidence level, what cross sections can we exclude for different particle masses? We can answer these questions using a common statistical procedure known as hypothesis testing.

In hypothesis testing, we have two hypotheses, denoted as the null hypothesis, $H_0$, and the alternative hypothesis, $H_1$. We begin by assuming that the null hypothesis is true, and then ask what the probability is that the null hypothesis is true given the present data. Before we look at the data, we predefine a critical value, $\alpha$, and if the probability that the null hypothesis is true is less than this critical value, we reject the null hypothesis in favor of the alternative hypothesis. The confidence level is the probability that we have rejected the null hypothesis when it is indeed false, mathematically defined as $1 - \alpha$. For this analysis, we choose a critical value of 0.05, corresponding to a confidence level of 95%.

In hypothesis testing, we assume that the null hypothesis is true until it proven false, comparable to the American justice system where a defendant is considered innocent until proven guilty. Failing to reject the null hypothesis...
does not mean that we have proven that the null hypothesis is true, for it merely implies that we do not have enough evidence or data to reject the null. Therefore, we have to be very judicious in our choice of the null hypothesis depending on what conclusion we are trying to make. If we are hoping to discover a new particle, our null hypothesis would be that this particle does not exist. However, if there is not enough data to disprove this null hypothesis, we can turn the question around to find what limits the data allows us to put on the expected mass of the new particle. For this “limit setting” version of hypothesis testing, our null hypothesis is that we do have a signal characterized by a strength \( \mu \). This allows us to disentangle the data into its separate signal and background components.

For example, we can consider the histogram representing the data as the sum of two histograms composed of our background and signal respectively. For a histogram with \( N \) bins, this means that we can find the probability that an event will land in the \( i^{th} \) bin as

\[
E[n_i] = \mu s_i + b_i
\]

where \( \mu \) is the signal strength and \( s_i \) and \( b_i \) are the probabilities for a signal or background event to be in the \( i^{th} \) bin.

To find these probabilities \( s_i \) and \( b_i \), we can utilize the probability density functions (pdfs) for the signal and background obtained from theoretical calculations of the RSG theory and the Standard Model, respectively, as illustrated in Eqs 5 and 6

\[
s_i = s_{tot} \int_{bini} f_s(x; \theta_s) \, dx
\]

\[
b_i = b_{tot} \int_{bini} f_b(x; \theta_b) \, dx
\]

where \( s_{tot} \) and \( b_{tot} \) are the total number of expected signal and background events and \( f_s(x; \theta_s) \) and \( f_b(x; \theta_b) \) are the pdfs for signal and background, respectively.

In the analysis, we cannot completely calculate the pdfs solely theoretically beforehand, as there are some variables called “nuisance parameters” which are not known a priori. Therefore, we fit the values for these nuisance parameters from the data, although this increased flexibility does lead to a loss in sensitivity \([3]\). In Eqs 5 and 6, the nuisance parameters are \( \theta_s \), \( \theta_b \), and \( b_{tot} \), but we can combine these parameters into one nuisance parameter, \( \theta \), i.e., \( \theta = (\theta_s, \theta_b, b_{tot}) \). To reduce the uncertainty introduced by these nuisance parameters, we can make additional measurements on other kinematical variables besides the ones that we are considering to allow us to
additionally constrain the values that our nuisance parameters can take. For example, we can study a kinematic variable in a region where we only expect background events to learn about the shape and number of background events [3]. Eq 7 can allow us to constrain the values of theta based on the expected events in each of the bins based on the distribution $u_i$ which depends on the value of $\theta$.

$$E[m_i] = u_i(\theta) \quad (7)$$

Since we are counting the number of events in each bin, this means that the number of events in each bin follows a Poisson distribution. Therefore, we can find the probability for our entire data set to look the way that it does by examining the Likelihood function. Since the number of events in each bin is independent, this means that we can find this likelihood function simply by multiplying the Poisson distributions over the number of bins each of the histograms, as shown in Eq 8.

$$L(\mu, \theta) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} e^{-u_k} \quad (8)$$

The likelihood function depends on both the signal strength and the nuisance parameters, and the most likely values for $\mu$ and $\theta$ are those which maximize the likelihood function.

### 2.2 Test Statistic $\tilde{q}_\mu$

Given the likelihood function, we can now define the test statistic, $\tilde{q}_\mu$, that we use to quantify the degree to which the data diverges from the null hypothesis. To test a certain signal strength, we might consider initially just taking the ratio of the likelihood functions as shown in Eq 9

$$\lambda(\mu) = \frac{L(\mu, \hat{\theta})}{L(\hat{\mu}, \hat{\theta})} \quad (9)$$

where $\hat{\theta}$ represents the values of the nuisance parameters that maximize the likelihood function for a given $\mu$, while $\hat{\mu}$ and $\hat{\theta}$ correspond to the $\mu$ and $\theta$ values that together maximize the likelihood function. Note that since the likelihood function is always positive, $\lambda$ can only take on values between 0 and 1, while smaller $\lambda$ values imply a less likely strength for the given $\mu$.

The multiplication of all the terms in the likelihood functions is very computationally expensive. We can get around these pesky products by looking instead at the logarithm of the test statistic, given by Eq 10.
\[ t_\mu = -2 \ln(\lambda(\mu)) \]  \hspace{1cm} (10)

Now the test statistic \( t_\mu \) goes to infinity for less likely values of \( \mu \) and 0 for the \( \mu \) that yields perfect agreement with data.

As currently defined, the likelihood function could be maximized for any possible \( \hat{\mu} \) value, even if \( \hat{\mu} \) is negative. Since a negative signal strength does not make sense in our null hypothesis that there is a new particle, we want the test statistic to reflect this. Therefore, if the likelihood is maximized for a negative \( \mu \) value, we will assume that a more realistic version of the likelihood function should have its maximal value for \( \hat{\mu} = 0 \).

Also, if the signal strength that maximizes the likelihood function, \( \hat{\mu} \), is greater than the signal strength, \( \mu \) that we are testing, then we can assume that the data supports a signal strength at least as large as this \( \mu \) and set the test statistic to zero to indicate this agreement.

Eq 11 quantifies the last two paragraphs to show how we get from the test statistic \( t_\mu \) to the one that we use for the analysis, \( \tilde{q}_\mu \). Even though the test statistic \( \tilde{q}_\mu \) is defined piecewise, it is still continuous.

\[
\tilde{q}_\mu = \begin{cases} 
-2 \ln \frac{L(\mu, \hat{\theta}(\mu))}{L(0, \hat{\theta}(0))} & \hat{\mu} < 0 \\
-2 \ln \frac{L(\mu, \hat{\theta}(\mu))}{L(\hat{\mu}, \hat{\theta})} & 0 \leq \hat{\mu} \leq \mu \\
0 & \hat{\mu} > \mu 
\end{cases} \hspace{1cm} (11)
\]

2.3 The Toys Method

Different masses for the RSG particle correspond to different likelihood functions since they have different signal strengths. Therefore, first we need to calculate the \( \tilde{q}_\mu \) value that we have for each of the potential mass points that we’re looping over using the method described in Section 2.2.

From the test statistic observed in nature, we want to find what this can tell us about the signal strength of the proposed graviton. To find out the true \( \mu \) value, we use simulations, called “toys” to simulate ATLAS’s results for each of the RSG mass points, for each of the potential signal strengths that we are scanning over. We can then look at the probability density function distribution for \( \tilde{q}_\mu \) with the given signal strength overlaid with the distribution of \( \tilde{q}_\mu \) under the background only hypothesis where \( \mu = 0 \). This generates a plethora of histograms, but a representative one is shown in Figure 4 and for the mass point of 1.0 TeV and a signal strength of 2.1. For each of these histograms we can calculate the \( CL_s \) value for each of these values, as defined in Eq 12.
Figure 4: Example histogram for calculating the $CL_s(\mu)$ value for the mass point of 1.0 TeV and $\mu = 2.1$

$$CL_s = \frac{p_s}{1 - p_b}$$  \hspace{1cm} (12)

In Eq 12, $p_s$ is the probability that the $\bar{q}_\mu$ could be as extreme, or more extreme given $\bar{q}_\mu^{\text{obs}}$ from the data, or the right-sided tail for the red histogram, while $p_b$ is the probability that the $\bar{q}_\mu$ would not be as extreme as the data, or the left sided tail on the blue histogram. Since these histograms represent pdfs with areas equal to 1, the denominator of the $CL_s$, $1 - p_b$, is therefore the blue shaded right sided tail. As we increase the signal strength that we are testing, this will decrease the probability, $p_\mu$, that we could find a $\bar{q}_\mu$ value this extreme, thus decreasing our $CL_s$ value. Therefore, as we increase the signal strength, we decrease the $p_\mu$ and make it easier to reject the null hypothesis. To illustrate this point, Figure 5 shows the pdfs and expected values for the expected $\bar{q}_\mu$ for the same mass point at 1.0 TeV, but with the signal strength doubled to 4.2.

We then take these $CL_s$ values for each of the $\mu$'s that we scanned over and plot them, as illustrated for the 1.0 TeV mass point in Figure 6. As you can see from the picture, increasing signal strengths correspond to the decreasing the $CL_s$ values, just as we ascertained from Figures 4 and 5. Since we predefined our critical value at 5%, we want to find the largest $\mu$ that the data allows us to be sensitive to and still preclude the signal from the data.

What the code that runs through this actually does is find the upper and lower limits for $\mu$ and then interpolate between them to give the current $\mu$.
Figure 5: Example histogram for calculating the $CL_s(\mu)$ value for the mass point of 1.0 TeV and $\mu = 4.2$.

Figure 6: The $CL_s(\mu)$ values for mass point 1.0 TeV for the given $\mu$'s scans.
value. The error on this $\mu$ value is related to how many $\mu$ scans that we run. To get better precision, we could decrease the intervals between our $\mu$’s that we are scanning over, even though increasing the number of $\mu$ scans increases the program’s run time. The code that I was running the limit setting with would iteratively let me know where to add more $\mu$ scans to increase the precision until the error was less than 1.0%. The error on the $\mu$ value is given by Eq 13.

$$err = \frac{\mu_{upper} - \mu_{lower}}{\mu_{current}} \cdot 100\%$$ (13)

Finally, once we have the $\mu$ that yields a $CL_s$ of 0.05, we can then use this information to find what cross sections we can exclude with 95% confidence. We can find this cross section by multiplying the signal strength by a scaling factor. We need to find the cross section for each of the mass points, and this final result of the “toys” method is summarized Figure 7, affectionately named a “Brazil Plot” because of the graph’s colors. The red line shows us the coupling that corresponds to an RSG graviton with a coupling of 1.0, while the dashed line shows the cross-section that we can exclude.
2.4 Asymptotics Approximation

Determining the limits that we can set on the cross-section using the toys is computationally intensive, having to loop through not only all of the mass points, but also through a series of $\mu$ scans for each of the mass points to find the one $\mu$ that corresponds to a $CL_s = 0.05$. Running through this code takes a lot of time, but fortunately, there is another approach that we can use to determine the limit that we can set on the cross section.

If we have enough events, then because of the Central Limit Theorem we can assume that the signal strength that maximizes the likelihood function, $\hat{\mu}$, follows a Gaussian distribution with a mean $\mu'$ and standard deviation $\sigma$. This can allow us to use an asymptotics approximation to find the Brazil plot just by using a formula instead of looping through all of the iterations and obtain the Brazil Plot such as the one shown in Figure 8 much more quickly.
3 Toys vs. Asymptotics Comparisons

Although in theory the asymptotics and toys should yield the same results, we wanted to check if this was indeed the case. Therefore, my first task was to overlay the expected values from the asymptotics and the toys on top of each other to compare them, as illustrated in Figure 9. However, instead of the lines overlaying on top of each other, the toys set a better limit than the asymptotics. Figure 10 illustrates the extent to which these curves diverged through plotting \((\frac{\text{toys}}{\text{asymptotics}} - 1) \cdot 100\%\). For perfect agreement, we would expect a horizontal line at zero. However, the signal that we are testing has a discrepancy between 50% and 20%, doing worse at the lower mass points.

3.1 Comparing the Number of Background Events in the Signal Region

At first we thought that maybe the discrepancy between the asymptotics and toys was due to the finite number of background events in the signal region. To test this hypothesis, I repeated the above analysis but with two
different MV1 cuts, an MV1 cut of 0.50 which lets in more of the background, and then again with a more restrictive MV1 cut of 0.90. The original MV1 cut of 0.75 had 25 background events. The MV1 cut of 0.50, on the other hand, included 68 background events, while the MV1 cut of 0.90 only had 18 background events. We were hoping that the discrepancy would clear up with the less restrictive MV1 cut of 0.50 and get worse for the tighter MV1 cut of 0.90. However, the discrepancy between the toys and asymptotics persisted for both cuts as shown in Figures 11 and 12.

Both of these MV1 cuts follow the same trend with the toys performing better than the asymptotics over all of the mass points that we considered. I next compared the ratio of the asymptotics result over the toys result, shown in Figure 13. The toys and asymptotics curves are basically identical, although for the fewer statistics with the MV1 cut of 0.90 the ratio line is more jumpy. Therefore, we concluded that the discrepancy was not due to the number of background events in the signal region.
Figure 11: Comparing the Asymptotics and Toys for an MV1 cut of 0.50
Figure 12: Comparing the Asymptotics and Toys for an MV1 cut of 0.90
Figure 13: The ratio of the asymptotics over the toys shown for MV1 cuts of 0.50 and 0.90
3.2 Comparing the Number of Signal Events in the Signal Region

Finally, we wondered if perhaps the asymptotics and toys were diverging because of the small number of expected signal events for some of the higher mass points. Therefore, the last test that we tried was to compare the asymptotics and toys approximation for the signal strength multiplied by 30. If the signal is multiplied by 30, this means that we should be sensitive to signal strengths that are around 30 times smaller, so therefore when I ran over the toys I scaled the $\mu$'s that I scanned over down by 30 as well. The comparison of the asymptotics and toys for an MV1 cut of 0.50 with the signal strength scaled by 30 is shown in Figure 14. For the mass points of 1.1 and 1.2 TeV, the toys and asymptotics appear to converge nearly perfectly. The other points do not appear to do as well, but perhaps we could remedy this if we had increased the signal strength even more.

Finally, we wanted to compare how the ratio of the asymptotics over the toys compared for whether or not we scaled the signal. The ratio of the asymptotics over the toys is illustrated in Figure 15. In this case, perfect
agreement with the toys and asymptotics would correspond to a horizontal line with a value of 1. The toys match the asymptotics almost perfectly for 1.1 and 1.2 TeV, but other than that, for mass points above 600 GeV, the toys always do a better job approximating the asymptotics when the signal is scaled. However, the scaled signal ratio plot is substantially more jumpy. We suspect that this is because of the errors on the number of expected RSG events. Furthermore, when we scale the signal, these errors just become more apparent. However, this graph seems to support the conclusion that the discrepancy with the toys and asymptotics is due to the tiny number of expected signal events for the analysis that we are considering.

Based on the comparisons that I did with the toys and the asymptotics, it appears that the asymptotics approximation is more valid for a larger number of signal events than we are looking at. Although this would seem to indicate that we should use the toys once we unblind the analysis, this may be impossible to do when running on all of the systematics, because my comparisons were for no systematics. Systematics allow us to determine the error caused by each of the nuisance parameters, but this involves looping over more iterations for each of the toys that we are considering, blowing up
the computation time. We still have to loop over the asymptotics, but since
we have a formula, we can avoid these nested loops. So although the toys
would be a more accurate option, we may have to stick with the asymptotics
anyways. Because of time restraints, I also used the asymptotics for the plots
with different $p_T$ cuts that are shown in the remainder of this paper.
4 Optimizing the $p_T$ cut

4.1 The Current Cuts

In our analysis, we are considering two large R jets, and we call the jet with the higher $p_T$ the “leading jet” and the one with the lower $p_T$ the “subleading jet.” Our goal in the analysis is to set limits on the mass points that we can exclude for the graviton, and so to do this, we want to optimize our cuts so that we keep the maximum amount of the signal while excluding most of our background. One of our larger backgrounds for this process is $t\bar{t}$ decaying into two b quarks. We can find what $p_T$ cut we should make to eliminate most of this background by using

$$\Delta R \approx \frac{2m}{p_T} \quad (14)$$

Since we are considering a jet with $\Delta R = 1.0$ and the mass of the top quark is 173 GeV, this means a $p_T$ cut of 350 GeV will remove most of the $t\bar{t}$ from our signal. Therefore, at the commencement of the summer, the analysis was utilizing a $p_T$ cut of 350 GeV on the leading jet to eliminate the background from $t\bar{t}$. For the subleading jet, we again utilized this relation and knew that since the mass of the Higgs is approximately 125 GeV, this means that the Higgs decaying into products confined within the $\Delta R = 1.0$ jet would need to have a $p_T$ greater than 250 GeV. If the jet from the Higgs boson has a higher $p_T$ value, it would have a smaller $\Delta R$ value. Jets with lower $p_T$ would not be confined within our $\Delta R$ requirement anyway. However, although this was the motivation for choosing these cuts initially, Allison Marsh’s and my task this summer was to try to discover if a different choice of $p_T$ cuts could lead to a more optimal analysis.

4.2 Finding the cuts for 90% signal efficiency

To start off trying to answer this question, we began by looking at the histograms for the distribution of the leading and subleading jets for each of the signal mass points. For the purposes of this exercise, these histograms had a cut of 250 GeV for both the leading and subleading jets. These distributions are shown in Figures 16 and 17, respectively.

Just a brief glance at these histograms reveals that the cut of 250 GeV really has only started cutting out signal for the subleading jet. But in general, the two graphs seem to follow the same basic trend. The distributions of the jet $p_T$ start to flatten out for the higher mass points, illustrative of the increasing mass width for these gravitons.
Figure 16: Distribution of the $p_T$ values for the leading jet

Figure 17: Distribution of the $p_T$ values for the subleading jet
Figure 18: Cuts on the leading jet to still maintain 90% efficiency

We wanted to determine what $p_T$ cut we could make for the leading or subleading jet to still maintain 90% of our signal. The graphs for these 90% signal efficiency cuts are shown below in Figures 18 and 19. These two graphs seem to follow the same basic trend as well. For larger mass points, we can implement higher $p_T$ cuts, up until some point around 1.5 TeV where we need to start decreasing our cut to maintain the signal efficiency because of the graviton’s increasing mass width. Since the leading jet is defined as the jet with the larger transverse momentum value, we can implement higher $p_T$ cuts on the leading jet than for the subleading jet. We maintain this 90% efficiency if we implement these cuts on the leading OR subleading jet, not both. If we cut on both, this would reduce our efficiency to a more undesirable 81%.
Figure 19: Cuts on the subleading jet to still maintain 90% efficiency
4.3 Implementing the $p_T$ cuts

Since examining higher $p_T$ cuts for the leading jet would allow us to eliminate more of the QCD background, we started by looking at the limits that we could exclude implementing the different $p_T$ cuts of 300, 350, 400, 450, and 500 GeV on the leading jet as depicted in Figure 20. As you can see, higher $p_T$ cuts allow us to set lower limits on the cross section, and once we get to $M_{RSG} = 900$ GeV, the 500 GeV $p_T$ cut sets the best limit of the five that we are considering. Furthermore, up until about an RSG mass of 1.6 TeV, increasing the mass point allowed us to set lower limits for the expected cross section. However, once we passed this threshold, all the lines that we have started to curve up, and this is because for masses this high, our jets become so boosted that our $\Delta R$ cut of 0.3 for our smaller calorimeter jets becomes too wide and these jets start overlapping, reducing our reconstruction ability. Therefore, for larger mass points, a $\Delta R$ cut of 0.25 or 0.2 might fare better for our smaller calorimeter jet resolution.

Furthermore, since initially our $p_T$ cut was at 350 GeV, we wanted to compare how much better (or worse) our limits were for these various mass points. Therefore, we looked at the ratio of our limit with the new $p_T$ cut
Comparing sensitivities with different pT cuts

Figure 21: Ratio of the limits that we can set for a new $p_T$ cut versus the nominal $p_T$ cut of 350 GeV.

versus the nominal $p_T$ cut, which is shown in Figure 21. This graph shows us that the 300 GeV cut sets the best limit for the lower mass points, but when we get to the 1.0 TeV region that we are interested in, the cut of 450 GeV would be more optimal.

5 Conclusions

Based on the toys and asymptotics comparisons, it appears that the toys do a better job at setting a lower limit than the asymptotics for the signal strengths that we are expecting to see. Even though the toys are the correct answer, and the asymptotics should only be used if they replicate the toys solutions, implementing the toys might not be feasible in the unblinded analysis.

Secondly, we can substantially improve the limits that we can set by implementing these different $p_T$ cuts for the different masses. For this analysis, we do not have enough events to afford to use these higher cuts and start cutting out our already tiny signal. However, these new mass dependent cuts can substantially improve our results for the second version of this analysis.
with the greater luminosity for run two of the LHC.

6 Acknowledgements

First off, I want to thank my supervisor, Dr. Emily Thompson, for providing a truly amazing summer research experience through her instruction and guidance. I also want to extend a thank you to Dr. Massimiliano Bellomo for helping me understand the Statical Analysis and answering my questions about the code or results. Finally, I am so grateful to Professor Parsons and the Columbia REU program for providing this opportunity for research in amazing particle physics at CERN.
References


