Published by Institute of Physics Publishing for SISSA
Received: *September 8, 2005*
Accepted: *November 11, 2005*
Published: *November 28, 2005*

Spin discrimination of new heavy resonances at the LHC

Robert Cousins, Jason Mumford, Jordan Tucker and Viatcheslav Valuev

University of California
Los Angeles, California 90095, U.S.A.
E-mail: cousins@physics.ucla.edu, mumford@physics.ucla.edu, tucker@physics.ucla.edu, slava.valuev@cern.ch

Abstract: If a new resonance is discovered at the LHC, the problem of discriminating between two spin hypotheses is in the category of *simple* hypothesis testing if the angular distributions of the decay products are fully specified by the models. We discuss the interpretation of the likelihood ratio statistic for discriminating between a Z' boson (spin 1) and a graviton (spin 2), contrasting this interpretation with previously published work. Inclusion of the possibility of a spin-0 resonance makes the unambiguous identification of either spin-1 or spin-2 resonances more difficult.

Keywords: Beyond Standard Model, Large Extra Dimensions.

JHEP11(2005)046

Contents

1. Introduction

If a new resonance is discovered at the Large Hadron Collider (LHC), the characterization of its spin and couplings will proceed via the traditional methods of measuring production and decay probabilities and distributions. For the simple topology of a narrow resonance decaying to a fermion-antifermion pair, much-discussed candidates include spin-1 Z' bosons [1] and spin-2 gravitons (G^*) [2]. Allanach et al. [3] have described in detail an algorithm for distinguishing spin-1 from spin-2 particles by constructing an unbinned likelihood ratio statistic incorporating the angles of the decay products. We find, however, that the interpretation of this likelihood ratio statistic in ref. [3] needs to be changed.

In the present paper, we discuss in some detail the statistical issues for comparing angular distributions corresponding to different spin hypotheses, including also spin 0. Assisted by a toy Monte Carlo simulation of the approximate acceptance of the ATLAS and CMS experiments [4], we illustrate the correct interpretation of the unbinned likelihood ratio statistic within the framework of Neyman-Pearson hypothesis testing, and give rough estimates for the level at which LHC experiments may be expected to classify the spin of new resonances. More detailed calculations, including detector simulation, reconstruction effects, and systematic uncertainties, are being pursued within the LHC collaborations [5] and are beyond the scope of this note. One may also use other observables, e.g., transverse momentum, rapidity, invariant mass, and resonance width. While the eventual likelihood functions will be more comprehensive, the validity of the statistical interpretation that we describe should remain unchanged.

JHEP11(2005)046

Channel	d-functions	Normalized density for $\cos\theta^*$
$q\bar{q} \to \gamma^*/\mathrm{Z}^0/\mathrm{Z}' \to f\bar{f}$	$\left\lvert d^1_{1,1}\right\rvert^2 + \left\lvert d^1_{1,-1}\right\rvert^2$	$P_1 = \frac{3}{8}\left(1+\cos^2\theta^*\right)$
$q\bar{q} \to \mathrm{G}^* \to f\bar{f}$	$\left\lvert d^2_{1,1}\right\rvert^2 + \left\lvert d^2_{1,-1}\right\rvert^2$	$P_q = \frac{5}{8}\left(1-3\cos^2\theta^* + 4\cos^4\theta^*\right)$
$gg \to \mathrm{G}^* \to f\bar{f}$	$\left\lvert d^2_{2,1}\right\rvert^2 + \left\lvert d^2_{2,-1}\right\rvert^2$	$P_g = \frac{5}{8}\left(1-\cos^4\theta^*\right)$

Table 1: Angular distributions for the decay products of spin-1 and spin-2 resonances, considering only even terms in $\cos\theta^*$ as explained in the text.

In section 2, we discuss the angular distributions expected for Z′ and G* decay to a lepton pair. In section 3, we review classical hypothesis testing for simple hypotheses, which we apply to spin discrimination at the LHC in section 4, treating the alternate hypotheses symmetrically. In section 5, we discuss a variant that treats the two spin hypotheses asymmetrically. Discrimination from spin 0 is described in section 6. We conclude in section 7.

2. Angular distributions

For purpose of illustration, we consider the decay of a new resonance to a light lepton pair $\ell^+\ell^-$, either e^+e^- or $\mu^+\mu^-$. To leading order, the subdiagram for Z′ formation is quark-antiquark ($q\bar{q}$) annihilation, while for a graviton there exist both $q\bar{q}$ annihilation and gluon-gluon (gg) fusion. One defines θ^* as the angle in the dilepton center-of-mass (CM) reference frame between the negative lepton ℓ^- and the incident quark or gluon. Parity-violating couplings for the Z′ can lead to a forward-backward asymmetry from a non-zero coefficient of $\cos\theta^*$ in the angular distribution; with requisite care at a proton-proton collider, this can provide important information about the properties of the resonance [6, 7]. In this paper, however, we focus on the information contained in terms with even powers of $\cos\theta^*$.

For the light fermions considered here, the helicity basis is appropriate. As illustrated in ref. [8], the decay angular distributions for a particle formed from incoming helicity eigenstates can be expressed in terms of Wigner d-functions for highly relativistic decay products, using angular momentum and helicity conservation. For the processes discussed in this paper, the relevant angular probability density functions in the absence of interference are in table 1, as also summarized in refs. [3, 9]. Except for the coefficients of possible additional parity-violating terms odd in $\cos\theta^*$ (not given in table 1), the coefficients do not depend on the couplings. For the spin-2 graviton, only the relative fractions of gluon fusion, $q\bar{q}$ annihilation, and background (predominantly from the Drell-Yan continuum) events are needed to arrive at a parameter-free form for the expected distribution. (For spin 1, the resonance and the Drell-Yan background have the same form.)

The fractions of *generated* events [10] arising from these processes are denoted by ϵ_q, ϵ_g, and ϵ_1, respectively, with $\epsilon_q + \epsilon_g + \epsilon_1 = 1$. For example, at the lightest graviton mass considered in this work, $M_{\ell\ell} = 1.5\,\mathrm{TeV}$, we have, in a typical range of masses considered ($\pm 120\,\mathrm{GeV}$ about $M_{\ell\ell}$), $\epsilon_q = 0.229$, $\epsilon_g = 0.762$, and $\epsilon_1 = 0.009$. While interference among $\gamma^*/\mathrm{Z}^0/\mathrm{Z}'$ amplitudes is included in the spin-1 event generator, interference in the $q\bar{q}$ graviton production is not; this does not effect the statistical issues in this paper, but should be investigated in the more complete analyses underway. Then the form of the probability density $P(\cos\theta^*)$ is

$$P(\cos\theta^*) = \epsilon_q P_q + \epsilon_g P_g + \epsilon_1 P_1\,. \tag{2.1}$$

In the lab frame at the LHC one has the symmetric proton beams (7 TeV on 7 TeV) and the 4-momenta of the final-state leptons that are reconstructed by the detector. Upon transforming to the dilepton CM, the proton beams are in general no longer symmetric or collinear, and the incoming parton directions must be inferred. The Collins-Soper frame [11], in which angles are measured with respect to the axis that bisects the target and beam axis in the dilepton CM frame, is generally regarded as the best frame for minimizing the effect of parton transverse momentum on the measurement of $\cos\theta^*$. We let $\theta^*_{\rm CS}$ denote the polar angle of the ℓ^- in this frame. Partial information on the sign of terms odd in $\cos\theta^*$ can be extracted with some care [6, 7]; however, in this note, the sign is random.

Thus, experimentally one will obtain a set of events with $\theta^*_{\rm CS}$ measured along with other quantities such as dilepton transverse momentum $p_{\rm T}^{\rm dil}$ and rapidity $y^{\rm dil}$. The toy simulation and analysis used for illustration in this paper contains only simple transverse momentum ($p_{\rm T}$) and pseudorapidity (η) cuts on the leptons as the crude approximation to detector acceptance (section 4), as well as the effect of $\theta^*_{\rm CS}$ as an approximation to the true θ^*. From this, one can construct the probability density $P_{\rm acc}(\cos\theta^*_{\rm CS})$ for events accepted (observed) in an experiment for each hypothesis H_i, where i labels the model such as Z$'$ or G*. In this study, we consider only the angular information and integrate over $p_{\rm T}^{\rm dil}$, $y^{\rm dil}$, and any other relevant quantities; if one has confidence that these quantities are well described by the event generators, more variables can be added to $P_{\rm acc}$. Since we do not add this information, $P_{\rm acc}$ for accepted events approximately factorizes:

$$P_{\rm acc}(\cos\theta^*_{\rm CS}|H_i) = P(\cos\theta^*_{\rm CS}|H_i)\overline{\Omega}(\cos\theta^*_{\rm CS})\,, \tag{2.2}$$

where $P(\cos\theta^*_{\rm CS}|H_i)$ is from eq. 2.1 with the ϵ_j set appropriately for the model considered (e.g. for the spin-1 hypothesis, we set $\epsilon_1 = 1$ and $\epsilon_q = \epsilon_g = 0$), and $\overline{\Omega}$ is the acceptance averaged over p_T, y, etc. The resolution smearing from using $\cos\theta^*_{\rm CS}$ instead of $\cos\theta^*$ is not important compared to other uncertainties; related studies modeling detector resolution indicate that the distribution of measured $\cos\theta^*_{\rm CS}$ is also not significantly distorted at the level of statistics considered here [7]. For the simple hypotheses discussed, the overall normalization of $P_{\rm acc}$ does not affect the results (since differences in log-likelihood ratios below are independent of normalizations), and hence need not be computed.

One can then explore, for sets of events generated according to the various spin hypotheses, which spin hypothesis is favored by synthetic data sets of various size, and with what significance. A key point is that eq. 2.2 has *no* free parameters if the fractions ϵ_q, ϵ_g, and ϵ_1 are considered to be fixed. For each observed event, one evaluates $P_{\rm acc}(\cos\theta^*_{\rm CS}|H_i)$ at the observed $\cos\theta^*_{\rm CS}$ to obtain the *likelihood* $\mathcal{L}(H_i)$ of that event under the given hypothesis. The combined likelihood of the data set under a hypothesis is then the product of the events' likelihoods; henceforth in this paper, $\mathcal{L}(H_i)$ refers to this product unless otherwise stated.

3. Hypothesis tests for simple hypotheses

3.1 Characterization of hypothesis tests prior to data collection

In the formal theory of classical hypothesis testing, there is an important distinction between *simple* hypotheses, which have no free parameters to be fit to the data, and *composite* hypotheses, which contain such parameters. The spin hypotheses that predict the dependence on $\cos\theta^*_{\rm CS}$ given by eq. 2.2 are *simple* hypotheses if, as assumed in this paper and in ref. [3], parameters such as ϵ_q are fixed using other inputs.

For testing a simple "null" hypothesis H_A against another simple "alternative" hypothesis H_B [12], one constructs a test statistic (function of the data) and rejects H_A if the measured value of the test statistic lies inside a region called the "critical" region; otherwise H_A is accepted. The extent of this region is determined by the "significance level" of the test, which is the probability α of rejecting H_A if it is true (the type-I error). Conversely, H_A is accepted if the test statistic lies inside the complementary region called the "acceptance" region. The probability β of accepting H_A if it is false (the type-II error) depends on H_B, and is the probability given H_B that the test statistic is in acceptance region for H_A. The probability of accepting H_B if it is true, $1-\beta$, is called the "power".

J. Neyman and E.S. Pearson [13] showed that for fixed α, the test statistic that maximizes the power against H_B is the likelihood ratio

$$\lambda = \frac{\mathcal{L}(H_A)}{\mathcal{L}(H_B)}\,; \tag{3.1}$$

the critical region is determined from the requirement that

$$\lambda \le \lambda^{\rm cut}, \tag{3.2}$$

with $\lambda^{\rm cut}$ adjusted to provide the desired α. In this traditional formulation of hypothesis testing, H_A and H_B are treated asymmetrically in that α is *fixed* and β then minimized.

If one wishes to treat the two simple hypotheses on a more equal basis (as might be the case for a new resonance), there is a modification to the theorem [15] in which one minimizes any given linear combination of α and β, and again λ is the optimal test statistic; so one may choose to minimize the symmetrical combination $\alpha+\beta$. In this paper, we instead require that $\alpha=\beta$, corresponding to the equal-tailed minimax most powerful test [16].

For investigating and summarizing which values of $\lambda^{\rm cut}$ correspond to which values of α and β, the quantity

$$-2\ln\lambda = 2\ln\mathcal{L}(H_B) - 2\ln\mathcal{L}(H_A) \tag{3.3}$$

is particularly useful. A potential point of confusion is that the asymptotic distribution of $-2\ln\lambda$ is different for comparing simple hypotheses than for comparing composite hypotheses. For simple hypotheses H_A and H_B, the central limit theorem implies that $-2\ln\lambda$ tends to a gaussian, as discussed in detail below.

JHEP11(2005)046

In contrast, for *composite* hypotheses in which one hypothesis has p additional parameters, Wilks [17] showed that $-2\ln\lambda$ is distributed asymptotically as a *chi-square* statistic with p degrees of freedom. Wilks's theorem is the basis for the usual methods for obtaining approximate confidence intervals or regions for parameters of fits to a likelihood function. For composite hypotheses, the theorem proving optimality of the likelihood ratio statistics is not applicable except in special cases (one-sided tests).

The values of $\lambda^{\rm cut}$ used in section 4 of ref. [3] correspond to assuming incorrectly that the asymptotic form of $-2\ln\lambda$ for two simple hypotheses is a *chi-square* with one degree of freedom [18].

3.2 Post-data reporting of results of hypothesis testing

While the above formalism is quite useful for characterizing a test before it is performed, it is still rather rigid in that once data is in hand, it is desirable to say more than, e.g., "Hypothesis H_A is rejected by a test with $\alpha = \beta = 0.07$". One would like a number that distinguishes between data near $\lambda^{\rm cut}$ and data far from it. Berger et al. [16] suggest that when H_A is rejected, one additionally reports the *conditional error probability* (CEP) $\alpha(\lambda) = \lambda/(1+\lambda)$, and likewise when H_A is accepted, one reports the CEP $\beta(\lambda) = 1/(1+\lambda)$. This is justified from several points of view, both frequentist and Bayesian. For frequentists, the CEP's $\alpha(\lambda)$ and $\beta(\lambda)$ are the type-I and type-II error probabilities, respectively, for that subset of an ensemble of experiments that obtains the value of λ actually observed. For Bayesians, they are the posterior probabilities of H_A and H_B, respectively, if they are assigned equal prior probabilities. The usefulness of such an approach is debated in ref. [19].

4. Spin hypothesis testing at the LHC

To illustrate the statistical technique, we used HERWIG 6.507 [4] to generate 200,000 events each of Z' (using the Sequential Standard Model (SSM) [20] couplings) and G^* resonances (from the Randall-Sundrum (RS) model [2]) at each of three masses ($M_{\ell\ell} = 1.5$, 2.0, and 2.5 TeV), with the appropriate level of Drell-Yan background. The Z' sample was generated with full $\gamma^*/\mathrm{Z}^0/\mathrm{Z}'$ interference. However, $\gamma/\mathrm{Z}^0/\mathrm{G}^*$ interference is not implemented in HERWIG; the effect of this remains to be investigated further. For the illustrations in this paper, we fix the RS ratio $k/\bar{M}_{\rm Pl}$ to be 0.1, at the upper end of the range usually considered. Thus the scale Λ_π controlling the coupling is set to 3.9 TeV in HERWIG for $M_{\ell\ell} = 1.5\,\mathrm{TeV}$, and scaled linearly with $M_{\ell\ell}$. For lower values of $k/\bar{M}_{\rm Pl}$, more study is required. No explicit K factors were added. We proceed on the assumption that a new resonance has been discovered (perhaps by a technique as in ref. [21]), and we wish to determine its spin.

For the crude approximation to detector acceptance, we require that the $\ell^+\ell^-$ decay products have lab pseudorapidity η constrained to $|\eta| < \eta^{\rm cut}$; values of $\eta^{\rm cut}$ in the range 2.4-2.5 (characteristic of LHC experiments) give similar results. A simple cut on the $p_{\rm T}$ of the decay products, $p_{\rm T} > 20\,\mathrm{GeV}$, is applied to take trigger selection efficiency into account (with negligible effect).

JHEP11(2005)046

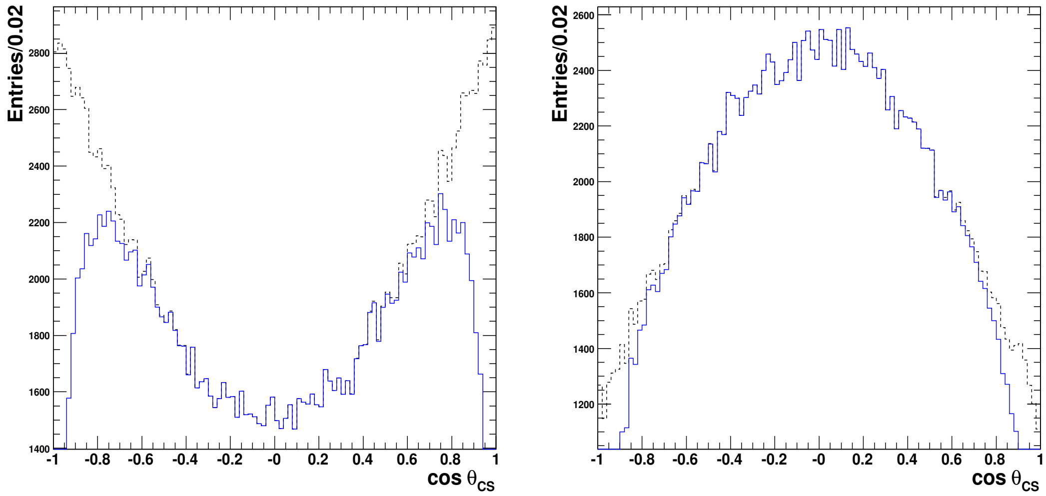

Figure 1: Histograms of $\cos\theta^*_{\mathrm{CS}}$ of individual events from the generated (dashed) and accepted (solid) samples of Z′ (left) and G* (right) for $M_{\ell\ell} = 1.5\,\mathrm{TeV}$.

The window in mass for selecting events in the resonance peak will be optimized by the respective experiments based on details of the mass resolution function. For the present study, in which the background is so low that there is little advantage to a narrow mass window, we use a window of $\pm 120\,\mathrm{GeV}$. Since the results below are quoted in terms of number of events needed to discriminate spins at a given significance level, and since events near the window edge carry similar angular information, the conclusions are essentially independent of this choice.

Events within a mass window thus chosen are selected, and $\cos\theta^*_{\mathrm{CS}}$ is calculated. Histograms of $\cos\theta^*_{\mathrm{CS}}$ for Z′ and G* events before and after acceptance cuts for $M_{\ell\ell} = 1.5\,\mathrm{TeV}$ are in figure 1.

Figure 2 contains the generated and accepted distributions for G* for $M_{\ell\ell} = 2.0$ and $2.5\,\mathrm{TeV}$. While the generated angular distribution for Z′ (considering only even terms in $\cos\theta^*$) does not change with $M_{\ell\ell}$, the corresponding distribution for G* becomes more peaked at higher masses owing to the increased contribution from the $q\bar{q}$ channel. The angular distribution for $q\bar{q}$ annihilation, P_q, also pulls up the overall G* distribution at high $|\cos\theta^*|$; this effect is only partially seen, however, because of acceptance.

The ratio λ of the likelihoods of the hypotheses is then calculated for each event, assigning spin 1 as the null hypothesis H_A and spin 2 as the alternative hypothesis H_B. In taking the ratio, the average acceptance in eq. 2.2 cancels to good approximation and one essentially recovers the ratios of the angular forms in eq. 2.1. (The additive constant due to normalizations is not included.) Figure 3 contains histograms of $-2\ln\lambda$ for the events at $M_{\ell\ell} = 1.5\,\mathrm{TeV}$. The peak at the right corresponds to events near $\cos\theta^*_{\mathrm{CS}} = 0$, where λ is most favorable to the spin-2 hypothesis. The long tail to the left extends to values of λ more favorable to the spin-1 hypothesis.

For each sample of events, a "simulated experiment" then corresponds to the selection of N events from that sample. The combined statistic $-2\ln\lambda$ for the experiment is then

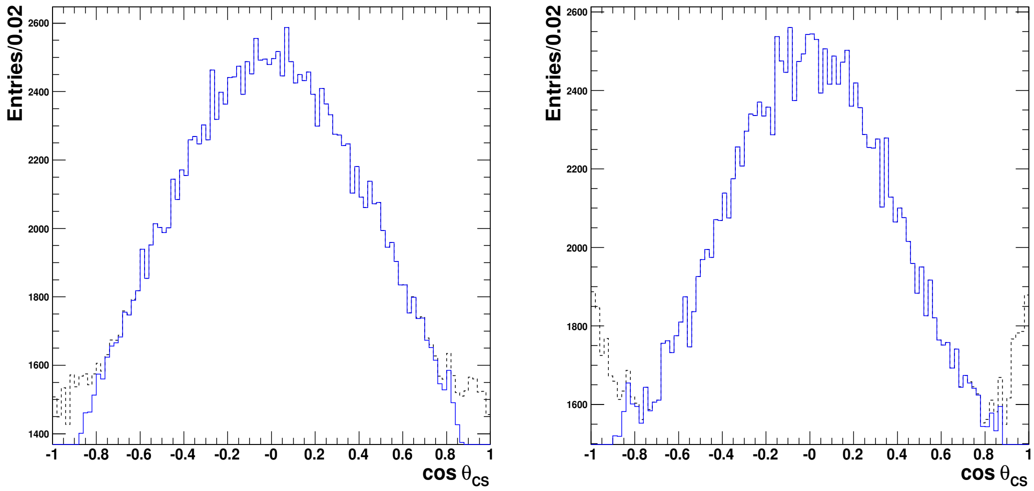

Figure 2: Histograms of $\cos\theta^*_{\rm CS}$ of individual events from the generated (dashed) and accepted (solid) samples of G* for $M_{\ell\ell} = 2.0\,$TeV (left) and $2.5\,$TeV (right).

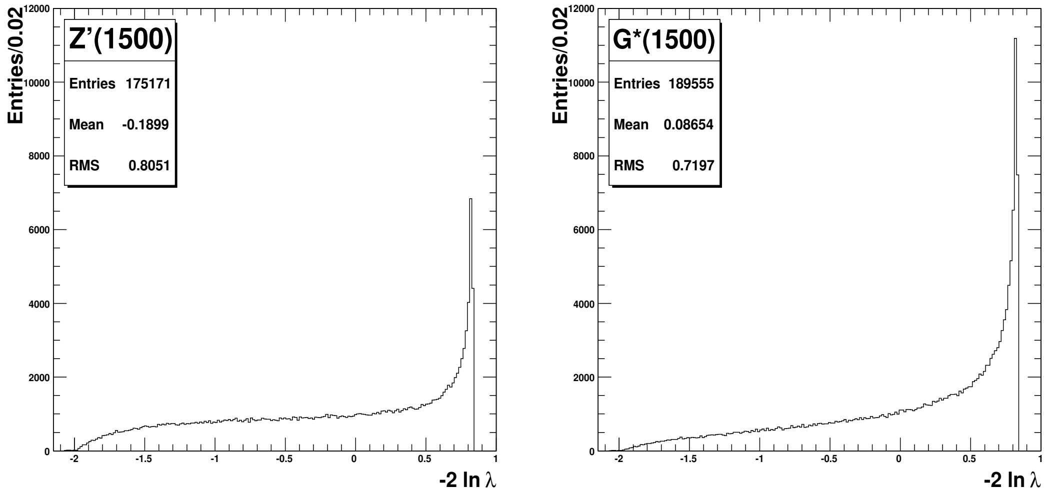

Figure 3: Histograms of $-2\ln\lambda = 2\ln\mathcal{L}(H_B) - 2\ln\mathcal{L}(H_A)$ for individual events from the accepted samples generated according to H_A (Z$'$) and H_B (G*) at $M_{\ell\ell} = 1.5\,$TeV.

the sum of $-2\ln\lambda$ for each of the selected events. Large numbers of experiments are simulated. In the spirit of the statistical bootstrap, the selection of N events is performed with replacement, i.e., an event selected from the sample is put back into the sample to be randomly selected again.

Figure 4 contains the resulting histograms for $N = 50$ and 200 events per experiment. In view of the asymmetries in figure 3, the convergence of the sums toward gaussians is quite striking. Superimposed on the histograms are gaussian curves that are not fits to the histograms (except for overall normalization). Rather, the means and rms deviations are fixed according to the central limit theorem by scaling the means and the rms deviations of the respective event histograms in figure 3 by N and $\sqrt{N}$, respectively.

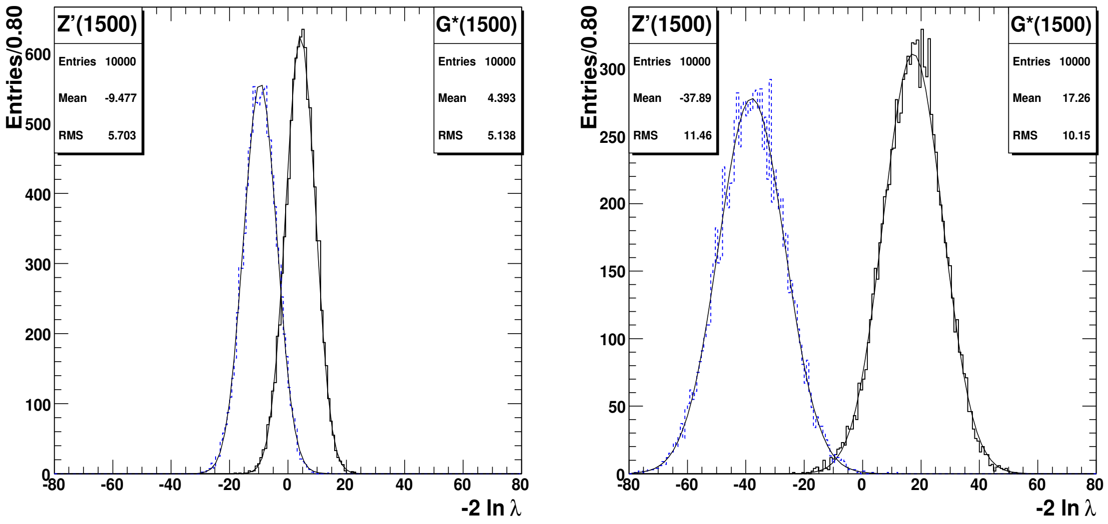

Figure 4: Histograms of $-2\ln\lambda$ for simulated experiments at $M_{\ell\ell} = 1.5\,\mathrm{TeV}$, each of which is a sum of N values of $-2\ln\lambda$ from Z$'$ and G* events in figure 3, for $N = 50$ (left) and $N = 200$ (right) events per experiment. The superimposed gaussian curves have means fixed at a factor of N greater than that of the events in figure 3, and rms deviations a factor of $\sqrt{N}$ greater.

From ensembles of experiments such as these, we determine, for each incremental number N of events per experiment, the value of λ^{cut} defining the critical region with $\alpha = \beta$ (tail probability for spin 1 above λ^{cut} is equal to tail probability for spin 2 below λ^{cut}), and the associated value of $\alpha = \beta$. Since gaussian tail probabilities are commonly expressed in equivalent number of standard deviations, we express α and β in terms of "σ" for *one*-tailed tests, e.g., for $\alpha = 0.159$ we report $\alpha = 1\sigma$, for $\alpha = 0.05$ we report $\alpha = 1.64\sigma$, and so on.

The values of α so obtained are displayed in a scatterplot vs N in figure 5. The expected scaling with $\sqrt{N}$ is evident; a fit to a power law over the range $N = 20$ to 280 gives an exponent of 0.5 to high precision, and is superimposed. We note that the results reported in table 4 of ref. [3] are not consistent with such scaling.

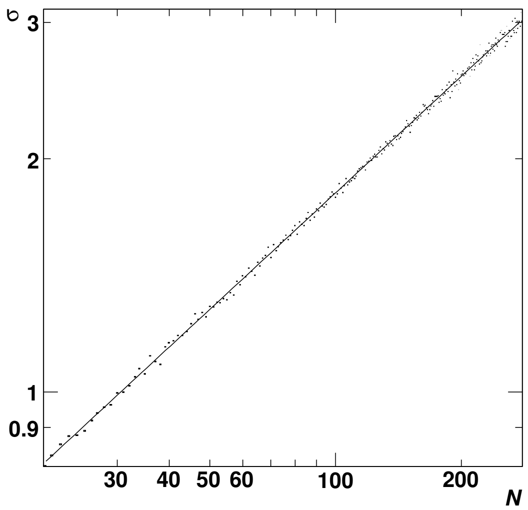

Figure 5: Scatterplot of α (expressed in number of σ) reached for the number of events N for simulated experiments at $M_{\ell\ell} = 1.5\,\mathrm{TeV}$, with superimposed power law fit discussed in the text.

Given the statistical fluctuations evident in figure 5, we report results using the fitted power law. The first column of table 2 contains these results for 1σ for the three mass values studied. The number of events N includes both signal and background events; however, the background contribution is less than 1% for all mass values considered.

JHEP11(2005)046

$M_{\ell\ell}$ (TeV)	N		N_p
	Tail integral	Gauss approx	Gauss approx
1.5	31	30.4 ± 0.6	8.5 ± 0.2
2.0	31	30.8 ± 0.6	8.2 ± 0.2
2.5	32	31.7 ± 0.6	8.1 ± 0.2

Table 2: Numbers of events N required to discriminate spin-1 and spin-2 hypotheses with $\alpha = \beta$ corresponding to 1σ (one-tailed) obtained by integration of the tails in the simulated experiments, and by the gaussian approximation in eq. 4.1. The last column contains numbers of events N_p for a test with α corresponding to 1σ allowing large β, as discussed in section 5. The number of events scales as the square of the number of σ desired, as discussed in the context of figure 5. The uncertainties shown are statistical only, from simple error propagation.

$M_{\ell\ell}$ (TeV)	$-2\ln\lambda^{\text{cut}}$	
	Tail integral	Gauss approx
1.5	−1.35	−1.34 ± 0.04
2.0	−2.37	−2.45 ± 0.06
2.5	−3.13	−3.24 ± 0.07

Table 3: Values of $-2\ln\lambda^{\text{cut}}$ found that discriminate spin-1 and spin-2 hypotheses with $\alpha = \beta$ corresponding to k=1 σ (one-tailed) obtained by integration of the tails in the simulated experiments, and by the gaussian approximation in eq. 4.2. The value of $-2\ln\lambda^{\text{cut}}$ scales as the square of the number of σ desired.

Since histograms such as those in figure 4 are so well-described as gaussian, the results we obtain can be well approximated by a simple calculation using the central limit theorem and the event histograms such as those in figure 3. For H_A (Z′) or H_B (G*), the mean in figure 3 is denoted by μ_i and the rms deviation by σ_i, for $i = A, B$. Then the experiments formed by sampling N events per experiment have $-2\ln\lambda$ approximately gaussian with mean $M_i = N\mu_i$ and rms deviation $S_i = \sqrt{N}\sigma_i$. In this approximation, the criterion $\alpha = \beta$ specifies that λ^{cut} is such that $-2\ln\lambda$ is the same number k of respective standard deviations away from each mean:

$$\begin{aligned} M_A + kS_A &= M_B - kS_B \\ N\mu_A + k\sqrt{N}\sigma_A &= N\mu_B - k\sqrt{N}\sigma_B \\ N &= k^2\left(\frac{\sigma_A+\sigma_B}{\mu_B-\mu_A}\right)^2 . \end{aligned} \tag{4.1}$$

The value for $-2\ln\lambda^{\text{cut}}$ is given by:

$$\begin{aligned} -2\ln\lambda^{\text{cut}} &= N\mu_A + k\sqrt{N}\sigma_A \\ &= N\mu_A + \sqrt{N}\,\frac{\mu_B-\mu_A}{\sigma_A+\sigma_B}\,\sqrt{N}\sigma_A \\ &= N\,\frac{\mu_B\sigma_A+\mu_A\sigma_B}{\sigma_A+\sigma_B} . \end{aligned} \tag{4.2}$$

Plugging in the appropriate values gives good agreement with the measured values of N and $-2\ln\lambda^{\text{cut}}$ as seen in table 2 and table 3, respectively.

JHEP11(2005)046

For the three masses in table 2, the number of events is independent of mass within the statistics generated. Of course, because the production cross section falls rather steeply with mass, the integrated luminosity needed for spin discrimination increases with mass. For RS gravitons, the numbers in table 2 correspond to less than $1\,\text{fb}^{-1}$ for $k/\bar{M}_{\text{Pl}} = 0.1$ at $M_{\ell\ell} = 1.5\,\text{TeV}$ and to less than $10\,\text{fb}^{-1}$ at $2.5\,\text{TeV}$ (for each dilepton decay mode). For Z′ production, the cross section is model-dependent; for the Sequential Standard Model Z′, the range is similar to that of RS gravitons with the above coupling.

5. Alternatives to the $\alpha = \beta$ criterion

In Neyman-Pearson hypothesis testing, the choice of the significance level α is a subjective choice that typically encodes something about the experimenter's relative degree of belief in the null and alternative hypotheses [14]. For example, if one is predisposed to interpret any new resonance as a Z′ and considers G* to be highly unlikely, then one may wish to alter the criterion of $\alpha = \beta$. For example, one might take the position attributed to R.A. Fisher [14] and consider the compatibility of the data with respect to H_A without reference to β. In this case, the number of events N_p to reach k_p standard deviations can be defined in analogy with eq. 4.1 as

$$\begin{aligned} M_B &= M_A + k_p S_A \\ N_p \mu_B &= N_p \mu_A + k_p \sqrt{N_p}\sigma_A \\ N_p &= \frac{k_p^2 \sigma_A^2}{(\mu_B - \mu_A)^2}\,. \end{aligned} \tag{5.1}$$

The values of N_p for k_p=1 σ so computed are in the last column of table 2. This approach requires fewer events for the same significance level ($N_p \approx N/4$ if $\sigma_A \approx \sigma_B$ and the use is also a one-tailed test) because the probability of a type-II error, β, is allowed to become large (0.5). For characterizing the spin discrimination for a new resonance discovered in forthcoming experiments, we believe that it is more meaningful to keep both type-I and type-II errors under control.

In the philosophy in which one only considers α, once data is in hand, one quotes the p-value, which is the smallest value of α for which H_A would be rejected based on that data set. (This is the common analog of the procedure recommended in section 3.2.) Even in this case, however, one should quote the power against H_B when such a well-specified alternative exists as is the case here (unlike, for example, typical goodness of fit tests).

6. Discrimination from spin 0

While the motivation of discriminating Z′ from G* has focused studies on discriminating spin 1 from spin 2, another possibility to be considered is spin 0 (which is uniform in $\cos\theta^*$). For accepted spin-0 events, the probability density for $\cos\theta^*_{\text{CS}}$ is somewhat in between the mostly concave-upward function for spin 1 and the predominantly concave-downward function for spin 2 in figures 1 and 2. (As noted above, in figure 2, the sharp upturn in $P(\cos\theta^*)$ at large $|\cos\theta^*_{\text{CS}}|$ for spin 2 at higher $M_{\ell\ell}$ is suppressed by acceptance.)

$M_{\ell\ell}$ (TeV)	N		N_p
	Tail integral	Gauss approx	Gauss approx
1.5	108	108.3 ± 4.1	26.1 ± 1.0
2.0	103	102.4 ± 3.7	24.7 ± 0.9
2.5	99	99.0 ± 3.5	24.0 ± 0.8

Table 4: Numbers of events N required to discriminate spin-0 and spin-1 hypotheses with $\alpha = \beta$ corresponding to k=1 σ (one-tailed) obtained by integration of the tails in the simulated experiments, and by the gaussian approximation in eq. 4.1. The last column contains the number of events N_p for a test with α corresponding to k_p=1 σ allowing large β. The number of events needed scales as the square of the number of σ desired, as discussed in the context of figure 5.

$M_{\ell\ell}$ (TeV)	N		N_p
	Tail integral	Gauss approx	Gauss approx
1.5	141	141.2 ± 6.0	38.0 ± 1.6
2.0	151	150.9 ± 6.6	38.9 ± 1.7
2.5	155	154.8 ± 6.8	38.9 ± 1.7

Table 5: Numbers of events N required to discriminate spin-0 and spin-2 hypotheses with $\alpha = \beta$ corresponding to k=1 σ (one-tailed) obtained by integration of the tails in the simulated experiments, and by the gaussian approximation in eq. 4.1. The last column contains the number of events N_p for a test with α corresponding to k_p=1 σ allowing large β. The number of events needed scales as the square of the number of σ desired, as discussed in the context of figure 5.

Thus, for example, one can ask how well spin 2 can be distinguished from spin 0, again based on the angular distributions alone.

Rather than using a custom production model for a spin-0 resonance, we simulated the uniform $P(\cos\theta^*)$ characteristic of generator-level spin-0 decay (whose particle is here denoted by S^0) by using a simple hit-or-miss Monte Carlo technique starting from the spin-1 Z′ sample. I.e., for each event in the Z′ sample, we obtained a random number R from a random-number generator whose output is uniform in the range 0 to 1, and put the event in the "spin 0" sample if and only if $R(1+\cos^2\theta^*) \leq 1$. The resulting unweighted sample is uniform in $\cos\theta^*$, and then modified by acceptance at large $|\cos\theta^*_{\mathrm{CS}}|$, as figure 6 shows. The sample has distributions in $p_{\mathrm{T}}^{\mathrm{dil}}$ and rapidity y^{dil} characteristic of Z′ production, which mildly affects the acceptance; however the conclusions remain the same if one fakes a spin-0 sample starting from a spin-2 sample. (Of course, if one wishes to use $p_{\mathrm{T}}^{\mathrm{dil}}$ and y^{dil} directly to discriminate spins, one needs to have available trusted models for their distributions.)

We then evaluate the likelihood of the spin-0 hypothesis and construct the likelihood ratio λ for spin-0 vs spin-1 and spin-0 vs spin-2 hypotheses for the relevant event samples. Tables 4 and 5 give the results for the number of events needed to test the two pairs of hypotheses.

As expected, discriminating either spin 1 or spin 2 from spin 0 requires significantly more events than discriminating spin 2 from spin 1.

JHEP11(2005)046

7. Conclusion

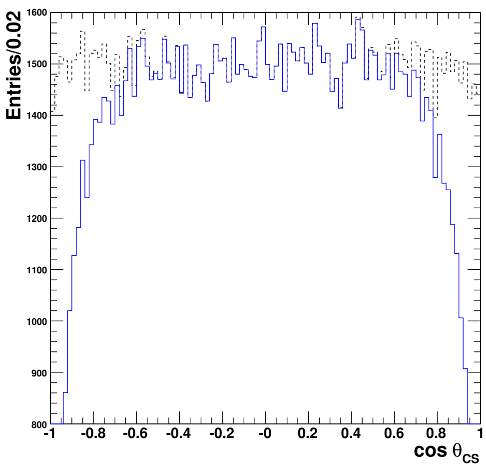

Figure 6: Angular distribution of the "spin-0" sample, produced by the hit-or-miss M.C. technique described in the text, for the generated (dashed) and accepted (solid) samples.

The conclusion of these studies is one of caution: even in the context of generator-level toy studies, the interpretation of a likelihood ratio of simple hypotheses requires some care. Indications are that, for the numbers of events required for spin discrimination at the LHC, $-2\ln(\mathcal{L}(H_A)/\mathcal{L}(H_B))$ for an experiment is well-approximated by gaussians, whatever the true spin. If one treats two spin hypotheses symmetrically, so as to keep both type-I and type-II errors under control, the number of events required is significantly greater (by about a factor of 4 for the examples above) than with the same type-I error fraction but with type-II errors uncontrolled; thus one must be clear about which test one is describing.

As this paper is limited primarily to comments on the statistical technique using only a crude simulation of acceptance, the numerical results are illustrative only and will be refined within the LHC collaborations using realistic simulation and reconstruction programs. Almost certainly, one will want to add more observables into the likelihood function, after checking that the simulation is faithful to the physics. The general procedure described in this paper for interpreting the likelihood ratios remains the same. Inclusion of systematic uncertainties requires additional care, appealing to the variety of techniques in the statistical literature for handling nuisance parameters.

References

[1] E.g., M. Cvetič and P. Langacker, *Implications of abelian extended gauge structures from string models*, *Phys. Rev.* **D 54** (1996) 3570; *New gauge bosons from string models*, *Mod. Phys. Lett.* **A 11** (1996) 1247 [hep-ph/9602424].

[2] L. Randall and R. Sundrum, *A large mass hierarchy from a small extra dimension*, *Phys. Rev. Lett.* **83** (1999) 3370 [hep-ph/9905221].

[3] B.C. Allanach, K. Odagiri, M.A. Parker and B.R. Webber, *Searching for narrow graviton resonances with the ATLAS detector at the large hadron collider*, *JHEP* **09** (2000) 019 [hep-ph/0006114].

[4] To facilitate comparison with ref. [3], we use the HERWIG event generator: G. Corcella et al., *Herwig 6: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, *JHEP* **01** (2001) 010 [hep-ph/0011363].

[5] P. Traczyk, private communication. For earlier work using another statistic, see C. Collard, M.-C. Lemaire, P. Traczyk, G. Wrochna, *Prospects for study of Randall-Sundrum gravitons in the CMS experiment*, CMS Note 2002/050.

[6] M. Dittmar, *Neutral current interference in the TeV region: the experimental sensitivity at the LHC*, *Phys. Rev.* **D 55** (1997) 161 [hep-ex/9606002].

[7] CMS collaboration results, Z′ *forward-backward asymmetry and model distinguishability*, http://cmsdoc.cern.ch/cms/PRS/results/susybsm/susybsm.html.

[8] D.H. Perkins, *Introduction to High Energy Physics*, 4th ed. Cambridge University Press, Cambridge 2000.

[9] P. Burikham, T. Figy and T. Han, *TeV-scale string resonances at hadron colliders*, *Phys. Rev.* **D 71** (2005) 016005 [hep-ph/0411094].

[10] The authors of ref. [3] express their likelihood function L in terms of fractions of *accepted* events from each process, which perhaps obscures the point that the overall normalization of L is irrelevant to the simple hypothesis test.

[11] J.C. Collins and D.E. Soper, *Angular distribution of dileptons in high-energy hadron collisions*, *Phys. Rev.* **D 16** (1977) 2219.

[12] The two hypotheses are usually denoted by H_0 and H_1. In this paper, they are denoted by H_A and H_B, respectively, in order to avoid potential confusion with spins 0, 1, and 2.

[13] J. Neyman and E.S. Pearson, *On the problem of the most efficient tests of statistical hypotheses*, *Phil. Trans. R. Soc. Series A* **231** (1933) 289, reprinted in *Breakthroughs in statistics*, ed. by S. Kotz and N. Johnson, Vol. 1, Springer-Verlag, 1992;
see also discussions in G. Upton, I. Cook, *Oxford dictionary of statistics*, Oxford University Press, Oxford 2002 and ref. [14].

[14] A. Stuart, K. Ord, and S. Arnold, *Kendall's advanced theory of statistics*, Volume 2A, 6th ed., Arnold, London 1999 and earlier editions by Kendall and Stuart.

[15] R. Bartoszyński and M. Niewiadomska-Bugaj, *Probability and statistical Inference*, Wiley 1996.

[16] J.O. Berger, L.D. Brown, and R.L. Wolpert, *A unified conditional frequentist and bayesian test for fixed and sequential simple hypothesis testing*, *Ann. of Stat.* **22** (1994) 1787.

[17] S.S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, *Annals of Math. Stat.* **9** (1938) 60.

[18] For example, in section 4 of ref. [3], a threshold of 1.92 on $-\ln\mathcal{L}$ is used for 95% confidence, i.e., in our notation $\lambda^{\rm cut} = 3.84$. 3.84 is the correct value of $\Delta(-2\ln\mathcal{L})$ *with respect to the best-fit value* in order to obtain an approximate 95% C.L. confidence interval on the parameter in a one-parameter fit.

[19] J.O. Berger, *Could Fisher, Jeffreys and Neyman have agreed on testing?*, *Stat. Sci.* **18** (2003) 1. See also commentary following by other statisticians.

[20] G. Altarelli, B. Mele and M. Ruiz-Altaba, *Searching for new heavy vector bosons in $p\bar{p}$ colliders*, *Z. Physik* **C 45** (1989) 109.

[21] R. Cousins, J. Mumford, V. Valuev, *Detection of* Z′ *gauge bosons in the dimuon decay mode in CMS*, CMS Note 2005/002.