This Dissertation

entitled

SEARCH FOR LEPTON FLAVOR VIOLATING DECAYS

OF HIGGS BOSONS

WITH THE CMS EXPERIMENT

typeset with NDdiss2$_\varepsilon$ v3.2017.2 (2017/05/09) on April 26, 2019 for

Nabarun Dev

SEARCH FOR LEPTON FLAVOR VIOLATING DECAYS
OF HIGGS BOSONS
WITH THE CMS EXPERIMENT

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy
in
Physics

by

Nabarun Dev

Colin Philip Jessop, Director

Graduate Program in Physics
Notre Dame, Indiana
April 2019

SEARCH FOR LEPTON FLAVOR VIOLATING DECAYS OF HIGGS BOSONS WITH THE CMS EXPERIMENT

Abstract

by

Nabarun Dev

DEDICATED TO

To my family

CONTENTS

FIGURES

TABLES

PREFACE

Long time ago in a galaxy far far away....(preface is optional)

ACKNOWLEDGMENTS

I would like to acknowledge the light side of the force, Master Kenobi and Grand Master Yoda.

SYMBOLS

c	speed of light
m	mass
e	elementary charge
E	energy

CHAPTER 1

SIGNAL EXTRACTION AND SYSTEMATIC UNCERTAINTIES

1.1 Introduction

The analysis is in essence a sophisticated counting experiment. The presence of a signal is indicated by an excess of events over the predicted background, in the distribution of a signal variable. For our analyses the signal variables are collinear mass or BDT output, as described in Sections **??** and **??**. Given that there are several uncertainties, both experimental and theoretical and also due to the innate randomness in the process, it is possible that an excess is observed when there is no signal. So, when an excess is observed, a p-value which represents the probability that the excess is due to statistical fluctuations is computed. A very low p-value is taken to indicate that the excess corresponds to an observed signal and not merely a statistical fluctuation. Conversely, if no excess is observed (upper exclusion) limits are set on the product of branching fraction and production cross-section. A 95% CL (confidence level) is taken as a requirement for ruling out a signal at or above a certain value known i.e. upper exclusion limit. The first part of this chapter describes the statistical methods used, that very closely follow the procedure used for LHC Higgs boson search and described in [1].

Several sources of systematic uncertainties need to be considered when making the above measurement. The sources of these uncertainties can be theoretical, experimental or purely statistical in nature. Further, they can effect only the overall scale of the distributions (used to make the measurement), or effect there shape i.e.

change the scale differently in each bin of the distribution. All the uncertainties used in the analyses and their sources are described in the secon part of this chapter.

1.2 Statistical methods for signal extraction

In the following section, the expected signal event yields are denoted by s, and backgrounds by b. The parameter μ that appears is the signal strength modifier, which changes the signal production cross-sections of all the production mechansims by exactly the same scale μ.

1.2.1 Likelihood function

The Poisson distribution is an appropriate model for n, the number of times an event occurs in an interval if the following assumptions are true [2].

- The occurrence of one event does not affect the probability that a second event will occur. That is, events occur independently.
- The rate at which events occur is constant. The rate cannot be higher in some intervals and lower in other intervals. This rate is the average number of events in the interval. λ.
- Two events cannot occur at exactly the same instant; instead, at each very small sub-interval exactly one event either occurs or does not occur.

The poisson probablity of distribution is then given by:

$$P(n_events) = \frac{e^{-\lambda}\lambda^{n}}{n!} \tag{1.1}$$

For a counting experiments such as ours, the above conditions approximately hold. The expected number of events is $\mu \cdot s + b$. The likelihood function $\mathcal{L}(data|\mu)$ is then given by:

$$\mathcal{L}(data|\mu) = \prod_{i=1}^{bins} \frac{(\mu \cdot s_i + b_i)^{n_i}}{n_i!} e^{-\mu \cdot s_i - b_i} \tag{1.2}$$

,where n_i is the number of events observed in the bin i of the distribution, and s_i and b_i are expected number of signal and background events in that bin respectively.

1.2.2 Treatment of systematic uncertainties

All systematic uncertainties are handled by introducing them as nuisance parameters. Nuisance parameters are parameters that influence the model but are not of interest in our measurement, e.g., if we are interested in knowing only the mean of a population that is expected to be distributed as a gaussian, the standard deviation becomes a nuisance parameter for the model that we fit. In our experiment, the nuisance parameters are embedded into the likelihood function. In order for the likelihood function to have a clean factorised form [1], all sources of uncertainties considered are considered 100%-corrrelated or uncorrelated. If an uncertainty is partially correlated, it is either separated into 100%-corrrelated or uncorrelated components, or considered 100%-corrrelated or uncorrelated, depending on whichever is a more conservative estimate. The full suite of nuisance parameters is represented as θ. These effect the expected signal and backgeound yields which are now represented as $s(\theta)$ and $b(\theta)$. Each component of θ is associated with a default value $\tilde{\theta}$, reflecting our degree of belief on the real value of θ. The pdf (probablity distribution function) $\rho(\theta|\tilde{\theta})$ can then be interpreted as a posterior distribution from measurements of $\tilde{\theta}$. Using Bayes' theorem:

$$\rho(\theta|\tilde{\theta}) = \rho(\tilde{\theta}|\theta) \cdot \pi_\theta(\theta), \tag{1.3}$$

where the priors $\pi_\theta(\theta)$ are taken as flat distributions representing no prior knowledge of θ. This reformulation allows us to use the pdf of $\tilde{\theta}$ instead, i.e. $\rho(\tilde{\theta}|\theta)$ to directly constrain the likelihood of the measurement. The likelihood function after the introduction of systematic uncertainties now becomes:

$$\mathcal{L}(data|\mu,\theta) = Poisson(data|\mu \cdot s(\theta) + b(\theta)) \cdot \rho(\tilde{\theta}|\theta) \tag{1.4}$$

Systematic uncerainties that effect only the overall scale of the distributions, correspond to a mutliplicative factor in the signal and/or background yields, and are described by log-normal pdfs. Log-normal pdfs are characterised by the width κ, and are well-suited for positively valued observables. The log-normal distribution looks like:

$$\rho(\theta|\tilde{\theta}) = \frac{1}{\sqrt{2\pi}\ \ln(\kappa)} \exp\ (\frac{\ln(\theta/\tilde{\theta})^2}{2(\ln\ \kappa)^2})\frac{1}{\theta} \tag{1.5}$$

Systematic uncertainties that effect the scale of the distribution differently in each been have the effect of altering its shape along with its scale. Such uncertainties are called shape uncertainties [3], and are modeled using a linear extrapolation method [4]. In practice, two alternate distributions obtained by varying the nuisance by ±1 standard deviation are used, and a parameter is added to the likelihood that smoothly interpolates between these shapes.

1.2.3 Calculation of exclusion limits

The CL$_\mathrm{s}$ method [5–7] is used to set upper exclusion limits when no excess of data over background is observed. The test statistic used generally for hypothesis testing in searches at the LHC, uses profiling of nuisances as described above, and is based on the likelihood ratio [8], which by the Neyman-Pearson lemma is known as the most powerful discriminator. This is denoted by $\tilde{q_\mu}$, and is given by:

$$\tilde{q_\mu} = -2\ \ln\frac{\mathcal{L}(\mathrm{data}|\mu,\hat{\theta_\mu})}{\mathcal{L}(\mathrm{data}|\hat{\mu},\hat{\theta})},\ \text{with } 0 \leq \mu \leq \hat{\mu} \tag{1.6}$$

,where $\hat{\theta_\mu}$ refers to the conditional maximum likelihood estimators of θ, i.e. the set of nuisances parameters that maximize the likelihood for a given signal strength μ, while $\hat{\mu}$ and $\hat{\theta}$ refer to the global maximum likelihood estimators for μ and θ. The lower constraint on $\hat{\mu}$ i.e., $\hat{\mu} \geq 0$ ensures that the signal rate cannot be negative, while the upper constraint that $\hat{\mu}$, which is the global maximum value, cannot be less than the value of μ under consideration is imposed to guarantee that upward fluctuations of data such that $\hat{\mu} \geq \mu$ are not considered as evidence against the signal hypothesis,i.e., a signal of strength μ.

Now, using equation 1.6, the observed value of the test statistic,$\tilde{q_\mu}^{obs}$, is calculated for the signal strength μ. Also, maximum likelihood estimators for the nuisance parameters, for the background-only($\mu = 0$) and signal-plus-background(current $\mu > 0$ under consideration) hypotheses are calculated. They are denoted by $\hat{\theta_0}^{obs}$ and $\hat{\theta_\mu}^{obs}$ respectively, and are used to generate toy Monte carlo pseudo-datasets. These pseudo datasets are used to construct pdfs, using equation 1.6, of test statistics $f(\tilde{q_\mu}|0, \hat{\theta_0}^{obs})$ and $f(\tilde{q_\mu}|\mu, \hat{\theta_\mu}^{obs})$ by treating them as they were real data. Example of these distributions are shown in Fig. 1.1.

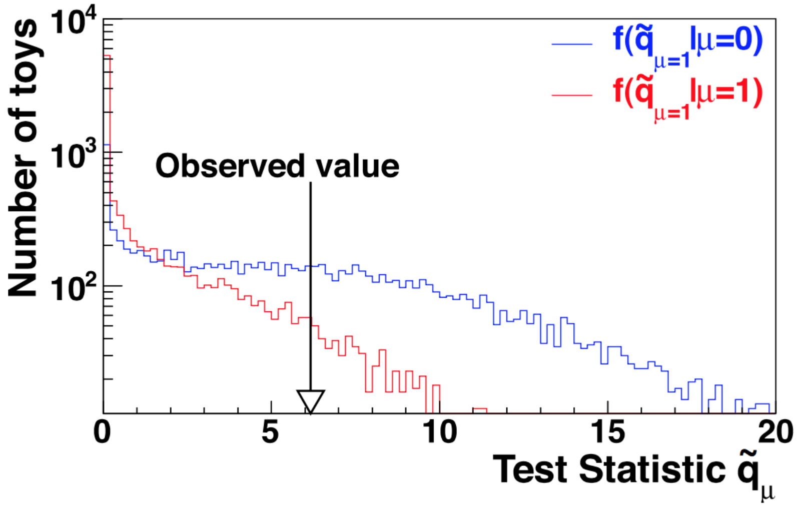

Figure 1.1: Test statistic distributions for ensembles of pseudo-data generated for signal-plus-background (red) and background-only (blue) hypotheses. [1]

Having constructed the above pdfs, it is now possible to calculate the probabilities of the observations under both hypotheses. The first quantity that we calculate is:

$$p_\mu = P(\tilde{q_\mu} \geq \tilde{q_\mu}^{obs}|\text{signal-plus-background}) = \int_{\tilde{q_\mu}^{obs}}^{\text{inf}} f(\tilde{q_\mu}|\mu, \hat{\theta}_\mu^{obs}) d\tilde{q_\mu} \tag{1.7}$$

The above quantity corresponds to $\mathrm{CL_{s+b}}$ and measures the incompatibilty of data with signal-plus-background hypothesis. This quantity alone is not adequate for hypothesis testing in situations when the signal is so small that both hypotheses are compatible with the observation and a downward fluctuation of the background can lead to an inference of signal.

The second quantity we calculate is:

$$1 - p_b = P(\tilde{q_\mu} \geq \tilde{q_\mu}^{obs}|\text{background-only}) = \int_{\tilde{q_\mu}^{obs}}^{\text{inf}} f(\tilde{q_\mu}|0, \hat{\theta}_0^{obs}) d\tilde{q_\mu} \tag{1.8}$$

This quantity corresponds to $\mathrm{CL_b}$ and measures the incompatibilty of data with the background. The incompatibilty of the data with background-only hypothesis alone doesn't tell us that it is indeed compatible with the signal, and so is not considered a good test of the signal hypothesis.

The ratio of the two quatities referred to as $\mathrm{CL_s}$ [5–7] helps deal with both situations above well, and is given by:

$$\mathrm{CL_s} = \frac{p_\mu}{1 - p_b} \tag{1.9}$$

The 95% CL is then arrived at by iterating over μ until we have $\mathrm{CL_s} = 0.05$. And the amount of signal or above, given by that μ, denoted as $\mu^{95\%CL}$, is said to be excluded at 95% CL.

1.2.4 Median expected Limits

Upper exclusion limits calculated using toy datasets of background-only expectation, are called expected limits. A large set of background-only pseudo-data is generated, and $\mathrm{CL_s}$ and $\mu^{95\%CL}$ is calculated for each of them. The median expected limit is calculated by integrating over this distribution until the 50% quantile is reached. The $\pm 1\sigma$ bands are calculated similary by integrating the distribution to the appropiate quantiles are reached. The calculation of median expected limits does not involve using the observed data and hence can be calculated when the analyses is blinded to prevent experimenter's bias (as mentioned in Section **??**). This can be use to maximize the sensitivity of the search, as described in Sections **??** and **??**. A more stringent(lower) median limit corresponds to a more sensitive search.

1.2.5 Quantifying an excess of events

In case an excess of data over background is observed, it is necessary to make sure beyond a reasonable doubt that the excess is not merely a fluctuation. This is quantified using the background-only p-value, which is the probability for the background to fluctuate and give an excess of events as large or larger than that observed. The same test statistic as equation 1.6 is used with the signal stength set to 0 to correspond to the background-only hypothesis:

$$\tilde{q_0} = -2\,\ln \frac{\mathcal{L}(\mathrm{data}|0,\hat{\theta}_0)}{\mathcal{L}(\mathrm{data}|\hat{\mu},\hat{\theta})},\ \text{with } 0 \leq \hat{\mu} \tag{1.10}$$

The constraint on $\hat{\mu}$ being greater than 0 is required so that a deficit of events in observed data is not interpreted in the same manner as we would an excess. In other words a departure from the background hypothesis in the form of deficit of events is not considered in favour of the signal hypothesis. Following the same procedure as calculation of observed limits 1.2.3 and generating pseudo-data, the distribution

$f(\tilde{q_0}|0, \hat{\theta_0}^{obs})$ is constructed. The p-value is then given by:

$$p_0 = P(\tilde{q_0} \geq \tilde{q_0}^{obs}) = \int_{\tilde{q_0}^{obs}}^{\inf} f(\tilde{q_0}|0, \hat{\theta_0}^{obs}) d\tilde{q_0} \tag{1.11}$$

The p-value can be converted to significance $\mathcal{Z}_0$, which is an equivalent way of quantifying an excess and is related to the p-value by the following:

$$p_0 = \int_{\mathcal{Z}_0}^{\inf} \frac{1}{\sqrt{2\pi}} exp(-x^2/2) dx \tag{1.12}$$

Broadly, the signficance corresponds to how far into the tail of the distribution (i.e., away from the most probable value), assuming background hypothesis, the test statistic value corresponding to the observed data lies. The farther it is, the less likely it is to have been a fluctuation. The conventional standard in high energy physics to be able to claim observation of a process is a significane of 5σ, which corresponds to a p-value of 2.8×10^{-7}.

1.2.6 Experminetal uncertainties

1.2.7 Signal extraction

1.3 Heavy Higgs Analysis

1.3.1 Theoretical uncertainties

1.3.2 Experminetal uncertainties

1.3.3 Signal extraction

CHAPTER 2

INTERPRETATION OF RESULTS

CHAPTER 3

CONCLUSION

APPENDIX A

BOOSTED DECISION TREES

A.1 Introduction

BIBLIOGRAPHY

1. CMS and ATLAS Collaborations. Procedure for the lhc higgs boson search combination in summer 2011. Technical report, August 2011.

2. Wikipedia. Poisson distribution. Website. `https://en.wikipedia.org/wiki/Poisson_distribution`.

3. J. S. Conway. Incorporating nuisance parameters in likelihoods for multisource spectra. In *PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN, Geneva*, January 2011.

4. A. L. Read. Linear interpolation of histograms. *Nucl. Instrum. Meth*, 425, April 1999.

5. A. L. Read. Presentation of search results: The $\mathrm{CL_s}$ technique. *Journal of Physics G*, 28, September 2002.

6. A. L. Read. Modified frequentist analysis of search results (The $\mathrm{CL_s}$ method). In *Workshop on confidence limits, CERN, Geneva, Switzerland, 17-18 Jan 2000*, August 2000.

7. Thomas Junk. Confidence level computation for combining searches with small statistics. *Nuclear Instruments and Methods A*, 434, September 1999.

8. Eilam Gross Glen Cowan, Kyle Cranmer and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *European Physics Journal C*, 71, February 2011.

This document was prepared & typeset with pdfLaTeX*, and formatted with* NDdiss2$_\varepsilon$ *classfile (v3.2017.2[2017/05/09])*