TECHNISCHE UNIVERSITÄT DRESDEN

Application of the Profile Likelihood Method to constrain the parameter space of MSSM Higgs bosons' cross section

Bachelor-Arbeit
zur Erlangung des Hochschulgrades
Bachelor of Science
im Bachelor-Studiengang Physik

vorgelegt von

Christian Lorenz
geboren am 01.09.1989 in Dresden

Institut für Kern- und Teilchenphysik
Fachrichtung Physik
Fakultät Mathematik und Naturwissenschaften
Technische Universität Dresden

2012

Eingereicht am 21.Mai 2012

1. Gutachter: Prof. M. Kobel
2. Gutachter: Prof. K. Zuber

Abstract

The subject of this thesis is a basic introduction to statistical methods which are frequently used in various scientific fields. Concepts of parameter estimation and hypothesis tests are discussed and finally applied to particle physics.
An upper limit on cross-section times branching fraction of a generic MSSM Higgs boson ϕ as a function of the Higgs boson mass m_A in the m_h^{max} scenario is presented. The considered di-τ final state is $\tau\tau \to \mu + \tau_{had}$. The data, recorded by the ATLAS-detector in 2011 at the LHC at a center-of-mass energy of 7 TeV, corresponds to an integrated luminosity of 4.7 fb^{-1}.

Kurzdarstellung

Diese Arbeit gibt zunächst eine grundlegenede Einführung in die wichtigsten statistischen Methoden zur Datenanalyse: Parameterschätzung und Hypothesentests. Abschließend wird dies auf eine Problemstellung der Teilchen Physik angewandt.
Mithilfe der vom ATLAS-Detektor bei einer Schwerpunktsenergie von 7 TeV aufgenomen Daten aus dem Jahr 2011 wird ein oberes Limit auf den Wechselwirkungsquerschnitt der MSSM-Higgsboson Produktion im m_h^{max} Szenario berechnet. Die Datenmenge entspricht einer integrieten Luminosität von $4,7$ fb^{-1}. Betrachtet wird lediglich der $\tau\tau \to \mu + \tau_{had}$ Endzustand.

Contents

1. Introduction

In modern science the concept of statistics became indispensable as the observed effects are on a more and more smaller scale. The development of the theory of probability has made an enormous progress in the last century. Important persons are for example Andrej Nikolaevič Kolmogorov who stated an axiomatic formulation of probability in 1933, Jerzy Neyman and Egon Pearson who i.a. developed the concept of hypothesis tests further.
This theoretical foundations are well understood but in practice scientists are often unsure about the accurate usage of statistical tools and the interpretation of their results. This work provides an overview of the foundation of some commonly used methods of parameter estimation in chapter 2. In chapter 3 they are illustrated using the currently at CERN (and elsewhere) used frameworks ROOT and ROOFIT.
The theoretical achievements of chapter 2 are used in chapter 4 to obtain an upper limit on the cross section of a generic MSSM Higgs boson.

2. Theoretical Foundations

2.1. Interpretation of Probability: bayesian and frequentist

Since in statistics the concept of probability plays an important role, it is useful to mention two different interpretations of probability: the bayesian and the frequentistic interpretation.
The ("classical") *frequentistic* probability defines the probability of some event "X" by the fraction of experiments with outcome "X" in the limit of an infinitely often repeated experiment.
The ("subjective") *bayesian* probability is a "degree of believe" which is rated on a scale from 0 to 1.
It is important to distinguish between these two probabilities, as both definitions lead to the same mathematical concept of probability, but to different statements.
To get a better idea of the difference between bayesian and frequentistic probability imagine the following situation:
Two men are playing a game. In the game one rolls a die; Player 1 wins if the outcome is "6" and Player 2 wins if the outcome is "1". Let's assume they played half an hour, rolled the die 100 times and finally Player 2 won 54 times and Player 1 only 11 times.

Now Player 1 could think of a loaded die. If he is a "Frequentist" he would expect, assuming the die is not loaded, every number from one to six appearing an equal number of times. Based on that he can think of how likely it is to get this very outcome of the game, if they would repeat the game again and again. If it is to unlikely he may say the die is loaded.

If he is a "Bayesian" he does basically the same, but furthermore he takes into account some prior probability that the die is loaded at all. If he is playing with a friend he might think: "Well, Player 2 is my best friend and I am sure he doesn't use a loaded die. He was just lucky". But if he goes into a casino, plays the same game not with a friend but with a croupier and gets the same outcome as before, his thought could be like: "I do not trust those Casinos. They are likely to use loaded dice!". He combines his prior probability with the outcome and makes then a statement about the probability of a loaded die. But this statement depends on the prior probability, hence it is subjective.

To summarize this, a Frequentist makes a statement about the probability of an experiment's outcome, assuming a hypothesis. If the probability of an actual outcome is below a certain value he rejects the hypothesis.
A Bayesian "updates" his prior probability for a hypothesis on the basis of an experiment's outcome. He rejects the hypothesis if its probability is below a certain value.
In science both interpretations are used. The bayesian approach is useful to involve prior knowledge but there are also *non informative prior probabilities* that have the less impact on the final result. The choice which approach to use should always be based on the actual problem.
In this thesis the focus is on the frequentistic approach.

2.2. Parameter Estimation

In physics one often has theoretical models derived from theory that describe physical phenomena. But the model may depend on free parameters μ that cannot be derived from theory. Nevertheless, from the model it is possible to predict the expected distribution of the measured observable x_m in the limit of a large sample space, assuming different values for the parameters μ. This distribution is called *probability density function (pdf)* of x : $P(x|\mu)$. To check whether the model is useful to describe nature and to make inferences on the free parameters μ, measurements from experiments are "compared" with the possible hypothesis $P(x|\mu_h)$ via statistical methods. Moreover it happens that the pdf depends not only on μ but also on parameters that take into account experimental uncertainties. As the experimenter is not really interested in the true values of these parameters, they are called nuisance parameters and often denoted as θ. The pdf can then be written as: $P(x|\mu, \theta)$.
To keep things more clear we disregard the nuisance parameters in the first instance. In section 2.3 a possible method is described how to handle nuisance parameters.

2.2.1. Estimators

One way to make an inference about a parameter is to estimate the parameter's value itself (*point estimate*) by using an estimator. An estimator $\hat{\mu}$ in general is a function of a set measured values $\{x_i\}$ that gives an approximation of the true value μ' of a parameter μ. He is called consistent if he gives the true value μ' of μ in the limit of an infinite sample-space. Furthermore an estimator is unbiased (biased) if $E(\hat{\mu}) - \mu = 0$ ($\neq 0$), where $E(\hat{\mu})$ is the expected value of $\hat{\mu}$. Usually estimators are consistent and at least unbiased in the limit of a large sample-space.

An example is the *maximum likelihood estimator (MLE)*:
The MLE uses the *likelihood function* that is defined by:

$$\mathcal{L}(\mu|X) = \prod_{i=1}^{N} P_i(x_i|\mu) \tag{2.1}$$

The outcome $X = \{x_i\}$ are all measurements that are taken into account and are parameters, whereas μ is the variable of interest. Hence $\mathcal{L}$ is the probability to measure X if μ is realized in nature.
Note that the set $\{x_i\}$ does not necessarily contain measurements of the same quantity but of different quantities that all depend somehow on μ. Hence, they may have different pdf's P_i.
The MLE $\hat{\mu}$ for μ' is defined by the values of μ which maximise $\mathcal{L}$. The maximised likelihood function is called *maximum likelihood function.*

2.2.2. Confidence Intervals

Besides point estimates it is possible to do an *interval estimate* such as a *confidence interval.* This is an interval $[\mu_1, \mu_2]$ as a function of the measured value that contains the true (but unknown) μ' in a fraction α (often referred to as *Confidence Level CL*) of experiments:

$$p(\mu' \epsilon [\mu_1, \mu_2]) = \alpha\ . \tag{2.2}$$

First one constructs for each hypothesis μ_h an *acceptance region* (also *acceptance interval*) $[x_1, x_2]$ [1] that holds the condition:

$$\int_{x_1}^{x_2} P(x|\mu_h)\, dx = \alpha\ . \tag{2.3}$$

The criteria for the construction of acceptance intervals are arbitrary (except for condition 2.3), but in order to achieve consistent results the construction of all the acceptance intervals must be based on the same criteria and must not be influenced by the measured data afterwards.
The set of all acceptance intervals is called the *confidence belt.* After performing an experiment the desired confidence interval is the collection of all μ_h's whose

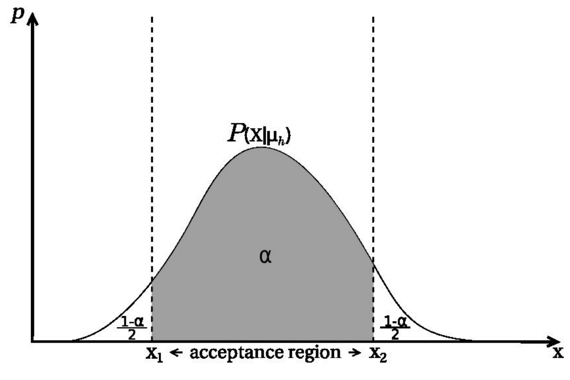


Figure 2.1.: scheme of an acceptance region

[1]the acceptance region does not necessarily have to be a connected interval

acceptance intervals include the measured value x_m.
By construction, the measured value x_m caused by μ' is in a fraction α an element of the acceptance region $[x_1, x_2]$ of μ' (in a frequentistic context). Thus the resulting confidence interval covers in a fraction α the true μ'.
Notice the construction's frequentistic character: it does not say the true value μ' lies with a probability of α inside that interval, but the interval covers μ' with the probability α.

2.2.3. Neyman Construction

The Neyman Construction [1] is a simple method for constructing confidence belts in order to estimate central confidence intervals or upper limits on the parameter of interest μ. This construction defines the acceptance regions for central confidence intervals (figure 2.1) by equation 2.4 and for upper limits by equation 2.5:

$$\int_{-\infty}^{x_1} P(x|\mu_h)\, dx = \int_{x_2}^{\infty} P(x|\mu_h)\, dx = \frac{1-\alpha}{2} \tag{2.4}$$

$$\int_{x_1}^{\infty} P(x|\mu_h)\, dx = \alpha\ . \tag{2.5}$$

Under strictly formal aspects this construction of confidence intervals would meet the demands. The problems come with boundary conditions that are physically motivated. In particle physics the parameter of interest μ represents the *signal strength* of the researched phenomena. In other words, it represents the number of detected events caused by the researched phenomena in addition to the number of events caused by other (well known) phenomena (also called background). Since the background-phenomena is assumed to describe the physics already in a good manner, the new phenomena can cause only more (not less!) events and thus μ is expected to be positive.
But due to statistical fluctuations it is not unlikely to observe a x_m which leads to a confidence interval that contains only (or at least some) negative μ's. An experimenter may not publish negative limits because they seem to be physically wrong and therefore he publishes just an upper limit (as defined in equation 2.5) or something else that sounds reliable. Nevertheless by publishing the "modified" confidence interval he implicitly uses different methods to construct one confidence belt (the one he chooses before and the other one he chooses after the outcome of the experiment - he is "flip-flopping"). This leads to acceptance intervals that cover less than the given CL α and the whole construction does not work any more.

Feldman & Cousins developed a method *Unified Approach* to choose the acceptance regions, that avoids the "flip-flopping" but gives account to physically motivated boundary conditions like $\mu \geqq 0$.

2.2.4. Unified Approach

The Unified Approach [2], developed by Feldman & Cousins, is an alternative method to the Neyman Construction. It uses the likelihood ratio R (2.6) to specify which possible outcomes X are an element of the acceptance interval.

$$\mathrm{R} = \frac{\mathcal{L}(\mu_h|X)}{\mathcal{L}(\hat{\mu}|X)} \tag{2.6}$$

It is the ratio of the likelihood function at μ_h (numerator) and the maximum likelihood function as described in section 2.2.1 (denominator). Evidently the relation $0 < \mathrm{R} \leq 1$ holds.
To illustrate the concept of the Unified Approach let us think of a counting experiment with the outcome of $X = n$ counts. Therefore the pdf is given by a Poisson distribution with mean ν:

$$\mathcal{P}(n|\nu) = \frac{\nu^n}{n!}\mathrm{e}^{-\nu} \,. \tag{2.7}$$

Applied to particle physics the mean is the sum of signal $\mu \cdot s$ and background b events with μ as signal strength. As already explained in section 2.2.3, the signal strength is assumed to be positive. Therefore, if the best estimator $\hat{\mu}$ appears to be negative the best physically allowed value of $\hat{\mu}$ is zero. To get the acceptance interval of hypothesis μ_h one constructs first for each n the likelihood ratio R; then one adds n with respect to the value of R in decreasing order to the acceptance interval until the sum of their probabilities $P(n|\mu)$ reaches the desired CL α.
For large n the confidence intervals are approximately a central limit as the Neyman construction would provide. But for small n the confidence intervals become an upper limit on μ. The construction for continuous pdf's and the issues that come due to the discreteness of the Poisson distribution are well described in the paper of Feldman & Cousins [2].

2.3. Profile likelihood

In the previous chapters nuisance parameters have been neglected. But when analysing data from an experiment such parameters can appear and often their number is somewhat large. Then the likelihood function depends in addition to the parameter of interest μ on the nuisance parameters: $\mathcal{L}(\mu, \theta|X)$.
A possible treatment of nuisance parameters is to maximize the likelihood function for each value of μ with respect to the nuisance parameters θ. The new likelihood function $\mathcal{L}(\mu, \hat{\theta}_\mu|X)$ is called *profile likelihood function*, as the nuisance parameters have been

"profiled out".
Notice that $\hat{\theta}_\mu$ is the best estimator for θ for a given μ. With that the likelihood ratio becomes:

$$R = \frac{\mathcal{L}(\mu_h, \hat{\theta}_{\mu_h}|X)}{\mathcal{L}(\hat{\mu}, \hat{\theta}_{\hat{\mu}}|X)} \,. \tag{2.8}$$

2.4. Hypothesis Testing and Test Statistics

Assume two different hypothesis H_0 (null hypothesis) and H_1 (alternative hypothesis), represented by two different pdf's $P_0(x)$ and $P_1(x)$ which may have some overlap (as shown in figure 2.2). H_0 is the hypothesis to be tested. Note that $P_0(x)$ and $P_1(x)$ can be mathematically completely different pdf's but more often $P_1(x)$ is the same pdf as $P_0(x)$ just with different parameters, e.g.
$P_0(x) = P(x|\mu_0)$ and $P_1(x) = P(x|\mu_1)$.

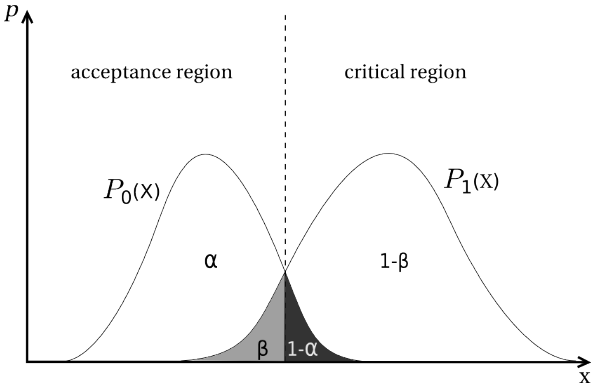


Figure 2.2.: scheme of hypothesis test's properties

To draw any conclusion from measured data (i.e. decide whether H_0 is true or not) it is necessary to define a critical region so, that if the measured value lies in this region, H_0 is rejected at the desired CL α. As one can see in figure 2.2 there is still a probability of $1-\alpha$ to reject H_0 although H_0 is true (this is called *error of the first kind*) and there is a probability of β to accept H_0 although H_0 is false (*error of the second kind*).

2.4.1. Test Statistic and Neyman Pearson Lemma

The power of a hypothesis test is defined by $1-\beta$, which is the probability to reject H_0 if H_0 is in fact false. Obviously it is desirable to increase the power of a test without decreasing the confidence level α.
Test statistics are a useful tool to achieve that. A test statistic is a function $t(X)$ that maps the measured data X to a single value. Hence, if the pdf $P_0(x)$ is known, it is possible to calculate the corresponding pdf of the test statistic $g_0(t)$.
After a measurement, the null hypothesis H_0 is rejected if the probability to obtain this or less likely data is smaller than $1-\alpha$. This probability is called the *p-value*. The Neyman Pearson Lemma [3] states that the most powerful test statistic is given by 2.9.

$$t(X) = \frac{\mathcal{L}_0(X)}{\mathcal{L}_1(X)} \tag{2.9}$$

Often one wants to test H_0 without having a particular alternative hypothesis H_1. In this case the maximum likelihood function (see section 2.2.1) is used instead of $\mathcal{L}_1$. Accordingly the not exclusively chosen alternative hypothesis H_1 becomes the hypothesis that fits best to the data. With this, equation 2.9 becomes the likelihood ratio as defined by equation 2.6 (respectively by equation 2.8 if nuisance parameters are involved). As already mentioned in section 2.2.4 the relation $0 < \mathrm{t}(X) \leq 1$ is evident. If $t(X)$ is close to 1 the hypothesis is more compatible with data.

Unfortunately, the calculation of $g_0(t)$ is often hard to do analytically and rather done with Monte-Carlo simulations. But those are very CPU intensive. Nevertheless, for some test statistics asymptotic formulae have been found.

2.4.2. Asymptotic Formulae

In this thesis the following test statistic is used:

$$t_\mu(X) = -2ln\left(\frac{\mathcal{L}(\mu, \hat{\theta}_\mu | X)}{\mathcal{L}(\hat{\mu}, \hat{\theta}_{\hat{\mu}} | X)}\right) . \tag{2.10}$$

It makes use of the Neyman Pearson Lemma 2.4.1 but transforms equation 2.9 in a way that $0 < t_\mu(X) < \infty$. If data and hypothesis are in good agreement then the value that comes out of the test statistic is close to 0. Note that μ is the parameter of interest and the test statistic depends on μ.
If we want to set an upper limit on μ, all $\mu < \hat{\mu}$ are regarded as part of the acceptance region. Thus, in this case it is reasonable to set $t_\mu = 0$. This leads to the test statistic q_μ used for upper limits:

$$q_\mu(X) = \begin{cases} -2ln\left(\frac{\mathcal{L}(\mu, \hat{\theta}_\mu | X)}{\mathcal{L}(\hat{\mu}, \hat{\theta}_{\hat{\mu}} | X)}\right) \\ 0 \end{cases} . \tag{2.11}$$

Cowan, Cranmer, Gross and Vitells found asymptotic formulae for the distribution of t_μ, q_μ and similar test statistics. The details of their work can be found in their paper [4]. Important for this thesis is formula 2.12 which gives the upper limit:

$$\mu_{up} = \hat{\mu} + \sigma\Phi^{-1}(\alpha) . \tag{2.12}$$

(Φ^{-1} is the inverse cumulative density function of the standard Gaussian)

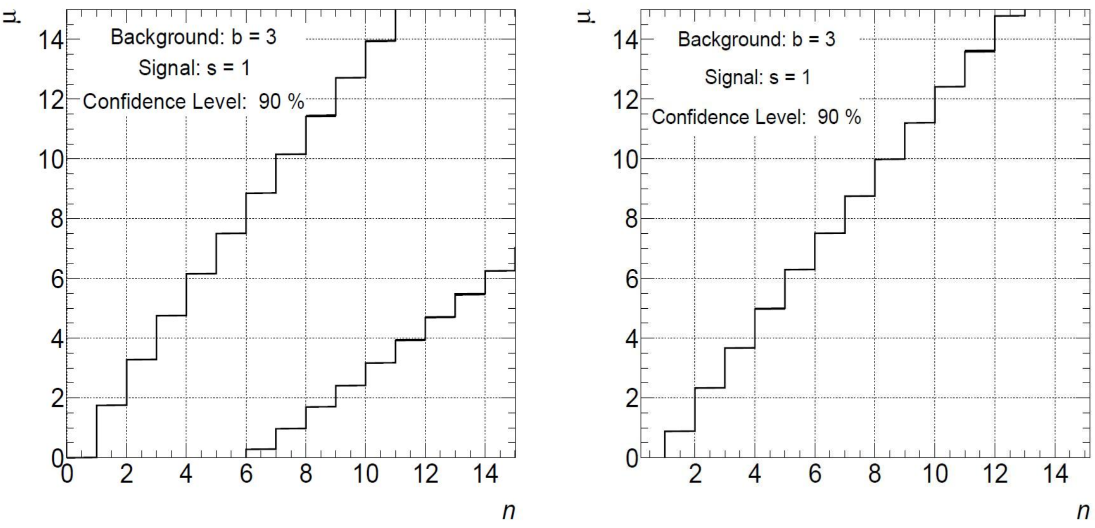
Summarizing, the maximum value of μ that is compatible with the data of a confidence level α can be determined by equation 2.12 using the MLE $\hat{\mu}$ and its standard variation σ.

3. Computational Implementation

All macros used to obtain the results of this thesis are written in the programming language C++ using the ROOT framework [5],in particular ROOFIT. In order to get familiar with the statistics, ROOT and ROOFIT I tried to reconstruct Feldman & Cousins' results they published in their paper [2].
The model (pdf) I used is a simple Poisson distribution with signal, background and signal strength μ (see equation 3.1) because it is similar to the model Feldman & Cousins used and to achieve the limit (see section 4) I use the same model with nuisance parameters.

$$\mathcal{P}(n|s\mu + b) = \frac{(s\mu + b)^n}{n!} \mathrm{e}^{-(s\mu+b)} \tag{3.1}$$

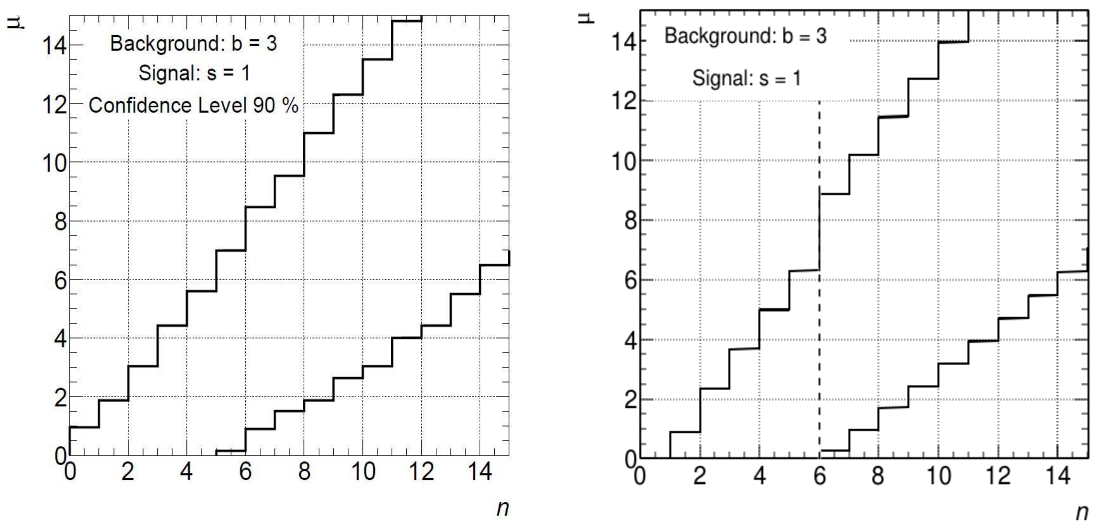
Using the Unified Approach to construct the confidence belt (figure 3.2a) leads to the exact same result as presented in the paper [2]. As well the results using the Neyman construction (see figure 3.1) are identical with Feldman & Cousins' results.
Computing the confidence belt via Neyman construction requires simulated toy data for each hypothesis to evaluate easily x_1 and x_2 according to equation 2.4 (or respectively equation 2.5). Those simulations are extensive in computational power but if not enough data have been simulated some artefacts could arise due to statistical fluctuations. An advantage of the unified approach is that there is no need for toy data and therefore needs less computational power.
In figure 3.2b the "flip-flopping" scenario (see also 2.2.3) is shown where the experimenter chooses the criteria for the acceptance region based on the data. If $n < 6$ he will publish an upper limit, if $n > 6$ he will publish a central limit. Then the acceptance regions of $0 \leq \mu \leq 8$ do not cover any more the 90% confidence level but less.



(a) central limit

(b) upper limit

Figure 3.1.: confidence belt using the Neyman Construction



(a) Unified Approach

(b) "flip-flopping" scenario

Figure 3.2.: (a) confidence belt using the Unified Approach; (b) the implicit confidence belt when the criteria are chosen based on the data

Furthermore, I intended to approve that the asymptotic formulae are correct and give a good approximation of the test statistic's distribution. 10^8 events have been simulated according to formula 3.1 in order to get also events that have a p-value of 10^{-6}. Unfortunately, it seems that something goes wrong when calculating the value of test statistic if this value is above a certain value.
I compared the distribution of the p-value which is shown in figure 3.3. For small values both, the asymptotic formula and the Monte Carlo simulation, match but for larger values some sort of discreteness occur. The same happens if the background b is set to very high numbers (10^7) where the discreteness of the Poisson distribution should have less impact. Hence this should not cause the issue. The reasons for this behaviour is not found yet. There are hints that the fit-algorithm or the event generation cause the problem.

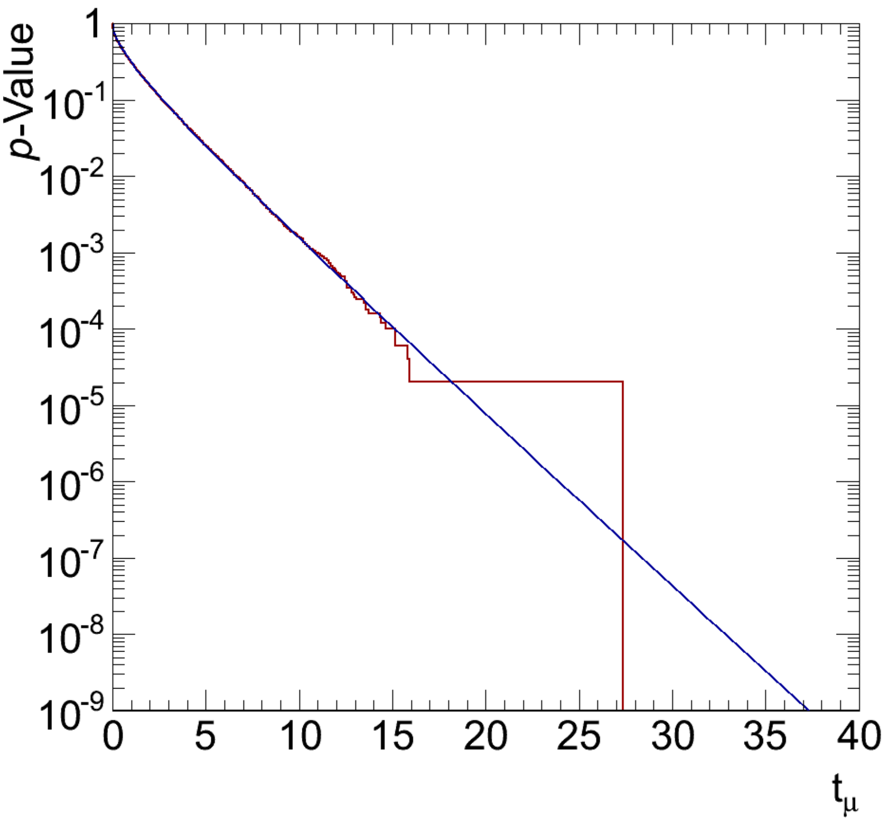


Figure 3.3.: comparison of asymptotic formula (blue) and Monte Carlo simulations (red)

4. Application to Particle Physics

Main objective of this thesis is the calculation of an upper limit on $\sigma \times BR(\phi \to \tau\tau)$ of a generic MSSM Higgs boson ϕ with the data obtained by the ATLAS experiment at the LHC.
This can be achieved by using the asymptotic formulae as discussed in section 2.4.2. Most important is formula 2.12. The pdf that is needed to evaluate the MLE $\hat{\mu}$ and its standard deviation is discussed in section 4.2.3. Uncertainties that are involved are described in section 4.2.2.

4.1. Physics Background

To describe the properties of particles and the properties of their interactions several Quantum Field Theories have been developed. Currently the Standard Model (SM) of Particle Physics is the best tested theory. Due to its concept it predicts the existence of the Higgs boson which is intensively searched for at the LHC at CERN with the detectors ATLAS and CMS.
Nevertheless the SM can not explain all phenomena that have been observed (e.g. Dark Matter). One possible suggestion is the introduction of supersymmetry. The Minimal Supersymmetric Standard Model is the smallest possible extension of the Standard Model in supersymmetric way.

4.1.1. Standard Model of Particle Physics (SM)

The SM is a gauge theory that describes the interactions between the "elementary" particles (and their antiparticles) via the electroweak interaction (a unified description of the electromagnetic and weak interaction) and the strong interaction. Interactions of particles are described by the exchange of gauge bosons.
The elementary particles can be classified according to their spin S into fermions with half-integer spin (see table 4.1) and bosons with integer spin (see table 4.2).
Fermions themselves are divided into leptons (Electron e, Myon μ, Tauon τ and their Neutrinos ν_x), which interact via the eletroweak interaction and quarks which interact in addition via strong interaction. Both, leptons and quarks, are divided into three generations where each generation contains two particles with different flavour (see table 4.1).

Table 4.1.: Fermions

	generation I flavour	generation II flavour	generation III flavour
leptons	e	μ	τ
	ν_e	ν_μ	ν_τ
quarks	u (up)	c (charm)	t (top)
	d (down)	s (strange)	b (bottom)

Quarks cannot be observed directly but they will always form a new particle by interaction with other quarks. The resulting particles are called *hadrons*. An example is the proton p that consists of two up quarks and one down quark.

The three forces are represented by three symmetry groups. If two particles are charged under the same symmetry group they can interact by exchanging a gauge boson of this symmetry group. The interactions and their gauge bosons are listed in table 4.2.

Table 4.2.: SM gauge bosons

interaction	gauge boson	symmetry group
electro weak	γ	U(1) $\otimes$ SU(2)
	$W^\pm$	U(1) $\otimes$ SU(2)
	Z	U(1) $\otimes$ SU(2)
strong	g	SU(3)

All the above listed particles have been observed in experiments, but in conflict with the SM, especially the $W^\pm$ and Z bosons have a mass, while the SM predicts them to be massless. Therefore the Higgs mechanism has been introduced.
It describes the mass of the bosons by interaction with a "Higgs field". The Higgs boson represents the excitation of this Higgs field. The SM already defines all properties of the Higgs boson except its mass.

4.1.2. Minimal Supersymmetric Standard Model (MSSM)

The MSSM embeds the SM into a larger symmetry group, where every particle and field of the SM gets a super symmetric partner. These partners are called *sleptons*, *squarks*, *gauginos* and *Higgsinos*. The interactions are already described by the SM symmetry groups.
This leads i.a. to five Higgs bosons whereof two are charged ($H^\pm$) and three are neutral

(A,h,H; also denoted as ϕ). Their properties are defined by two (unknown) parameters (at Born-level): $\tan\beta$ and the mass m_A of the Higgs boson A. In comparison, the decay modes of the SM Higgs boson into gauge bosons ($H_{SM} \to ZZ$ or $H_{SM} \to WW$) are dominant over a wide mass range whereas these decay modes are suppressed or, in case of the A boson, absent in the MSSM.
Decays into third generation fermions (see table 4.1) are dominant in the MSSM. Most interesting is the decay of a neutral Higgs boson into $\tau\tau$ final states. A τ lepton itself decays either into a lepton and two neutrinos or hadronically (denoted with τ_{had}). Decays into hadrons cause a bunch of various particles, called *jet*, which is then measured in the detector. Furthermore the production of Higgs bosons is classified into gluon-fusion ("ggA") and b-quark associated production ("bbA").

4.1.3. Large Hadron Collider and ATLAS Detector

The LHC (Large Hadron Collider), built by the European Organization for Nuclear Research (CERN), is the worlds largest particle collider. It is located near by Geneva, 100 m under ground in a tunnel of 27 km circumference. First a particle beam consisting of either protons or lead ions is accelerated in several pre-accelerators, then split into two opposing beams and finally injected into the LHC where the beams reach their maximum energy. [1]
The two beams intersect at four locations where the actual particle collisions take place. At these four locations the detectors ATLAS, CMS, ALICE and LHCb are established to identify and analyse the particles resulting from interactions of two particles.
Its aim is to explore new particles in order to test the predictions made by different theories of particle physics and to measure properties of some already well understood particles and interactions.
The probably most popular aim is the "hunt" for the Higgs boson.

The ATLAS detector ("A Toroidal LHC Apparatus") is especially designed to identify charged and neutral particles. It consists of basically four sub detectors: inner detector, electromagnetic calorimeter, hadronic calorimeters and muon chamber.
The inner Detector is of cylindrical shape with 7 m in length and a diameter of 2.3 m. Information for the reconstruction of charged particles' tracks are gathered using three different systems namely pixel detector, semiconductor tracker and transition radiation tracker. This environment is set up in a magnetic field of 2 T to bend the track of charged particles in order to reconstruct their momenta.
The next outer layer is the calorimeter system that measures the energy of photons, electrons and hadrons. Photons and electrons deposit their energy in the electromagnetic calorimeter by interacting with liquid argon (sensitive component) and lead (absorber). Hadrons pass the electromagnetic calorimeter and deposit their energy in the hadronic

[1] Currently the final energy of accelerated protons is 4 TeV, but it is planned to increase the energy per beam to 7 TeV

calorimeter that consists e.g. of scintillators and steal (absorber material) in the barrel part. The muon chambers are located outer most and provide an precise measurement of the momentum of Muons.
Neutrinos are not detectable since they interact rarely with matter.

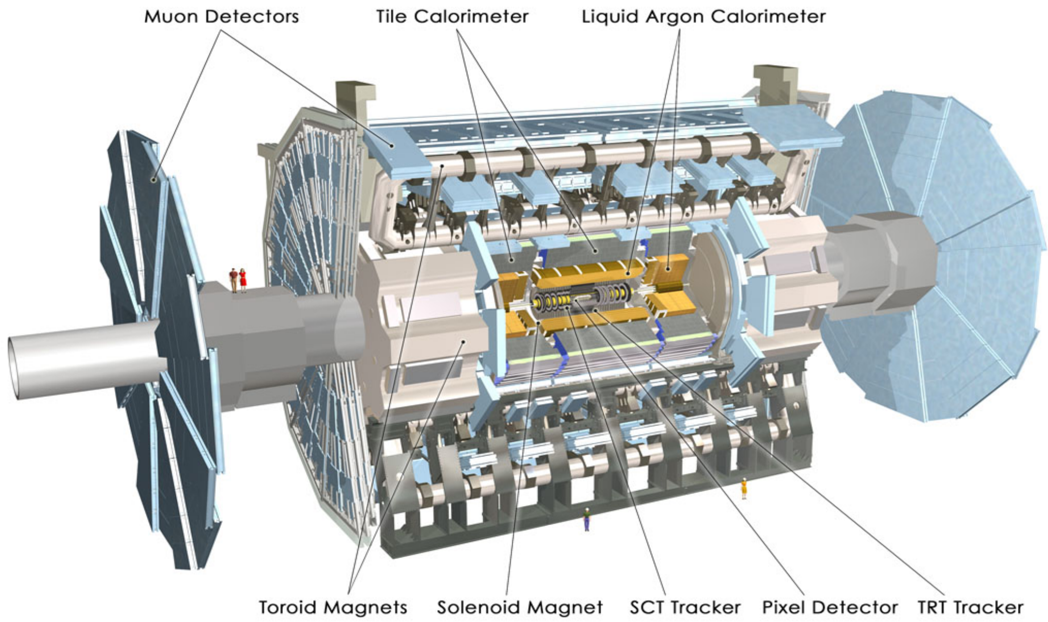


Figure 4.1.: The ATLAS detector [6]

At the LHC the rate of proton-proton interactions is roughly 40 MHz but the actual data storage rate is only 200 Hz. A trigger system of three trigger levels selects, based on predefined criteria, only those events that are interesting (e.g. for the Higgs boson search).

4.2. Limit on Cross Section of MSSM Higgs Boson

The upper limit is evaluated for 9 hypothetical masses m_A (model independent). Furthermore, gluon-fusion ("ggA") and b-quark associated production ("bbA") is studied separately.
The cross section of a certain signal process is defined by formula 4.1.

$$\sigma = \frac{N_{\text{signal}}}{\text{L}} \tag{4.1}$$

The luminosity L is measured directly whereas the number of signal events N_{signal} is the actual "test object".

4.2.1. Input data

The input data where provided by Stefanie Langrock [7] and in this section just a brief description of how the data have been obtained is given.
In this thesis the reviewed channel is $pp \to \phi \to \tau^+\tau^- \to \mu + hadron$, with data taken in 2011, corresponding to an integrated luminosity of $L = 4,7\,\text{fb}^{-1}$.
If $\tan\beta$ and m_A is given it is possible to derive the cross section for $\phi \to \tau\tau$. With that the hypothetical number of signal events is predicted.
The main background processes [2] are:

- W and Z boson production in association with jets
  $Z \to \tau\tau$ is an irreducible background as it has the same final states as $\phi \to \tau\tau$. Hence it is the process with largest impact on the background.
- multijet events (referred to as "QCD")
- $t\bar{t}$ and single t production

The event selection for $\tau\tau \to \mu + \tau_{had}$ final states requires a muon signature with high transversal momentum p_T and is done via multivariate methods (boosted decision tree (BDT).
Background estimations are mostly done by Monte Carlo simulations. The QCD background as well as the normalisation of the $W + jets$ background has been estimated from the data using control regions.
With these estimations some uncertainties come along that are briefly described in section 4.2.2.
The simulation of the expected signal events involves the cross sections for $\tan\beta = 20$ derived from theory: $\sigma_{theo} \times \text{BR}(\tau\tau \to lepton + \tau_{had})$.
These data are provided in form of histograms m_{viss} vs N_{b_i}. The visible mass m_{viss} is the reconstructed invariant mass of the visible τ decay products. For each hypothetical Higgs boson mass m_A the expected distribution of the reconstructed visible mass can be estimated. For the hypothesis test just an interval called *visible mass window* of one standard deviation centred around the Higgs bosons visible mass peak is considered (see table A.1), as a significant signal is expected only in this region.
Important to mention is that the observed event yield is below the background estimation.

[2] Processes, which have the same signature when detected, as the actually researched signal process $\phi \to \tau^+\tau^-$

4.2.2. Systematic Uncertainties

In table 4.3 all systematic uncertainties are listed and their effect on the different background sources is given in form of relative errors of 1 standard deviation[3]. The uncertainty of QCD background is treated differently as it is estimated from the data, not by MC-simulation. Hence the following systematic uncertainties influence only the background estimated by MC-simulation.

Luminosity Uncertainty of the luminosity measurement, which affects all background estimations in the same way.

Cross section The cross sections of the background processes is derived from theory and have some uncertainties that differs from process to process.

Muon smear The measurement of the transverse momentum p_T has just a limited resolution. This resolution differs between simulation and data and have to be corrected. With that an additional uncertainty appears which affects all backgrounds differently.

Muon scaling The identification efficiency of muons also differs between simulation and data. Therefore the Monte Carlo event yield is corrected by a scaling factor that has a certain uncertainty.

Muon p_T scale The p_T scale also differs between simulation and data and is corrected by a scaling factor that has a certain uncertainty.

Tau ID-efficiency The algorithm for τ identification cannot identify all τ particles or may identifies a τ although there is non. The uncertainty of the Tau-ID-algorithm's efficiency affects all background sources in the same manner.

Tau energy scale The Energy scale is corrected by a factor that is determined by comparing data and Monte-Carlo simulation and has a systematic uncertainty.

$W + jets$ **normalisation** After W-boson event selection the event yield differs between Monte-Carlo sample and data-sample for not yet resolved reasons. This is corrected by a scaling factor.

Statistical uncertainty As background estimation is an estimation of a poisson distributed variable, each background has some (uncorrelated) statistical uncertainty.

[3]asymmetric errors have been averaged

Table 4.3.: Systematic uncertainties $\epsilon_{i,j}$ of background estimation (relative errors in %)

Background	$Z \to \tau\tau$	$Z + jet$	$W + jet$	single t + $t\bar{t}$	QCD
Luminosity	3.9	3.9	3.9	3.9	3.9
Cross Section	5	5	5		
Muon Smear	0.28	0.77	0.85	0.13	
Muon Scaling	1.36	1.45	1.44	1.58	
Muon p_T	2.60	1.12	0.65	1.29	
Tau ES	4.48	2.78	7.51	8.53	
Tau-ID Efficiency	9.9	9.9	9.9	9.9	
W-boson scaling factor			9.88		
ABCD-method					13.20
Statistics	1.91	5.52	6.82	3.91	5.61

4.2.3. Model

The experiment is a counting experiment. Hence the model consists in the first instance of a poisson distribution with mean $\nu = \mu \mathrm{N}_{\text{signal}} + \mathrm{N}_{\text{background}} \equiv \mu s + b$.

$$\mathcal{P}(n|s\mu + b) = \frac{(s\mu + b)^n}{n!} \mathrm{e}^{s\mu+b} \tag{4.2}$$

The background consists of several processes as discussed in section 4.2.1, that themselves are subject to some measurements or estimations with uncertainties (see section 4.2.2). To illustrate how this is taken into account let us assume just two background processes b_1 and b_2 with only two uncertainties: luminosity L (correlated) and statistical uncertainty (uncorrelated). Then we write:

$$b = b_1(1 + \epsilon_{1,\mathrm{L}} \cdot \tilde{\theta}_{\mathrm{L}} + \epsilon_{1,stat} \cdot \tilde{\theta}_{1,stat}) + b_2(1 + \epsilon_{2,\mathrm{L}} \cdot \tilde{\theta}_{\mathrm{L}} + \epsilon_{2,stat} \cdot \tilde{\theta}_{2,stat}) \tag{4.3}$$

where $\epsilon_{i,j}$ denotes the relative uncertainty of b_i caused by the systematic uncertainty j (see table 4.3) and $\tilde{\theta}_{i,j}$ is the corresponding nuisance parameter. When moving $\tilde{\theta}_{i,j}$ to ± 1 the background b_i is varied by one standard deviation $b_i \cdot \epsilon_{i,j}$ of the systematic uncertainty j.

In this example three nuisance parameters are needed. Both backgrounds have different statistical uncertainties, that is why each variation is described by its own nuisance parameter $\tilde{\theta}_{i,stat}$. If the luminosity is varied, it affects both backgrounds in the same manner. Thus it is described by just one nuisance parameter $\tilde{\theta}_{\mathrm{L}}$.

To constrain the parameter space of the nuisance parameters, the Poisson (equation 4.2) is multiplied by one standard Gaussian $\mathcal{G}_{i,j}$ per nuisance parameter, referred to as constrained term:

$$\mathcal{G}_{i,j}(\tilde{\theta}_{i,j}, 0, 1) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{\tilde{\theta}_{i,j}}{2}) \; . \quad (4.4)$$

This treatment of background uncertainties is equivalent to the expectation, that the outcome of the background estimation follows a Gaussian with mean b_i and standard deviation $b_i \cdot \epsilon_{i,j}$ when the experiment is repeated infinitely often.

When denoting different backgrounds with i, nuisance parameters of correlated systematics with k and nuisance parameters of uncorrelated systematics with j, the final model can be written as:

$$\mathcal{L}(\mu, \tilde{\theta}|n) = \mathcal{P}(n|s\mu + b) \cdot \prod_{i,j} \mathcal{G}_{i,j}(\tilde{\theta}_{i,j}, 0, 1) \cdot \prod_{k} \mathcal{G}_k(\tilde{\theta}_k, 0, 1) \; . \quad (4.5)$$

The background b can be written as:

$$b = \sum_i b_i (1 + \sum_j (\epsilon_{i,j} \cdot \tilde{\theta}_{i,j}) + \sum_k (\epsilon_{i,k} \cdot \tilde{\theta}_k)) \; . \quad (4.6)$$

The whole likelihood function, the systematics $\epsilon_{i,j}$ and parameters are implemented using the RooWorkspace which is part of ROOFIT.

4.2.4. Limit

To obtain the upper limit for one hypothetical Higgs boson mass I used the asymptotic formulae (see section 2.4.2), in particular equation 2.12:

$$\mu_{up} = \hat{\mu} + \sigma \Phi^{-1}(\alpha) \; .$$

To get the MLE $\hat{\mu}$ and its standard deviation σ I fitted the likelihood function 4.5 to the obtained data using ROOFIT [4] . ROOFIT also calculates the covariance matrix. From that the standard deviation σ is deduced. The CL is set to $\alpha = 0.95 \mathrel{\widehat{=}} 95\%$ which leads to $\Phi^{-1}(\alpha) \simeq 1.64$.
The expected upper limit is the upper limit one would theoretically observe assuming the hypothetical cross section (hence the number of signal events N_{signal}) is true. Therefore one creates the so called *asimov dataset* by setting $n = s+b$ (which is equivalent to $\mu = 1$ that expresses the assumption of a true cross section). This *asimov dataset* n is now used as "measurement" and the same procedure is performed just as described above. To get the 1 and 2-σ error interval of the expected upper limit on simply adds $\pm N\sigma$:

[4] The default fitting tools MINUIT and MIGRAD are used

$$\mu_{up,exp\pm N\sigma} = \hat{\mu} + \sigma(\Phi^{-1}(\alpha) \pm N)\,.$$

The parameter μ can be transformed into a cross section via formula 4.7.

$$\sigma_{cs} = \mu \cdot \sigma_{theo} \tag{4.7}$$

By changing the hypothetical Higgs mass m_A the final limit plot with observed upper limit,expected upper limit and 1 and 2-sigma error band is created. Only difference between the calculation of two limits is the used data, as the visible mass window and the prediction of signal events changes with the Higgs mass.

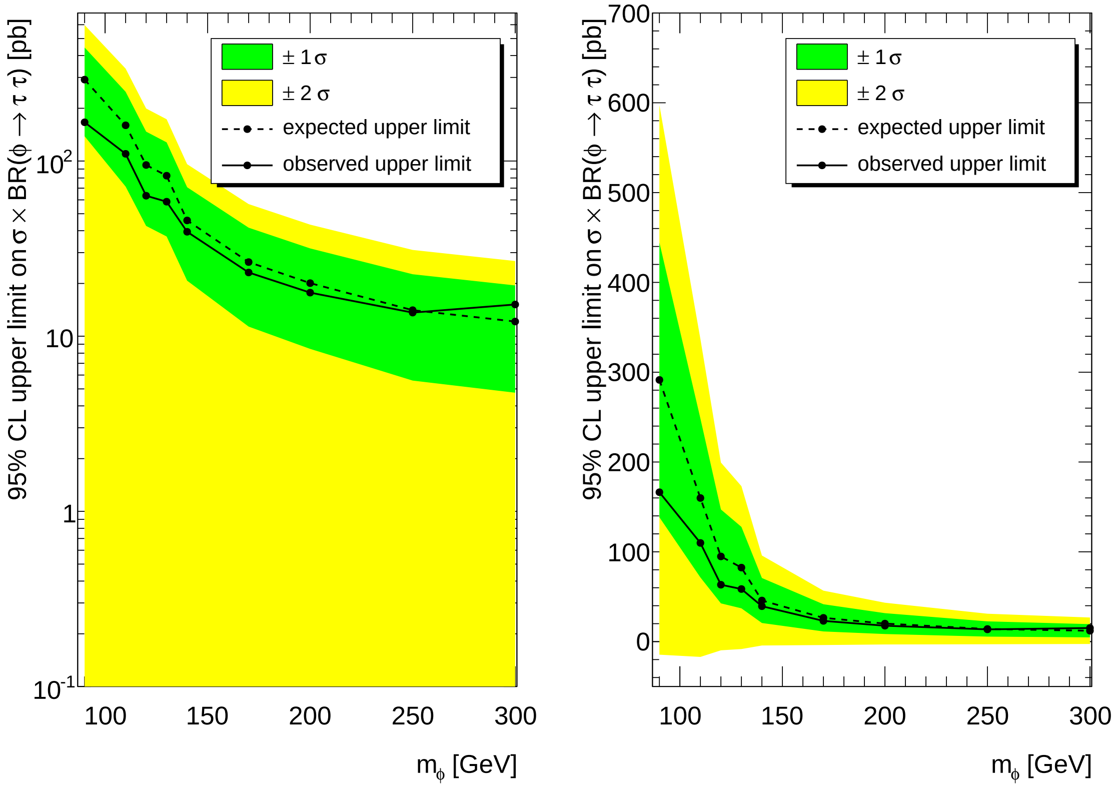


Figure 4.2.: Limit on gluon-fusion cross section

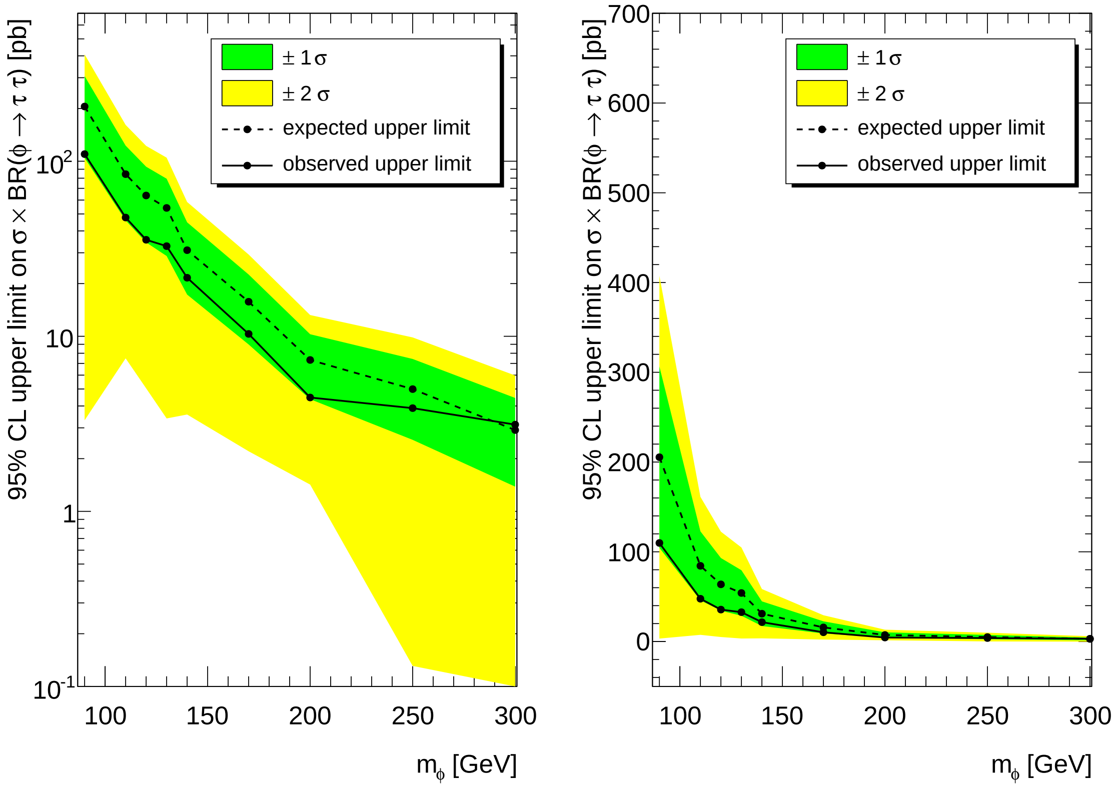


Figure 4.3.: Limit on b-quark associated cross section

5. Summary and Outlook

The search of the Higgs boson at CERN is one of the tasks in physics with top priority. To find the Higgs boson would be another prove of the actual understanding of how particle physics work. Statistical methods have been specially developed to find any possible signal out of the data.

In this thesis the foundation of these methods have been described. Emerging from the Neyman construction, that introduces the concept of acceptance regions, the Unified Approach introduces the likelihood ratio into the construction of acceptance regions. Finally, the profile likelihood allows us to deal with systematic uncertainties.
All the discussed methods have been implemented using the ROOT and ROOFIT framework which is the standard tool to analyse the data from the LHC.

Based on the analysis of $\tau\tau \to \mu + \tau_{had}$ final states [7] these statistical methods were used to obtain a model independent upper limit on $\sigma \times BR(\phi \to \tau\tau)$ of a generic Higgs boson. Correlations of different background event yields due to systematic uncertainties have been taken into account.

The sensitivity of this analysis is primarily limited by the Tau-ID efficiency's systematic uncertainty.
To get a more meaningful result more data is necessary or the systematic uncertainties, especially the Tau-ID efficiency's uncertainty, must be improved. More data can be achieved by recording more data or by analysing not only $\mu + \tau_{had}$ final states but also $e + \tau_{had}$ or even $\mu + e + 4\nu$ and $\tau_{had}\tau_{had}$ final states. Moreover it could be useful to take account to the asymmetry of the systematic uncertainties. But to modify the model in such a way more knowledge of the coherence of systematic uncertainties and their influence on background estimations is needed.
To assure that the algorithm works one could also repeat the calculation not using the asymptotic formulae but instead an approximation done with Monte-Carlo-simulations. First attempts have been made but as described in chapter 3 there are problems with the calculations of the Monte Carlo dataset.
Notice that this problem does not affect the presented upper limit as the test statistic has not to be calculated to evaluate the upper limit regarding equation 2.12.

A. Visible Mass Window

Table A.1.: 1σ visible mass window

m_A	Visible mass m_{viss} peak	lower bound	upper bound
90	62.5	52.1	72.9
110	72.5	61.46	83.54
120	75	62.29	87.71
130	77.2	62.46	91.94
140	82.5	67.23	97.73
170	95	76.84	113.16
200	105	84.49	125.51
250	115	87.58	142.42
300	125	91.3	158.7

Bibliography

[1] J. Neyman, *Outline of a theory of statistical estimation based on the classical theory of probability*, Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences **236** (1937) no. 767, 333–380.

[2] G. Feldman and R. Cousins, *Unified approach to the classical statistical analysis of small signals*, Physical Review D **57** (1998) no. 7, 3873.

[3] J. Neyman and E. Pearson, *On the problem of the most efficient tests of statistical hypotheses*, Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character **231** (1933) 289–337.

[4] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, The European Physical Journal C-Particles and Fields **71** (2011) no. 2, 1–19.

[5] R. Brun and F. Rademakers, *ROOT—an object oriented data analysis framework*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **389** (1997) no. 1, 81–86.

[6] G. A. e. a. The ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, Journal of Instrumentation **3** (2008) no. 08, S08003. `http://stacks.iop.org/1748-0221/3/i=08/a=S08003`.

[7] *Optimisation of the search for MSSM Higgs Bosons in the decay channel* $h/H/A \to \tau^+\tau^- \to \mu + \tau_{had} + \nu_\mu + 2\nu_\tau$ *with the ATLAS detector*, tech. rep., TU Dresden, Institut für Kern- und Teilchenphysik, Dresden.

Erklärung

Hiermit erkläre ich, dass ich diese Arbeit im Rahmen der Betreuung am Institut für Kern- und Teilchenphysik ohne unzulässige Hilfe Dritter verfasst und alle Quellen als solche gekennzeichnet habe.

Christian Lorenz

Dresden, Mai 2012