A study of the sPlot technique for application in the ATLAS B-Physics data analysis

Masterarbeit

zur Erlangung des akademischen Grades
Master of Science
(M.Sc.)

dem Fachbereich Physik der

Universität Siegen

vorgelegt von
Subrahmanya Siddardha Chelluri

September 2019

Contents

1 Introduction

The current theory of particle physics is summarized in the Standard Model [1–5] of particle physics. It is however widely considered insufficient and therefore various theories are being developed, in search of New Physics (NP), collectively known as Beyond the Standard Model (BSM) theories. So far none have been confirmed experimentally. The experiments at the Large Hadron Collider (LHC) [6] at CERN in Geneva are designed to probe matter with a center of mass energy of up to 14 TeV to address the open questions in particle physics. As not all the information about events in the collider experiments can be observed or known, the proof or disproof of a theoretical model can be achieved only on a purely statistical basis (by rejecting a hypothesis). But in order to obtain conclusive statements from statistical processes, sufficiently large data samples are needed. An important task in the research of High Energy Physics (HEP) is to analyse these large data sets to explore NP. Therefore statistical data analysis techniques play a crucial role in HEP research.

In particular, the study of the rare decays of B_s^0 and B^0 to two oppositely charged muons from the ATLAS experiment [7] involves MC data validation. For these studies the MC distributions are compared to the signal distributions of data for the Boosted Decision Tree (BDT) input variables. In order to obtain the signal distributions, background is statistically subtracted from the data for the BDT input variables using a sideband subtraction technique. However, the aim of this thesis is to study and explore an alternative technique, namely the sPlot technique [8], by taking an example with a possible application to the rare B meson decay analysis in the ATLAS experiment.

A brief review of the Standard Model is given in the second chapter. The LHC at CERN and the ATLAS detector are described in chapters three and four. The theory and data analysis of the rare decays $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$, the need for statistical data analysis in the environment of HEP and the motivation for studying the sPlot technique are discussed in chapter 5. The sPlot technique and its application on an example is studied in chapter 6. The sideband subtraction technique and in particular a new procedure to distinguish the signal and background distributions in case of a significant number of signal events leaking into the sidebands is discussed in chapter 7. In order to quantify the agreement between distributions obtained from the sPlot technique and the generated MC distributions, Maximum Likelihood Estimation (MLE) is performed using the TMinuit package in ROOT and the results are shown in chapter 8. Since the weighted MLE by TMinuit yields incorrect covariance matrix elements, a correction [9] has to be applied. The theory and implementation of the correction [9] to obtain the accurate covariance matrix elements for this particular example are described in chapter 9. Conclusions and further ideas are discussed in chapter 10.

2 Standard Model of particle physics

The Standard Model [1–5] of particle physics describes the elementary particles and their interactions via three (out of four) fundamental forces namely strong, weak and electromagnetic force. The fourth force, gravity, is not addressed in the SM. In the SM the electromagnetic interactions are described by Quantum Electrodynamics (QED) [10], strong interactions by Quantum Chromodynamics (QCD) [11] and weak interactions by Quantum Flavordynamics (QFD) [2–4]. All the above theories are within the framework of Quantum Field Theory (QFT). According to the SM the known universe (excluding dark matter and dark energy) consists of the elementary particles as shown in the Figure 2.1.

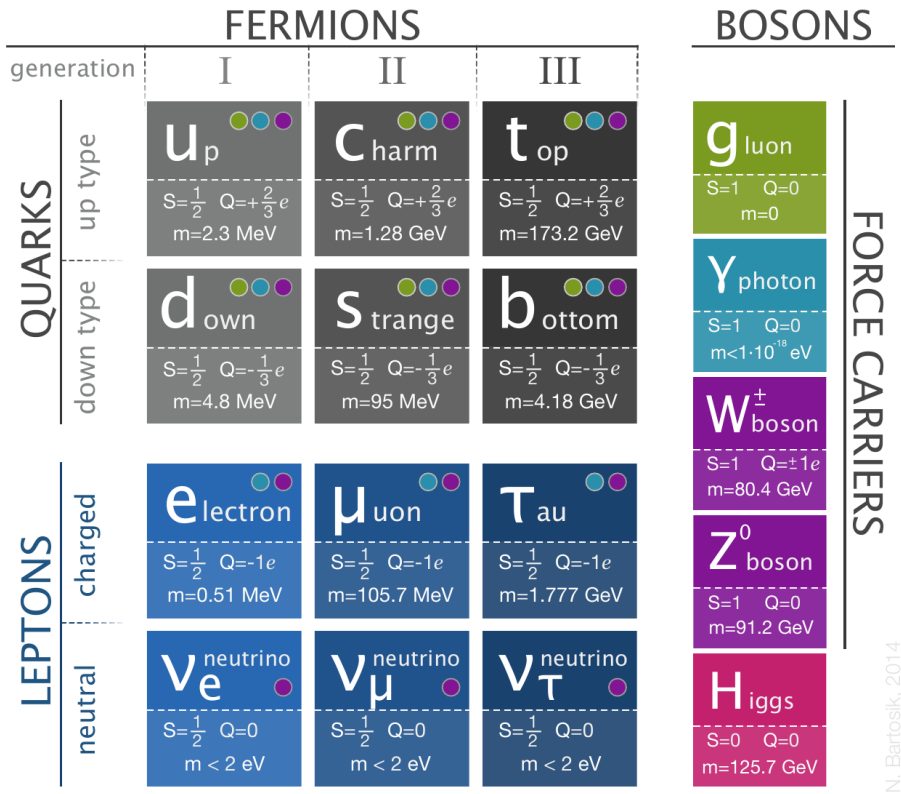


Figure 2.1: Standard Model of particle physics [12]

All elementary particles are classified into two groups based on the spin quantum number namely fermions and bosons. Fermions are particles with half-integer spin, follow

Fermi-Dirac statistics and obeys Pauli's exclusion principle. Bosons are integral spin particles and follow Bose-Einstein statistics.

The fermions in the Standard Model are the building blocks of matter (also anti-matter) and are classified into two groups namely quarks and leptons based on the *color* charge. Quarks are particles that possess color charge, hence undergo strong interactions where as leptons do not experience strong force. There are six quarks and six leptons ordered in three generations of two each as shown in figure 2.1. Only the first generation particles are stable. Unlike leptons, free quarks do not exist. One quark and one anti-quark form a meson and three quarks form a baryon. All the visible matter in the universe is made up of one lepton (electron) and two quarks (u and d).

According to QFT, the interactions of the elementary particles are mediated by gauge or vector bosons. Photon is the mediator for electromagnetic interaction, self interacting gluon for the strong interaction and the heavy W,Z bosons for the weak interaction.

The spin zero Higgs boson [13–16] corresponds to a scalar field namely the Higgs field which is responsible for the mass of the elementary particles.

The Standard Model is a successful theory but not complete. For instance, neutrinos are assumed to be massless according to the SM but experimentally it was found that neutrinos do posses mass. Moreover, gravity is not included in the SM and also it does not explain about dark matter and dark energy. There are a few theories like Super-Symmetry, String theory, Loop gravity etc proposed by various scientists but none has been confirmed experimentally.

3 Large Hadron Collider

3.1 Introduction to LHC

Large Hadron Collider (LHC) [6] is the world′s largest particle accelerator located at CERN (The European Organization for Nuclear Research) near Geneva, Switzerland. The LHC, in which protons are accelerated, is installed in an underground tunnel with circumference 26.7 km, at a depth of 40 to 170 m. Two beams of protons are accelerated simultaneously in opposite directions and are brought to collision at four interaction points inside the four main detectors ATLAS [17], CMS [18], LHCb [19] and ALICE [20] with an instantaneous luminosity of 10^{34} $cm^{-2}s^{-1}$. The design center of mass energy is 14 TeV. Each beam in the LHC consists of around 2556 bunches of protons with approximately $1.5x10^{11}$ protons per bunch.

The ATLAS and CMS detectors are general purpose detectors built to address all open questions of particle physics. LHCb is dedicated to B-physics and CP violation studies. ALICE is designed to study heavy-ion collisions at LHC.

3.2 Path of the protons

The protons, that collide in the LHC [6], begin their journey from a hydrogen bottle. Hydrogen atoms are ionized and the remaining protons are driven into the linear accelerator LINAC2 (LInear ACceleator) where the protons gain an energy of 50 MeV and then are accelerated up to 1.4 GeV in the Proton Synchroton Booster (PSB) and subsequently accelerated up to 26 GeV in the Proton Synchroton (PS). In the fourth stage they are accelerated up to 450 GeV in the Super Positron Synchroton (SPS) and finally accelerated to 6.5 TeV in the LHC, before they collide at the interaction points. It takes around 20 minutes for protons in the LHC to be accelerated to their maximum speed and energy. The CERN accelerator complex is shown in Figure 3.1.

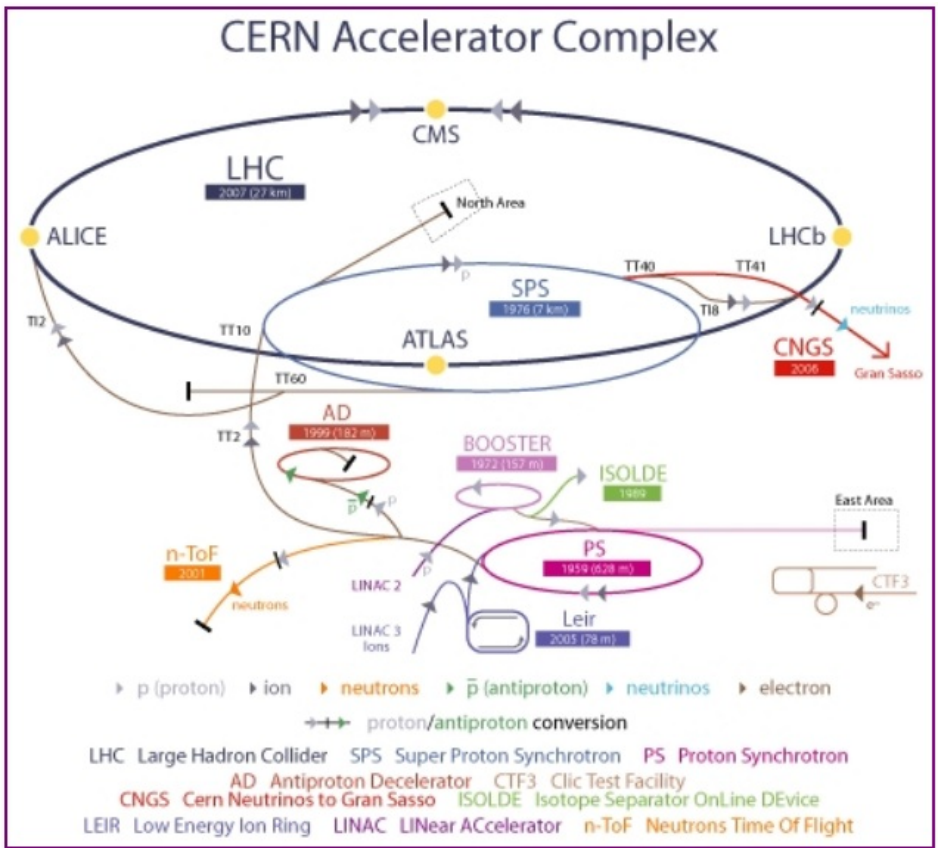


Figure 3.1: CERN accelerator complex [21]

4 ATLAS (A Torodial LHC AppartuS)

ATLAS [17] is one of the four major detectors at the LHC. It is cylindrical in shape measuring 46 m in length and 25 m in diameter and is located at point 1 of the LHC. It weighs around 7000 tons and is the largest detector ever constructed for a particle collider.

The subdetectors in ATLAS can be classified into two types, tracking detectors and calorimeters. Tracking detectors are used for measuring the paths of the charged particles and their vertices. The calorimeters are used for measuring energy deposits of the particles. ATLAS has several layers of various detectors arranged concentrically. The layout is shown in figure 4.1

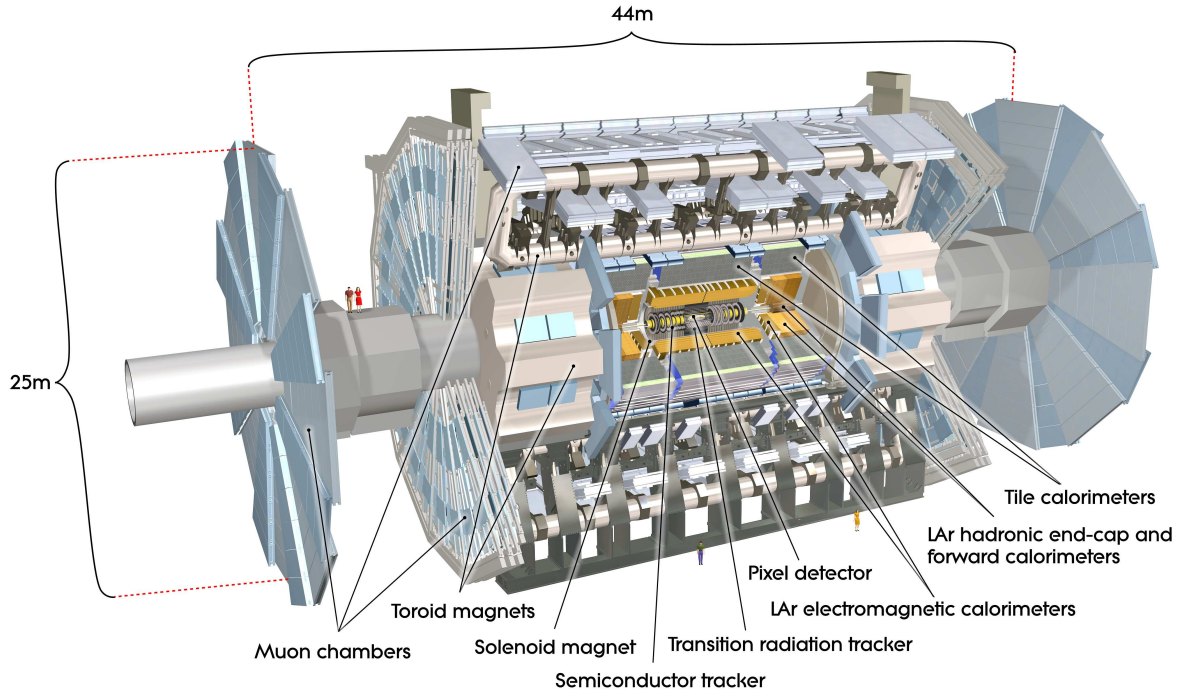


Figure 4.1: ATLAS detector layout [17]

4.1 ATLAS coordinate system

The nominal interaction point located at the center of the detector is defined as the origin, the direction of the beam defines the z-axis and the transverse plane defines the

$x-y$ plane, which is perpendicular to the z-axis. The positive y-axis is pointing upwards perpendicular to the plane of the LHC ring and the positive x-axis is defined as pointing from the interaction point to the center of the LHC ring. The observables defined in the transverse plane such as momentum and energy are usually named transverse momentum p_T and transverse energy E_T, etc. The side-A of the detector is defined as that with positive z and side-C is that with negative z [17]. The azimuthal angle ϕ is the angle measured in the $x-z$ plane ranging from 0 to 2π and the polar angle is the angle from the beam axis, which also defines pseudorapidity as:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \tag{4.1}$$

and the distances in the η-ϕ space are defined as :

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}\,. \tag{4.2}$$

4.2 Inner Detector system

The ATLAS Inner Detector (ID) system [17, 22] is surrounding the beam pipe measuring 7 m in length and 115 cm in radius. It provides high resolution momentum and vertex measurements. The ID has three subdetectors, namely the pixel detector, Semi-Conductor Tracker (SCT) and Transition Radiation Tracker (TRT). The ID system is surrounded by a solenoid magnet, which provides a field strength of 2 T. The cross-sectional view of the ID system and the schematic diagram of the subdetectors in the ID are shown in figures 4.2 and 4.3.

4.2.1 Pixel detector

The pixel detector [17,22,23] is the innermost detector and is crucial for tracking charged particles. It has four layers in the barrel region and three layers on each side of the endcaps and tracking is possible up to $|\eta| \leq 2.5$. The innermost Insertable B-Layer (IBL) [23] was installed during the shutdown of the LHC in 2013-2014.

There are a total of 1744 identical pixel modules (excluding the IBL layer). Each pixel module consists of a 250 μm thick silicon sensor, a flexible polyimide printed-circuit board and Front End (FE) electronics chips. The sensor and FE chips are connected by bump bonding. Each sensor has 47232 pixels. The nominal pixel size is 50x400 μm^2 (about 90 % of the pixels) and the remaining pixels measure about 50x600 μm^2. There are a total of 46080 readout channels per sensor.

The spatial resolution of the pixel detector is 10 μm in the transverse plane and 115 μm along the z-axis in both the barrel and the endcap region. It has very high granularity

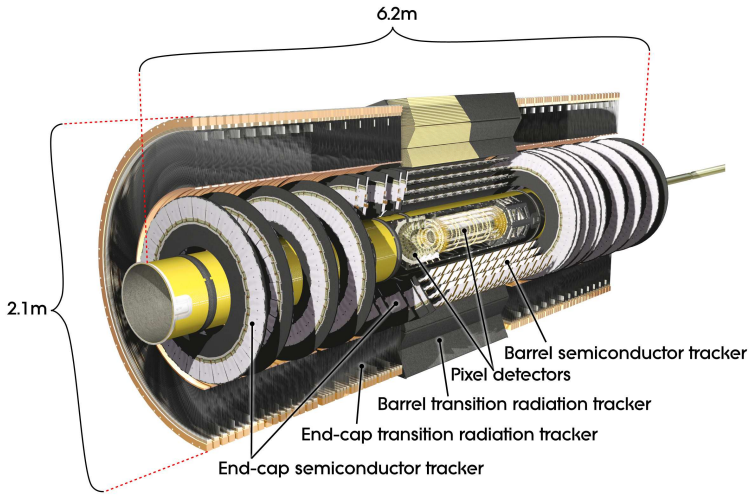


Figure 4.2: ATLAS Inner Detector cross-section view [17]

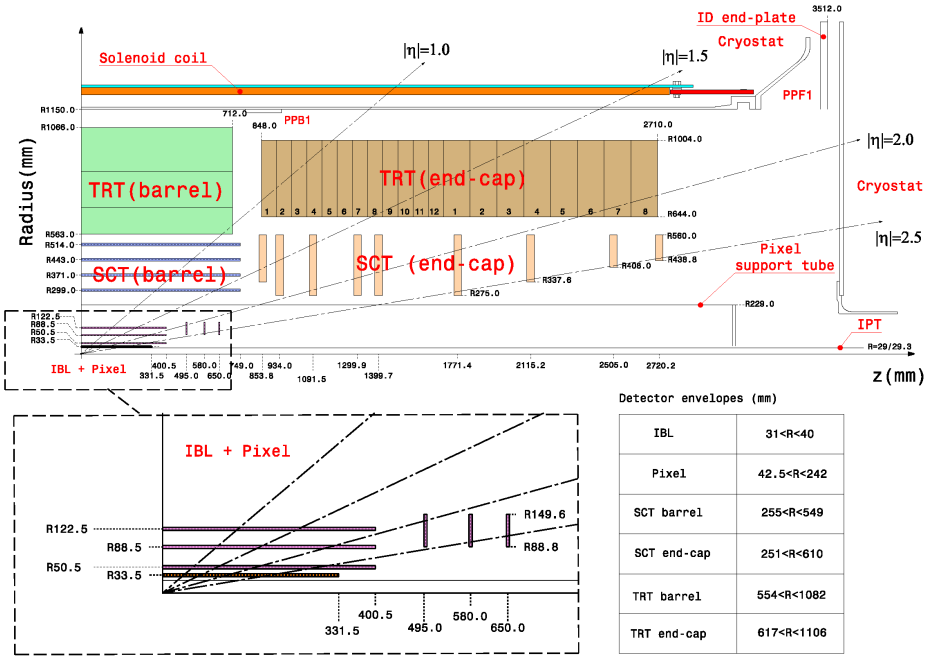


Figure 4.3: ATLAS Inner Detector layout [23]

and precise vertex resolution capabilities which are crucial for detecting short lived particles such as B mesons and τ leptons.

4.2.2 Semi-Conductor Tracker (SCT)

The SCT [17, 22] comprises four barrel layers and nine disks in each end-cap region as shown in Figure 4.3. Each silicon sensor measures 6.36 x 6.40 cm^2 with 768 readout

strips and each strip has 80 μm pitch. On each side of the module two sensor chips are wire-bonded together and further two such detector pairs are glued back to back at a 40 mrad stereo angle to provide two dimensional space points. These space points are later used for track reconstruction. In total the SCT detector covers 61m^2 with approximately 6.3 million readout channels. With the SCT, tracks separated by approximately 200 μm can be distinguished. It has a spatial resolution of 16 μm in $R-\phi$ and 580 μm in z, both in the barrel and the endcap region.

4.2.3 Transition Radiation Tracker (TRT)

The TRT [17,22] is the outermost detector of the ID system (as shown in figure 4.3) and consists of gas filled Polyimide drift straw tubes within the region of $|\eta| \leq 2$. Each straw is 4 mm in diameter containing a gold plated tungsten wire of diameter 31 μm. There are several layers of TRT modules resulting 50000 straws in total in the barrel region and a total of 32000 radial straws in the endcap region. The straws are arranged parallel to the beam pipe in the barrel region and radially in a wheel formation in the endcap region. The tubes are filled with a non-flammable gas mixture composed of 70% Xe, 27% CO_2 and 3% O_2. With its 351000 read-out channels, the TRT provides a spatial resolution of 130 μm in the $R-\phi$ plane and approximately 36 hits per track.

The working principle of the TRT is as follows: Charged particles moving at relativistic speeds emit X-rays (transition radiation) as they traverse the straws and gas mixture. This radiation is suppressed by a factor $1/m$ where m is the mass of the particle. As a result, electron identification is possible as electrons, being the lightest charged particles, leave a larger number of hits per track compared to pions or muons.

Due to a high particle flux, significant heat is generated within the straws by the ionization current in the gas. It is estimated to be 10 to 20 mW per straw assuming the LHC is operated at its design luminosity. They are cooled using a CO_2 cooling circuit.

4.3 Calorimetry

Calorimeters are used to measure the energies of particles and the location of the energy deposits. A particle entering into a calorimeter interacts with the dense material and if the energy of the incoming particle is high, new particles are created. These new particles interact with the same material again and more new particles are created resulting in a cascade. This process continues until the energies of the particles created are too low to create any new particles. The energy deposited by particles in this shower, which can be detected in the form of charge or light, serves as a measurement of the energy of the incident particle.

There are two types of particle showers: Particle showers induced by electromagnetic (EM) interaction and strong interaction. Therefore the ATLAS calorimeter system [17] consists of EM calorimeters and hadronic calorimeters (Fig 4.4) covering a range of

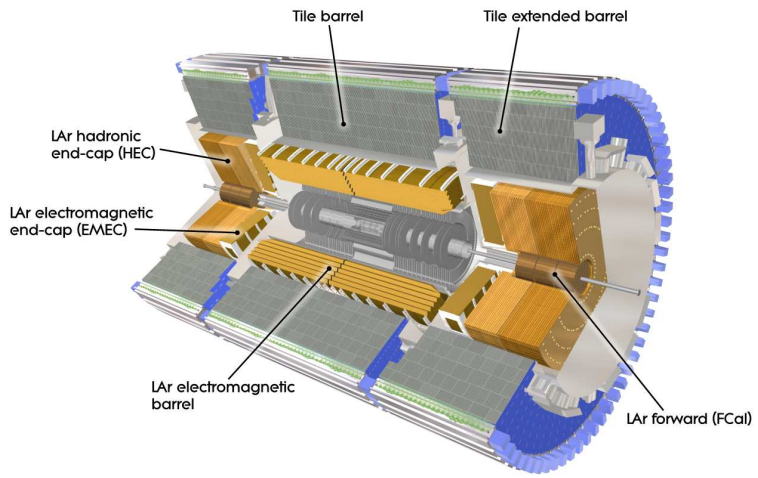


Figure 4.4: ATLAS calorimeter system [17]

$|\eta| \leq 4.9$. Muons are minimum ionizing particles. Therefore, the majority of them pass through calorimeters without absorption.

4.4 Muon Spectrometer

The Muon spectrometer (MS) [17] is the outermost component of the ATLAS detector. It is designed to precisely measure the momenta of muons penetrating through the calorimeters. The superconducting air-core toroid magnets provide the required intense magnetic field ($\sim$ 4 T) to bend the tracks of the muons. Since many processes trigger on muons, for example, $B_s^0 \to \mu^+\mu^-$, the MS is also a part of the trigger system and it is possible to trigger in the range $|\eta| \leq 2.4$. The MS has four sub-detector components, namely Monitored Drift Tubes (MDT), Cathode Strip Chambers (CSC), Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC). Figure 4.5 shows the muon system.

The MDTs are used for the precision measurement of the muon momenta and their tracks. The chambers consist of three to eight drift tubes. Each tube is 30 mm in diameter, filled with 93% AR and 7% CO_2. They operate at 3 atm pressure. In the center of each tube there is a 50 μm tungsten-rhenium wire acting as anode. The MDT resolution is 80 μm per tube and 35 μm per chamber.

The high granularity CSCs are used in the forward region ($2.0 < |\eta| < 2.7$), where the particle flux is high. CSCs are multi-wire proportional chambers with two cathodes segmented into strips. They provide (η, ϕ) coordinate information at 60 μm resolution in the bending plane and 5 mm in the transverse plane.

The RPCs and TGCs are parts of the trigger system. RPCs are installed in the barrel region ($|\eta| < 1.05$) and TGCs in the endcap region ($1.05 < |\eta| < 2.4$). The RPC

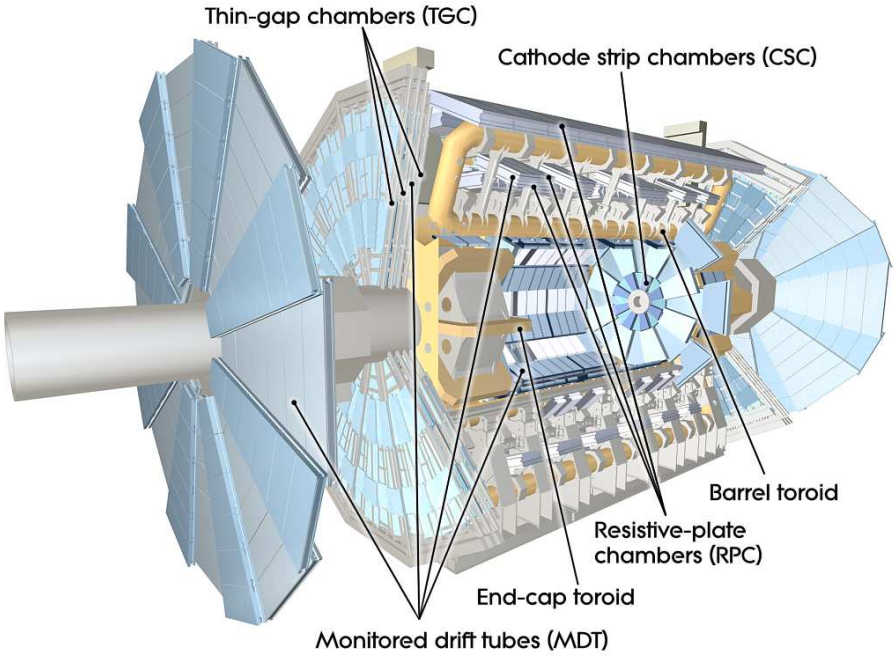


Figure 4.5: ATLAS muon system [17]

is a gaseous parallel electrode-plate detector without wires. Two resistive phenotic-melaminic plastic laminate plates are held parallel to each other separated by insulating spacers 2 mm wide. It contains a gaseous mixture of 94.7% $C_2H_2F_4$, 5% Iso-C_4H_{10} and 0.3% SF_6. The RPCs are operated at a high drift field (4.9 kV/mm) in order to induce avalanches in the gaseous mixture when an ionized particle traverses. The signals are read out by metallic strips, which are installed on the outer side of the resistive plates. A coincidence signal from a system of RPCs is used as a trigger. Using RPCs, it is possible to trigger on tracks in a wide p_T range, from 6 to 35 GeV.

TGCs are multi-wire proportional chambers having smaller wire-to-cathode distance (1.4 mm) than wire-to-wire distance (1.8 mm). Using a high quenching gas mixture (CO_2 and n-C_5H_{12}), these chambers are able to reach quasi-saturation mode. The TGCs provide good time resolution due to high electric fields ($\sim$ 2900 V) around the wires and the small distance between them. The RPCs and TGCs have a typical response time of 15 to 25 ns, which suffices to trigger events at every 25 ns (LHC bunch spacing).

4.5 Magnet system

A Magnet system is used for measuring the momenta of charged particles. The tracks of charged particles are bent in the magnetic field and the radius of curvature of a particle's track is directly proportional to its momentum. The ATLAS detector consists of two types of magnet systems [17] as shown in figure 4.6.

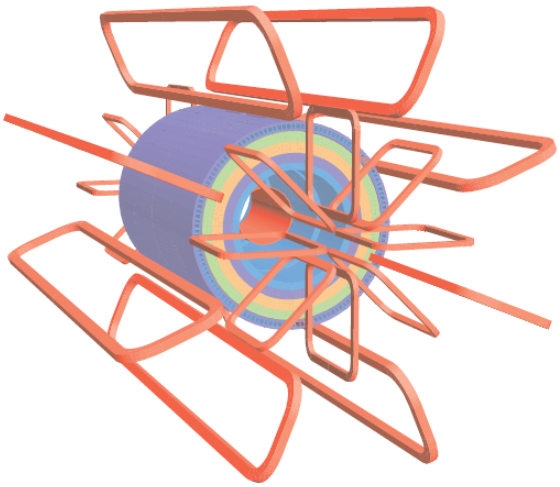


Figure 4.6: ATLAS magnet system [17]

The superconducting solenoid surrounds the ATLAS Inner Detector system. It is 5.3 m long and 2.5 m in diameter, aligned with the beam axis. It provides a homogeneous axial field of 2 T for bending the tracks of charged particles in the transverse plane.

The second system consists of superconducting air-core toroid magnetic systems, one in the barrel region and one on each side of the endcaps. Each toroid has 8 coils. In the barrel region the bending power is 1.5 to 5.5 Tm in the pseudorapdity range of $|\eta| < 1.4$. In the endcaps the bending power is 1 to 7.5 Tm in the pseudorapidity range of $1.6 < |\eta| < 2.7$.

4.6 Trigger System

Due to a very high collison rate (40 MHz), saving the data of every event is not feasible. Therefore a sophisticated trigger system is needed. The ATLAS trigger system [17, 24] is used to decide whether or not to keep the data of an event for later study. The trigger system in run 2 consisted of a hardware based trigger known as Level 1 (L1) and a software based trigger known as High Level Trigger (HLT).

The L1 trigger is used for selecting high p_T charged particles, photons, jets and missing energy. The L1 trigger is also used to define Regions of Interest (ROIs) for each event, which are physical locations (in (η, ϕ)) of those regions inside the detector which have interesting features. The triggering is done on the basis of the trigger menu which contains various criteria for selecting interesting events. It has a maximum latency of 2.5 μs. Using the L1 trigger the event rate is reduced to 100 KHz.

The HLT trigger is a software trigger, which further processes the events obtained from the L1 trigger using (partial) reconstruction algorithms. The HLT is used for track

reconstruction, finding vertices, analyzing ROIs etc. The decision time for the HLT is 200 μs and the event rate is further reduced to 1 KHz.

5 Statistical data analysis of rare decays

5.1 Rare decays and new physics

Transitions mediated by $W^{\pm}$ bosons are called weak charged current transitions and those mediated by Z bosons are weak neutral current transitions. Weak interactions are not necessarily flavor conserving. The amplitudes for a change of quark flavor in cross generations, via weak charged current transitions, are given by the Cabibbo–Kobayashi–Maskawa (CKM) matrix elements [25, 26].

However, unlike weak charged current transitions, Flavor Changing weak Neutral Current (FCNC) transitions are forbidden at tree level (for example $s \to b + Z^0$) and can take place only indirectly via the exchange of $W^{\pm}$ (such as $b \to c + W^- \to s + Z^0$), resulting in a loop diagram. This was explained by the GIM mechanism [4]; with the probability for such an FCNC decay being proportional to the mass of the quark involved. As a result the FCNC decays are highly suppressed in the SM.

The decay $B_s^0 \to \mu^+\mu^-$ is one such example of FCNC transitions. Feynman diagrams are shown in Figure 5.1. As a B_s^0 particle is spin-less, both the final state muons should be either left handed or right handed. Therefore this decay is further suppressed by the helicity factor.

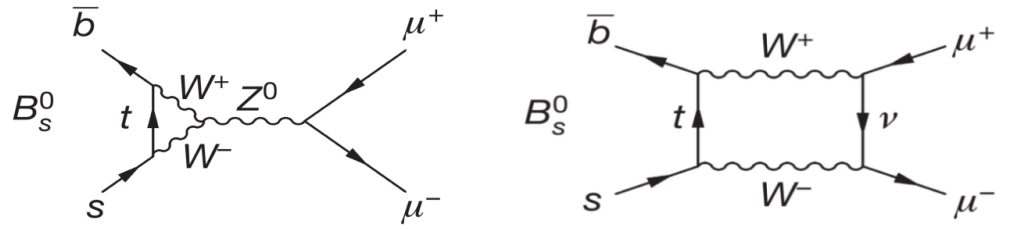
The theoretical formula [27–29] and prediction value (from the SM) of the branching ratio ($\mathcal{B}$) of $B_s \to \mu^+\mu^-$ is given by:

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-)_{SM} = \frac{G_F^4 m_W^4 sin^4\theta_W}{8\pi^5}\tau_B f_B^2 m_B m_\mu^2 \sqrt{1-\frac{4m_\mu^2}{m_B^2}}|C_{10}V_{tb}^* V_{ts}|^2 \,,$$

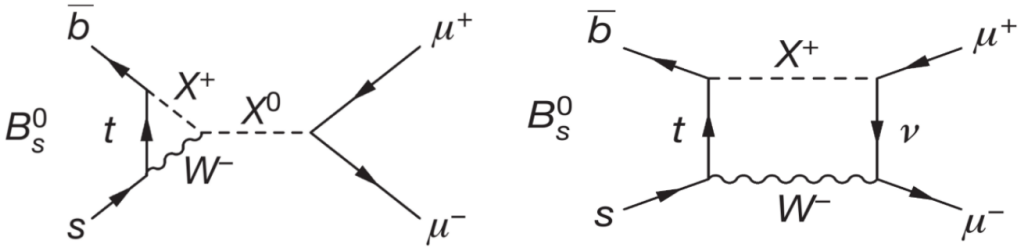
where G_F is the Fermi constant, M_W the W boson mass, θ_W the weak mixing angle, τ_B the B_s^0 lifetime, f_B^2 the lifetime decay constant, m_μ the muon mass, m_B the B_s^0 mass, C_{10} the Wilson coefficient and both the V_{tb}^* and V_{ts} are the CKM matrix elements. The SM branching fraction is predicted to be [30]:

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-)_{SM} = 3.65 \pm 0.23 \times 10^{-9} \,.$$

However, if there are new heavy particles, e.g as predicted by the Minimal Supersymmetery Standard Model (MSSM) theory, like A_0, χ_0 (Figure 5.1b), it will alter the FCNC



(a) Standard Model



(b)Additional contributions predicted by the MSSM

Figure 5.1: Decay of $B_s^0 \to \mu^+\mu^-$ [31]

transition amplitudes. Precisely measuring the decay rates of $B_s^0 \to \mu^+\mu^-$ is therefore necessary not only to find deviations from the SM but also to set limits to the BSM theories, in particular to the MSSM. Hence rare decays of hadrons containing a heavy beauty quark allow us to explore New Physics.

5.2 The need for statistics in data analysis

The final states corresponding to the desired decay are called signal events (or signal). Background events (or background) are events that appear to have the same final state as signal events but not originating from the decay searched for. The main reasons for background are partially reconstructed events, misidentification of particles and uncorrelated decays mimicking signal.

As the FCNC decays are very rare, studies of these decay modes require a large data sample to observe sufficient relevant particles. The ratio of the number of signal events to the number of total events (signal fraction) is so low, that the detection of the signal is not trivial. Moreover, it is impossible to determine whether a single event is background or signal on an event by event basis. Therefore only with statistical techniques is it possible to obtain precise results. The signal fraction is improved by statistically rejecting the background.

The vital role, that statistical techniques play in High Energy Physics (HEP) data

analysis, especially in background rejection, is the motivation for this thesis.

5.3 Statistical techniques in the data analysis of $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$

In the ATLAS run 2 analysis (2015-16), the branching fractions of the rare decays $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ are measured relative to the reference decay modes $B^\pm \to J/\psi(\to \mu^+\mu^-)K^\pm$, which is more abundant and has a well measured branching fraction.

The analysis is performed using multivariate techniques [7]. Due to the rarity of the signal events a completely data-driven analysis is not possible. Therefore events are simulated using Monte Carlo (MC) methods to improve statistics. However, a MC data set has to be validated against data to use it in the analysis. In order to validate the MC data set, signal distributions in the data are compared to the MC signal distributions for Boosted Decision Tree (BDT) input variables in the reference channels $B^\pm \to J/\psi\ K^\pm$ and $B_s^0 \to J/\psi\ \phi$.

In order to obtain the distributions for signal from the experimental data set, a sideband subtraction technique is used (explained in detail in chapter 7). Briefly, the idea is to use the B^+ mass variable to separate signal and background statistically. As a result, the signal distributions of the BDT input variables are obtained. Hence the B^+ mass is called the discriminating variable. From the MC studies it is concluded that the signal Probability Density Function (PDF) is a double Gaussian and a decaying exponential is used as background PDF. In this technique the discriminating variable distribution is divided into two regions by applying cuts. The region within the cuts is called the signal region and the rest sideband region. The cuts are applied at 5180 MeV and 5380 MeV in a total range of 5080-5480 MeV. The task is to estimate the background events in the signal region from the sidebands and to statistically subtract those events from the total number of events in the signal region to obtain the signal-only distribution. A simultaneous binned maximum likelihood fit is performed on the B^+ mass (discriminating) variable. From the fit the signal fraction (f_s = signal events/total events) is known. Further, this f_s is used to obtain the signal distributions for the BDT input variables such as pointing angle, isolation and $\chi^2_{PV,DVxy}$ (see table 1 in [7]). The plots in Figure 5.2 show the MC and data distributions for some BDT input variables for both the reference channels.

The signal distribution plots of the BDT input variables are compared to the MC data set. These comparison plots are needed for the validation of the MC samples. As the MC samples consist of only signal events, it is extremely important to separate signal from background in the data in order to compare it with the MC sample. As explained above, a sideband subtraction technique is used for this purpose. However, an alternative tech-

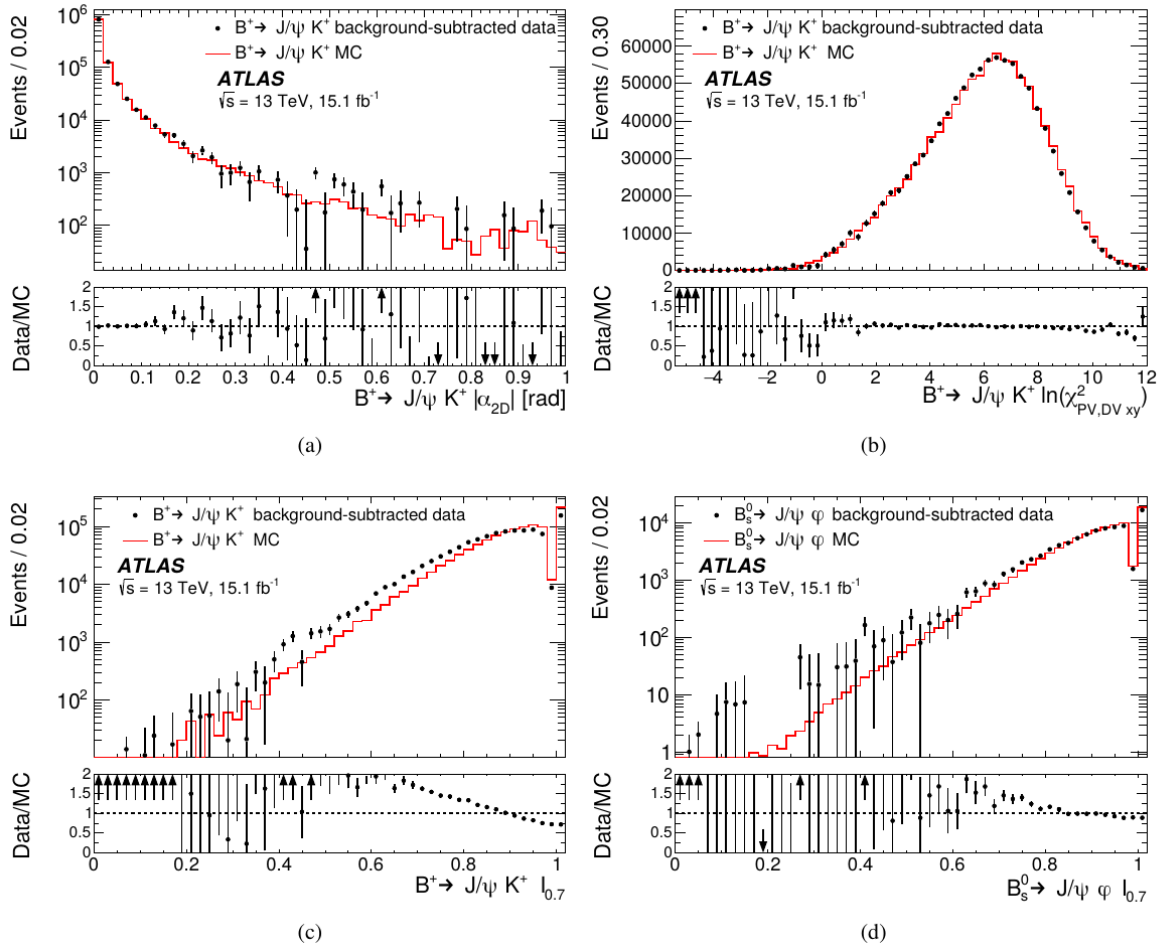


Figure 5.2: Data and MC distributions (for signal) in $B^+ \to J/\Psi K^+$ for discriminating variables: (a) $|\alpha_{2D}|$, (b) $\ln(\chi^2_{PV,DVxy})$, (c) $I_{0.7}$. The variable $I_{0.7}$ is also shown in (d) for $B^0_s \to J/\psi\phi$ [7]

nique, namely sPlot technique, can be used for the same purpose; i.e. to separate the signal from background in a mixed data set. Unlike the sideband subtraction technique, the sPlot technique does not need signal region cuts. Therefore it is advantageous to use the sPlot technique in cases where signal and background distributions are not cleanly separated.

The focus of this thesis is to understand and explore the sPlot technique [8] (explained in the next chapter). This technique can also be used to estimate the background and signal distributions in a mixed data set, which in turn helps to improve the signal to background ratio. This technique is studied with the aim of applying it to produce MC - data comparison plots for the $B^0_s \to \mu^+\mu^-$ data analysis.

6 Studying the sPlot technique

6.1 Introduction

The sPlot technique, introduced by M. Pivk and F.R. Le Diberder, is used to unfold data distributions [8]. The idea of the sPlot techinque is explained in the first two sections of this chapter and a detailed study of the technique with an example is discussed in the rest of the chapter.

In general, the data contains events from different species (for example, background and signal) merged in one sample. In the data analysis, it is necessary to know the underlying distribution for a particular species of events (such as the signal distribution) and the sPlot technique can be used in order to separate the distributions statistically. The variables, which constitute the data, can be divided into two classes. The first class consists of a set of variables for which the distributions (the form of the underlying PDFs) for each species is known. These are called discriminating variables. The second class are called control variables for which it is not necessary to know the underlying forms of the distributions for all species. The sPlot technique estimates the distributions for each species of the control variables using only the known distributions of discriminating variables. In this way, the distributions of the control variables are reconstructed without the use of any prior knowledge of the control variables. In the context of particle physics data analysis, sPlot is a statistical technique, which allows to separate signal and background distributions and keeping track of the statistical uncertainties per bin. The crucial assumption for sPlot to work is that the distributions of control variables are not correlated to the distributions of discriminating variables.

6.2 sPlot technique execution

The objective is to use the sPlot technique to reconstruct the distributions of all control variables for all species without using any prior knowledge about them but using solely the distributions of the discriminating variables. For simplicity, the execution of the technique is discussed by considering only one variable as the discriminating variable.

In order to meet the objectives sWeights are calculated (see explanation below) for each species. Then, to obtain the distribution for a specific species, histogram bins are filled with the sWeights of the corresponding species as weighting factors for each event for the control variables.

To illustrate: If the data contains two species (signal and background), two sets of sWeights are calculated; one for the signal and one for the background. Then, to obtain the distribution for a control variable of a specific species, e.g signal, the histogram is filled with signal sWeights as weighting factors to determine the signal distribution for that control variable.

The following symbols are used:

N is the total number of events,

N_{s} is the number of species of events populating the data sample,

N_i is the number of events expected on the average for the i^{th} species,

y is the set of discriminating variables,

f_n is the PDF of discriminating variables for the n^{th} species,

$f_i(y_e)$ denotes the value taken by the PDFs f_i for event e,

x denotes the set of control variables,

$V_{n,j}$ are covariance matrix elements.

The formula to calculate an sWeight is given by the equation:

$$ {}_sP_n(y_e) = \frac{\sum_{j=1}^{N_s} V_{nj} f_j(y_e)}{\sum_{k=1}^{N_s} N_k f_k(y_e)} \tag{6.1} $$

where ${}_sP_n(y_e)$ is the sWeight for discriminating variable y, species n and event e.

The denominator in (6.1) is the normalization factor. In order to use this formula two things must be calculated first; the number of events per species and the covariance matrix elements.

Firstly, to know the total number of events in each species, a simultaneous maximum likelihood fit is performed to the discriminating variable (the method of maximum likelihood fit is explained in chapter 8). From the fit, the parameter values for the underlying PDFs $f_n(y)$ and the total number of events N_i in each species are determined.

The covariance matrix elements are calculated for the parameters of discriminating variables. The second derivative of the likelihood function $\mathcal{L}$ with respect to the parameters at the central values from the fit yields the inverse covariance matrix elements. Equation 6.2 is used to determine the inverse covariance matrix elements V_{nj}^{-1} and then the covariance matrix elements are calculated by inverting the matrix V^{-1}:

$$V_{nj}^{-1} = \frac{\partial^2(-\mathcal{L})}{\partial N_n \partial N_j} = \sum_{n=1}^{N_s} \frac{f_n(y_e) f_j(y_e)}{(\sum_{k=1}^{N_s} N_k f_k(y_e))^2} \,. \tag{6.2}$$

A set of sWeights for each species are provided by (6.1). They are used in filling the histogram bins to obtain the distributions of the control variables for each species. The sum of the weights for all events of a particular species is equal to the number of events for that species provided by the fit.

The uncertainties per bin are calculated by the following formula:

$$\sigma[N_n \, {}_s\tilde{M}_n(x)\delta x] = \sqrt{\sum_{e \subset \delta x} ({}_sP_n)^2} \tag{6.3}$$

where $\tilde{M}$ is the x distribution and the sum $\sum_{e \subset \delta x}$ runs over the $N_{\delta x}$ events for which x_e (i.e. the value taken by the variable x for event e) lies in the x bin centered on $\bar{x}$ and of total width δx.

6.3 Studying the sPlot technique through an example

The idea is to study how the sPlot technique works by considering an example inspired by a situation encountered in particle physics data analysis. In this example, there are two variables in total: one discriminating, namely mass (m) and one control variable, namely v. The data set contains two species of data, signal and background. The goal is to estimate the sub-distributions for v i.e, signal and background, assuming the forms of the underlying PDFs of the mass variable m for both the signal and background are known.

The following steps are taken:

1. Data simulation (using the Monte Carlo method),
2. Calculation of sWeights,
3. Filling the v histograms with sWeights for both signal and background.

The following symbols are used:

N_{tot} is the number of total events,

N_{sig} is the number of signal events,

N_{bg} is the number of background events,

y is the discriminating variable (mass),

$f_{\text{sig}}(y)$ and $f_{\text{bg}}(y)$ are the PDFs of the discriminating variable (mass) for signal and background,

$f_{\text{sig}}(y_e)$ and $f_{\text{bg}}(y_e)$ are the values for PDFs of the discriminating variable m for signal and background for an event e.

6.3.1 Data simulation

In this example data is simulated using a Monte Carlo (MC) method. The data is simulated in ROOT using the TRandom3 random number generator.

The data is generated using two different functional forms namely, Gaussian and Exponential. In general, Gaussian and Exponential functions do not have a finite range. But in this example, as the data is generated in a fixed range, the normalization is done accordingly. The exact functional forms used in this example are given by:

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma \cdot [\text{erf}(\frac{b-\mu}{\sqrt{2}\sigma}) - \text{erf}(\frac{a-\mu}{\sqrt{2}\sigma})]} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} , \tag{6.4}$$

$$h(x) = \frac{1}{\tau \cdot (e^{-\frac{a}{\tau}} - e^{-\frac{b}{\tau}})} e^{-\frac{x}{\tau}} . \tag{6.5}$$

The total number of events N_{tot} is a random number obtained from a Poissonian with a mean of 100 000. The parameter values of the function and range of each variable are given for each species in Table 6.1.

Variable	Variable type	Signal	Background	Range
m	discriminating	Gaussian ($\mu = 60, \sigma = 4$)	Exponential ($\tau = 40$)	[0,150]
v	control	Exponential ($\tau = 80$)	Exponential ($\tau = 100$)	[0,350]
signal fraction $f_s = 0.2$				

Table 6.1

The generated data is shown in figures 6.1 and 6.2. The generated data for v is also plotted in log scale (Figure 6.3) to visualize the different slopes of signal and background. The reason, that variable m is designed to comprise a Gaussian signal with an exponential background, is because this is similar to the B^+ mass distribution in $B^+ \to J/\psi\ K^+$. Similarly, the reason that v comprises two decaying exponentials separated by slightly different decay rates, is because this mimics the problem of distinguishing the lifetimes of two elementary particles. The idea is to use the sPlot technique to obtain the distributions of signal and background for v, represented by the red dotted and black line histograms in Figure 6.2, from the variable m.

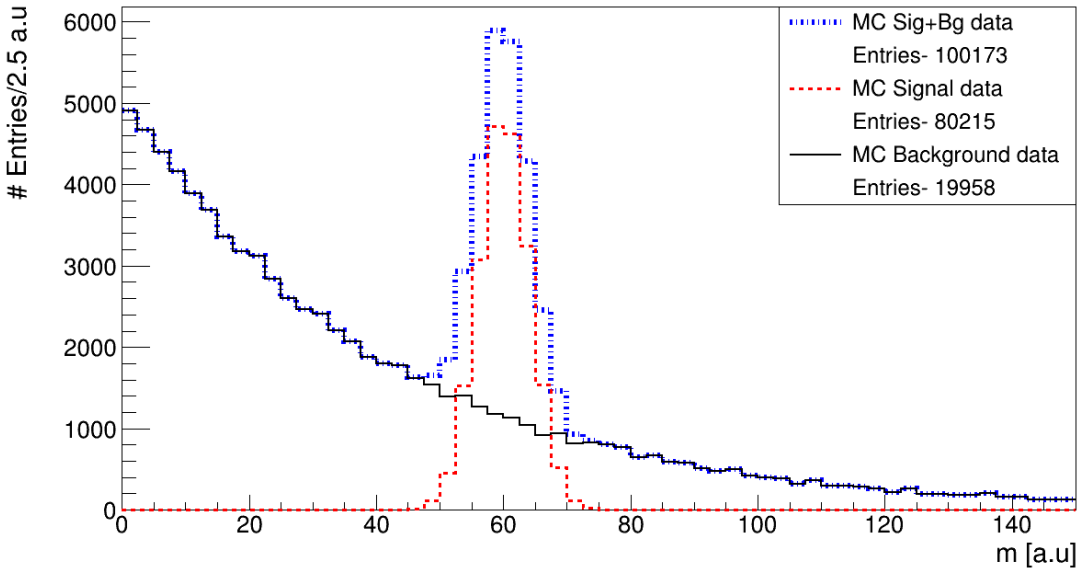


Figure 6.1: MC generated data for the discriminating variable mass.

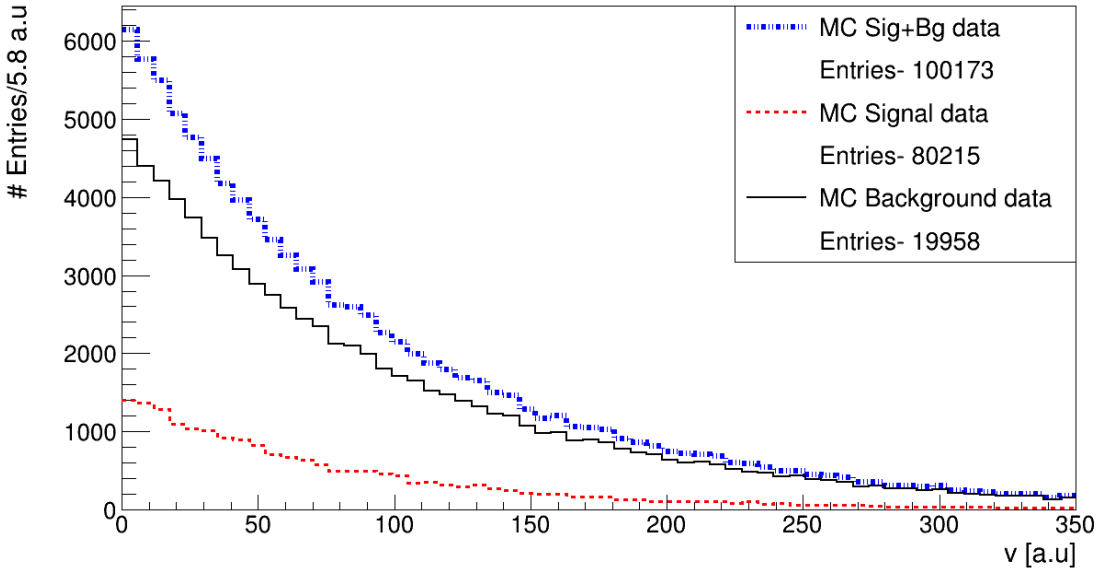


Figure 6.2: MC generated data for the control variable v.

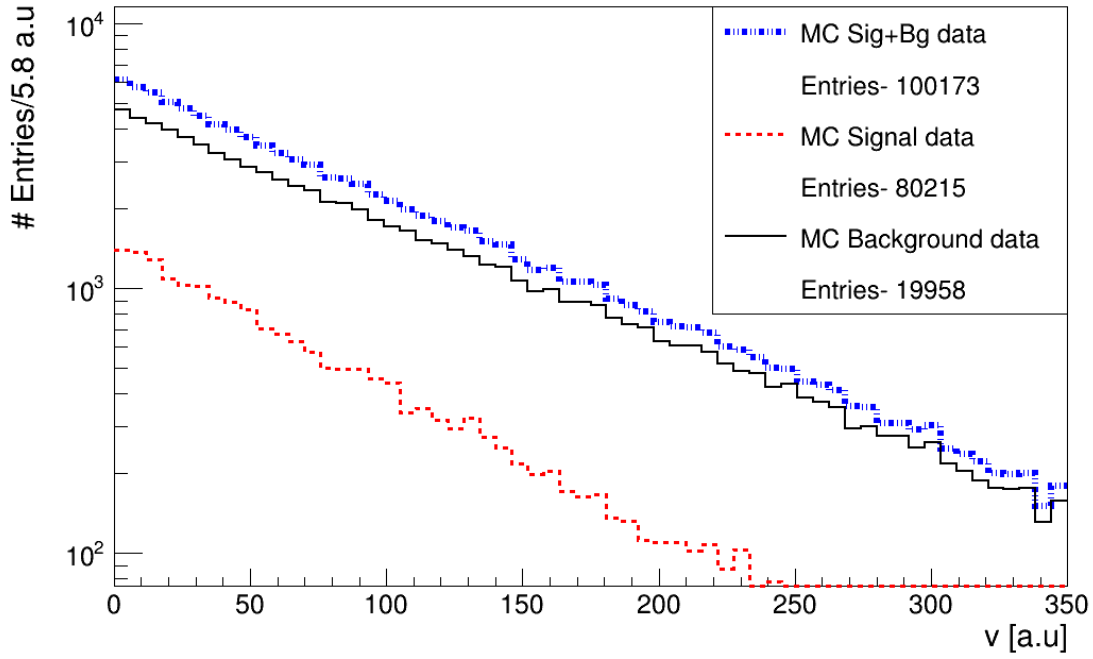


Figure 6.3: MC generated data for the control variable v in log scale.

6.3.2 Calculation of sWeights

As there are two species, namely signal and background, in this example there are two sets of sWeights, one for signal and one for background. In order to calculate the sWeights the number of events in each species should be known. A simultaneous un-binned maximum likelihood fit (red dotted line in Figure 6.4) is performed for the discriminating variable m and the signal fraction and parameter values are obtained. Then the number of signal and background events is calculated from the equations:

$$N_{\rm sig} = f_s \times N_{\rm tot}\,, \tag{6.6}$$

$$N_{\rm bg} = N_{\rm tot} - N_{\rm sig}\,. \tag{6.7}$$

The parameter values of the discriminating variable fit are used in the calculation of sWeights.

The next step is to calculate the covariance matrix elements. As there are two species the covariance matrix is a symmetric 2×2 matrix. The expressions for the matrix elements are given by:

$$V_{11}^{-1} = V_{\rm sig,sig}^{-1} = \frac{\partial^2(-\mathcal{L})}{\partial N_{\rm sig}\partial N_{\rm sig}} = \sum_{e=1}^{N_{tot}} \frac{f_{\rm sig}(y_e) f_{\rm sig}(y_e)}{(N_{\rm sig}\cdot f_{\rm sig}(y_e) + N_{\rm bg}\cdot f_{\rm bg}(y_e))^2}\,, \tag{6.8}$$

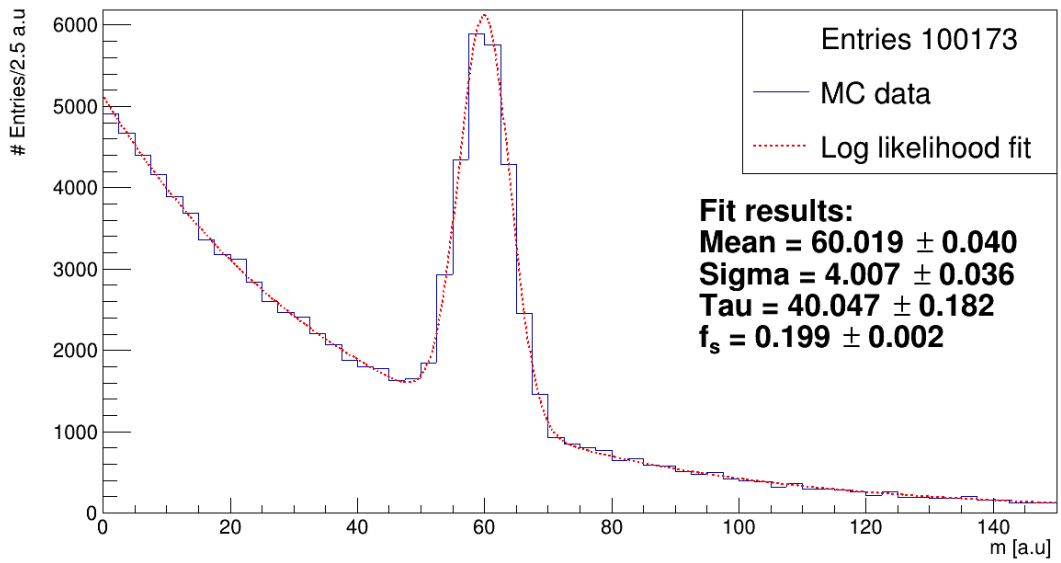


Figure 6.4: Simulatnous maximum likelihood fit for the discriminating variable mass

$$V_{22}^{-1} = V_{\mathrm{bg,bg}}^{-1} = \frac{\partial^2(-\mathcal{L})}{\partial N_{\mathrm{bg}} \partial N_{\mathrm{bg}}} = \sum_{e=1}^{N_{\mathrm{tot}}} \frac{f_{\mathrm{bg}}(y_e) f_{\mathrm{bg}}(y_e)}{(N_{\mathrm{sig}} \cdot f_{\mathrm{sig}}(y_e) + N_{\mathrm{bg}} \cdot f_{\mathrm{bg}}(y_e))^2}\,, \tag{6.9}$$

$$V_{12}^{-1} = V_{12}^{-1} = V_{\mathrm{sig,bg}}^{-1} = \frac{\partial^2(-\mathcal{L})}{\partial N_{\mathrm{sig}} \partial N_{\mathrm{bg}}} = \sum_{e=1}^{N_{\mathrm{tot}}} \frac{f_{\mathrm{sig}}(y_e) f_{\mathrm{bg}}(y_e)}{(N_{\mathrm{sig}} \cdot f_{\mathrm{sig}}(y_e) + N_{\mathrm{bg}} \cdot f_{\mathrm{bg}}(y_e))^2}\,. \tag{6.10}$$

The next step is to calculate the sWeights. The $N_{\mathrm{sig}}, N_{\mathrm{bg}}$ calculated above and the covariance matrix element values V_{ij} along with the function value of the discriminating variable m for both signal $f_{\mathrm{sig}}(y_e)$ and background $f_{\mathrm{bg}}(y_e)$ are used in order to obtain the sWeights for each event for the corresponding species:

$${}_sP_{sig}(y_e) = \frac{V_{\mathrm{sig,bg}} \cdot f_{\mathrm{sig}}(y_e) + V_{\mathrm{sig,sig}} \cdot f_{\mathrm{sig}}(y_e)}{N_{\mathrm{sig}} \cdot f_{\mathrm{sig}}(y_e) + N_{\mathrm{bg}} \cdot f_{\mathrm{bg}}(y_e)}\,, \tag{6.11}$$

$${}_sP_{bg}(y_e) = \frac{V_{\mathrm{bg,bg}} \cdot f_{\mathrm{bg}}(y_e) + V_{\mathrm{bg,sig}} \cdot f_{\mathrm{sig}}(y_e)}{N_{\mathrm{sig}} \cdot f_{\mathrm{sig}}(y_e) + N_{\mathrm{bg}} \cdot f_{\mathrm{bg}}(y_e)}\,. \tag{6.12}$$

6.3.3 sPlot histograms and results

The next step is to fill the histograms for the control variable v using the corresponding sWeights. Thus, the sPlot technique produces weighted histograms. Figure 6.5 show the distributions of sPlot histograms for v for both species.

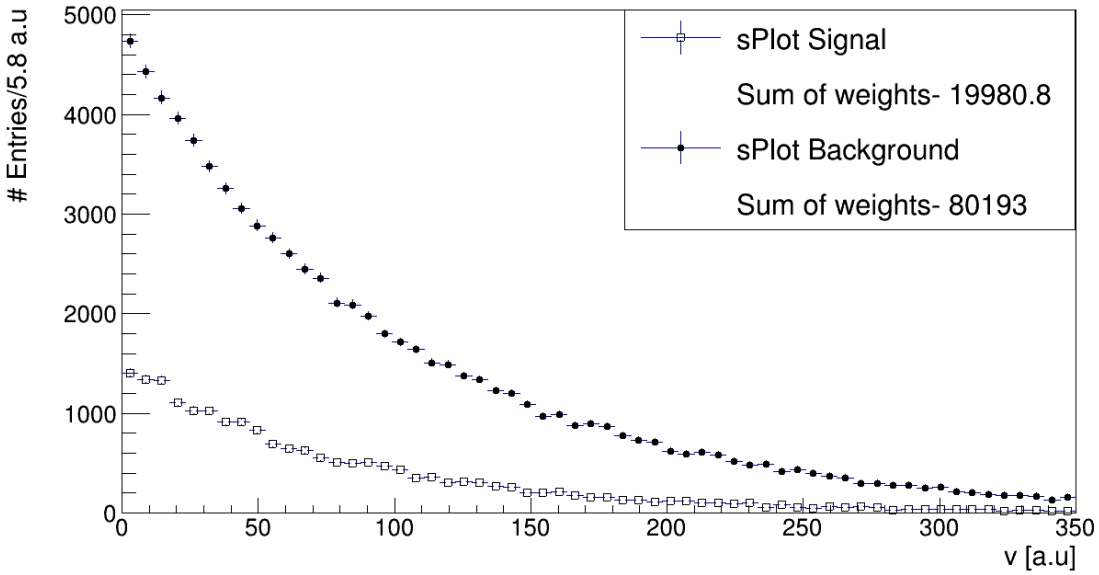


Figure 6.5: sPlot distribution for v for both signal and background

For a visual comparison the generated (true) histograms of the two control variables for both species are plotted as an overlay to the histograms obtained by the sPlot technique in Figure 6.6.

The shapes of the distributions obtained from the sPlot technique match with the generated distributions. In order to quantify the agreement, maximum likelihood fits are performed for the distributions obtained from this technique and compared with the true parameters. This is discussed in Chapter 8.

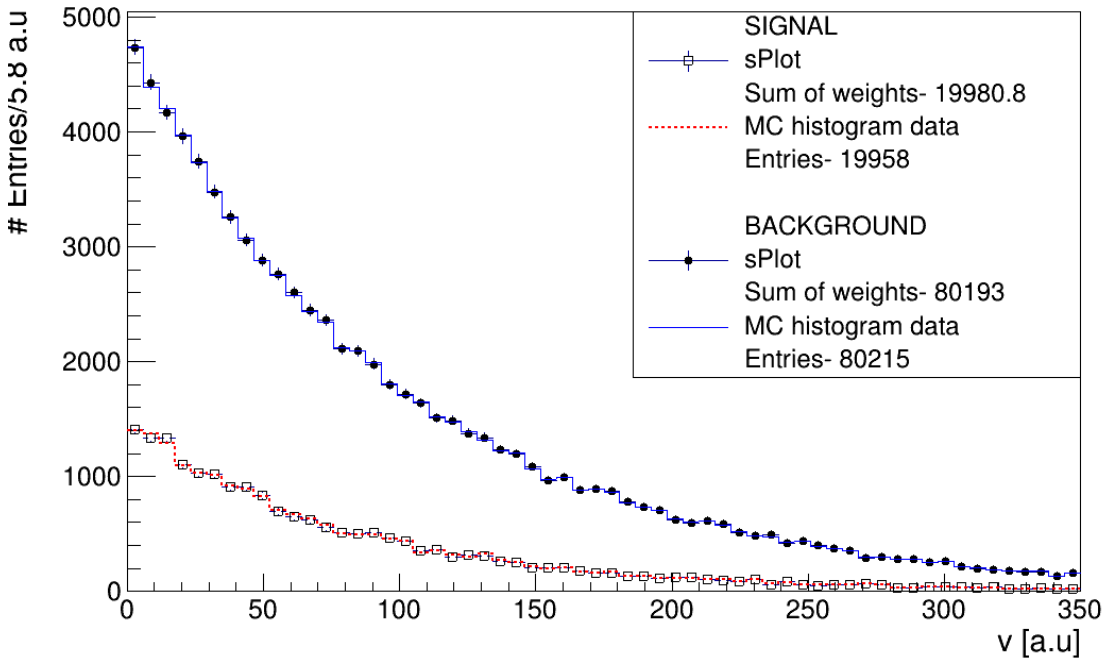


Figure 6.6: sPlot distributions with the generated data overlay for both signal and background for v

7 Sideband subtraction technique

7.1 Introduction

Sideband subtraction is another statistical technique to separate different species of data (for example, signal from background) for the control variables. As a part of this thesis this technique was also studied in detail, along with the sPlot technique, by considering the same example as described in the previous chapter.

7.2 Sideband subtraction technique execution

The sideband subtraction technique is illustrated by a diagram, as shown in Figure 7.1. It is the plot of the discriminating variable (mass) distribution for this example. In this technique, the data plot is divided into two regions by applying cuts to the discriminating variable. The region between the cuts is the signal region (SR) and the rest are the sideband regions (SB). The cuts are determined by the user such that the sidebands contain only (or mostly) background events.

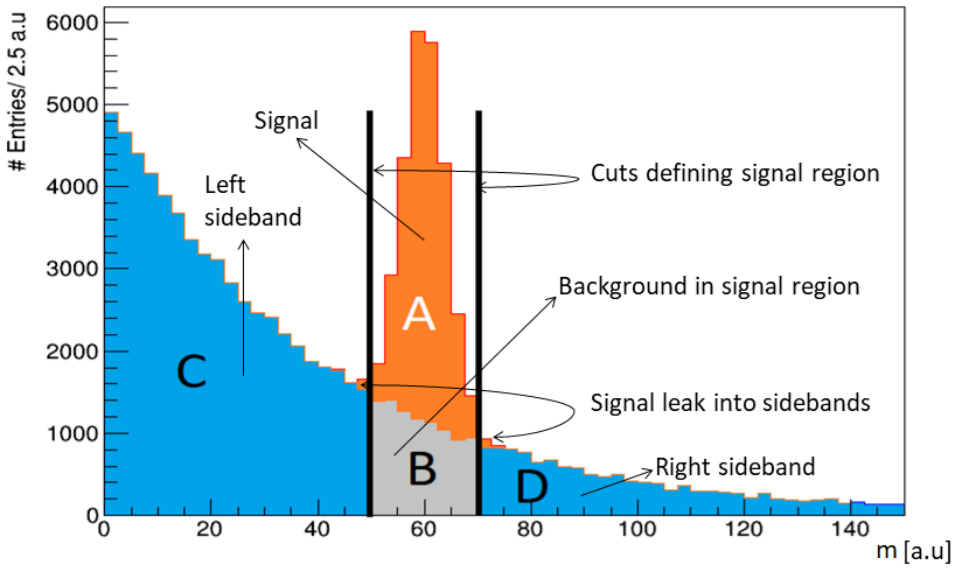


Figure 7.1: Plot illustrating the sidebands and the signal region with cuts applied at mass values 50 a.u and 70 a.u.

As shown in Figure 7.1, the cuts are chosen in such a way that the number of signal events leaking into the sidebands is negligible. In section 7.4 a case with a significant number of signal events leaking into the sidebands will be discussed.

The goal is to extract the signal and background distributions for the control variables using the distribution of the discriminating variable. For simplicity, the method is explained to obtain the signal and background distributions for one control variable only.

To obtain the background distribution for the control variable, the events from the sidebands (region C and D in Figure 7.1) of the discriminating variable are used to obtain the shape of the background distribution for the control variable.

To obtain the signal distributions for the control variable, the background events are to be subtracted from the total entries per bin in the signal region.

The following steps are performed.

1. A maximum likelihood fit is performed for the discriminating variable to obtain the signal fraction f_s and subsequently N_{sig} and N_{bg}.

2. Cuts are defined on the discriminating variable which define the signal region and the sidebands.

3. For the events in the sidebands, the distribution of the control variable is plotted. Then the distribution is scaled with the total number of background events N_{bg}. This scaled distribution gives the background distribution for the control variable. This is labeled as plot 1.

4. For the events in the signal region, the distribution for the control variable is plotted. This is labeled as plot 2.

5. The number of background events in the signal region is calculated by:

$$N_{\mathrm{bg}}^{\mathrm{SR}} = N_{\mathrm{bg}} - N_{\mathrm{bg}}^{\mathrm{SB}} \,. \tag{7.1}$$

6. Plot 1 is scaled to the total number of background events in the signal region $N_{\mathrm{bg}}^{\mathrm{SR}}$ by using the following scaling factor:

$$W_{\mathrm{sig}} = \frac{N_{\mathrm{SR}}^{\mathrm{bg}}}{N_{\mathrm{tot}}} \,. \tag{7.2}$$

This scaled plot is labeled as plot 3.

7. By subtracting events bin by bin in plot 2 from plot 3, the signal distribution for the control variable is obtained.

7.3 Example with negligible signal leak into sidebands

The task is to obtain the signal and background distributions for the control variable v using the discriminating variable m as described in section 6.3.

As in this case the leak of signal events into the sidebands is negligible, all the events in the sidebands are considered to be background events and thus used to obtain the background distribution.

To obtain signal and background distributions for control variable v the following steps are performed.

1. In this example the cuts are applied for the discriminating variable m at 50 a.u and 70 a.u.

2. For the events in the sidebands, the distribution for v is plotted and scaled to total number of background events N_{bg} as shown as black dotted points in Figure 7.2 (plot 1). This is the background distribution for control variable v.

3. For the events in the signal region, the distribution for v is plotted (plot 2). This is shown in Figure 7.3 and for better illustration of the problem the true MC generated data is also shown in Figure 7.4. As shown in Figure 7.4 the events for v corresponding to the signal region in m contains background.

4. The number of background events in the signal region $N_{\text{bg}}^{\text{SR}}$ is calculated by (7.1).

5. Plot 1 is scaled to $N_{\text{bg}}^{\text{SR}}$ (plot 3). This is shown in Figure 7.5.

6. By subtracting events bin by bin in plot 2 from plot 3, the signal distribution for the control variable v is obtained (white boxed points in Figure 7.2).

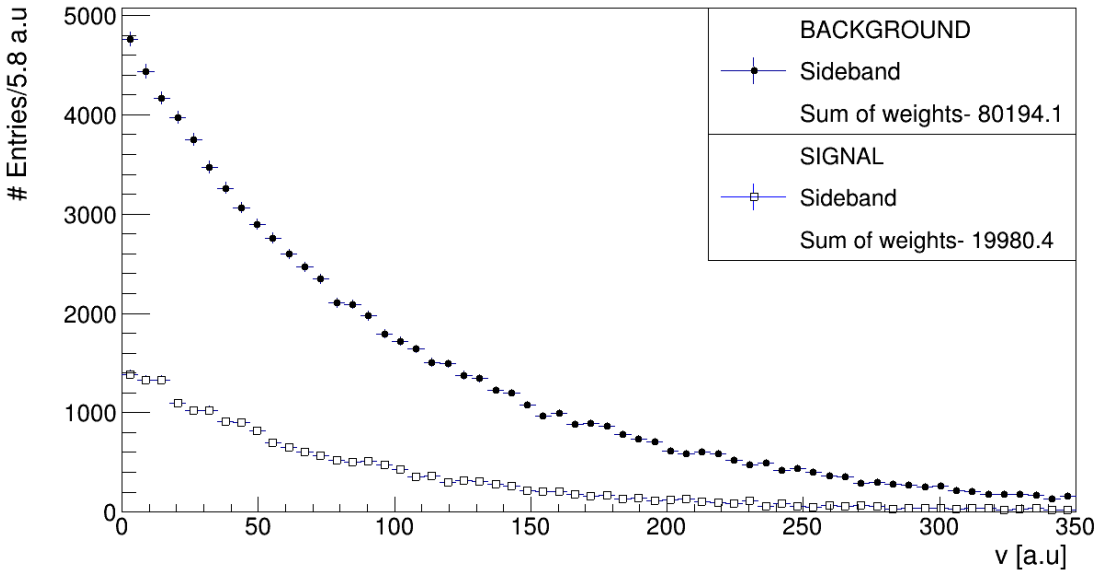


Figure 7.2: Black dotted points - distribution of v for the events in the sidebands, scaled to N_{bg}. White square points - distribution of v for the events in the signal region after subtracting the background.

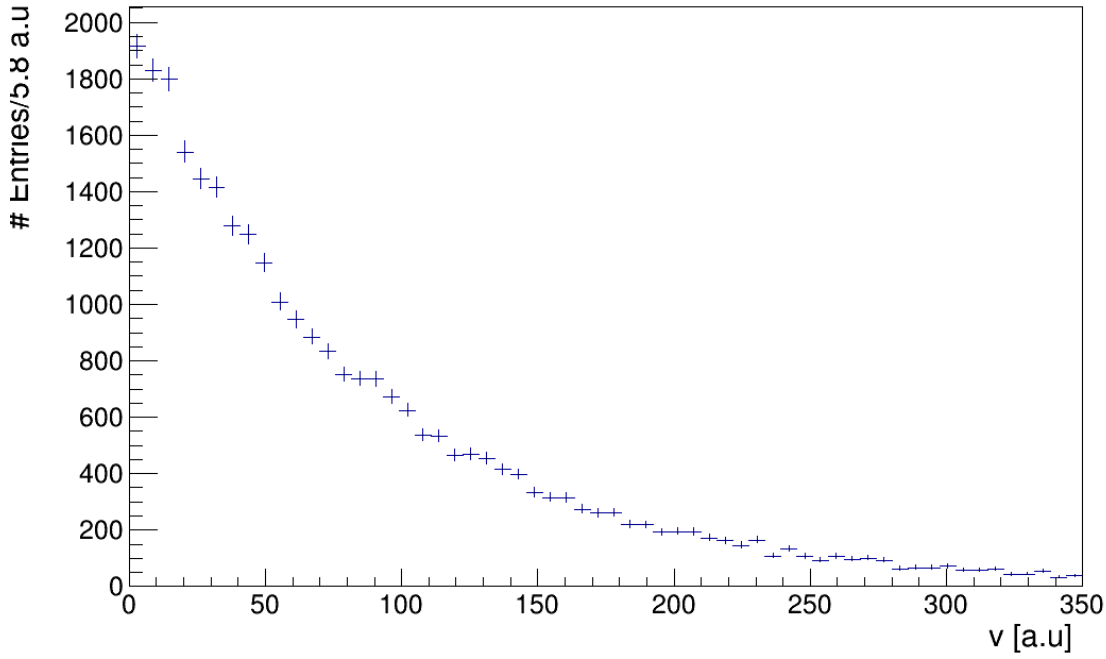


Figure 7.3: Distribution of v for the events in the signal region.

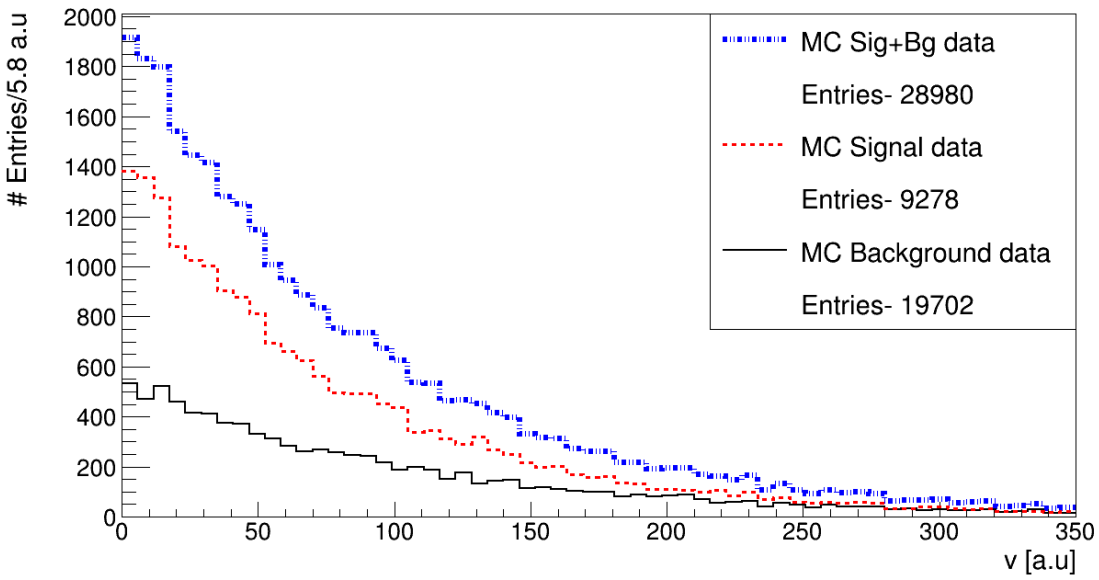


Figure 7.4: MC generated (true) data distribution of v for the events in the signal region. The black line shows the background contamination in the signal region which should be subtracted bin by bin from the total number of events (blue dot dash line) in order to obtain the signal distribution(red dotted line)

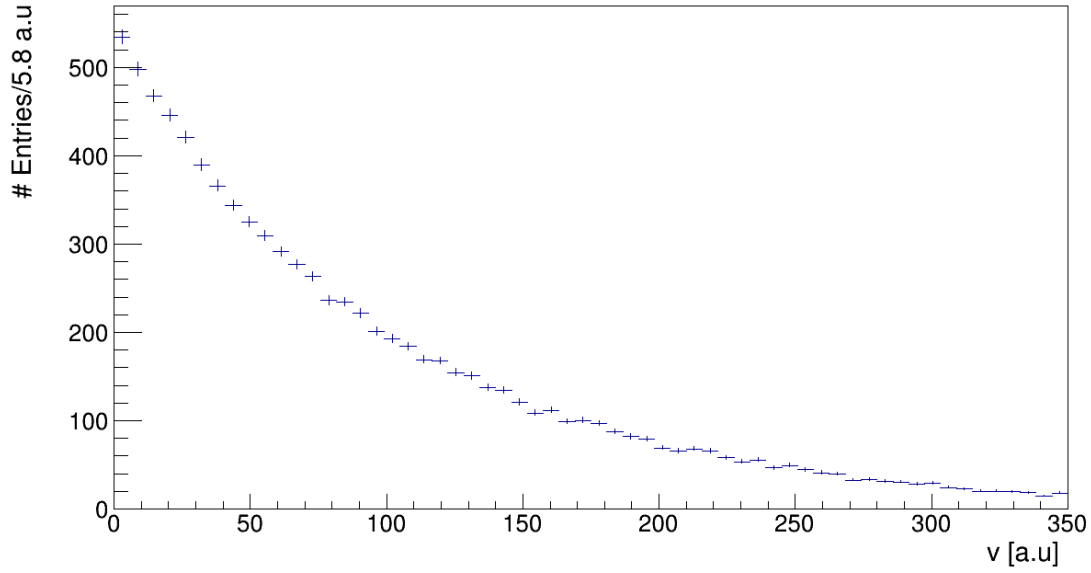


Figure 7.5: Distribution of v for the events in the sidebands, scaled to $N_{\text{bg}}^{\text{SR}}$.

7.4 Sideband-subtraction technique with a significant number of signal events leaking into the sidebands

An example of the case, where signal leaking into the sidebands for the discriminating variable is significant and cannot be neglected, is now considered: regions E and F in Figure 7.6 illustrate the signal leak into the sidebands.

In order to have reasonable statistics for the signal leak, the width σ of the Gaussian for m is increased from 4.0 to 7.0 during the MC data generation.

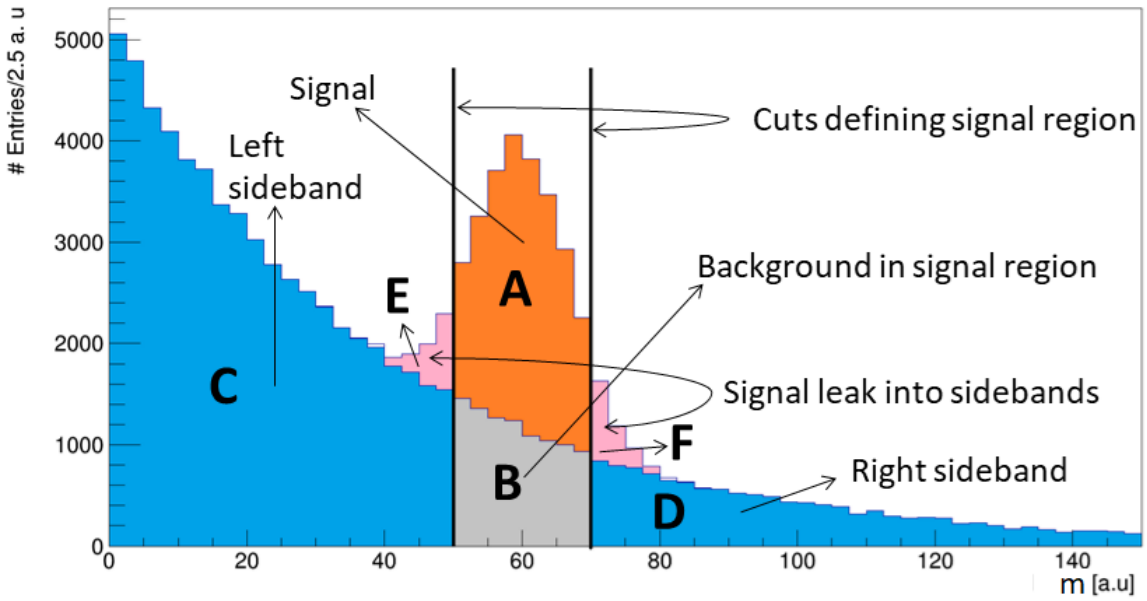


Figure 7.6: Illustrating sideband cuts with a larger signal leak into the sidebands.

The goal here is the same: to obtain the signal and background distributions for v using the distribution of the discriminating variable m.

The problem in this case is a bit more complicated than the previous case. As there is a significant leak of signal events into the sidebands, the sidebands consist of both signal and background events and the same is true for the signal region. A simplified picture of the data distributions according to species for v is shown in Figure 7.7. P^{SR} is the signal region histogram and P^{SB} is the sidebands histogram. The task is to find two weights for each region such that when filled with those weights only signal for v remains, which is labeled as P^{SS}.

Two weights are calculated, namely $W_{\mathrm{sig}}^{\mathrm{SR}}$ and $W_{\mathrm{sig}}^{\mathrm{SB}}$, such that:

$$W_{\mathrm{sig}}^{\mathrm{SR}}\mathrm{P}_i^{\mathrm{SR}} + W_{\mathrm{sig}}^{\mathrm{SB}}\mathrm{P}_i^{\mathrm{SB}} = \mathrm{P}_i^{\mathrm{SS}} \quad (i = \mathrm{A,B}) \tag{7.3}$$

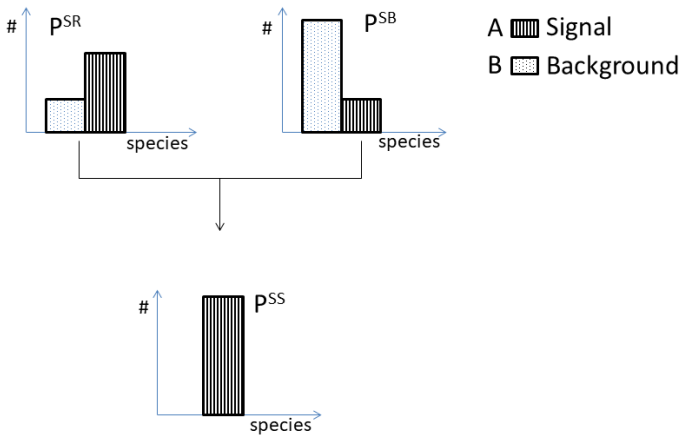


Figure 7.7: Naive model of the signal and background distributions in two bins for the control variable. P^{SR} is the distribution in the SR, P^{SB} is the distribution in the SB and P^{SS} is the sideband-subtracted distribution.

where P_i^{SR}, P_i^{SB} and P_i^{SS} indicate the number of elements in the plots P^{SR},P^{SB} and P^{SS} for species i (Figure 7.7).

By replacing A and B with signal and background and writing both the equations, two equations are obtained:

$$W_{\text{sig}}^{\text{SR}} N_{\text{sig}}^{\text{SR}} + W_{\text{sig}}^{\text{SB}} N_{\text{sig}}^{\text{SB}} = N_{\text{sig}}^{\text{SR}} + N_{\text{sig}}^{\text{SB}} \,, \tag{7.4}$$

$$W_{\text{sig}}^{\text{SR}} N_{\text{bg}}^{\text{SR}} + W_{\text{sig}}^{\text{SB}} N_{\text{bg}}^{\text{SB}} = 0 \tag{7.5}$$

with the constraint $N_{\text{sig}}^{\text{SR}} + N_{\text{sig}}^{\text{SB}} = N_{\text{sig}}$.

Equation (7.4) imposes the condition for N_{sig} to be maximum, whereas (7.5) provides the condition for N_{bg} to be zero.

Solving the system of equations 7.4 and 7.5 results in:

$$W_{\text{sig}}^{\text{SR}} = \frac{N_{\text{sig}} N_{\text{bg}}^{\text{SR}}}{N_{\text{sig}}^{\text{SR}} N_{\text{bg}}^{\text{SB}} - N_{\text{sig}}^{\text{SB}} N_{\text{bg}}^{\text{SR}}} \,, \tag{7.6}$$

$$W_{\text{sig}}^{\text{SB}} = -\frac{N_{\text{sig}} N_{\text{bg}}^{\text{SR}}}{N_{\text{sig}}^{\text{SR}} N_{\text{bg}}^{\text{SB}} - N_{\text{sig}}^{\text{SB}} N_{\text{bg}}^{\text{SR}}} \,. \tag{7.7}$$

As shown in (6.6) and (6.7), the N_{sig} and N_{bg} are obtained from the discriminating variable fit shown in Figure 7.8.

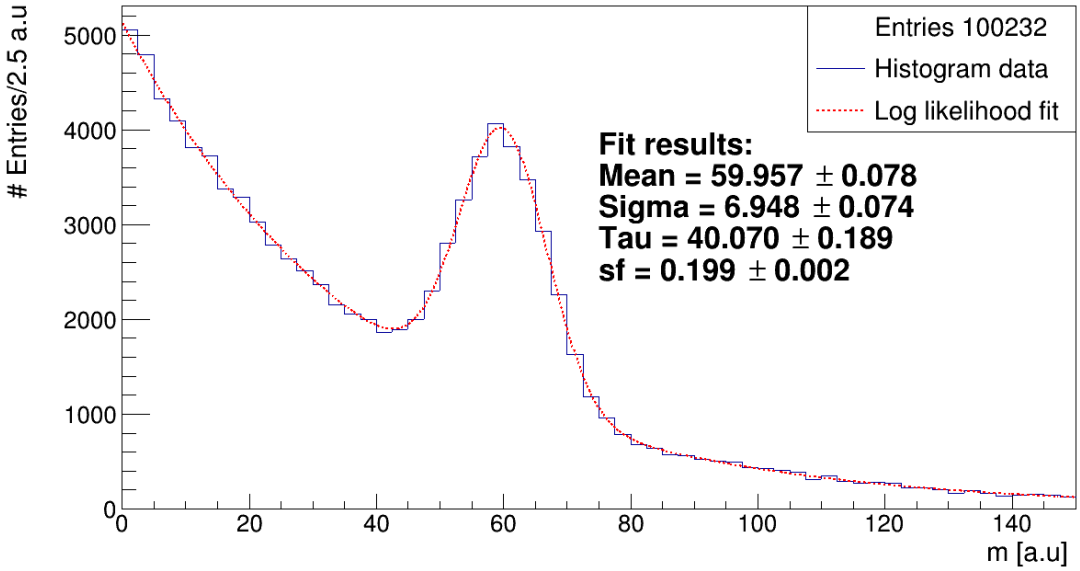


Figure 7.8: Maximum likelihood fit for discriminating variable mass. The width (σ) is increased to 7.0 a.u in this case.

The procedure to obtain the v signal distribution from the weights calculated in (7.6) and (7.7) is shown in Figure 7.9.

Similarly, in order to obtain the weights for the background distribution of the control variable v the following simultaneous equations are to be solved:

$$W_{\text{bg}}^{\text{SR}} N_{\text{bg}}^{\text{SR}} + W_{\text{bg}}^{\text{SB}} N_{\text{bg}}^{\text{SB}} = N_{\text{bg}}^{\text{SR}} + N_{\text{bg}}^{\text{SB}} \,, \tag{7.8}$$

$$W_{\text{bg}}^{\text{SR}} N_{\text{sig}}^{\text{SR}} + W_{\text{bg}}^{\text{SB}} N_{\text{sig}}^{\text{SB}} = 0 \,. \tag{7.9}$$

with the constraint $N_{\text{bg}}^{\text{SR}} + N_{\text{bg}}^{\text{SB}} = N_{\text{bg}}$.

This results in:

$$W_{\text{bg}}^{\text{SR}} = \frac{N_{\text{bg}} N_{\text{sig}}^{\text{SB}}}{N_{\text{sig}}^{\text{SB}} N_{\text{bg}}^{\text{SR}} - N_{\text{sig}}^{\text{SR}} N_{\text{bg}}^{\text{SB}}} \,, \tag{7.10}$$

$$W_{\text{bg}}^{\text{SB}} = -\frac{N_{\text{bg}} N_{\text{sig}}^{\text{SR}}}{N_{\text{sig}}^{\text{SB}} N_{\text{bg}}^{\text{SR}} - N_{\text{sig}}^{\text{SR}} N_{\text{bg}}^{\text{SB}}} \,. \tag{7.11}$$

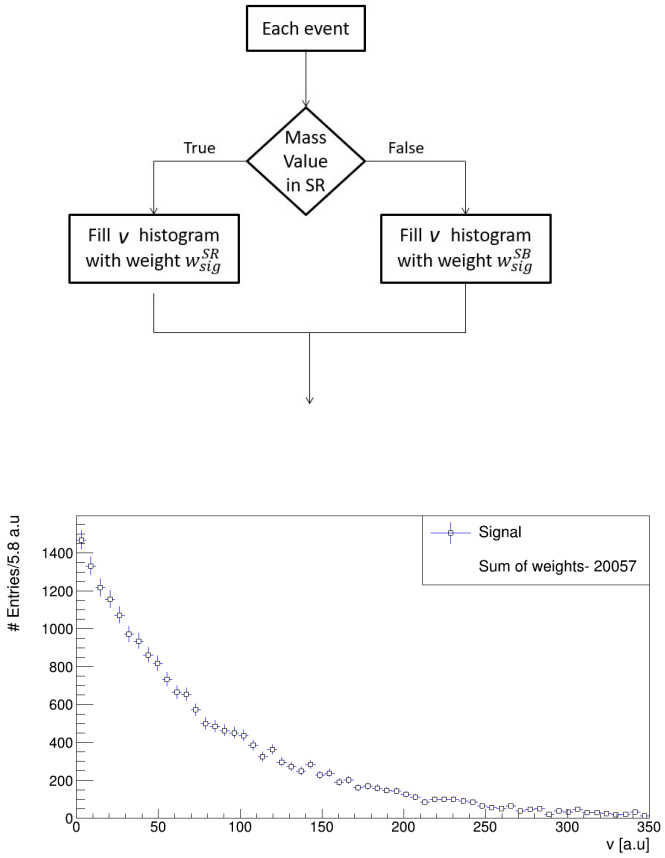


Figure 7.9: Procedure to obtain the signal distribution for v using the weights from (7.6) and (7.7).

These weights are used to obtain the background distribution for the control variable v in a similar way the signal distribution for v has been obtained. Figure 7.10 shows the signal and background distributions for v.

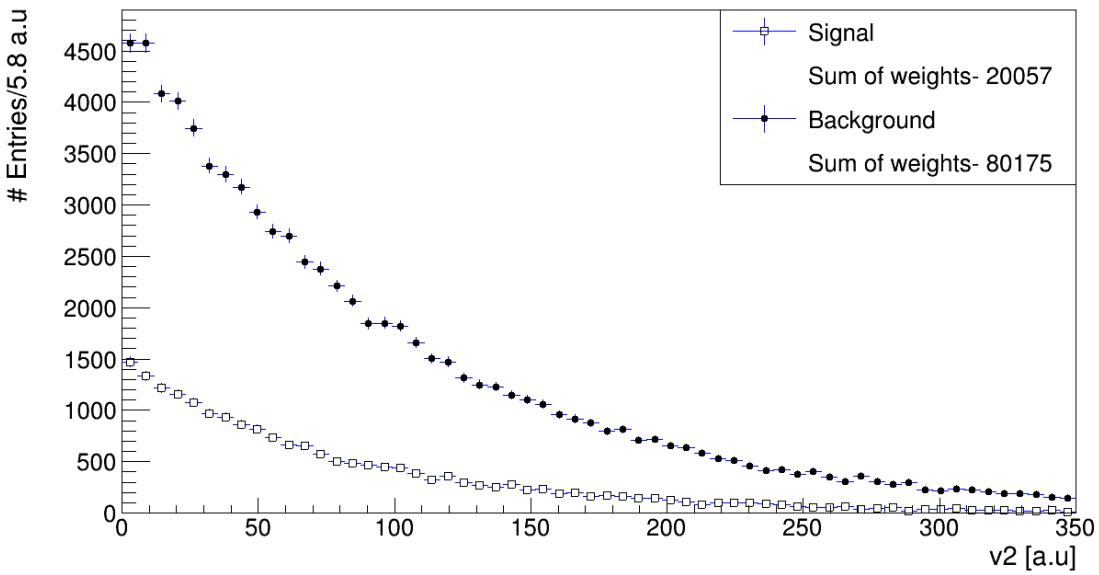


Figure 7.10: Distributions of the signal and the background for v obtained from executing the sideband subtraction technique. (Case in which there is a significant leak of signal events into the sidebands.)

7.5 Comparisons and summary

For comparison, the generated (true) distributions of two control variables for both species are plotted as an overlay to the distributions obtained by the sideband subtraction technique for negligible signal leak case in Figure 7.11. Subsequently, Figure 7.12 illustrates the comparison between the distributions obtained by the sideband subtraction technique with the larger leak and generated distributions. The shapes of the distributions obtained by this technique for v match the shapes of the generated distributions for signal and background.

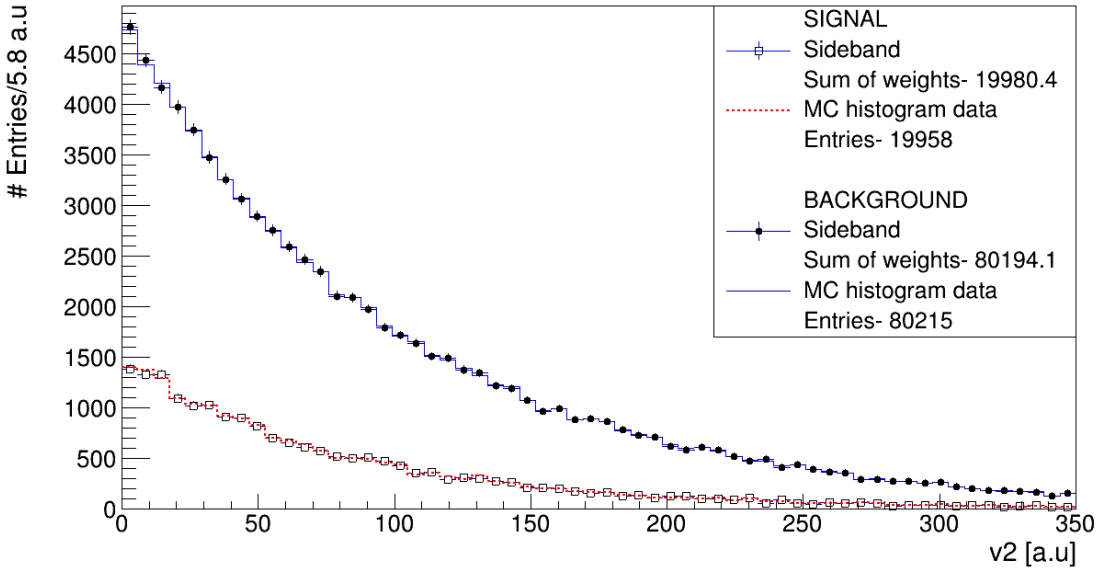


Figure 7.11: Distributions of the signal and the background for v obtained from executing the sideband subtraction technique with generated data overlayed (a case with negligible leak of signal events into the sidebands).

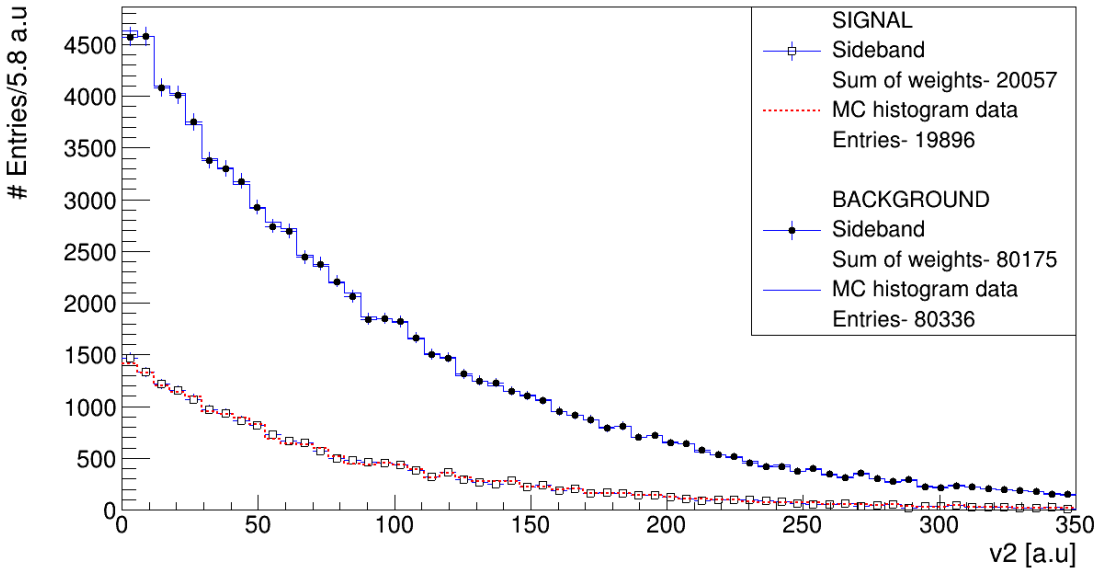


Figure 7.12: Distributions of the signal and the background for v obtained from executing the sideband subtraction technique with generated data overlayed (a case with significant leak of signal events into the sidebands).

8 Maximum likelihood estimation

The Maximum Likelihood Estimation (MLE) is a method to estimate the parameter values of the PDF in a statistical model, given the data set. In application, based on the problem, the MLE method can be employed in two ways; with and without weights. These are discussed in sections 8.1 and 8.2.

8.1 Maximum likelihood estimation without weights

8.1.1 Likelihood function

Consider a random variable x distributed according to a PDF $f(x;\vec{\theta})$ ($\vec{\theta}$ is a set of parameters). It is assumed that the form of the PDF is known but the value(s) of the parameters $\vec{\theta}$ are not known. The idea of the Maximum Likelihood (ML) method is to estimate the value(s) of these parameters given a finite data set so that the model describes the given data as well as possible.

Assuming that the data points are independent of each other and that all the data points are distributed according to the same PDF, the combined probability of these data points, given the PDF, can be written as the product of the probabilities of each data point:

$$f(x_1, x_2, ...x_n|\vec{\theta}) = f(x_1;\vec{\theta}) \cdot f(x_2;\vec{\theta}) \cdot ... \cdot f(x_n;\vec{\theta}) = \prod_{i=1}^{n} f(x_i;\vec{\theta})\,. \tag{8.1}$$

Then the likelihood function is defined as:

$$\mathcal{L}(\vec{\theta}|x_1, x_2, .., x_n) = \prod_{i=1}^{n} f(x_1, x_2, .., x_n|\vec{\theta})\,. \tag{8.2}$$

The probability of the distribution of the data points given the PDF's parameters is the same as the likelihood of the PDF given the data set. Therefore, maximizing the likelihood function yields the parameter values which best model the data set.

8.1.2 Maximum likelihood estimator

The task is to estimate the values of the parameters θ_i by maximizing the likelihood function $\mathcal{L}$ such that the PDF is the best description of the data set. The values of θ_i,

that maximize the likelihood $\mathcal{L}$, are the maximum likelihood estimators for the parameters.

Maximizing the likelihood $\mathcal{L}$ is given by:

$$\frac{\partial \mathcal{L}}{\partial \theta_i} = 0; \quad (i = 1, 2, 3, .., n)\,. \tag{8.3}$$

It is often quite difficult to maximize the likelihood function $\mathcal{L}$ directly, however maximizing the $\log \mathcal{L}$ is simpler because the product of logarithms can be expressed as a sum. Logarithm is a monotonic function and therefore it ensures that the maximum value of the logarithm of the likelihood ($\log \mathcal{L}$) occurs at the same point as the original likelihood function:

$$\frac{\partial \log \mathcal{L}}{\partial \theta_i} = 0; \quad (i = 1, 2, 3, .., n)\,, \tag{8.4}$$

$$\frac{\partial \log \mathcal{L}}{\partial \theta_1} + \frac{\partial \log \mathcal{L}}{\partial \theta_2} + ... + \frac{\partial \log \mathcal{L}}{\partial \theta_n} = 0; \quad (i = 1, 2, 3, .., n)\,. \tag{8.5}$$

8.1.3 Discriminating variable fit

As explained in section 6.3, the discriminating variable (mass) fit is a maximum likelihood fit without weights. The form of the PDF is given by:

$$p(x) = f_s\, g(x) + (1 - f_s) h(x) \tag{8.6}$$

where $g(x)$ and $h(x)$ are given in (6.4) and (6.5) and f_s given in Table 6.1 .

The log likelihood is given by:

$$\log \mathcal{L} = \sum_{i=1}^{N} \log p(x_i)\,. \tag{8.7}$$

The fit is performed in ROOT using the TMinuit class. To quantify the goodness of the fit, a bin by bin pull plot is drawn along with the fitted distribution as shown in Figure 8.1.

The bin by bin pull values are calculated by:

$$\text{Pull} = \frac{z_i - f(x_i)}{\sigma_i} \tag{8.8}$$

where x_i is the value of the x coordinate at the center of the ith bin, $f(x_i)$ the value of the PDF at x_i, z_i the bin contents and σ_i the error of the ith bin.

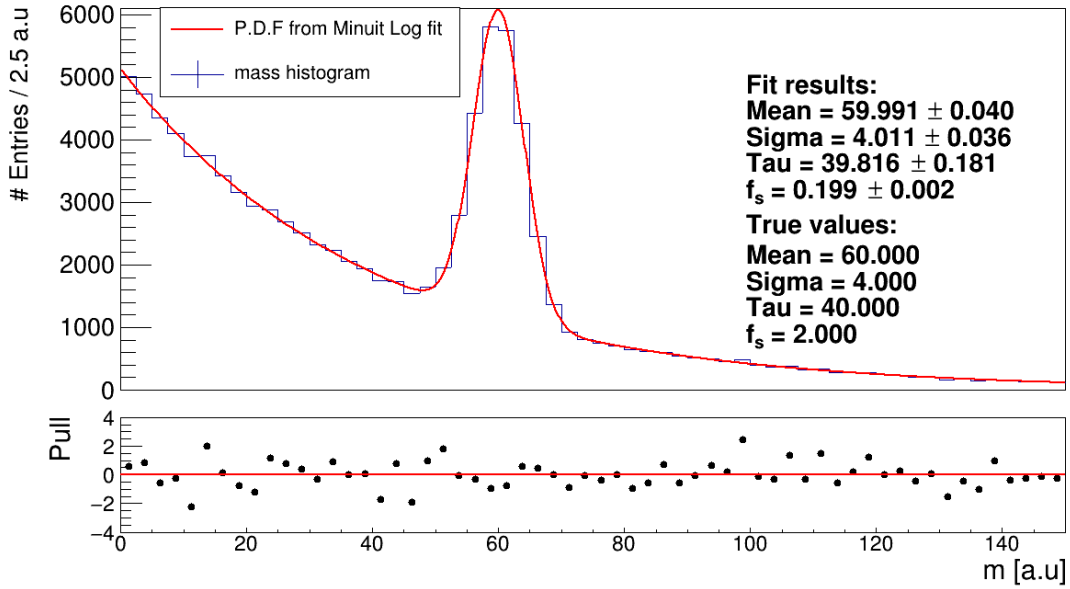


Figure 8.1: Unbinned fit for the discriminating variable (mass) along with the pull plot

8.2 Maximum likelihood fit with weights

As explained in section 6.3 the histograms obtained by the sPlot technique are weighted histograms. Therefore weighted MLEs are performed.

The likelihood function for the weighted case is given by:

$$\mathcal{L}(\vec{\theta}|x_1, x_2, .., x_n) = \prod_{i=1}^{n} f(x_i; \vec{\theta})^{w_i} \tag{8.9}$$

where w_i is the weight for the ith event.

The log likelihood $\log \mathcal{L}$ is given by:

$$\log \mathcal{L} = \sum_{i=1}^{N} w_i \cdot \log f(x_i)\,. \tag{8.10}$$

The weights w_i are expressed in the power as seen in (8.9), because, for example, an event filled with weight two is equivalent to filling the same event twice.

The Figures 8.2 and 8.3 show the weighted fit results for the sPlot histograms with pull distributions for the control variables v.

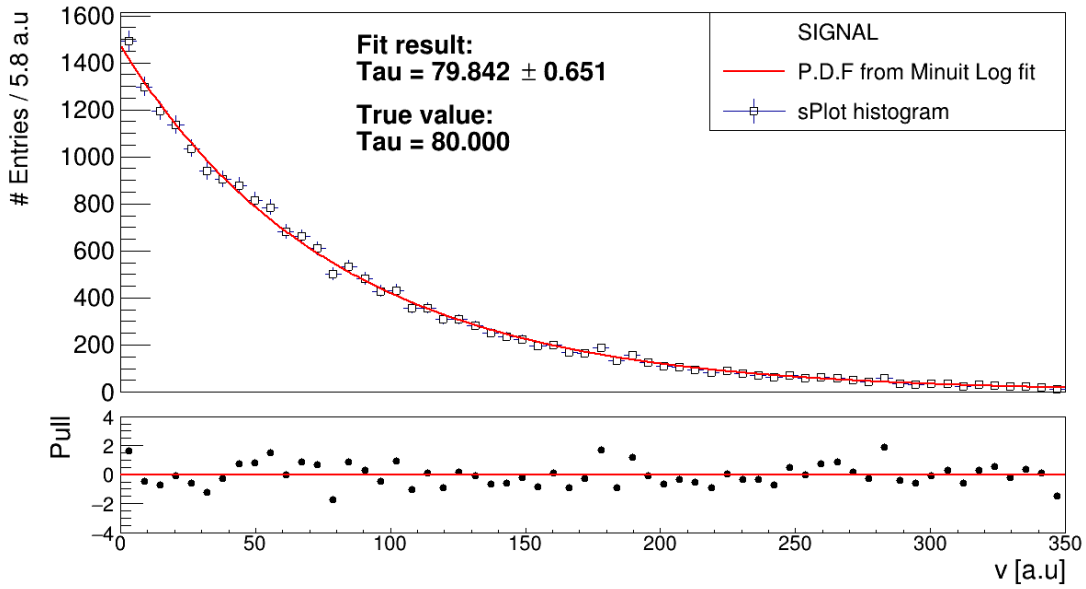


Figure 8.2: Unbinned weighted maximum likelihood fit for the v signal distribution along with the pull distribution.

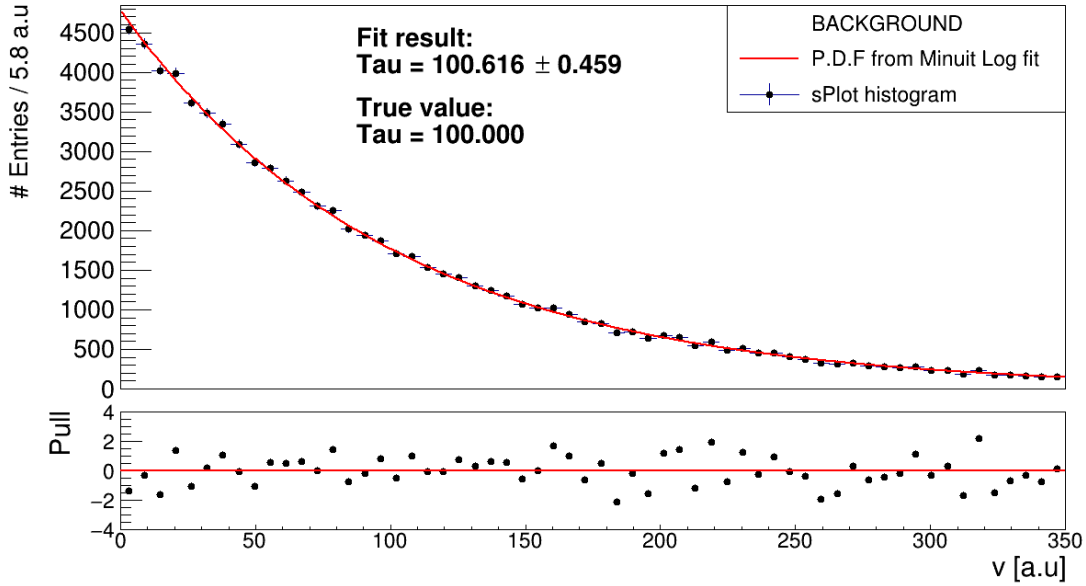


Figure 8.3: Unbinned weighted maximum likelihood fit for the v background distribution along with the pull distribution.

8.3 Summary

From the fit values and the pull distributions it is clear, that the central values of the fit results for the signal and background PDFs of the control variable are within 3σ of the true values (Table 6.1). Therefore, it is concluded, that by using the sPlot technique the distributions of the two species of the control variable are obtained accurately.

In addition, high statistics tests have been performed in order to confirm that the values of the errors obtained from TMinuit fits are accurate. The mechanism of the tests performed and results are discussed in the next chapter.

9 Correction of uncertainties for weighted maximum likelihood estimation

9.1 Closure tests for MLE

In the process of executing the sPlot technique the MLE method is used for the discriminating variable fit and a weighted MLE method is used for fitting the control variable distributions obtained by the sPlot technique. In order to verify that the MLE using TMinuit is accurate, a pseudo experiment cycle consisting of data generation, obtaining sPlot distributions and then fitting the sPlot distributions using the MLE method is performed 10 000 times to obtain high statistics. Then the pull distributions for all the fitted parameters are drawn and analyzed. The pull distribution values are calculated by:

$$\text{Pull} = \frac{\theta_i - \theta_{tr}}{\sigma_{\theta_i}} \tag{9.1}$$

where θ_i is the central value of the parameter θ and σ_{θ_i} the error on θ for ith trial. θ_{tr} is the true value of θ. (In the example, the discriminating variable fit has 4 parameters and control variable v fit has only one parameter τ.)

The expected results for unbiased fits (for parameter θ) are, that
The mean value of the pull distribution $P_\mu^\theta = 0$, and
The width of the pull distribution $P_\sigma^\theta = 1$.

9.1.1 Closure test results of MLE for the discriminating variable

The pull distribution for the parameter μ of the discriminating variable is shown in Figure 9.1. The pull distributions for all the other parameters are shown in Appendix B but the results are summarized in Table 9.1.

As shown in Table 9.1, the mean value of the pull distribution (P_μ^θ) is zero within the statistical uncertainty and the width (P_σ^θ) is equal to one for all the parameters. Therefore it is concluded that the MLE for the discriminating variable using TMinuit is accurate.

Parameter	θ_{tr}	$\bar{\theta}_{fit}$	P^{θ}_{μ}	P^{θ}_{σ}
μ	60.0	59.9997 ± 0.0003	-0.0086 ± 0.0100	1.0050 ± 0.0071
σ	4.0	4.0000 ± 0.0003	0.0003 ± 0.0099	0.9904 ± 0.0070
τ	40.0	40.0001 ± 0.0018	0.0005 ± 0.0100	1.0030 ± 0.0069
f_s	0.2	0.2000 $\pm 1.6 \times 10^{-5}$	0.0058 ± 0.0098	0.9872 ± 0.0069

Table 9.1: θ_{tr} is the true value of the parameter θ, $\bar{\theta}_{fit}$ the mean of the central value of the fit parameter, P^{θ}_{μ} is the mean value of the pull distribution for parameter θ and P^{θ}_{σ} the width of the pull distribution for parameter θ.

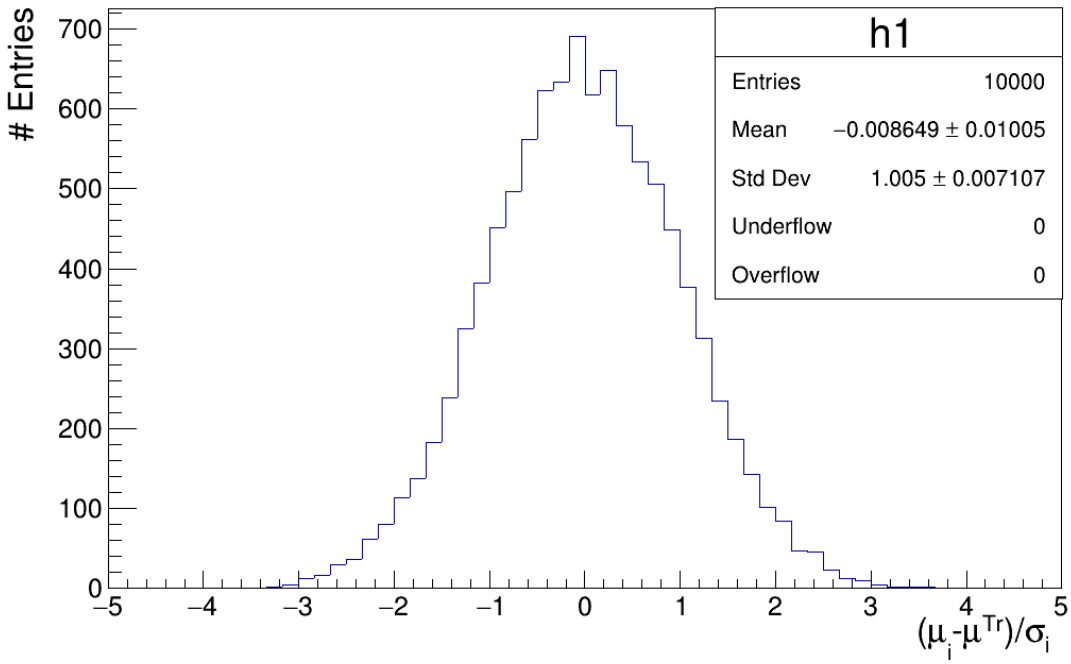


Figure 9.1: Pull distribution for parameter μ for the discriminating variable.

9.1.2 Closure test results of MLE for the control variable

The pull distributions for the parameter τ corresponding to the control variable v for both the signal and the background are shown in Figures 9.2 and 9.3. The results are summarized in Table 9.2. The mean value is zero within the statistical uncertainty. However, the width (P^{θ}_{σ}) is not equal to one. Hence it is concluded that the errors obtained from MLE using TMinuit for the case with weights are not accurate and therefore the uncertainties reported by the MLE need to be corrected.

Parameter	θ_{tr}	$\bar{\theta}_{fit}$	$\bar{\sigma}_{TM}$
τ_{sig}	80.0	80.0005 ± 0.0080	0.6519 ± 9.5 $\times 10^5$
τ_{bg}	100.0	99.9981 ± 0.0047	0.4539 ± 3.4 $\times 10^5$

Parameter	P^{θ}_{μ}	P^{θ}_{σ}
τ_{sig}	-0.0164 ± 0.0123	1.2396 ± 0.0087
τ_{bg}	-0.0120 ± 0.0105	1.0541 ± 0.0074

Table 9.2: θ_{tr} is the true value of the parameter θ, $\bar{\theta}_{fit}$ the mean of the central value of the fit parameter, $\bar{\sigma}^{\theta}_{TM}$ is the mean of the uncertainty on the central value of parameter θ (before correction), P^{θ}_{μ} is the mean value of the pull distribution for parameter θ and P^{θ}_{σ} the width of the pull distribution for parameter θ.

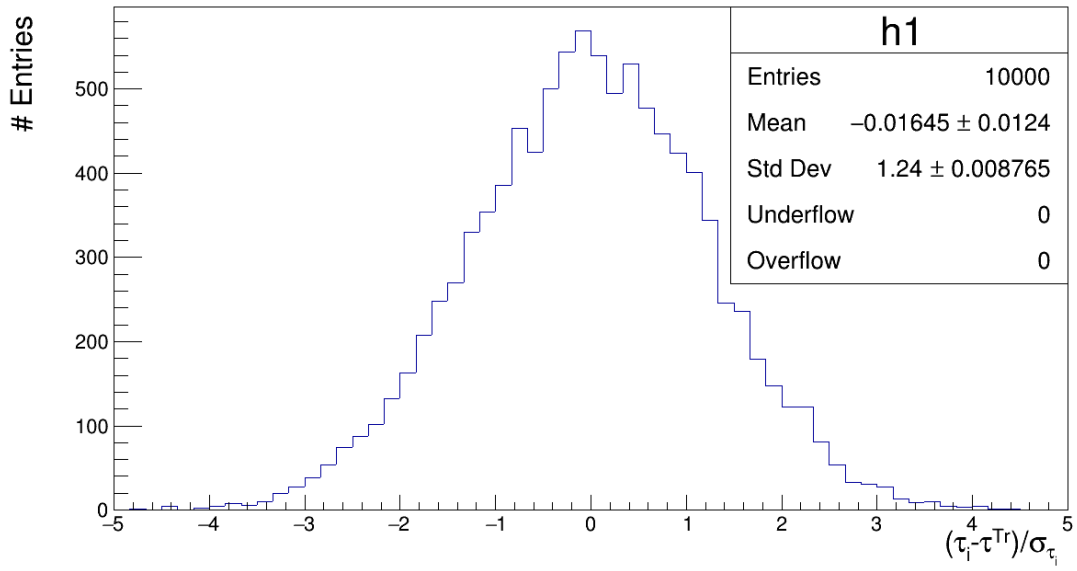


Figure 9.2: Pull distribution for τ signal

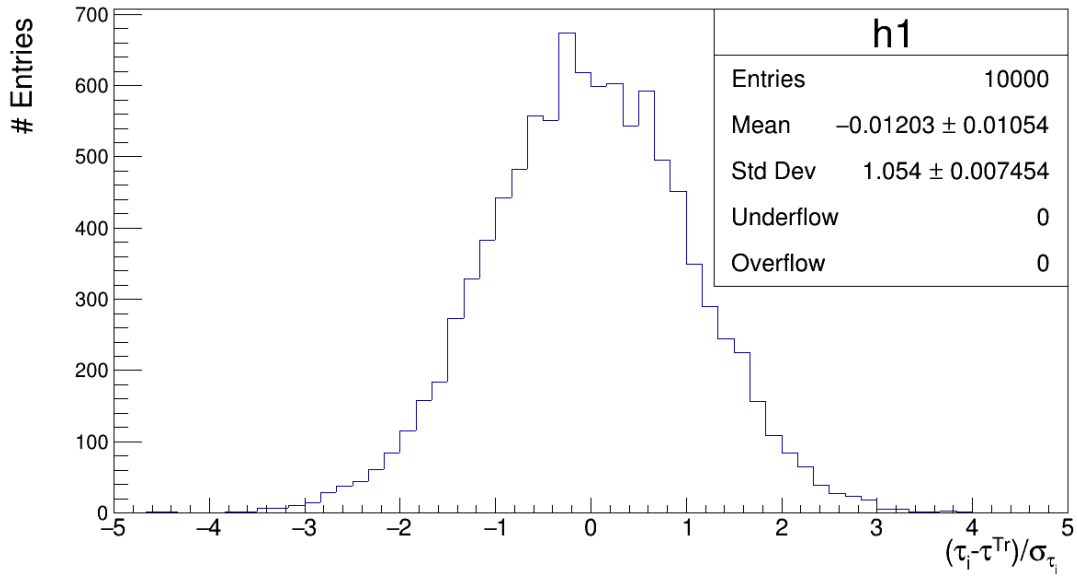


Figure 9.3: Pull distribution for τ background

9.2 Error correction for weighted maximum likelihood estimation

The log likelihood ($\log \mathcal{L}$) for the weighted case is given by:

$$\log \mathcal{L} = \sum_{i=1}^{N} w_i \cdot \log f(x_i) \tag{9.2}$$

where w_i are the weights and N the total number of events.

The central value of the parameter estimation is asymptotically distributed according to a normal distribution about the true value θ_{Tr}. However, the statistical uncertainties on the parameters are biased as shown in the pull distributions in Figures 9.2 and 9.3.

The reason for the miscalculation of the statistical uncertainties in TMinuit is the following. The sum of the weights does not scale with the total number of events. So the sum of weights is larger or smaller than N:

$$\sum_{i=1}^{N} w_i > N \quad \text{or} \quad \sum_{i=1}^{N} w_i < N \,. \tag{9.3}$$

Therefore a correction for the statistical uncertainties of the fitted parameters to the weighted MLE is necessary.

The uncertainties are calculated by taking the square root of the Hesse covariance matrix elements. However TMinuit first calculates the inverse Hesse covariance matrix as shown below [9]:

$$G_{mn} = \sum_{i=1}^{N} w_i \, \frac{1}{(f(x_i))^2} \, \frac{df_i}{d\theta_m} \, \frac{df_i}{d\theta_n} \,. \tag{9.4}$$

Therefore, a correction should be applied to TMinuit's inverse Hesse covariance matrix. The Hesse covariance matrix is the inverse of G.

The correction term is given by [9]:

$$F_{mn} = \sum_{i=1}^{N} w_i^2 \, \frac{1}{(f(x_i))^2} \, \frac{df_i}{d\theta_m} \, \frac{df_i}{d\theta_n} \,. \tag{9.5}$$

The corrected covariance matrix is given by:

$$G^{-1}_{\text{corrected}} = G^{-1} F G^{-1} \,. \tag{9.6}$$

In Appendix A the derivation [9] of this result is retraced for a PDF with one parameter only.

In the case when the weights are equal to one:

$$G^{-1}_{\text{corrected}} = G^{-1} \tag{9.7}$$

Therefore in the case of the MLE without weights, a correction to the covariance matrix is not needed.

9.3 Correction factor for the control variable fit

The error correction discussed in the previous section is performed for the control variable v in the example. As the exponential function has only one parameter (τ) the covariance matrix is a 1×1 matrix.

The form of the PDF is given by:

$$h(x) = \frac{1}{\tau \cdot (e^{-\frac{a}{\tau}} - e^{-\frac{b}{\tau}})} e^{-\frac{x}{\tau}} \tag{9.8}$$

where a and b are the lower and upper limits of the domain of the function.

The F matrix term is given by:

$$F_{11} = \sum_{i=1}^{N} (w_i)^2 \left(\frac{1}{h(x_i)} \frac{\partial h(x_i;\tau)}{\partial \tau} \right)^2 \tag{9.9}$$

$$\frac{1}{h(x_i)} \frac{\partial h(x_i;\tau)}{\partial \tau} = \frac{(\frac{x-a}{\tau} - 1)e^{-\frac{a}{\tau}} - (\frac{x-b}{\tau} - 1)e^{-\frac{b}{\tau}}}{\tau \cdot (e^{-\frac{a}{\tau}} - e^{-\frac{b}{\tau}})} \,. \tag{9.10}$$

By applying the above correction term to the variance of τ we obtain the new Hesse covariance matrix, which is given by (9.6).

9.4 Closure test results for the control variable after correction to covariance matrix

The closure tests as disscussed in section 9.1 are performed again with the corrected uncertainties and the corresponding pull distributions are shown in Figures 9.4 and 9.5. The results are summarized in Table 9.3.

The mean values of the pull distributions for both the signal and the background for v are within 2 σ of the expected value. The width (P^{θ}_{σ}) of the pull distributions for both the signal and the background for v are also within 2 σ of the expected value.

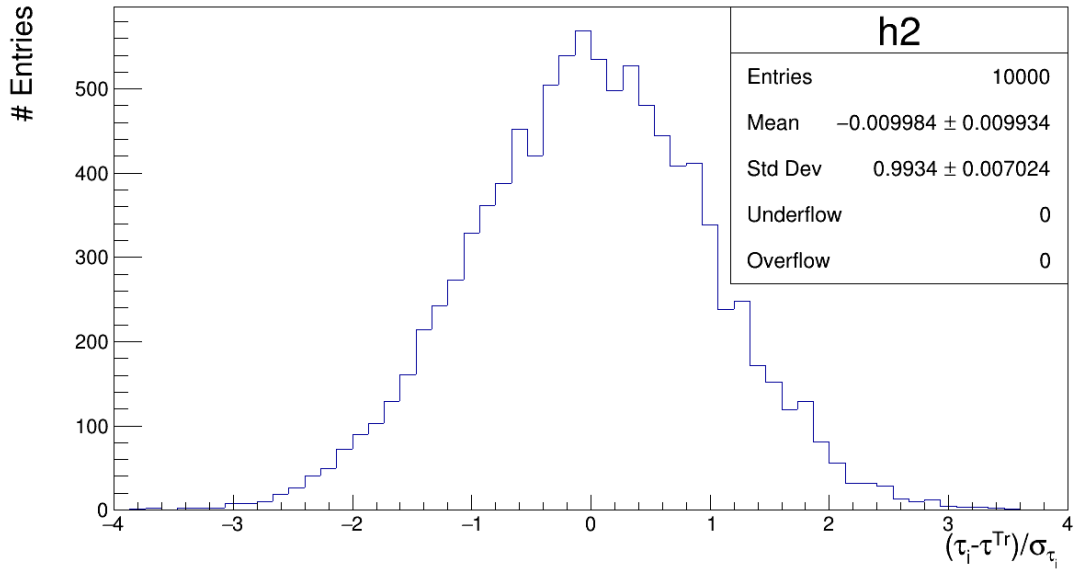


Figure 9.4: The pull distribution for v signal using corrected errors

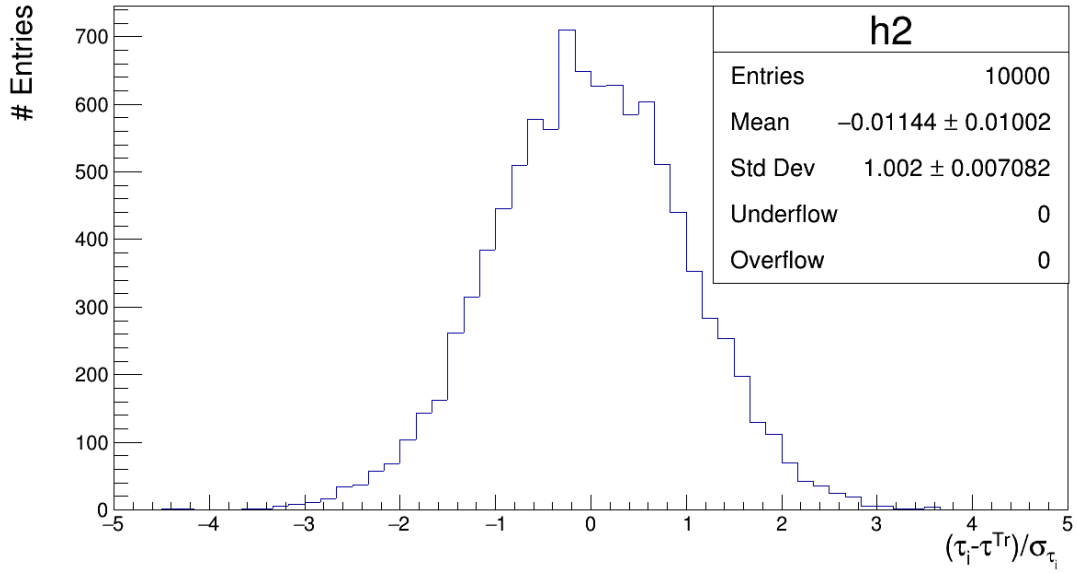


Figure 9.5: The pull distribution for v background using corrected errors

Parameter	θ_{tr}	$\bar{\theta}_{fit}$	$\bar{\sigma}_{cor}$
τ_{sig}	80.0	80.0005 $\pm$ 0.0080	0.8132 $\pm$ 0.0001
τ_{bg}	100.0	99.9981 $\pm$ 0.0047	0.4777 $\pm$ 3.84 $\times 10^{-5}$

Parameter	P_{μ}^{θ}	P_{σ}^{θ}
τ_{sig}	-0.0099 $\pm$ 0.0099	0.9933 $\pm$ 0.0070
τ_{bg}	-0.0114 $\pm$ 0.0100	1.0016 $\pm$ 0.0070

Table 9.3: θ_{tr} is the true value of the parameter θ, $\bar{\theta}_{fit}$ the mean of the central value of the fit parameter, $\bar{\sigma}_{cor}^{\theta}$ is the mean of the uncertainty on the central value of parameter θ (after correction), P_{μ}^{θ} is the mean value of the pull distribution for parameter θ and P_{σ}^{θ} the width of the pull distribution for parameter θ.

10 Conclusions and Outlook

The aim of this thesis was to explore a statistical technique, namely the sPlot technique [8], in the context of data analysis for high energy physics. The motivation to study this technique came from the study of rare decays of B_s^0 and B^0 mesons from the ATLAS experiment [7] at CERN.

The sPlot technique was applied, as an example application, to a MC pseudo data set. The data was generated with two variables, discriminating variable and control variable, and two species, signal and background, with 100 000 events. The discriminating variable consists of a Gaussian signal with a decaying exponential background. The control variable is modeled by two decaying exponential distributions with different lifetimes, one for the signal and one for the background.

The objective was to extract the signal and background distributions of the control variable by employing the sPlot technique using the known PDF form of the discriminating variable. The implementation of the technique was divided into three steps. Firstly, a Maximum Likelihood Estimation (MLE) was performed for the discriminating variable. A set of sWeights for each species were calculated in the second step and lastly the control variable for each of the two species was filled in histogram bins with corresponding sWeights to obtain the desired distribution.

In order to quantify the agreement between the distributions obtained from the sPlot technique and the generated MC distributions, weighted MLEs were performed for the sPlot distributions and the parameters obtained from the MLE were compared with the parameters of the generated MC distributions. This comparison was not made only once, but repeated 10 000 times in order to obtain statistically meaningful results. For each run, this three step cycle of data generation, implementing the sPlot technique and weighted MLE for the sPlot distributions was executed. This produced 10 000 independent results; a sufficiently large statistics to perform closure tests.

It was found in the closure tests that the MLE executed by the TMinuit package in ROOT for the discriminating variable (during the sWeights calculation) was accurate and not biased. However, the MLE for the sPlot distributions did not yield accurate statistical uncertainties on the central value of the parameters. The reason is that the distribution of the discriminating variable is not weighted but the control variable distributions obtained by the sPlot technique are. The calculation of the Hesse covariance matrix elements by TMinuit does not take the weights properly into account. Therefore a correction to the covariance matrix [9] was applied for the weighted MLEs.

For comparison, the same example was also studied using the sideband subtraction technique. In this technique, by applying two cuts, the whole range of the discriminating variable is divided into signal region and sidebands, such that the sidebands contain mostly background. The most general case with signal leaking into sidebands was studied and the problem was solved with an idea to calculate weights according to the species (section 7.3).

The advantage of the sPlot technique over the sideband subtraction technique is that it does not need signal region cuts. Therefore it is preferable over the sideband subtraction technique. The sPlot technique is also preferred when the data contains more number of species. Subsequent to this thesis it will be interesting to study the technique more thoroughly with new variables and different distributions. The sPlot technique can be used in the data analysis of rare B meson decays. Specifically, it can be used as a replacement for the sideband subtraction technique in the "Data - Monte Carlo comparisons" in the reference channel $B^{\pm} \to J/\psi\ K^{\pm}$ in order to validate the MC data set for $B_s^0 \to \mu^+\mu^-$ analysis.

11 Appendix A

In this appendix the derivation of the corrected covariance matrix [9] for the weighted maximum likelihood estimation is retraced to the most simple case of a PDF with one parameter only (chapter 9, (equation 9.6)).

Let the variables be divided into two groups x and y, x depending directly on the parameters $\vec{\theta}$ to be estimated and y depending indirectly on x:

$$P(x,y|\theta) = P(x|\theta)\,Q(y|x) \text{ with normalization } \int P(x|\theta)dx = \int Q(y|x)dy = 1\,. \quad (11.1)$$

The PDF can be written as

$$f(x,y|\theta) = \frac{P(x|\theta)Q(y|x)e(x,y)}{\int P(x|\theta)Q(y|x)e(x,y)dxdy} \quad (11.2)$$

with weights $w = \frac{1}{e}$.

The normalization condition and definition of estimator E[.] is given by

$$\int\int f(x,y;\theta)dx\,dy = 1\,, \quad (11.3)$$

$$E[g(x)] = \int g(x)f(x,y;\theta)dx \quad (11.4)$$

with $f(x,y) = \frac{PQe}{\int PQe}$. For simplicity we assume that the PDF has only one parameter and the likelihood function is given by:

$$L = \prod_i^N \frac{P(x_i|\theta)Q(y_i|x_i)e(x_i,y_i)}{\int P(x|\theta)Q(y|x)e(x,y)dxdy}\,. \quad (11.5)$$

The log likelihood is given by:

$$\log L = \sum_i \left(\frac{1}{e_i}\log P_i\right) \quad (11.6)$$

with $w_i = \frac{1}{e_i}$.

The variance is given by (section 7.3.2 [9]):

$$\sigma_{\hat{\theta}}^2 = \frac{E\left[\zeta(\theta_0)^2\right]}{E\left[\frac{\partial\zeta(\theta)}{\partial\theta}\right]^2} \tag{11.7}$$

with:

$$\zeta(\theta) = \frac{1}{N}\frac{\partial \log L}{\partial\theta} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{e_i}\frac{P_i^{'}}{P_i} \tag{11.8}$$

and:

$$\frac{\partial\zeta(\theta)}{\partial\theta} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{e_i}\left(\frac{P_i^{''}}{P_i} - \frac{(P_i^{'})^2}{P_i^2}\right). \tag{11.9}$$

Calculating the estimation for (11.9) in order to obtain the denominator term in (11.7):

$$E\left[\frac{\partial\zeta(\theta)}{\partial\theta}\right] = \frac{1}{N}\sum_{i=1}^{N}\left(E\left[\frac{1}{e_i}\frac{P_i^{''}}{P_i}\right] - E\left[\frac{1}{e_i}\frac{(P_i^{'})^2}{P_i^2}\right]\right). \tag{11.10}$$

As all the terms in the summation are the same we obtain:

$$E\left[\frac{\partial\zeta(\theta)}{\partial\theta}\right] = \left(E\left[\frac{1}{e_i}\frac{P_i^{''}}{P_i}\right] - E\left[\frac{1}{e_i}\frac{(P_i^{'})^2}{P_i^2}\right]\right). \tag{11.11}$$

The first term in the above equation is zero because:

$$E[\frac{1}{ep}P''] = E[\frac{1}{ep}\frac{\partial^2 P}{\partial\theta^2}], \tag{11.12}$$

$$E[\frac{1}{ep}\frac{\partial^2 P}{\partial\theta^2}] = \int\frac{P''}{eP}\frac{PQe}{\int PQe} = \frac{\int P''Q}{\int PQe}, \tag{11.13}$$

$$\frac{\int P''Q}{\int PQe} = \int\frac{\partial^2}{\partial\theta^2}P(x|\theta)Q(y|x)dxdy, \tag{11.14}$$

$$\frac{\int P''Q}{\int PQe} = \frac{\partial^2}{\partial\theta^2}\left(\int P(x|\theta)Q(y|x)dx\right), \tag{11.15}$$

$$\frac{\int P''Q}{\int PQe} = \frac{\partial^2}{\partial\theta^2}(1) = 0. \tag{11.16}$$

Therefore we obtain:

$$E\left[\frac{\partial\zeta(\theta)}{\partial\theta}\right] = -E\left[\frac{1}{e}\frac{(P^{'})^2}{P^2}\right]. \tag{11.17}$$

Similarly, the numerator term in (11.7) can be obtained by estimating (11.10):

$$E\left[\zeta(\theta_0)^2\right] = \frac{1}{N^2} E\left[\left(\sum_{i=1}^{N} \frac{1}{e_i}\frac{P_i^{'}}{P_i}\right)^2\right], \tag{11.18}$$

$$E\left[\zeta(\theta_0)^2\right] = \frac{1}{N^2} E\left[\left(\sum_{i=1}^{N} \frac{1}{e_i^2}\frac{P_i^{'}}{P_i}\right)^2 + \sum_{i,j=1}^{N} \frac{1}{e_i e_j}\frac{P_i^{'}P_j^{'}}{P_i P_j}\right]. \tag{11.19}$$

By simplifying we get:

$$E\left[\zeta(\theta_0)^2\right] = \frac{1}{N} E\left[\frac{1}{e^2}\left(\frac{P^{'}}{P}\right)^2\right]. \tag{11.20}$$

Substituting (11.17) and (11.20) in (11.7), we get:

$$\sigma_{\hat{\theta}}^2 = \frac{\frac{1}{N} E\left[\frac{1}{e^2}\left(\frac{P^{'}}{P}\right)^2\right]}{E\left[\frac{1}{e}\frac{(P^{'})^2}{P^2}\right]}. \tag{11.21}$$

Further the variance can be expressed as:

$$\sigma_{\theta}^2 = \frac{D_2^{'}}{N D_1^{'2}} \tag{11.22}$$

with:

$$D_2^{'} = E\left[\left(\frac{P^{'}}{Pe}\right)^2\right] \tag{11.23}$$

and:

$$D_1^{'} = E\left[\frac{1}{e}\left(\frac{P^{'}}{P}\right)^2\right]. \tag{11.24}$$

Further, by using (11.4) $D_1{}'$ and $D_2{}'$ can be expressed as:

$$D_1{}' = \frac{\int \frac{P'^2 Q}{P}}{\int PeQ}, \tag{11.25}$$

$$D_2{}' = \frac{\int \frac{P'^2 Q}{Pe}}{\int PeQ}. \tag{11.26}$$

The task now is to find the $D_1^{'}$ and $D_2^{'}$ which should be substituted in (11.22).

Recall that the the log likelihood is given by:

$$\log L = \sum_i \left(\frac{1}{e_i} \log P_i \right) . \tag{11.27}$$

We define:

$$\hat{D}_1^{'} = -\frac{\partial^2}{\partial \theta^2} \left[\frac{1}{N} \sum_{i=1}^{N} \frac{\log P_i}{e_i} \right] , \tag{11.28}$$

$$\hat{D}_2^{'} = -\frac{\partial^2}{\partial \theta^2} \left[\frac{1}{N} \sum_{i=1}^{N} \frac{\log P_i}{e_i^2} \right] . \tag{11.29}$$

By differentiating (11.28) twice we obtain:

$$\hat{D}_1^{'} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{e_i} \left(\frac{P_i^{'}}{P_i} \right)^2 - \frac{1}{N} \sum_{i=1}^{N} \frac{1}{e_i} \left(\frac{P_i^{''}}{P_i} \right) , \tag{11.30}$$

$$E\left[\hat{D}_1^{'} \right] = E\left[\frac{1}{N} \sum_{i=1}^{N} \frac{1}{e_i} \left(\frac{P_i^{'}}{P_i} \right)^2 - \frac{1}{N} \sum_{i=1}^{N} \frac{1}{e_i} \left(\frac{P_i^{''}}{P_i} \right) \right] , \tag{11.31}$$

$$E\left[\hat{D}_1^{'} \right] = \frac{1}{N} \sum_{i=1}^{N} E\left[\frac{1}{e_i} \left(\frac{P_i^{'}}{P_i} \right)^2 \right] - \frac{1}{N} \sum_{i=1}^{N} E\left[\frac{1}{e_i} \left(\frac{P_i^{''}}{P_i} \right) \right] . \tag{11.32}$$

The second term in above equation is 0 because of (11.15) and (11.16).
So we obtain:

$$E\left[\hat{D}_1^{'} \right] = E\left[\frac{1}{e} \left(\frac{P^{'}}{P} \right)^2 \right] , \tag{11.33}$$

$$E\left[\hat{D}_1^{'} \right] = D_1^{'} . \tag{11.34}$$

Similarly:

$$E\left[\hat{D}_2^{'} \right] = D_2^{'} . \tag{11.35}$$

Thus the estimation of $D_1^{'}$ and $D_2^{'}$ given in (11.24) and (11.23) can be found by using $\hat{D}_1^{'}$ and $\hat{D}_2^{'}$. Once $D_1^{'}$ and $D_2^{'}$ is known, the variance is found from (11.22).

12 Appendix B

The following figures are the pull distributions from the discriminating variable fit.

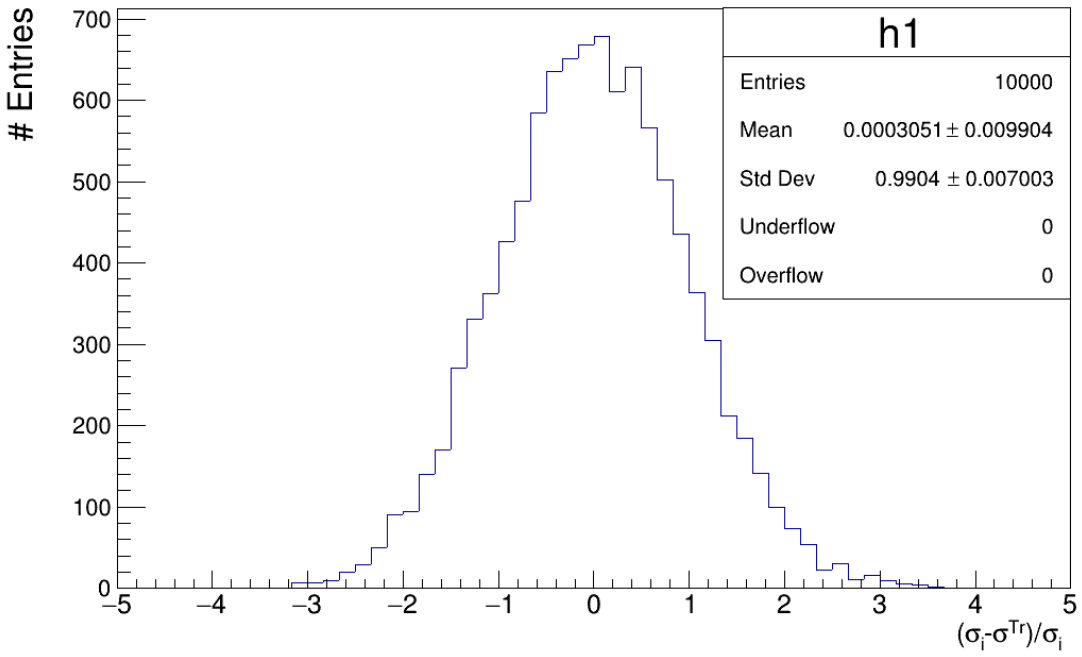


Figure 12.1: The pull distribution for parameter σ in the discriminating variable fit.

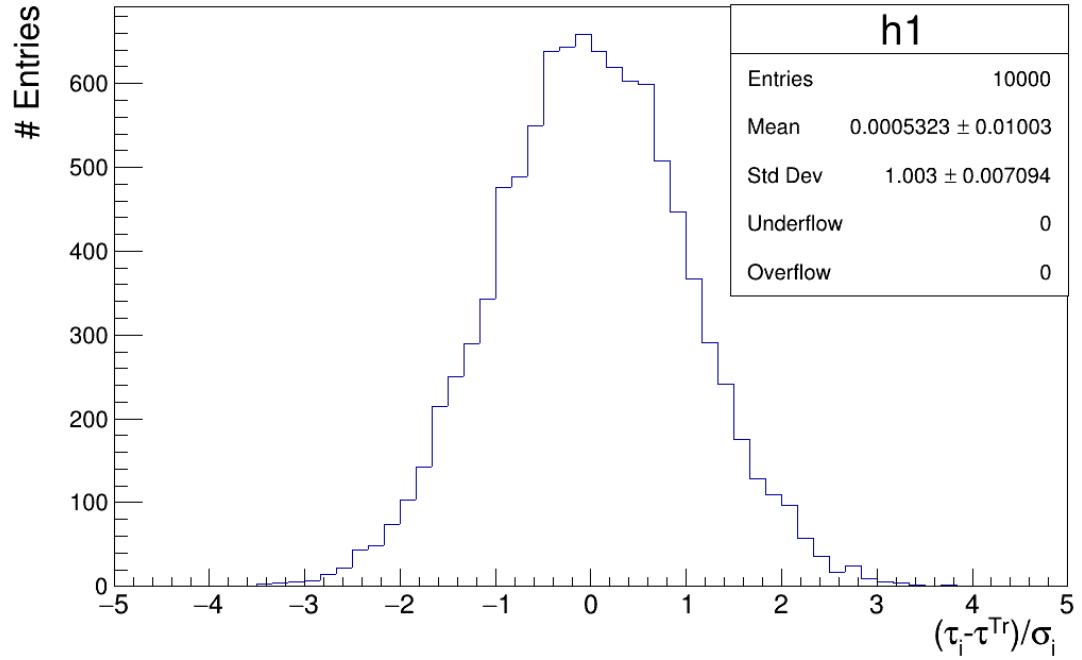


Figure 12.2: The pull distribution for parameter τ in the discriminating variable fit.

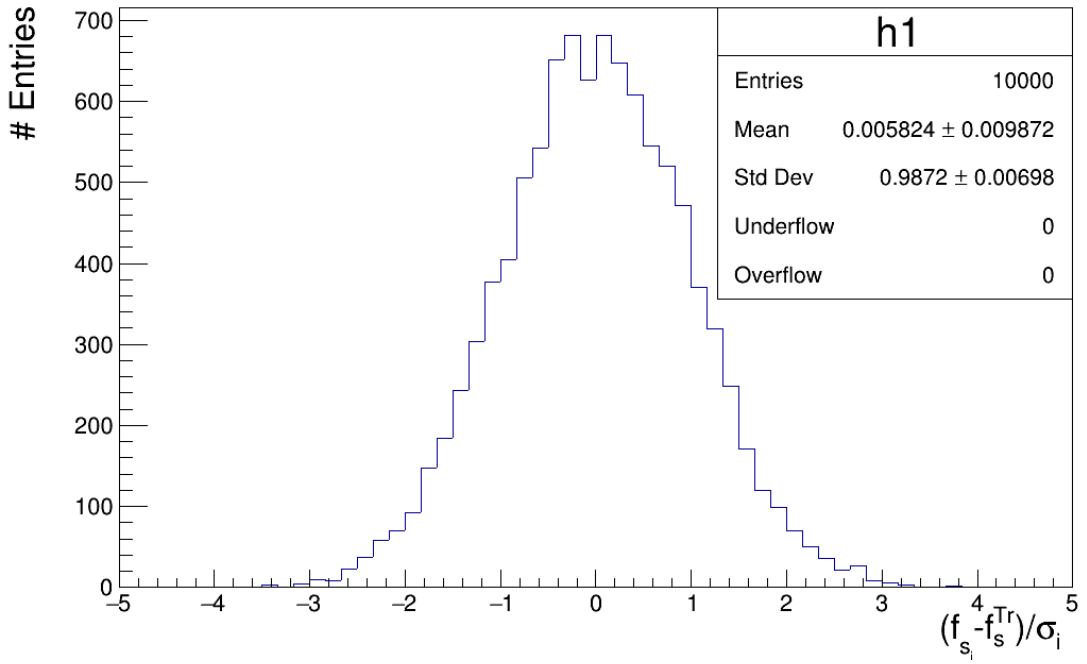


Figure 12.3: The pull distribution for f_s in the discriminating variable fit.

Bibliography

[1] S. L. Glashow, Partial Symmetries of Weak Interactions, *Nucl. Phys.*, 22:579–588, 1961.

[2] Abdus Salam, Weak and Electromagnetic Interactions, *Conf. Proc.*, C680519:367–377, 1968.

[3] Steven Weinberg, A Model of Leptons, *Phys. Rev. Lett.*, 19:1264–1266, 1967.

[4] S. L. Glashow, J. Iliopoulos, and L. Maiani, Weak Interactions with Lepton-Hadron Symmetry, *Phys. Rev.*, D2:1285–1292, 1970.

[5] Mary K. Gaillard, Paul D. Grannis, and Frank J. Sciulli, The Standard model of particle physics, *Rev. Mod. Phys.*, 71:S96–S111, 1999.

[6] Lyndon Evans and Philip Bryant, LHC Machine, *JINST*, 3:S08001, 2008.

[7] Morad Aaboud et al., Study of the rare decays of B_s^0 and B^0 mesons into muon pairs using data collected during 2015 and 2016 with the ATLAS detector, *JHEP*, 04:098, 2019.

[8] Muriel Pivk and Francois R. Le Diberder, sPlot: A Statistical tool to unfold data distributions, *Nucl. Instrum. Meth.*, A555:356–369, 2005.

[9] W.T Eadie et al., Statistical Methods in Experimental Physics, *American Elsevier Pub. Co*, 1971.

[10] R. P. Feynman, Mathematical formulation of the quantum theory of electromagnetic interaction, *Phys. Rev.*, 80:440–457, 1950, [,198(1950)].

[11] Harald Fritzsch and Murray Gell-Mann, Current algebra: Quarks and what else?, *eConf*, C720906V2:135–165, 1972.

[12] N. Bartosik, Associated top-quark-pair and b-jet production in the dilepton channel at $\sqrt{s} = 8$ TeV as test of QCD and background to tt+Higgs production, 2014.

[13] Peter W. Higgs, Broken symmetries, massless particles and gauge fields, *Phys. Lett.*, 12:132–133, 1964.

[14] Peter W. Higgs, Spontaneous Symmetry Breakdown without Massless Bosons, *Phys. Rev.*, 145:1156–1163, 1966.

[15] F. Englert and R. Brout, Broken Symmetry and the Mass of Gauge Vector Mesons, *Phys. Rev. Lett.*, 13:321–323, 1964, [,157(1964)].

[16] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, Global Conservation Laws and Massless Particles, *Phys. Rev. Lett.*, 13:585–587, 1964, [,162(1964)].

[17] G. Aad et al., The ATLAS Experiment at the CERN Large Hadron Collider, *JINST*, 3:S08003, 2008.

[18] S. Chatrchyan et al., The CMS Experiment at the CERN LHC, *JINST*, 3:S08004, 2008.

[19] A. Augusto Alves, Jr. et al., The LHCb Detector at the LHC, *JINST*, 3:S08005, 2008.

[20] K. Aamodt et al., The ALICE experiment at the CERN LHC, *JINST*, 3:S08002, 2008.

[21] C. Lefevre, The cern accelerator complex, Technical report, 2008.

[22] ATLAS collaboration, ATLAS Inner Detector: Technical Design Report. Vol. 1, 1997.

[23] B. Abbott et al., Production and Integration of the ATLAS Insertable B-Layer, *JINST*, 13(05):T05008, 2018.

[24] M. zur Nedden, The LHC Run 2 ATLAS trigger system: design, performance and plans, *JINST*, 12(03):C03024, 2017.

[25] Nicola Cabibbo, Unitary Symmetry and Leptonic Decays, *Phys. Rev. Lett.*, 10:531–533, 1963, [,648(1963)].

[26] Makoto Kobayashi and Toshihide Maskawa, CP Violation in the Renormalizable Theory of Weak Interaction, *Prog. Theor. Phys.*, 49:652–657, 1973.

[27] Andrzej J. Buras, Jennifer Girrbach, Diego Guadagnoli, and Gino Isidori, On the Standard Model prediction for BR(Bs,d to mu+ mu-), *Eur. Phys. J.*, C72:2172, 2012.

[28] Gerhard Buchalla and Andrzej J. Buras, The rare decays $K \to \pi\nu\bar{\nu}$, $B \to X\nu\bar{\nu}$ and $B \to l^+l^-$: An Update, *Nucl. Phys.*, B548:309–327, 1999.

[29] Gerhard Buchalla and Andrzej J. Buras, Qcd corrections to rare k- and b-decays for arbitrary top quark mass, *Nuclear Physics B*, 400(1):225 – 239, 1993.

[30] Christoph Bobeth, Martin Gorbahn, Thomas Hermann, Mikolaj Misiak, Emmanuel Stamou, and Matthias Steinhauser, $B_{s,d} \to l^+l^-$ in the Standard Model with Reduced Theoretical Uncertainty, *Phys. Rev. Lett.*, 112:101801, 2014.

[31] Vardan Khachatryan et al., Observation of the rare $B_s^0 \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data, *Nature*, 522:68–72, 2015.

Acknowledgements

Firstly, I would like to thank Prof. Dr. Peter Buchholz for the opportunity to work in his group, for supporting me throughout the period of thesis. I have learnt much from this research experience.

I am grateful for the technical expertise of Dr. Wolfgang Walkowiak, co-guide for my thesis, and for his invaluable guidance.

I would also like to thank Prof. Dr. Ivor Fleck, not only for accepting to be a second reviewer for my thesis, but also for his wise advice and support throughout.

I would also like to thank all the members in the group: Dr. Iskander Ibragimov, for explaining the subtle details in the rare B-Physics analysis strategy and also for the support during my stay at CERN. Tim-Philip Hücking and Dr. Angel Compoverde for their help in coding.

For proof reading of this whole thesis I would like to thank Elizabeth Hale.

I am also grateful to Dr. Qader Dorsti and Reimund Bayerlein for their support during my student job.

I must thank Prof. Hans Dieter Dahmen for his kindness and mentoring throughout my two year masters.

I am indebted to my teachers, B.V. Ramarao and R. Ananda Kumar, in school and college who inspired me in my pursuit of physics.

I also extend my thanks to all my friends, colleagues and in particular to the group of Elite Seven.

I would like to acknowledge my parents and sister, for their incredible encouragement and my grandfather, C.A. Sastri, for his support and guidance.

Lastly, I must thank Isha Malhotra, a sister to me, for her amazing motivation throughout.

Erklärung

Hiermit versichere ich, dass ich die vorliegende schriftliche Masterarbeit selbstständig verfasst und keine anderen als die von mir angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die anderen Werken dem Wortlaut oder dem Sinne nach entnommen sind, wurden in jedem Fall unter Angabe der Quellen (einschließlich des World Wide Web und anderer elektronischer Text- und Datensammlungen) kenntlich gemacht. Dies gilt auch für die beigegebenen Zeichnungen, bildlichen Darstellungen, Skizzen und dergleichen. Mir ist bewusst, dass jedes Zuwiderhandeln als Täuschungsversuch zu gelten hat, der die Anerkennung der Masterarbeit ausschließt und weitere angemessene Sanktionen zur Folge haben kann.

Siegen,

(Vorname Nachname)