

# Masterthesis: analysis of $B \rightarrow K^* \mu \mu$ decay

Supervisors: Prof. Dr. Nicola Serra, Dr. Marcin Chzaszcz

Author: Oliver Dahme

Here comes the abstract

## Contents

<b>1</b>	<b>Motivation</b>	<b>2</b>
<b>2</b>	<b>The <math>B \rightarrow K^* \mu \mu</math> decay</b>	<b>2</b>
2.1	Kinematics	2
2.2	Operators for $B \rightarrow X_s l^+ l^-$ decays	4
2.3	Wilson coefficients	4
<b>3</b>	<b>The LHCb Experiment</b>	<b>6</b>
3.1	The LHCb Detector	7
3.2	The LHCb trigger system	9
<b>4</b>	<b>Contribution to the Analysis</b>	<b>10</b>
4.1	Classification	10
4.1.1	classifiers test	10
4.2	SPlot	14
4.2.1	Likelihood method	14
4.2.2	$inPlot$ technique	14
4.2.3	$sPlot$ technique	15
4.3	Reweighting	17
<b>5</b>	<b>RooMCMakovChain</b>	<b>17</b>
5.1	Monte Carlo Markov Chain (MCMC)	17
5.2	Metropolis Hastings	18
5.3	Features	18
	<b>Bibliography</b>	<b>21</b>

# 1 Motivation

In the Standard Model of particle physics the different leptons, Electron, Muon and Tauon only differ in their masses. Therefore if these leptons have a high energy (TeV) where the mass becomes negligible the leptons all behave the same. This phenomenon is called lepton universality. For a long time it was one of the ground pillars of the Standard Model. But recent experimental results of the LHCb collaboration [1] suggest a violation of the lepton universality. They measured the branching fraction of four  $B_0$  decays with a  $b$  quark to  $s$  quark transition and two leptons in the final state. To reduce systematic uncertainties the following double ratio of these branching fractions has been considered a well defined test of lepton universality. Since leptons at high energies should behave the same, this double ratio is expected to be equal to one.

$$R_{K^*0} = \frac{\mathcal{B}(B^0 \rightarrow K^{*0} \mu^+ \mu^-)}{\mathcal{B}(B^0 \rightarrow K^{*0} J/\psi(\rightarrow \mu^+ \mu^-))} \bigg/ \frac{\mathcal{B}(B^0 \rightarrow K^{*0} e^+ e^-)}{\mathcal{B}(B^0 \rightarrow K^{*0} J/\psi(\rightarrow e^+ e^-))} , \quad (1)$$

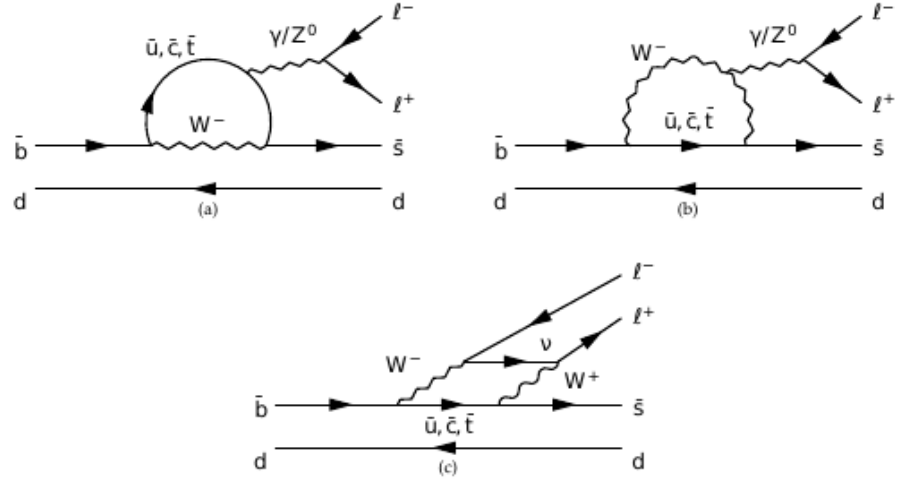
$$R_{K^*0} = \begin{cases} 0.66^{+0.11}_{-0.07}(\text{stat}) \pm 0.03(\text{syst}) \text{ for } 0.045 < q^2 < 1.1 \text{ GeV}^2/c^4, \\ 0.69^{+0.11}_{-0.07}(\text{stat}) \pm 0.05(\text{syst}) \text{ for } 1.1 < q^2 < 6.0 \text{ GeV}^2/c^4. \end{cases}$$

The measurement shows a 2.1-2.3 and 2.4-2.5  $\sigma$  deviation from the Standard Model in the two  $q^2$  regions, respectively. To investigate further the same measurement is performed with new data from the LHCb detector. The goal of this thesis is to contribute to this new measurement.

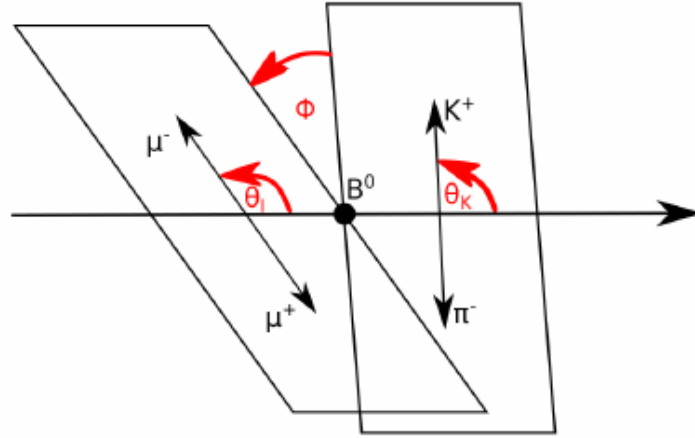
## 2 The $B \rightarrow K^* \mu \mu$ decay

### 2.1 Kinematics

In this section the decay itself and its kinematics are explained: The Decay is a flavor changing neutral current (FCNC) with four charged particles in the final state. The FCNC is a current, which changes the flavor of a fermion without changing its electric charge. The four particles in the final stage are: The  $K^+$  and  $\pi^-$  from the  $K^*$  decay and two leptons from the loop or box diagrams:

Figure 1: Feynman diagrams for decay  $B_d \rightarrow \mu^+ \mu^-$  at lowest order

The kinematics of the decay are defined by the three angles  $\theta_K$ ,  $\theta_L$  and  $\phi$ , shown in figure 2 and the invariant di-muon mass square  $q^2$ .

Figure 2: kinematic variables of the decay  $B^0 \rightarrow K^{*0} \mu \mu$ 

From them one can obtain the differential decay rate for the  $B^0$  Meson:

$$\frac{d^4\Gamma}{d\cos\theta_L d\cos\theta_K d\phi dq^2} = \frac{9}{32\pi} I(q^2, \theta_L, \theta_K, \phi)$$

with:

$$I(q^2, \theta_L, \theta_K, \phi) = I_1^S \sin^2(\theta_K) + I_1^C \cos^2\theta_K + \left( I_2^S \sin^2\theta_K + I_2^C \cos^2\theta_K \right) \cos^2\theta_L$$

$$+ I_3 \sin^2\theta_K \sin^2\theta_L \cos 2\phi + I_4 \sin 2\theta_K \sin 2\theta_L \cos \phi$$

$$+ I_5 \sin 2\theta_K \sin \theta_L \cos \phi$$

$$+ \left( I_6^S \sin^2\theta_K + I_6^C \cos^2\theta_K \right) \cos \theta_L + I_7 \sin 2\theta_K \sin \theta_L \sin \phi$$

$$+ I_8 \sin 2\theta_K \sin 2\theta_L \sin \phi + I_9 \sin^2\theta_K \sin^2\theta_L \sin 2\phi$$
(2)

This decay rate is defined as the probability per unit time that the particle will decay.

## 2.2 Operators for $B \rightarrow X_s l^+ l^-$ decays

In this section all the operators of the decay are derived from the effective Lagrangian. the operators in Quantum field theory are used to create or destroy particles, by applying them to a quantum field. This section should just give an overview if you need more detailed information consider reading [2].

The effective Lagrangian for  $B \rightarrow X_s l^+ l^-$  decays has the form:

$$\begin{aligned} \mathcal{L}_{eff} = & \mathcal{L}_{QCD,QED}(u, d, s, c, b, e, \mu, \tau) \\ & + \frac{4G_F}{\sqrt{2}} [V_{us}^* V_{ub} (C_1^c P_1^\mu + C_2^c P_2^\mu) + V_{cs}^* V_{cb} (C_1^c P_1^c + C_2^c P_2^c)] \\ & + \frac{4G_F}{\sqrt{2}} \sum_{i=3}^{10} [(V_{us}^* V_{ub} + V_{cs}^* V_{cb}) C_i^c + V_{ts}^* V_{tb} C_i^t] P_i. \end{aligned} \quad (3)$$

The first term in equation 3 contains the kinetic terms of the light SM particles as well as their QCD and QED interactions. The remaining two terms consist of  $\Delta B = -\Delta S = 1$  local operators of dimension ( $d \leq 6$ ), which contain those light fields. The mass of the s quark can be neglected in comparison with the b mass. One gets the following operators:

$$\begin{aligned} \mathcal{O}_1^u &= (\bar{s}_L \gamma_\mu T^a u_L)(\bar{u}_L \gamma^\mu T^a b_L), & \mathcal{O}_6 &= (\bar{s}_L \gamma_{\mu_1} \gamma_{\mu_2} \gamma_{\mu_3} T^a b_L) \sum_q (\bar{q} \gamma^{\mu_1} \gamma^{\mu_2} \gamma^{\mu_3} T^a q), \\ \mathcal{O}_2^u &= (\bar{s}_L \gamma_\mu u_L)(\bar{u}_L \gamma^\mu b_L), & \mathcal{O}_7 &= \frac{e}{g^2} m_b (\bar{s}_L \sigma^{\mu\nu} b_R) F_{\mu\nu}, \\ \mathcal{O}_1^c &= (\bar{s}_L \gamma_\mu T^a c_L)(\bar{c}_L \gamma^\mu T^a b_L), & \mathcal{O}_8 &= \frac{1}{g} m_b (\bar{s}_L \sigma^{\mu\nu} T^a b_R) G_{\mu\nu}^a, \\ \mathcal{O}_2^c &= (\bar{s}_L \gamma_\mu c_L)(\bar{c}_L \gamma^\mu b_L), & \mathcal{O}_9 &= \frac{e^2}{g^2} (\bar{s}_L \gamma_\mu b_L) \sum_l (\bar{l} \gamma^\mu l), \\ \mathcal{O}_3 &= (\bar{s}_L \gamma_\mu b_L) \sum_q (\bar{q} \gamma^\mu q), & \mathcal{O}_{10} &= \frac{e^2}{g^2} (\bar{s}_L \gamma_\mu b_L) \sum_l (\bar{l} \gamma^\mu \gamma_5 l), \\ \mathcal{O}_4 &= (\bar{s}_L \gamma_\mu T^a b_L) \sum_q (\bar{q} \gamma^\mu T^a q), \\ \mathcal{O}_5 &= (\bar{s}_L \gamma_{\mu_1} \gamma_{\mu_2} \gamma_{\mu_3} b_L) \sum_q (\bar{q} \gamma^{\mu_1} \gamma^{\mu_2} \gamma^{\mu_3} q), \end{aligned} \quad (4)$$

where  $\sum_q$  and  $\sum_l$  denote the sums over light quarks and all leptons, respectively.

## 2.3 Wilson coefficients

In this section the Wilson coefficients are derived from the effective Hamiltonian of the decay. The effective Hamiltonian for  $b \rightarrow s \mu^+ \mu^-$  transitions can be written as:

$$H_{eff} = -\frac{4G_F}{\sqrt{2}} (\lambda_t H_{eff}^{(t)} + \lambda_u H_{eff}^{(u)}) \quad (5)$$

The  $\lambda_i$  can be expressed with CKM combinations  $\lambda_i = V_{ib} V_{is}^*$ .

$$H_{eff}^{(t)} = C_1 \mathcal{O}_1^c + C_2 \mathcal{O}_2^c + \sum_{i=3}^6 C_i \mathcal{O}_i + \sum_{i=7,8,9,10,P,S} (C_i \mathcal{O}_i + C_i' \mathcal{O}_i') \quad (6)$$

$$H_{eff}^{(u)} = C_1 (\mathcal{O}_1^c - \mathcal{O}_1^u) + C_2 (\mathcal{O}_2^c - \mathcal{O}_2^u). \quad (7)$$

The contribution of  $H_{eff}^{(u)}$  has a double Cabibbo suppression and is therefore usually dropped. It is kept here since it is sensitive to complex phases of decay amplitudes. The operators  $P_{i \leq 6}$  are the same as for general  $B \rightarrow X_s l^+ l^-$  decays, see equation 4. The remaining ones are given by:

$$\begin{aligned}
\mathcal{O}_7 &= \frac{e}{g^2} m_b (\bar{s} \sigma_{\mu\nu} P_R b) F^{\mu\nu}, & \mathcal{O}'_7 &= \frac{e}{g^2} m_b (\bar{s} \sigma_{\mu\nu} P_L b) F^{\mu\nu}, \\
\mathcal{O}_8 &= \frac{1}{g} m_b (\bar{s} \sigma_{\mu\nu} T^a P_R b) G^{\mu\nu a}, & \mathcal{O}'_8 &= \frac{1}{g} m_b (\bar{s} \sigma_{\mu\nu} T^a P_L b) G^{\mu\nu a}, \\
\mathcal{O}_9 &= \frac{e^2}{g^2} (\bar{s} \sigma_\mu P_L b) (\bar{\mu} \gamma^\mu \mu), & \mathcal{O}'_9 &= \frac{e^2}{g^2} (\bar{s} \gamma_\mu P_R b) (\bar{\mu} \gamma^\mu \mu), \\
\mathcal{O}_{10} &= \frac{e^2}{g^2} (\bar{s} \gamma_\mu P_L b) (\bar{\mu} \gamma^\mu \gamma_5 \mu), & \mathcal{O}'_{10} &= \frac{e^2}{g^2} (\bar{s} \gamma_\mu P_R b) (\bar{\mu} \gamma^\mu \gamma_5 \mu), \\
\mathcal{O}_S &= \frac{e^2}{16\pi^2} m_b (\bar{s} P_R b) (\bar{\mu} \mu), & \mathcal{O}'_S &= \frac{e^2}{16\pi^2} m_b (\bar{s} P_L b) (\bar{\mu} \mu), \\
\mathcal{O}_P &= \frac{e^2}{16\pi^2} m_b (\bar{s} P_R b) (\bar{\mu} \gamma_5 \mu), & \mathcal{O}'_P &= \frac{e^2}{16\pi^2} m_b (\bar{s} P_L b) (\bar{\mu} \gamma_5 \mu),
\end{aligned} \tag{8}$$

where  $m_b$  denotes the running b mass in the  $\overline{MS}$  scheme and  $g$  is the strong coupling constant and  $P_{L,R} = (1 \pm \gamma_5)/2$ . In the Standard Model the primed Operators with opposite chirality to the unprimed operators vanish or are highly suppressed as are the  $\mathcal{O}_S$  and  $\mathcal{O}_P$ . The contributions of  $\mathcal{O}_{1,2,3,4,5,6}$  are neglected, since they are either heavily constrained or their impact turns out to be generically very small. For example in the left-right symmetric models or throughout gluino contributions in a general Minimal Supersymmetric Standard Model.

The  $C_i$  coefficients in the equations 6 and 7 are called Wilson coefficients. They encode short-distance physics and New Physics effects. For the calculation a matching scale  $\mu = m_W$  is chosen, in a perturbative expansion in powers of  $\alpha_s(m_W)$ . Then the Wilson coefficients are evolved down to scales  $\mu = m_b$  according to the solutions of the renormalization group equations. Contributions by New Physics enter through  $C_i(m_W)$ , while the low scales are determined by the Standard Model. To allow a more organized expansion of the Wilson coefficients in perturbation theory the factors  $16\pi^2/g^2 = 4\pi/\alpha_s$  are included into the definitions of the operators  $\mathcal{O}_{i \geq 7}$ . All the  $C_i$  expand as:

$$C_i = C_i^{(0)} + \frac{\alpha_s}{4\pi} C_i^{(1)} + \left(\frac{\alpha_s}{4\pi}\right)^2 C_i^{(2)} + \mathcal{O}(\alpha_s^3) \tag{9}$$

where  $C_i^{(0)}$  is the tree-level contribution, which is equal to zero for all operators except  $\mathcal{O}_2$  and  $C_i^{(n)}$  denotes the n-loop contributions. Before discussing the Wilson coefficients in details, let's look at the Operators again; the operators  $\mathcal{O}'_S$  and  $\mathcal{O}'_P$  are given in terms of conserved currents. They carry no scale-dependence. They do not mix with other operators and their Wilson coefficients are at the matching scale.  $\mathcal{O}_9$  is also given by conserved currents. It mixes with  $\mathcal{O}_{1,2,3,4,5,6}$  via a virtual photon decaying into  $\mu^+ \mu^-$ . In addition there is a scale dependence from the factor  $1/g^2$ . This dependence is also present in  $C_{10}$  which otherwise would be scale independent.

In equation ?? one can see that  $C_7$  and  $C_9$  always appear in a particular combination with other Wilson

coefficients in matrix elements. Therefore effective coefficients are defined:

$$\begin{aligned}
 C_7^{eff} &= \frac{4\pi}{\alpha_s} C_7 - \frac{1}{3} C_3 - \frac{4}{9} C_4 - \frac{20}{3} C_5 - \frac{80}{9} C_6, \\
 C_8^{eff} &= \frac{4\pi}{\alpha_s} C_8 + C_3 - \frac{1}{6} + 20 C_5 - \frac{10}{3} C_6, \\
 C_9^{eff} &= \frac{4\pi}{\alpha_s} C_9 + \mathcal{Y}(q^2), \\
 C_{10}^{eff} &= \frac{4\pi}{\alpha_s} C_{10}, \\
 C_{7,8,9,10}^{eff} &= \frac{4\pi}{\alpha_s} C'_{7,8,9,10},
 \end{aligned} \tag{10}$$

$$\begin{aligned}
 \text{where } \mathcal{Y}(q^2) &= h(q^2, m_c) \left( \frac{4}{3} C_1 + C_2 + 6 C_3 + 60 C_5 \right) \\
 &\quad - \frac{1}{2} h(q^2, m_b) \left( 7 C_3 + \frac{4}{3} C_4 + 76 C_5 + \frac{64}{3} C_6 \right)_{env} \\
 &\quad - \frac{1}{2} h(q^2, 0) \left( C_3 + \frac{4}{3} C_4 + 16 C_5 + \frac{64}{3} C_6 \right) \\
 &\quad + \frac{4}{3} C_3 + \frac{64}{9} C_5 + \frac{64}{27} C_6.
 \end{aligned} \tag{11}$$

The function  $h(q^2, m_q)$  comes from the fermion loop and for completeness is presented in equation 12 below. If you need more details consider reading [3].

$$\begin{aligned}
 h(q^2, m_q) &= -\frac{4}{9} \left( \ln \frac{m_q^2}{\mu^2} - \frac{2}{3} - z \right) - \frac{4}{9} (2+z) \sqrt{|z-1|} \cdot \begin{cases} \arctan \frac{1}{\sqrt{z-1}} & z > 1 \\ \ln \frac{1+\sqrt{1-z}}{\sqrt{z}} - \frac{i\pi}{2} & z \leq 1 \end{cases} \\
 z &= \frac{4m_q^2}{q^2}
 \end{aligned} \tag{12}$$

### 3 The LHCb Experiment

In this section the experimental setup of the detector is presented, which recorded the data used in this thesis. The Large Hadron Collider beauty experiment (LHCb) is one of four large experiments based at the CERN laboratory near Geneva in Switzerland. It is part of the Large Hadron Collider (LHC), a proton-proton accelerator and collider located in a vast underground tunnel with 26.7 km circumference beneath the Swiss-French countryside. The other three experiments are CMS and ATLAS which are dedicated to a wide range of physics and have therefor very large detectors. ALICE investigates quark-gluon plasma and therefor needs heavy ion collisions, instead of proton collisions.

The protons in the LHC have a kinetic energy of 7 TeV, which allows a collision energy, in the LHCb detector, of 13 TeV. In the year 2016 the LHCb had a recorded luminosity of  $1906 \text{ pb}^{-1}$ . For this thesis  $2280 \text{ pb}^{-1}$  of data, collected at LHCb during the years 2011 to 2016 are used. LHCb is dedicated to favour physics. It therefor investigates rare decays and CP violation in beauty and charm hadrons.

CERN's Accelerator Complex

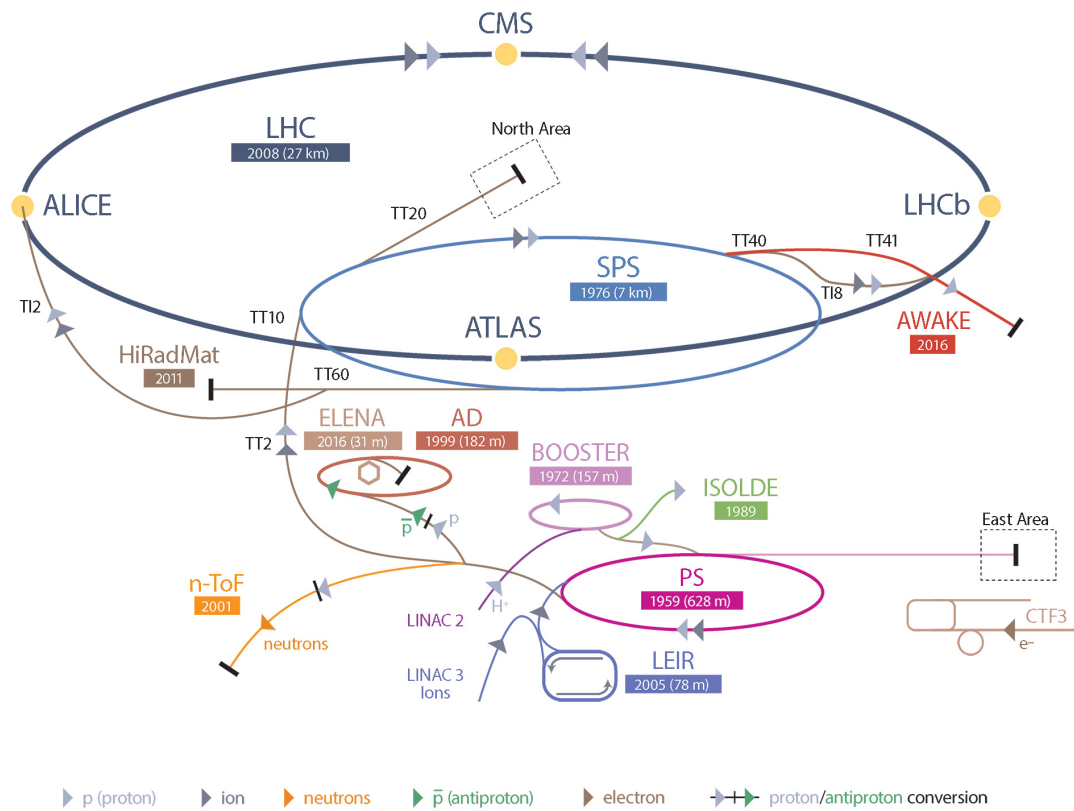


Figure 3: CERN's Accelerator Complex.[4] The protons get injected in the lineare accelerator LINAC2. Then they get pre-accelerated in 3 synchrotrons (BOOSTER,PS,SPS) where the protons reach a kinetic energy of 450 GeV. That is the entering energy of the LHC which accelerates them futher up to 7 TeV, before they collide at the four detectors: CMS, ATLAS, LHCb and ALICE.

### 3.1 The LHCb Detector

The LHCb Detector has a fix target geometry, because beauty hadrons are manily produced at small angeles with respect to the beam pipe.

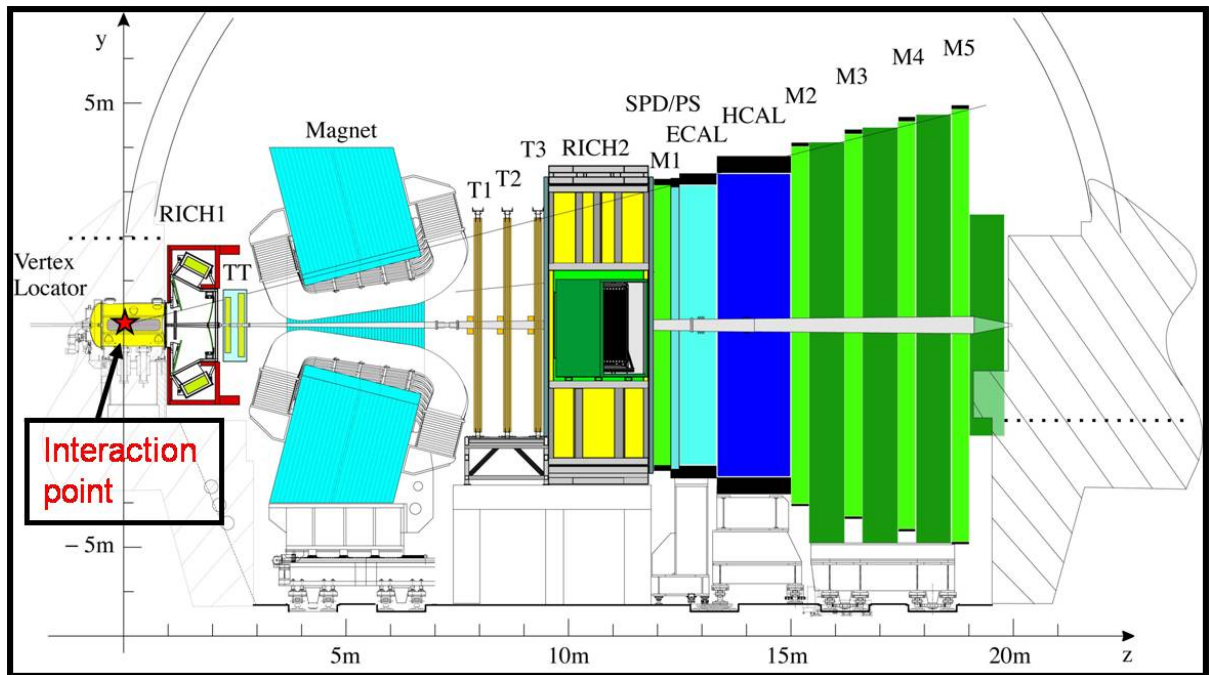


Figure 4: Basic layout of the LHCb detector [5]. The interaction point is inside the vertex detector and the beam pipe passes through the center. The different subdetectors are the two Ring Imaging Cherenkov Detectors (RICH1 and RICH2), the tracking stations (TT and T1 to T3), the scintillator pad detector (SPD), the preshower electromagnetic calorimeter (ECAL), the hadronic calorimeter (HCAL) and the muon stations (M1-M5).

**Vertex Locator (VELO) [6] :** Velo picks out B mesons from the multitude of other particles produced. This is a complex task since B mesons have very short lifetimes spent close to the beam. The VELO's silicon detector elements must be placed at a distance of just five millimetres to the interaction point. To prevent damage to the detector during beam injection and stabilization it is mechanically moved to a safe distance. Velo measures B mesons indirectly by detecting its decay particles, nevertheless it has a resolution of 10 microns. .

**Ring Imaging Cherenkov (RICH) detectors [7] :** The RICH detectors measure the emission of Cherenkov radiation, which happens when a charged particle passes through a medium faster than light does. It is a similar effect like the sonic boom an aircraft produces by breaking the sound barrier. The shape of the light cone depends on the particle's velocity, enabling the detector to determine its speed.

**Magnet [8] :** The big magnet of the LHCb experiment weighs 27 tonnes and is mounted inside a 1,450 tonne steel frame. This powerful magnet forces all charged particles to change their trajectory. By examining the curvature of the path one can calculate its momentum.

**Trackers [9] :** The LHCb's tracking system consists of a series of four large rectangular stations, each covering an area of  $40 \text{ m}^2$ . While flying through this area charged particles will leave a trace, therefore one can estimate the trajectory of a particle. The trajectory is used to link the signals left in other detector elements to the corresponding particle. In LHCb two different tracker technologies are used: The silicon tracker placed close to the beam pipe, uses silicon microstrips. If a charged particle passes such a stripe it collides with the silicon atoms, liberating electrons and creating an electric current, which is then recorded. The outer tracker situated further from the beam pipe consists of gas-filled tubes. The gas ionizes when a charged particle hits a gas molecule, producing electrons. These reach an anode wire situated in the centre of each tube. The position of the track is found by timing how long it takes



electrons to reach it.

**Calorimeters [10]** : Calorimeters stop particles as they pass through, measuring the amount of energy lost. In LHCb there are two different types: The electromagnetic calorimeter responsible for light particles like electrons and photons and the hadronic calorimeter responsible for heavier particles containing quarks. Both have a sandwich-like structure, with alternating layers of metal and plastic plates. If a particle hits a metal plate it produces a shower of secondary particles. These will excite polystyrene molecules in the plastic plates, which then emit ultraviolet light. The energy lost by the particle in the metal plate is proportional to the amount of UV light produced in the plastic plates.

**Muon System [11]** : The muon system consists of 5 rectangular stations, which cover an area of  $435 \text{ m}^2$ . Each station has chambers filled with three gases: carbon dioxide, argon and tetrafluoromethane. Passing muons react with the mixture and electrodes detect the result.

### 3.2 The LHCb trigger system

The rate of events at the LHCb interaction point is 40 MHz. But the rate to have a B meson contained in the detector is 15 kHz. But the offline computing power just allows 2 kHz to be recorded. The LHCb trigger system aims to 'fill' this 2 kHz with interesting B decays and important control decays like  $J/\psi$  decays. The trigger has two levels:

The **Level Zero (L0)** trigger reduces the beginning 40 MHz to 1 MHz. To get this high rate it can only rely on fast sub-detectors as the calorimeters and the muon system. The L0 trigger looks for events with high transverse momentum with respect to the particle beam axis (pT), because particles from a B decay have this attribute, since B Mesons are always produced almost parallel to the beam axis. In addition the L0 trigger performs a simplified vertex reconstruction with the signal of two silicon layers of the VELO to identify events with multiple proton-proton collisions. They are rejected because for this kind of events it is much more difficult to reconstruct B meson decays, since it is harder to distinguish primary and secondary vertex of the B decay.

The **High Level Trigger (HLT)** is an algorithm that runs on a farm of 1000 16-core computers. It has two stages: HLT1 which reduces the event rate to a few tens of kHz and HLT2 which reduces the rate to the 2 kHz which are recorded. HLT1 gets all the candidates of the L0 trigger and uses the full detector information on them to search for particles with a high impact parameter with respect to the proton-proton collision. These particles are most likely decay products from B mesons, because of its relatively long life-time. They typically fly 1 cm away from the collision before decaying resulting in a high impact parameter for the decay products. HLT2 does a complete reconstruction of the events. It starts with the track of the VELO and connects them to the tracks in the other sub-detectors. Most important are displaced vertices, since they are a strong indicator for B decays. The selection is divided into two parts. The inclusive selection searches for resonance decays like  $D^*$  or  $J/\psi$ . The exclusive selection is designed to provide the highest possible efficiency to fully reconstruct B decays of interest. It therefore uses all information available such as mass and vertex quality and intermediate resonances.

## 4 Contribution to the Analysis

This section presents the two contribution done for the analysis of the  $B \rightarrow K^* \mu \mu$  decay.

The first contributed to the elimination of combinatorial background. Combinatorial background is the background that is created due to the fact that in a hadron collider there are always several reactions happening at once. The different reactions combine to background. To separate the combinatorial background from the signal a classification is done. Classification is a term from the Machine Learning field and means identifying to which category an object belongs. For example, consider having 100 pictures of animals and you want to know which ones are cats, dogs, and cows. These are the categories and the pictures are the objects. The classification is performed by algorithms called classifiers. But before using them, they have to be trained. That is done by a training data set. It should look similar to the real data set, except that every object is labeled by its category. To get back to the animal pictures: One could create a training set by taking 5 pictures of a cat, 5 of a dog, and 5 of a cow. These 15 pictures are the training set for the classifier. Now that the classifier knows how a cat, a dog, and a cow looks like, it can identify them in a random collection of pictures.

### 4.1 Classification

In this section it is explained how to use the Machine Learning method Classification to separate combinatorial background from signal:

To do so, Monte Carlo simulated data which contains only  $B \rightarrow K^* \mu \mu$  decays is labeled with probability 1 to be signal. Then it is merged with real data and used to train the classifiers. But since the classification becomes naturally biased if the data to classify is the same as the training data, a technique called K-folding is used. K-folding separates the data Monte Carlo mix into several parts called folds. To classify one fold, all the other folds are used for training. After iterating over all the folds, one has a completely classified data set without any bias.

The contribution of this thesis is that several different classifiers are tested to find the one best suited for the analysis.

#### 4.1.1 classifiers test

First, the following list of classifiers were tested and compared in terms of performance:

- Ada Boost
- uGB + knnAda
- uBoost
- uGB + FI
- xgb
- sk\_bdtg
- sk\_bdt

The test was performed with 30000 events from the 2016 LHCb  $B \rightarrow K^* \mu \mu$  data and 10000 events from

the Monte Carlo simulation. The following list of parameters are used in brackets are the names of the parameters in the root files.

- Decay vertex location for reconstructed particles (ENDVERTEX)
- Primary vertex location (OWNPV)
- Impact parameter (IP\_OWNPV)
- Flight distance (FD\_OWNPV)
- The cosine of the angle between primary vertex and decay vertex and recorded momentum (DIRA\_OWNPV)

To compare the different classifiers the ROC curves and the correlations to the kinematic variables of the decay (see 2) are used. The receiver operating characterisitc (ROC) curve is a graphical plot that illustrates the ability of a classifier to correctly classify the data as its dicrimination threshold varies. Since the curves for the different folds all look alike only one is presented here 5. One can find the other nine curves in the appendix ?? . It turns out that all the classifiers classify the data correctly with just some minor variances. The ROC curve is in that case not a good tool to compare the different classifiers.

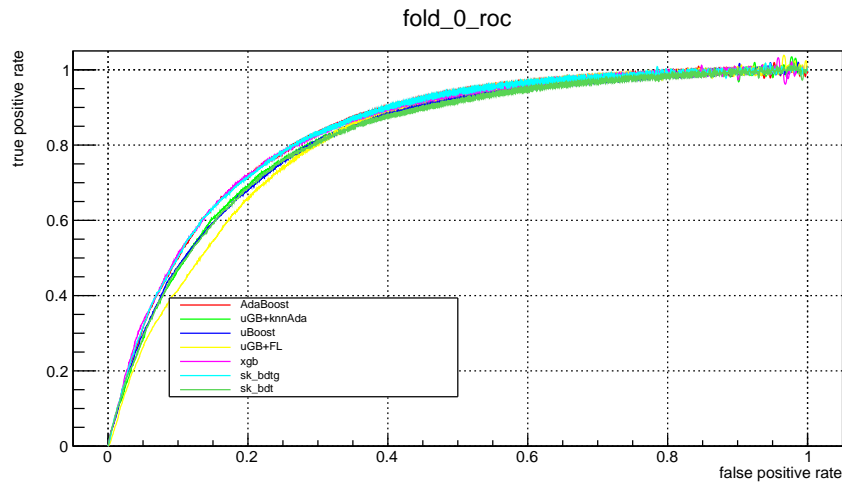


Figure 5: ROC curve of the first fold. One can see that all the classifiers are competitive in terms of classifying correctly

The next step is to check for unwanted correlations, only the correlations with the mass are shown here. The others can be found in appendix ??.

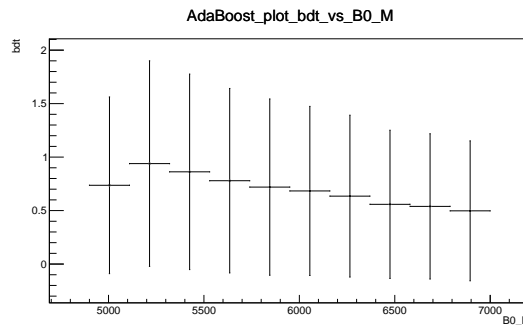


Figure 6: This plot shows the correltation between the bdt decision of the AdaBoost classifier and the  $B_0$  mass. There is clearly a correlation starting from the second bin from the left.

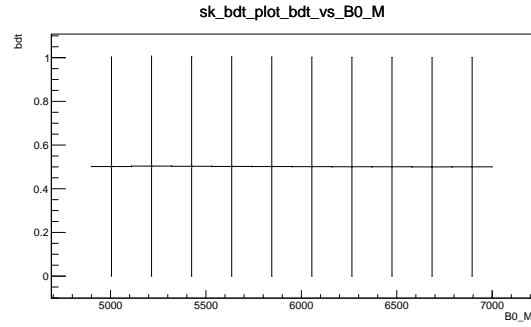


Figure 7: This plot shows the correlation between the bdt decision of the sk\_bdt classifier and the  $B_0$  mass. There is no correlation.

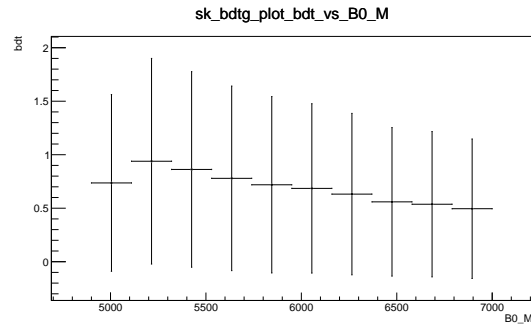


Figure 8: This plot shows the correlation between the bdt decision of the sk\_bdtg classifier and the  $B_0$  mass. There is clearly a correlation starting from the second bin from the left.

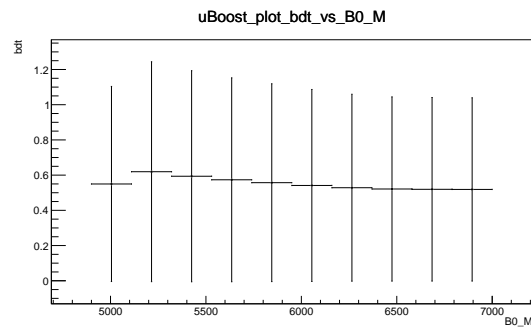


Figure 9: This plot shows the correlation between the bdt decision of the uBoost classifier and the  $B_0$  mass. There is just a very small correlation compared to other classifiers.

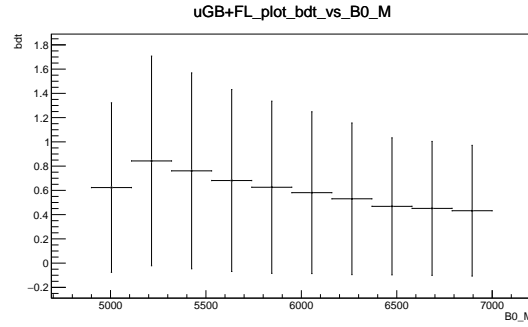


Figure 10: This plot shows the correlation between the bdt decision of the uGB+FL classifier and the  $B_0$  mass. There is clearly a correlation starting from the second bin from the left.

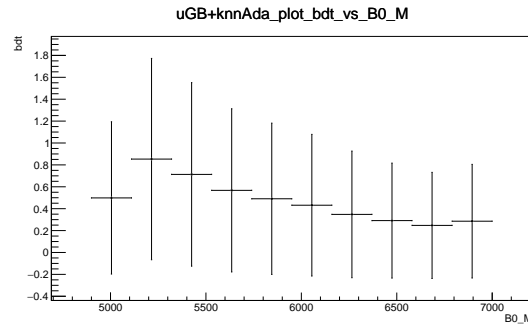


Figure 11: This plot shows the correlation between the bdt decision of the uGB+knnAda classifier and the  $B_0$  mass. There is clearly a correlation starting from the second bin from the left.

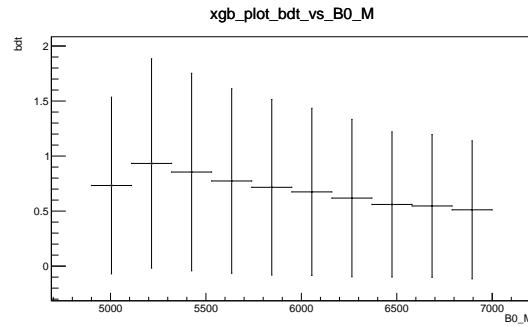


Figure 12: This plot shows the correlation between the bdt decision of the xgb classifier and the  $B_0$  mass. There is clearly a correlation starting from the second bin from the left.

The two classifiers with the least correlation are the sk\_bdt and the uBoost classifiers. Therefore they were used to do the separation between signal and background in the  $B \rightarrow K^* \mu \mu$  data.

## 4.2 SPlot

In this section the SPlot Technique to sperate two or more merged distributions. It has been used to get the inital weights for the Reweighting in chapter 4.3.

### 4.2.1 Likelihood method

Consider an analysis of a data sample, which consists of several types of events. These types represent signal components and background components, for example from different experiments. The log-likelihood of such a data sample is expressed as:

$$L = \sum_{i=1}^N \ln \left[ \sum_{j=1}^{N_S} N_j f_j(y_i) \right] - \sum_{j=1}^{N_S} N_j \quad (13)$$

- $N$  = total number of events
- $N_S$  = number of types
- $N_i$  = expected average number of events for type  $i$
- $y$  = set of diciminating variables
- $f_j$  = PDF of the  $i$ th type
- $f_j(y_i)$  = value of PDF for event  $y_i$
- $x$  = control variable, not a part of  $L$  by construction

The yields  $N_i$  and the free parameters of the PDF are obtained by maximizing the above log-likelihood (eq 13).

### 4.2.2 *inPlot* technique

Consider a varibale  $x$  which can be expressed as a function of the dicriminating variables  $y$  used in the fit. Furthermore a fit has been performed to determine the yields  $N_i$  for all types. From the knowledge of the PDF and the values of  $N_i$  a naive weight can be defined as:

$$P_n(y_i) = \frac{N_n f_n(y_i)}{\sum_{k=1}^{N_S} N_k f_k(y_i)} \quad (14)$$

which will lead to the  $x$ -distribution  $\tilde{M}_n$  defined by:

$$N_n \tilde{M}_n(\bar{x}) = \sum_{i \in \delta x} P_n(y_i) \quad (15)$$

where sum  $\sum_{i \in \delta x}$  contains alle events for which  $x_i$  lies in the interval centered on  $\bar{x}$  and of total width  $\delta x$ . Therefor  $N_n \tilde{M}_n(\bar{x}) \delta x$  is the  $x$ -distribution of the histogrammed events, using the weights of eq 14.

With this procedure one can on average reproduce the true distribution  $\mathbf{M}_n(x)$ . One can even replace the sum in eq 15 by an integral:

$$\left\langle \sum_{i \in \delta x} \right\rangle \rightarrow \int dy \sum_{j=1}^{N_s} N_j f_j(y) \delta(x(y) - \bar{x}) \delta x \quad (16)$$

Furthermore through identifying the number of events  $N_i$  from the fit one gets:

$$\langle N_n \rangle \tilde{M}_n(\bar{x}) = \int dy \sum_{j=1}^{N_s} N_j f_j(y) \delta(x(y) - \bar{x}) P_n(y) \quad (17)$$

$$= \int dy \sum_{j=1}^{N_s} N_j f_j(y) \delta(x(y) - \bar{x}) \cdot \frac{N_n f_n(y)}{\sum_{k=1}^{N_s} N_k f_k(y)} \quad (18)$$

$$= N_n \int dy \delta(x(y) - \bar{x}) f_n(y) \quad (19)$$

$$= N_n \mathbf{M}_n(\bar{x}) \quad (20)$$

One can see that the sum over events of the naive weight  $P_n$  provides a direct estimate of the  $x$ -distribution for the  $n$ th type. But this procedure has a major drawback, since  $x$  is correlated to  $y$ , the PDFs of  $x$  enter implicitly in the definition of the naive weight. Therefore the  $\tilde{M}_n$  distributions are a bad estimate for the quality of the fit, since these distributions are biased in a difficult way, when the PDFs  $f_i(y)$  are not accurate.

Consider for example a data sample where one of the types has events on the tail of the  $x$ -distribution. Such events require the true distribution to account for the tail. But since the events are averaged the weights on the tail are going to be very small missing those events in the estimated true distribution. Only the core of the  $x$ -distribution can be examined with *inPlots*.

#### 4.2.3 *sPlot* technique

In the previous section it was shown that if a variable  $x$  belongs to a set  $y$  of discriminating variables, one can reconstruct the expected  $x$  distribution. Consider now two sets of variables  $x$  and  $y$ , where  $x$  does not belong to  $y$  and which are uncorrelated, hence the total PDFs  $f_i(x, y)$  all factorize into products  $\mathbf{M}_i(x) f_i(y)$ . The equation 20 does not hold anymore because, when summing over the events the  $x$ -PDFs  $\mathbf{M}_j(x)$  appear:

$$\langle N_n \rangle \tilde{M}_n(\bar{x}) = \int \int dy dx \sum_{j=1}^{N_s} N_j \mathbf{M}_j(x) f_j(y) \delta(x - \bar{x}) P_n \quad (21)$$

$$= \int dy \sum_{j=1}^{N_s} N_j \mathbf{M}_j(\bar{x}) f_j(y) \frac{N_n f_n(y)}{\sum_{k=1}^{N_s} N_k f_k(y)} \quad (22)$$

$$= N_n \sum_{j=1}^{N_s} \mathbf{M}_j(\bar{x}) \left( N_j \int dy \frac{f_n(y) f_j(y)}{\sum_{k=1}^{N_s} N_k f_k(y)} \right) \quad (23)$$

$$\neq N_n \mathbf{M}_n(\bar{x}). \quad (24)$$

The correction term

$$N_j \int dy \frac{f_n(y) f_j(y)}{\sum_{k=1}^{N_s} N_k f_k(y)} \quad (25)$$

is not identical to the kroenecker delta  $\delta_{jn}$ . In fact the  $N_n \tilde{M}_n$  distribution obtained by the naive weight is a linear combination of the true distribution  $\mathbf{M}_j$ .

To go forward one has to realize that the correction term is related to the inverse of the covariance matrix, given by the second derivatives of  $-L$ , after the minimization.

$$\mathbf{V}_{nj}^{-1} = \frac{\partial^2(-L)}{\partial N_n \partial N_j} = \sum_{i=1}^N \frac{f_n(y_i) f_j(y_i)}{\left( \sum_{k=1}^{N_s} N_k f_k(y_i) \right)^2} \quad (26)$$

If one averages and is replacing the sum over events by intergals (eq 16) the varaince matrix reads:

$$\langle \mathbf{V}_{nj}^{-1} \rangle = \int \int dy dx \sum_{e=1}^{N_s} N_e \mathbf{M}_e(x) f_e(y) \frac{f_n(y) f_j(y)}{\left( \sum_{k=1}^{N_s} N_k f_k(y) \right)^2} \quad (27)$$

$$= \int dy \sum_{e=1}^{N_s} N_e f_e(y) \frac{f_n(y) f_j(y)}{\left( \sum_{k=1}^{N_s} N_k f_k(y) \right)^2} \cdot \int dx \mathbf{M}_l(x) \quad (28)$$

$$= \int dy \frac{f_n(y) f_j(y)}{\sum_{k=1}^{N_s} N_k f_k(y)} \quad (29)$$

Therefor equation 21 can be rewritten as:

$$\langle \tilde{M}_n(\bar{x}) \rangle = \sum_{j=1}^{N_s} \mathbf{M}_j(\bar{x}) N_j \langle \mathbf{V}_{nj}^{-1} \rangle. \quad (30)$$

To get the distribution of intrest one has to invert this matrix equation:

$$N_n \mathbf{M}_n(\bar{x}) = \sum_{j=1}^{N_s} \langle \mathbf{V}_{nj} \rangle \langle \tilde{M}_j(\bar{x}) \rangle \quad (31)$$

The true distribution of  $x$  can still be reconstructed using the naive weight (eq 14), through a linear combination of  $_{in}Plots$ . In other words: When  $x$  does not belong to the set  $y$ , the weights are not given by equation 14, they are given by a covariance-weighted quantity called  $sWeight$  defined by:

$$_sP_n(y_i) = \frac{\sum_{j=1}^{N_s} \mathbf{V}_{nj} f_j(y_i)}{\sum_{k=1}^{N_s} N_k f_k(y_i)} \quad (32)$$

With the  $sWeights$  on can obtain the distribution of the  $x$  variable by histogramming the  $_sPlot$ :

$$N_{ns} \tilde{M}_n(\bar{x}) \delta x = \sum_{i \in \delta x} _sP_n(y_i) \quad (33)$$

On average it reproduced the true distribution:

$$\langle N_{ns} \tilde{M}_n(x) \rangle = N_n \mathbf{M}_n(x) \quad (34)$$

In the case were  $x$  is significantly correlated with  $y$ , the  $_sPlots$  from equation 33 connt be compared with the pure distributions of the various types. To solve that problem one can perform a Monte Carlo simulation of the procedure and obtain the expected distributions to which the  $_sPlots$  should be compared with.

For more information on  $_sPlots$  consider reading [12].



### 4.3 Reweighting

The information needed in an analysis is not all contained inside the data. For example efficiencies can not be recorded by a detector. But they can be simulated by a Monte Carlo simulation of the particle decay. The problem is that the simulation has to be the same as data to get the correct answers. Since resimulating everything with different parameters until one hits the data distribution would take too much resources. One can simply reweight the simulation to match the data. As an initial weight the  $sWeights$  see equation 32 are used. The parameters used to reweight the MC samples are applied in the following order:  $'nTracks'$ ,  $'B_0 p_T'$  and the quality of the  $K\pi\mu\mu$  vertex. The new weights derived by the difference in data and simulation of these parameters are used to weight all the MC samples. For this analysis the MC and Data of the  $K^* \rightarrow J/\Psi K^0$  are used, because it is a very clean channel.

## 5 RooMCMkovChain

RooMCMkovChain is a new class for the ROOT Data Analysis Framework, which can be used to for fits and presents a alternative to the brodly used MINUIT algorithm. RooMCMkovChain uses a metropolis algorithm as a minimizer for the negative log likelihood function. The metropolis algorithm is based on a Monte Carlo Markov Chain and can therefore easily be scaled to multidimensional parameter space and moreover, such kind of algorithms can easily be parallelized.

### 5.1 Monte Carlo Markov Chain (MCMC)

A Markov Chain is a random process which undergoes several states. From each state there is a probability ditribution to change into another state or to stay. Most important i the asumption that every next step just depends on the current state. The figure 13 illustrates the behavior of such a Markov Chain.

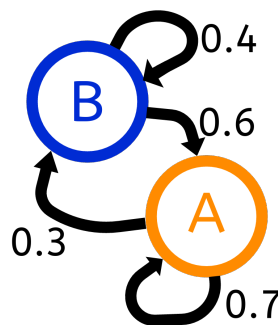


Figure 13: Illustration of states and the probability distribution to change or the stay in a state of a Monte Carlo Markov Chain. In state A there is a 0.3 probability to change into state B and a 0.7 probability to stay in state A.

## 5.2 Metropolis Hastings

Metropolis Hastings is a so called Monte Carlo updater. It updates the probabillity distribution when a change of states is proposed. That is very usefull in the cases where the states are not discrete like in figure 13 but continous. Suppose that a specified distribution has unnormalized density  $h$ . The Metropolis Hastings update does the following:

1. The current state  $x$  is proposed to move to another state  $y$  with a conditional probability given  $x$  denoted as  $q(x, \cdot)$ .
2. The Hastings ratio is calculated:

$$r(x, y) = \frac{h(y) \cdot q(y, x)}{h(x) \cdot q(x, y)} \quad (35)$$

3. The proposed move to  $y$  is accepted with the probability  $a(x, y)$ :

$$a(x, y) = \min(1, r(x, y)) \quad (36)$$

This principle is used in the RooMCMkovChain class to minimize the negative log likelihood function. In detail a robust adaptive Metropolis hastings process is implemented. It is defined as:

1. Compute next proposed state  $Y_n$  as a random change from the current state  $X_{n-1}$ :

$$Y_n \equiv X_{n-1} + S_{n-1} U_n$$

, where  $U_n$  is an independent random vector.

2. With the probability  $\alpha_n$  the proposal is accepted and  $X_n = Y_n$ . Otherwise the proposal is rejected and  $X_n = X_{n-1}$

$$\alpha_n \equiv \min\{1, r(X_{n-1}, Y_n)\}$$

, where  $r$  is the hastings ratio see equation 35.

3. compute the lower-diagonal matrix  $S_n$  with positive diagonal elements satisfying the equation:

$$S_n S_n^T = S_{n-1} \left( I + \eta_n (\alpha_n - \alpha_*) \frac{U_n Y_n^T}{||U_n||^2} \right) S_{n-1}^T \quad (37)$$

, where  $I \in R^{dx d}$  is the identity matrix.

## 5.3 Features

The features of the RooMCMkovChain class will be shown by fitting the following probability density function (pdf):

$$g(x) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} + f \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}} \quad (38)$$

It is so called double gaus pdf, which is just the sum of two gaussian pdfs with a fractional parameter  $f$ . The result of the fit is compared with the result of the Minuit algorithm:

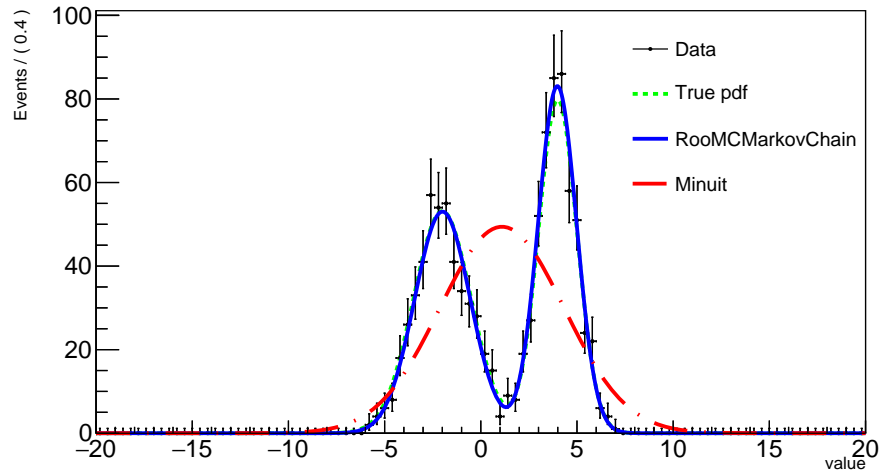


Figure 14: Fit of double gauss pdf (see equation 38) with RooMCMarkovChain and with Minuit. The true pdf has the following values:  $\mu_1 = 4, \sigma_1 = 1, \mu_2 = -2, \sigma_2 = 1.5$  and the fraction  $f = 0.5$

The corresponding terminal output gives the estimated parameters with error interval and correlation coefficients of each parameter pair.

```
RooFit v3.60 -- Developed by Wouter Verkerke and David Kirkby
Copyright (C) 2000-2013 NIKHEF, University of California & Stanford University
All rights reserved, please read http://roofit.sourceforge.net/license.txt

Starting Monte Carlo Markov Chain Fit with 12000 points and cutoff after 7000 points
1% 2% 3% 4% 5% 6% 7% 8% 9% 10% 11% 12% 13% 14% 15% 16% 17% 18% 19% 20% 21% 22% 23% 24% 25% 26%
% 33% 34% 35% 36% 37% 38% 39% 40% 41% 42% 43% 44% 45% 46% 47% 48% 49% 50% 51% 52% 53% 54% 55% 5
62% 63% 64% 65% 66% 67% 68% 69% 70% 71% 72% 73% 74% 75% 76% 77% 78% 79% 80% 81% 82% 83% 84% 85
91% 92% 93% 94% 95% 96% 97% 98% 99% 100%
NO.  NAME      VALUE      ERROR
1      frac      5.14084e-01  1.47307e-02
2      mean1      3.98243e+00  5.07205e-02
3      mean2     -1.99256e+00  7.51154e-02
4      sigma1      9.98988e-01  3.87089e-02
5      sigma2      1.44724e+00  5.86681e-02

CORRELATION COEFFICIENTS
NO.   1      2      3      4      5
1      1.000  -0.072 -0.043  0.095 -0.060
2     -0.072  1.000  0.131 -0.151  0.136
3     -0.043  0.131  1.000 -0.135  0.194
4      0.095 -0.151 -0.135  1.000 -0.171
5     -0.060  0.136  0.194 -0.171  1.000
```

Figure 15: Terminal output of the RooMCMarkovChain fit.

This gives a good initial overview of results after the fit has finished. The error calculation can be set to assume gaussian or non-gaussian errors.

In addition several other properties of the parameters can be obtained:

1. The 1 dimensional profile of the negative log likelihood function for a given parameter.

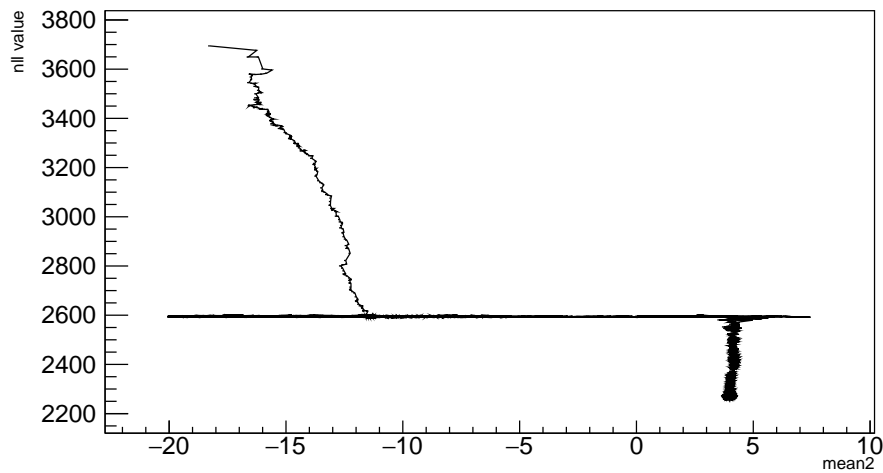


Figure 16: Profile of the negative log likelihood function for  $\mu_2$  in equation 38. There is a local minimum at the nll value of 2600.

From this plot on can see how the algorithm reached the minimum of the negative log likelihood function and if there are local minima.

2. The walk distribution of a given parameter.

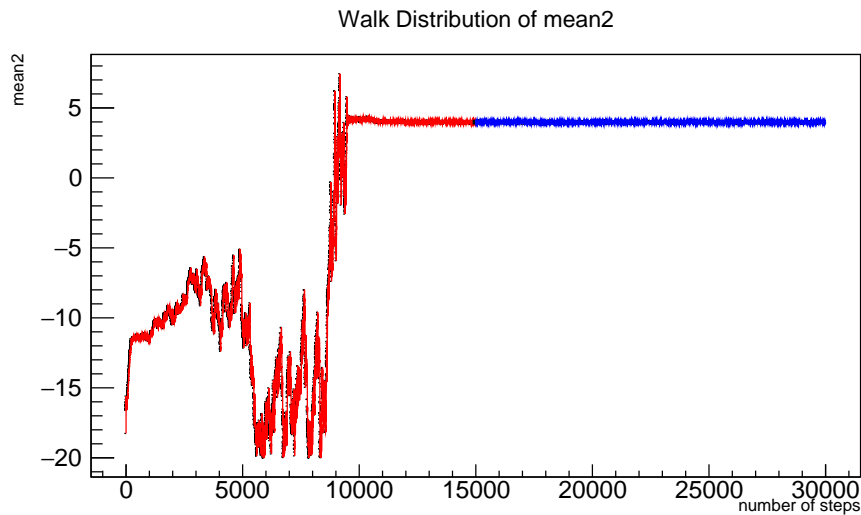


Figure 17: Walk distribution of  $\mu_2$  in equation 38. The red points are not considered for the error calculation.

This plot is very important for handling the RooMCMkovChain class. Since the error calculation is based on the variance of the walk distribution, cutting of points in the beginning greatly reduces this variance. The user has to choose how many points are cut off. Plotting the walk distribution of all parameters helps choosing the right amount.

3. The walk distribution of a given parameter as a histogram.

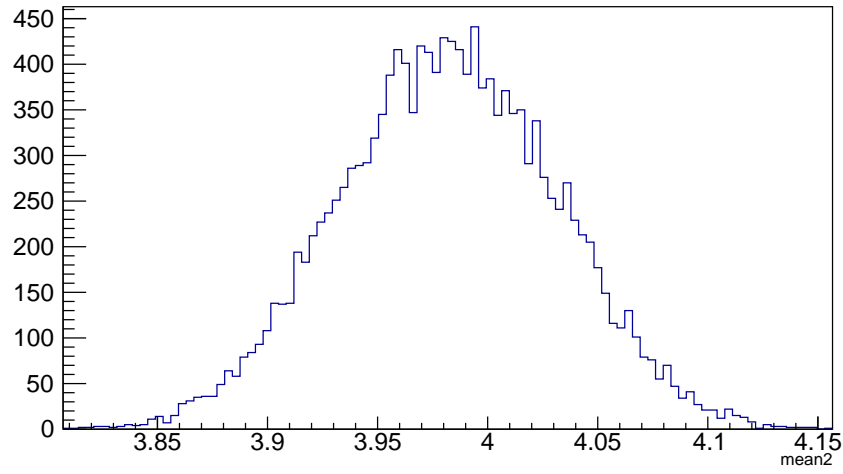


Figure 18: Walk distribution of  $\mu_2$  in equation 38 as a histogram.

This plot can be used to check, which error strategy should be used. If the walk distribution histogram is gaussian, gaussian errors can be assumed.

#### 4. The scatterplot between two given parameters.

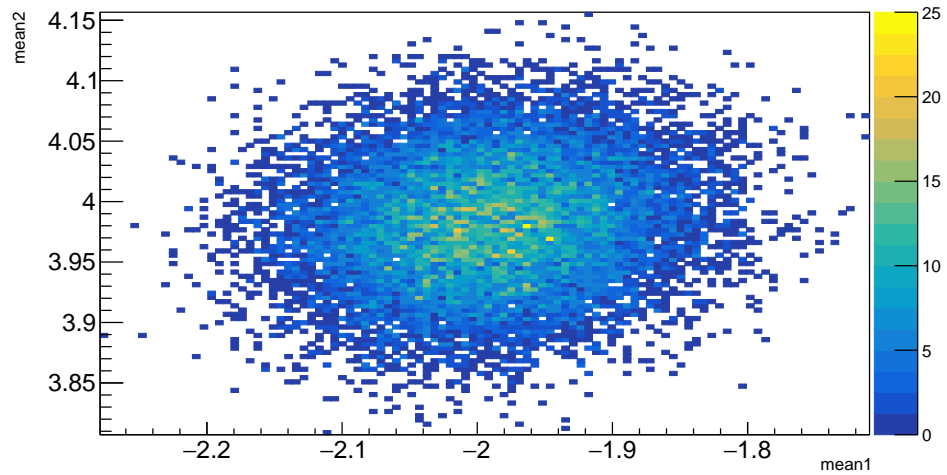


Figure 19: Scatterplot between  $\mu_1$  and  $\mu_2$  in equation 38.

This plot shows the correlation between two parameters graphically.

## References

- [1] JHEP08 (2017) 055
- [2] C. Bobeth, M. Misiak and J. Urban, Nucl. Phys. B 574 (2000) 291 [arXiv:hep-ph/9910220].

- [3] arXiv:0811.1214 [hep-ph]
- [4] CERN Accelerator Complex, <http://www.stfc.ac.uk/research/particle-physics-and-particle-astrophysics/large-hadron-collider/cern-accelerator-complex/>
- [5] Science and Technology Facilities Council article about LHCb , <https://www.ppd.stfc.ac.uk/Pages/LHCb.aspx>
- [6] VELO description, <http://lhcb-public.web.cern.ch/lhcb-public/en/Detector/VELO2-en.html>
- [7] RICH description, <http://lhcb-public.web.cern.ch/lhcb-public/en/Detector/RICH2-en.html>
- [8] Magnet description, <http://lhcb-public.web.cern.ch/lhcb-public/en/Detector/Magnet2-en.html>
- [9] Tracker description, <http://lhcb-public.web.cern.ch/lhcb-public/en/Detector/Trackers2-en.html>
- [10] Calorimeters description, <http://lhcb-public.web.cern.ch/lhcb-public/en/Detector/Calorimeters2-en.html>
- [11] Muon system description, <http://lhcb-public.web.cern.ch/lhcb-public/en/Detector/Muon2-en.html>
- [12] sPlot: a statistical tool to unfold data distributions, arXiv:physics/0402083 [physics.data-an]