

Search for the rare fully leptonic decay $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ at LHCb

Slavomira Stefkova

High Energy Physics

Blackett Laboratory

Imperial College London

A thesis submitted to Imperial College London
for the degree of Doctor of Philosophy

Abstract

This thesis reports the branching fraction measurement of the rare Cabibbo-suppressed decay $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$. The decay is observed for the first time with a 5.5σ deviation from the background-only hypothesis. This is the first observation of a $b \rightarrow d$ quark transition in the baryon sector. The dataset used for the measurement corresponds to 3 fb^{-1} of pp collisions collected at the LHCb experiment at CERN. The branching fraction is measured using $\Lambda_b^0 \rightarrow J/\psi (\rightarrow \mu^+\mu^-) p\pi^-$ as a normalisation channel and is measured as

$$\mathcal{B}(\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-) = (6.9 \pm 1.9 \pm 1.1_{-1.0}^{+1.3}) \times 10^{-8},$$

where the first error is the statistical uncertainty, the second is the systematic uncertainty and the third is the uncertainty on $\mathcal{B}(\Lambda_b^0 \rightarrow J/\psi p\pi^-)$. The measurement of $\mathcal{B}(\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-)$ can be combined with the branching fraction measurement for $\Lambda_b^0 \rightarrow pK^-\mu^+\mu^-$ to give constraints on the ratio of CKM matrix elements $|\frac{V_{td}}{V_{ts}}|$. Such a determination of $|\frac{V_{td}}{V_{ts}}|$ requires a theory prediction for the ratio of the relevant form factors.

This thesis also reports the ratio of tracking efficiencies, ϵ_{rel} , between data and simulation for $K_s^0 \rightarrow \pi^+\pi^-$ decays occurring within the LHCb detector acceptance. As K_s^0 particles are long-lived, their associated tracking efficiencies are less precisely determined compared to those of shorter-lived particles. The average value of ϵ_{rel} for $K_s^0 \rightarrow \pi^+\pi^-$ decays, where the K_s^0 has a flight distance of $\gtrsim 1 \text{ m}$, is found to be

$$\epsilon_{\text{rel}} = 0.70 \pm 0.02.$$

To perform this calibration measurement a novel technique was developed which has the potential to be used in measuring the value of ϵ_{rel} for other decays involving long-lived particles.

Contents

Declaration of originality	3
Copyright declaration	4
Acknowledgements	5
List of Figures	6
List of Tables	7
Glossary	8
1 Introduction	10
2 The LHCb detector	11
2.1 LHCb Layout	13
Bibliography	16
Appendices	18
A Boosted Decision Trees	18
B The <i>sPlot</i> technique	21

Declaration of originality

The work presented in this thesis is the result of collaborative work between members of the LHCb collaboration and myself. All the analysis work (chapters ??–??) presented in this thesis was performed by myself, with the exception of producing the simulation and data samples used in the analysis in ??. All work and plots presented in this thesis that were not the product of my own work are appropriately referenced.

This thesis has not been submitted for any other qualification.

Eluned Smith

September 2016

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the license terms of this work.

Acknowledgements

Firstly, I would like to thank my supervisor Ulrik Egede. I would like to thank you for all the support, time and article corrections that you have invested in me. You have given me the opportunity to grow as physicist as well as you taught me to appreciate solving problems differently. You have given me the freedom to follow many different activities such as conferences, hardware, as well as detector maintenance all leading to meeting many great people that I am happy to call my friends.

To you Patrick Owen, I owe a big thank you for showing me all the support throughout my entire analysis. Without your great insight and motivation that you have kept even when you changed institute, it would have been completely different journey.

I would also like to express special thanks to Mike McCainn, whose patience knows no limits, but apparently LHC's is 15 minutes. I learnt how handle many power-cycling buttons as well as croquet. Here I would also like to extend my gratitude to many of the RICH collaborators such as Antonis, Silvia, Roberta whom I have met while working with the upgrade.

And ofcourse all the Imperial crew: Dave, Andrei, Mitesh, Paula, Eluned, Will, Fede, Sophie, Matt, Felix and Malte, who all have their part in this work.

None of this would be possible without the STFC, the financial body on behalf of British government, who financially supported my PhD and the LHCb collaboration as a whole.

Lastly but most importantly, I would like to thank my family. I have never heard from you that something cannot be done. You have taught me to be curious, to be independent, to be strong. We have sacrificed the most precious commodity, time spent together, in order for me to follow my interests. And below is just one example it was all worth it.

List of Figures

2.1	Accelerator complex at CERN	12
2.2	Angular production and acceptance of LHCb's $b\bar{b}$ pair (in red) as well as General Purpose Detector (in yellow). LHCb covers region with highest production cross-section at 8 TeV. These plots were produced using PYTHIA8 [1] simulation. This plot was taken from [2].	13
2.3	The schematic slice of LHCb detector in y, z plane where z is defined to be the direction parallel to beamline, and x, y define the plane perpendicular to the beamline. θ , the opening angle in y-z plane with $\theta = 0$ along z -axis.	15
2.4	Integrated luminosity collected in different years of data-taking. As compared to ATLAS and CMS the integrated luminosity is much lower, due to pile-up conditions. In 2017, there were two pp collision energies at $\sqrt{s} = 13$ and 5 TeV.	15
A.1	An example decision tree. The S and B stand for 'Signal-like' and 'Background-like'. The β_i variables refer to the cut values chosen by the machine learning algorithm after the tree has been trained on signal and background samples. The blue ovals represent final nodes called leafs, which each leaf having an associated purity, i.e. the fraction of the weight of a leaf due to signal events.	19

List of Tables

2.1 14

Glossary

ALICE A Large Ion Collider Experiment.

ATLAS A Toroidal LHC ApparatuS.

CMS Compact Muon Solenoid.

LHCb The Large Hadron Collider beauty experiment.

PID Particle IDentification.

SM Standard Model.

To Nain, on behalf of all four of us

Chapter 1

Introduction

The Standard Model ([SM](#)) is an effective theory which describes fundamental particles and their interactions to an impressive precision.

bla

Chapter 2

The LHCb detector

In this section, overview of accelerator complex at CERN as well as physics motivation behind **LHCb** detector and its details will be described.

CERN built one of the most exciting laboratories to study elementary particle interactions. The complex set of particle accelerators and detectors is shown in [Figure 2.1](#). The process of accelerating protons starts with the source of protons. Protons are obtained from hydrogen gas bottle by applying and an electric field separates hydrogen into positively and negatively charged constituents. The first proton accelerator in the chain, Linac 2, accelerates the protons to the energy of 50 MeV. It is a tank composed of several chambers where the resonant cavity is tuned to a specific frequency which creates potential differences in the cavities making accelerate the protons. These are then injected into the Proton Synchrotron Booster (PSB). Here the protons are accelerated to 1.4 GeV. The next line is the Proton Synchrotron (PS) reaching energy of 25 GeV. Before either entering the LHC or North Area (mainly used as testing facility for experiment upgrades) Super Proton Synchrotron (SPS) is the last stop. Here proton acceleration to 450 GeV is achieved.

The Large Hadron Collider (LHC) is a complex machine which accelerates beams of protons in opposite directions in ~ 27 km circular tunnel. It is located 50-157 m below ground on the border of Switzerland and France. Once the desired energy is achieved proton-proton, pp , or ion, collisions happen at four distinct points, where different detectors with different physics focus are located. These are [ATLAS](#), [CMS](#), [ALICE](#) and **LHCb**.

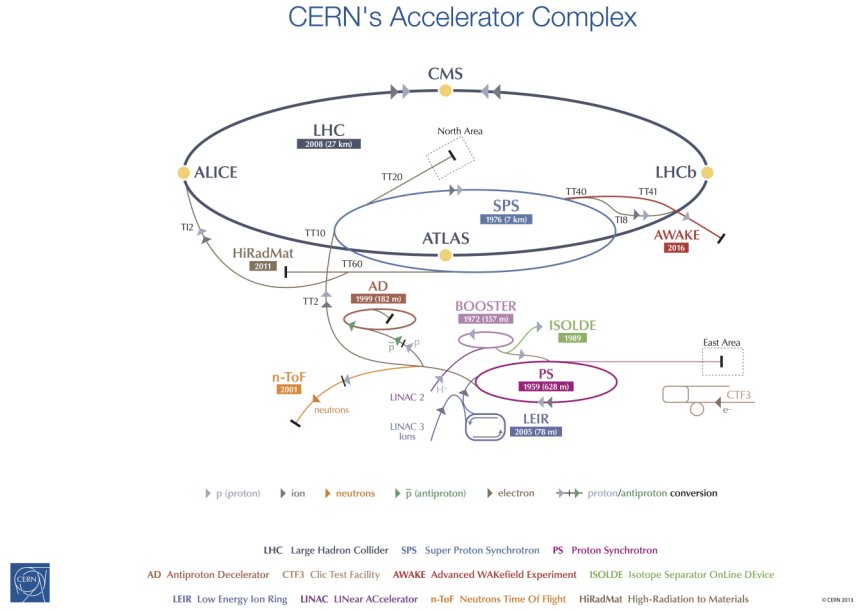


Figure 2.1: Accelerator complex at CERN

Study of $B^+ \rightarrow \mu^+ \mu^- \mu^+ \nu$ was performed using data obtained at **LHCb**.

2.1 LHCb Layout

LHCb differs from the other general purpose detectors on the LHC ring as its studies properties of heavy particles containing b or c quarks. This can be attributed to the geometrical acceptance and unique vertex resolution as well as excellent **PID**.

Contrary to the two general purpose detectors where the collisions are occurring in the centre of the detector, **LHCb**'s collision point is located at one end of the detector, hence its description as a forward single-arm spectrometer. This means that information about products outside of its scope are not known, meaning that there is no overall constraint on collision information, unlike in other flavour experiments. This is compensated by production mechanism of $b\bar{b}$ and $c\bar{c}$ in pp interactions, which occurs via gluon-gluon fusion. In this process, each gluon will carry part of proton's momentum. If the two gluons from two protons carry significantly different momentum, the $b\bar{b}$ system will be boosted with respect to the pp rest frame, either in forward or backward cone closely to the beamline, as can be seen in [Figure 2.2](#).

The angular coverage of LHCb is formally defined using pseudorapidity η ,

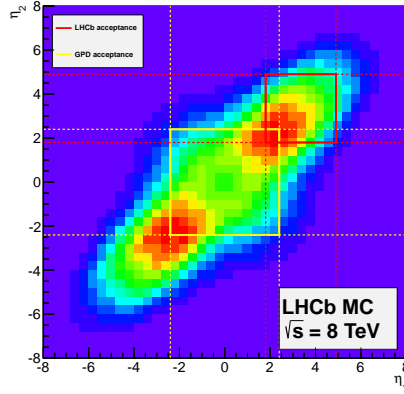


Figure 2.2: Angular production and acceptance of LHCb's $b\bar{b}$ pair (in red) as well as General Purpose Detector (in yellow). LHCb covers region with highest production cross-section at 8 TeV. These plots were produced using PYTHIA8 [1] simulation. This plot was taken from [2].

$$\eta = -\ln(\tan \frac{\theta}{2}) \quad (2.1)$$

where θ is defined in [Figure 2.3](#). [LHCb](#), hence, covers the region $2 < \eta < 5$ or 300 mrad. The production cross-section of the fundamental process $pp \rightarrow b\bar{b}X$, $\sigma(pp \rightarrow b\bar{b}X) = 75.3 \pm 5.4 \pm 13.0 \text{ } \mu\text{b}$ at 7 TeV [\[3\]](#) and $144 \pm 1 \pm 21 \text{ } \mu\text{b}$ at 13 TeV [\[4\]](#) has been measured.

To limit the background coming from especially soft QCD processes (due to hadronic collision enviroment) global event cuts, *GECs*, are put in the place. To limit the occupancy of the detector only events with 600 (in 7,8 TeV) and 450 (in 13 TeV) tracks and less are allowed to be processed. In order to achieve these occupancies, μ_{vis} , the average number of visible pp interactions per bunch-crossing is kept below 1, so that the pile-up, the visible number of pp interaction in the visible events, is limited. This LHCb-specific control of luminosity is achieved by *luminosity levelling*. This procedure achieves stable instantenous luminosity by controlling the transverse overlap of the beams at collision point. It limits the effects of luminosity decay, which can lead to trigger alterations during specific data taking run, resulting in systematic uncertainties.

So far, the detector has been running since 2010, with diferrent running conditions summarized in Table , and with collected integrated luminosity [Figure 2.4](#)

	\sqrt{s} [TeV]	Luminosity $\mathcal{L} [\times 10^{32} cm^{-2} s^{-1}]$	Number of Bunches	Bunch Spacing [ns]
2011	7	3.5	1300	50
2012	8	4	1300	50
2015	13	4	1.12	25
2016	13	4	1.12	25
2017	13	4	1.12	25

Table 2.1

are hence heavily exploited at LHCb, with $pp \rightarrow b\bar{b}X$ This is compensated by preferential decays of containing b particles in a specific angular region. This region is defined along the beam pipe ϕ z-axis

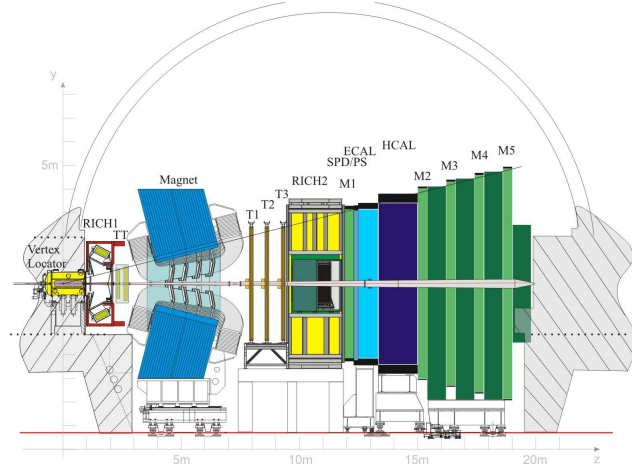


Figure 2.3: The schematic slice of LHCb detector in y, z plane where z is defined to be the direction parallel to beamline, and x, y define the plane perpendicular to the beamline. θ , the opening angle in y - z plane with $\theta = 0$ along z - axis.

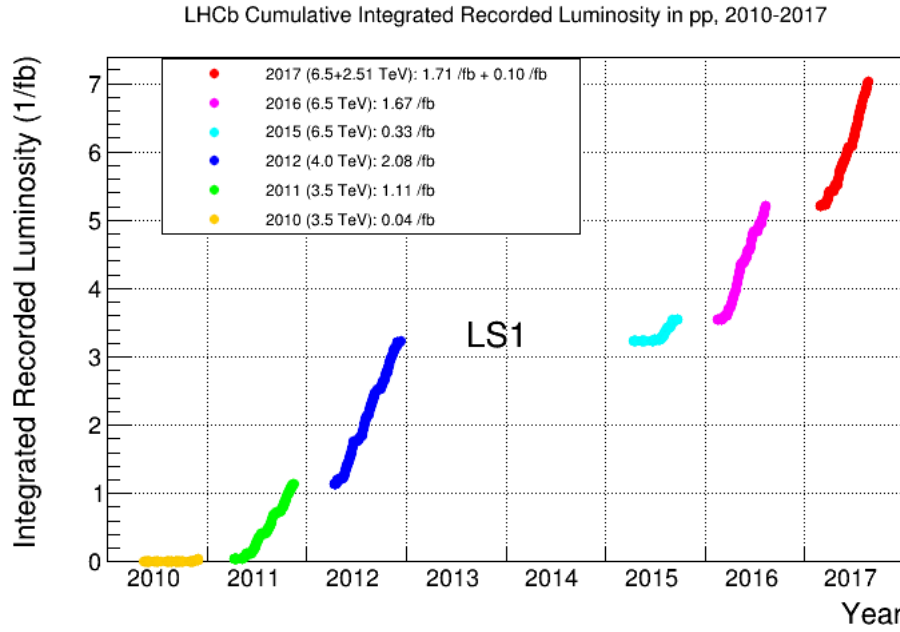


Figure 2.4: Integrated luminosity collected in different years of data-taking. As compared to ATLAS and CMS the integrated luminosity is much lower, due to pile-up conditions. In 2017, there were two pp collision energies at $\sqrt{s} = 13$ and 5 TeV.

Bibliography

- [1] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852, [arXiv:0710.3820](#).
- [2] LHCb Collaboration, *Image of the acceptance of using simulation taken from* https://lhcb.web.cern.ch/lhcb/speakersbureau/html/bb_productionangles.html, Accessed in 2017.
- [3] LHCb collaboration, R. Aaij *et al.*, *Measurement of $\sigma(pp \rightarrow b\bar{b}X)$ at $\sqrt{s} = 7$ TeV in the forward region*, *Phys. Lett. B* **694** (2010) 209, [arXiv:1009.2731](#).
- [4] LHCb collaboration, R. Aaij *et al.*, *Measurement of the b-quark production cross-section in 7 and 13 TeV pp collisions*, *Phys. Rev. Lett.* **118** (2017) 052002, Erratum *ibid.* **119** (2017) 169901, [arXiv:1612.05140](#).

Appendices

Appendix A

Boosted Decision Trees

Many rare decay analyses make extensive use of BDTs and they are important in the $\Lambda_b^0 \rightarrow p\pi^-\mu^+\mu^-$ analysis. Firstly, the concept of a decision tree is introduced followed by a brief explanation of boosted decision trees.

A decision tree, in the context of data mining, is a supervised machine learning method which allows for the prediction of the value of a target variable based on several input variables. In particle physics, the purpose of the decision tree is to classify an event as being either signal or background, based on the event's input variables. The input variables, $\{x_i\}$, are various physics parameters. Each cut point in the tree is referred to as a node and the final nodes are referred to as leaves. A very simple example is shown in [Figure A.1](#). The purity, P , of a leaf refers to the fraction of the weight of a leaf due to signal events, e.g. if a leaf had 20 signal events and 15 background events it would have a purity of 0.75. If a leaf has a purity larger than 0.5 it is deemed to correspond to signal and if lower, to background.

A decision tree is constructed by a process called training. For this, samples of known signal and background events are used. These samples could be either simulation or data. For each x_i the best dividing point is decided, that is, the cut that gives the best separation between signal and background. This optimum point is decided by using the Gini index defined as

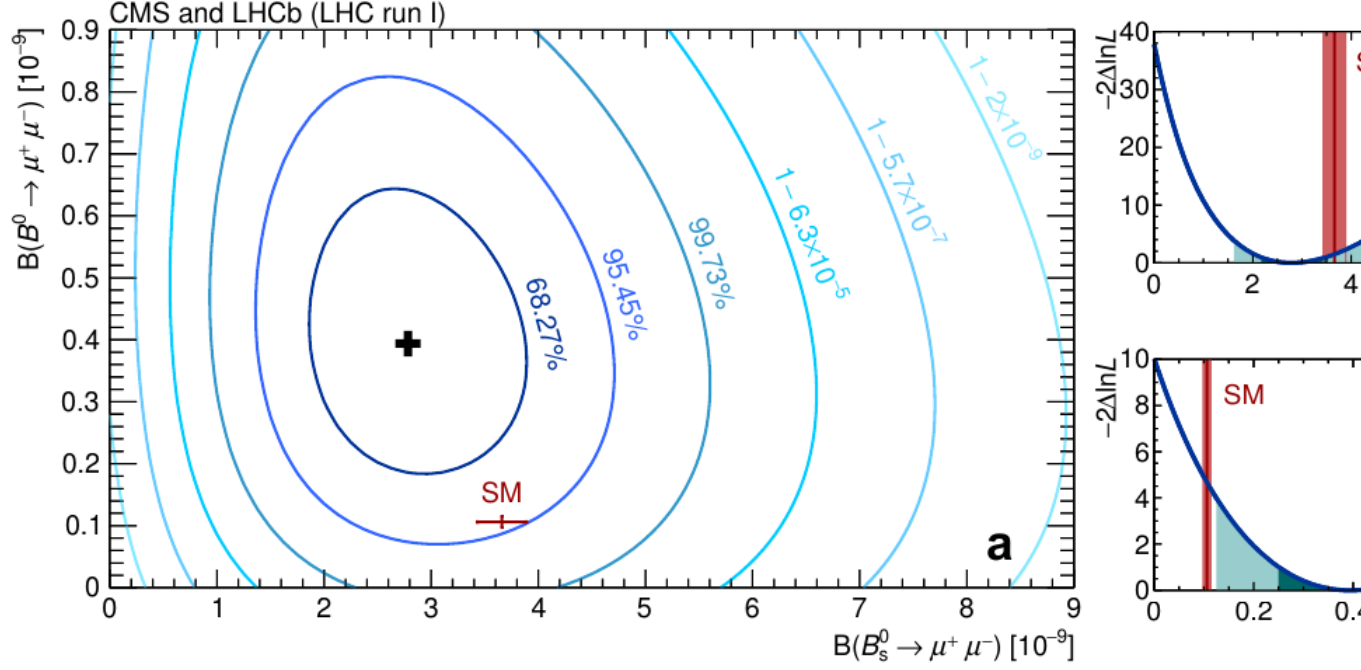


Figure A.1: An example decision tree. The S and B stand for ‘Signal-like’ and ‘Background-like’. The β_i variables refer to the cut values chosen by the machine learning algorithm after the tree has been trained on signal and background samples. The blue ovals represent final nodes called leafs, which each leaf having an associated purity, i.e. the fraction of the weight of a leaf due to signal events.

$$Gini = \sum_{i=1}^n W_i P(1 - P), \quad (\text{A.1})$$

where W_i is weight of the i^{th} event, which would generally be unity for the case of a non-boosted decision tree. The cutting point is then found by maximising the separation, Δ , between the Gini index of the parent node and the combined Gini index of the child nodes, as given in [Equation A.2](#)

$$\Delta = Gini_{parent} - Gini_{child_1} - Gini_{child_2}. \quad (\text{A.2})$$

The depth of a tree (the maximum number of cuts or nodes) is normally a number specified

before the training begins.

Boosting a decision tree involves training many trees ($\mathcal{O} \sim 1000$) and giving misclassified events a higher weight. A misclassified event is defined as a known signal event being placed on a background leaf and vice versa. By giving the events which are difficult to classify more weight, the next tree to be trained will effectively have to work harder in order to classify events correctly.

The total score on an event is deduced by following an event through from tree to tree and, for the algorithms used in this thesis, is simply given by the weighted sum of the scores over the individual trees.

Data sets are split into two (or more) sub samples, where one half is used for training the tree and the other is used for testing the tree, and the distributions of the event scores (the BDT output) for training and testing samples are compared for signal and background. Cases where the training sample performs better than the testing sample are referred to as over-trained trees, which is often due to the BDT becoming sensitive to the statistical fluctuations of the training sample.

The distribution of events scores for a given dataset can then be cut on in order to increase the fraction of signal events.

Appendix B

The *sPlot* technique

The *sPlot* technique is used extensively throughout this thesis. It is used in cases when there is a merged dataset which consists of data from different sources of data species, namely background and signal. These datasets are assumed to have two different sets of variables associated with the events they contain. Discriminating variables are those whose distributions are known for background and signal. Control variables are those whose distributions are unknown, or are assumed to be unknown.

The *sPlot* technique allows the distribution of the control variables for each data species to be deduced by using the species discriminating variable. This method relies on the assumption that there is no correlation between the discriminating variable and the control variable. The discriminating variable used in this thesis is always the mass distribution. The full mathematical description of the *sPlot* technique can be found in Ref [1] , the key points are outlined here.

An unbinned extended maximum likelihood analysis of a data sample of several species is considered. The log-likelihood is expressed as

$$\mathcal{L} = \sum_{e=1}^N \left\{ \ln \sum_{i=1}^{N_s} N_i f_i(y_e) \right\} - \sum_{i=1}^{N_s} N_i, \quad (\text{B.1})$$

where N is the total number of events considered, N_s is the number of species of event (i.e. two - background and signal), N_i is the average number of expected events for the i^{th}

species, y represents the set of discriminating variables, $f_i(y_e)$ is the value of the Probability Density Function (PDF) of y for event e for the i^{th} species and the control variable, x , does not appear in the expression of \mathcal{L} by definition.

For the simple (and not particularly practical) case of the control variable x being a function of y , i.e. completely correlated, one could naively assume that the probability of a given event of the discriminating variable y being of the species n would be given by

$$\mathcal{P}_n(y_e) = \frac{N_n f_n(y_e)}{\sum_{k=1}^{N_s} N_k f_k(y_e)}. \quad (\text{B.2})$$

The distribution for a control variable x for the n^{th} species, $M_n(x)$, can be deduced by histogramming in x and applying $\mathcal{P}_n(y_e)$ as a weight to event e . In this scenario the probability, $\mathcal{P}_n(y_e)$, would run from 0 to 1.

In the case considered in this thesis, where x is entirely uncorrelated with y , it can be shown that $\mathcal{P}_n(y_e)$ can be written as

$$\mathcal{P}_n(y_e) = \frac{\sum_{j=1}^{N_s} V_{nj} f_j(y_e)}{\sum_{k=1}^{N_s} N_k f_k(y_e)}, \quad (\text{B.3})$$

where V_{nj} is the covariance matrix between the species n and the j^{th} species. The inverse of this covariance matrix is given by the second derivative of $-\mathcal{L}$ in [Equation B.1](#).

The quantity in [Equation B.3](#) is donated as the sWeight. In this thesis the species, n , in [Equation B.3](#) is always the signal. Because of the presence of the covariant derivative the sWeight of an event can be both positive and negative. The more negative an event is, the more likely it is to be background and vice versa for positive sWeights. The signal distribution for the control variable x , $M_s(x)$, can again be deduced by histogramming events in x , applying the sWeight to each event.