

# CERN Summer Student Programme 2021 Report: $R_{K^*}$ Analysis and Lepton Flavour Universality

Ho Yin AU YANG

Supervisor: Prof. Luis Roberto FLORES-CASTILLO

The Chinese University of Hong Kong

## Abstract

The Standard Model predicts that charged leptons such as electrons and muons should have identical electroweak interaction strengths, which is known as lepton flavour universality. However, in some specific beauty quark decays, recent experimental results indicate that the ratio between branching fractions of electron and muon channels of the same decay may be deviant from unity, suggesting violations of the universality. In this work, we study the decay  $B_d^0 \rightarrow K^{*0} l^+ l^-$ , where  $l^-$  represents an electron or a muon, in preparation for the measurement of the ratio  $R(K^*)$ , which is the ratio of the branching fraction of the muon decay channel to that of the electron decay channel with normalization considering resonant decays that also produce the same particle products.

## 1 Introduction

The Standard Model of particle physics predicts that different charged leptons, namely electron, muon, and tau, should have the same electroweak interaction strength. This phenomenon is called Lepton Flavour Universality (LFU) and is respected by many particle decays [1]. However, growing evidence suggests possible LFU violation in some beauty quark decays. In the  $R_{K^*}$  analysis, we study the decay  $B_d^0 \rightarrow K^{*0} l^+ l^-$  of which there is a further decay of  $K^{*0} \rightarrow K^\pm \pi^\mp$ . Here  $l^- l^+$  can be an electron-positron pair or a muon-antimuon pair. Fig 1 shows the mechanisms of the decay allowed by the SM, both of which involve weak interactions.

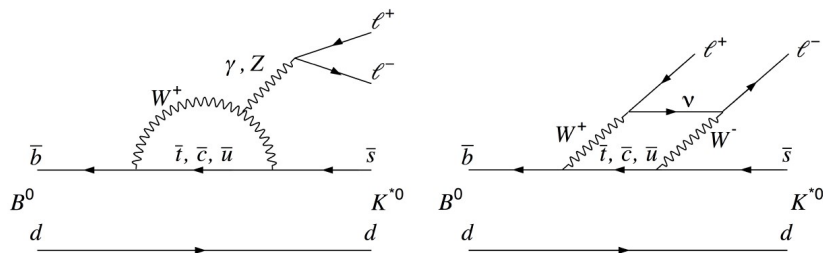


Figure 1: Standard Model allowed decay processes

The main idea is to measure the ratio between the branching fractions of the two decay channels, and if it deviates significantly from unity, LFU violation is implied. Naively, one may define the ratio by:

$$R_{K^{*0}} = \mathcal{B}(B_d^0 \rightarrow K^{*0} \mu^+ \mu^-) / \mathcal{B}(B_d^0 \rightarrow K^{*0} e^+ e^-) \quad (1)$$

But instead, we define  $R(K^*)$  as a double ratio:

$$R_{K^{*0}} = \frac{\mathcal{B}(B_d^0 \rightarrow K^{*0} \mu^+ \mu^-)}{\mathcal{B}(B_d^0 \rightarrow K^{*0} J/\psi(\rightarrow \mu^+ \mu^-))} \bigg/ \frac{\mathcal{B}(B_d^0 \rightarrow K^{*0} e^+ e^-)}{\mathcal{B}(B_d^0 \rightarrow K^{*0} J/\psi(\rightarrow e^+ e^-))} \quad (2)$$

The branching fractions of  $J/\psi$  channels are introduced to normalize both numerator and denominator above because the  $J/\psi \rightarrow l^+l^-$  decay is known to be consistent with LFU to within 0.4%, and this cancels out the systematic uncertainties due the difference in reconstruction efficiencies between electrons and muons. In the analysis we define the non-resonant and resonant channels by  $q^2$ , the di-lepton invariant mass squared, where the non-resonant channel is set at  $1.1 < q^2 < 6.0 \text{ GeV}^2/c^4$  and resonant at  $6.0 < q^2 < 11.0 \text{ GeV}^2/c^4$ . Resonant decays refer to  $J/\psi$  and other charmonium ( $c\bar{c}$ ) decay channels, and the resonant channel hosts the  $q^2$  region that corresponds to large magnitudes of such states being produced; for instance,  $J/\psi$  is produced resonantly at  $q^2 = 9.59 \text{ GeV}^2$ .

By SM predictions,  $R(K^*)$  should be close to unity, and previous measurements by LHCb have shown deviations with the ratio measured in 2017 at  $0.69 \pm_{0.07}^{+0.11} (\text{stat}) \pm 0.05 (\text{syst})$  in the non-resonant channel for  $1.1 < q^2 < 6.0 \text{ GeV}^2/c^4$  (Fig 2), which is 2.4-2.5 standard deviations from SM predictions [2].

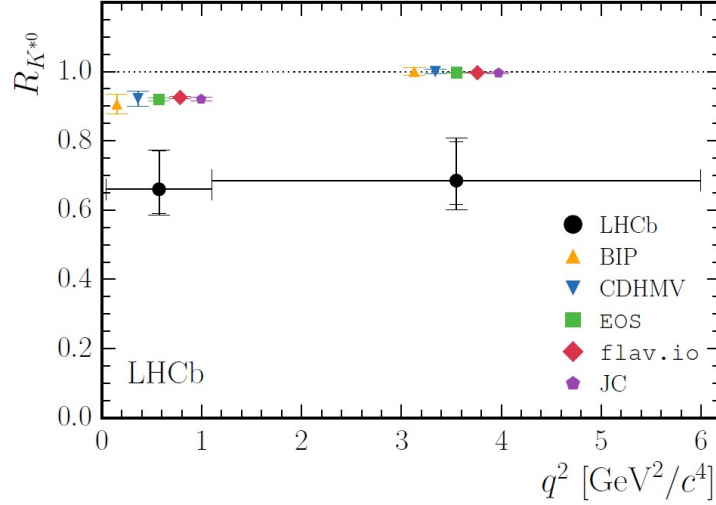


Figure 2: LHCb measurements of  $R(K^*)$  compared to SM predictions (coloured)

## 2 Current Progress

The  $R(K^*)$  analysis group in ATLAS consists of two groups: one focusing on the electron channel and one on the muon channel. The main goal is to devise efficient methods to identify the signal events of interest from the background events, such that further calculations can be performed. In this section, I will give a brief overview of the current progress in both groups.

### 2.1 Electron Channel

As one of the first steps of the event selection process, triggers select which events to keep or discard right after collision data is produced. In the electron channel, the BeeX trigger is used, which triggers on low electron  $p_T$  and di-electrons with a small angular separation. Trigger efficiencies and the corresponding uncertainties are being studied, which helps us understand how well triggers reject background and accept signal events, and the error of these processes. Trigger options for Run 3 are being studied as well.

Signal reconstruction is another key step being studied. Recall that the main decay process concerned is:

$$B_d^0 \rightarrow K^{*0} l^+ l^- \rightarrow K \pi l^+ l^- \quad (3)$$

Therefore, an event is identified as a  $B_d^0$  candidate if there is a pair of well-reconstructed, oppositely charged particles identified as electron-positron or muon-antimuon, together with two well-reconstructed,

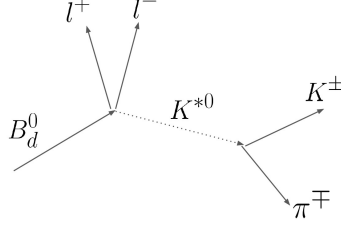


Figure 3: Schematics of the decay and the vertices

oppositely charged particles where one is identified as a kaon and the other as a pion. The outgoing particle candidates should be roughly consistent with their respective known mass values, and their tracks should reconstruct a good enough di-lepton vertex and 4-track vertex [2]. The vertex fit requirement at this stage is loose.

For event selection, a common and simple method is the cut-based approach, where events are selected if their parameters fall within certain ranges, e.g.  $4700 < m(B_d^0) < 5700 \text{ MeV}$  sets a cut based on the B mass. However, using only this method is not suitable here as it is estimated by extrapolation that only  $\approx 40$  signal events survive in the non-resonant channel, yet the background level is similar to the resonant channel, i.e. an overall low signal rate with basically no background reduction.

The current solution is to employ machine learning (ML), in particular neural networks (NNs), where a combination of signal Monte Carlo samples and data-driven background is used. First a very loose cut-based selection helps reduce the data volume, then the NNs are trained in non-resonant regions but then applied to both the non-resonant and resonant channels. Applying to data, currently there are improvements in signal efficiency but not enough background is removed. Other NN settings or combinations of parameters used in the NN features are some options being explored.

## 2.2 Muon Channel

The muon channel is less developed compared to the electron channel, but has very similar workflow and hence can borrow the tools such as ML codes from the electron channel. Currently, we are modifying electron channel ML codes to adapt to the muon channel, where for example, the electron and muon n-tuples have different sets of branch names, which requires checking and modification to ensure a smooth transition. We are also testing ML codes on muon data samples as part of the checking procedure.

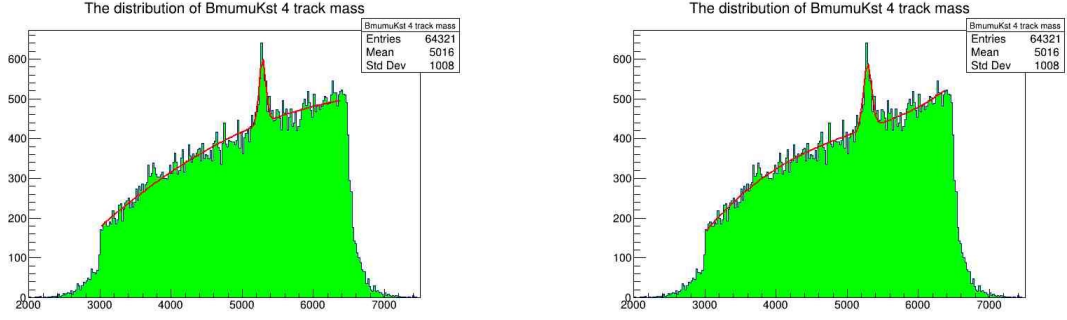
Regarding the muon channel data, currently the n-tuples are too large to be derived, and tighter selection schemes are being considered for production. Trigger studies are also ongoing to decide which triggers are suitable for the muon channel.

## 3 Practice Work Done

Before performing actual analysis, I have done practice work on sample data this summer to familiarize myself with plotting, fitting, and statistical analysis tools among others, which is summarized in this section.

### 3.1 Fitting histograms from test n-tuples

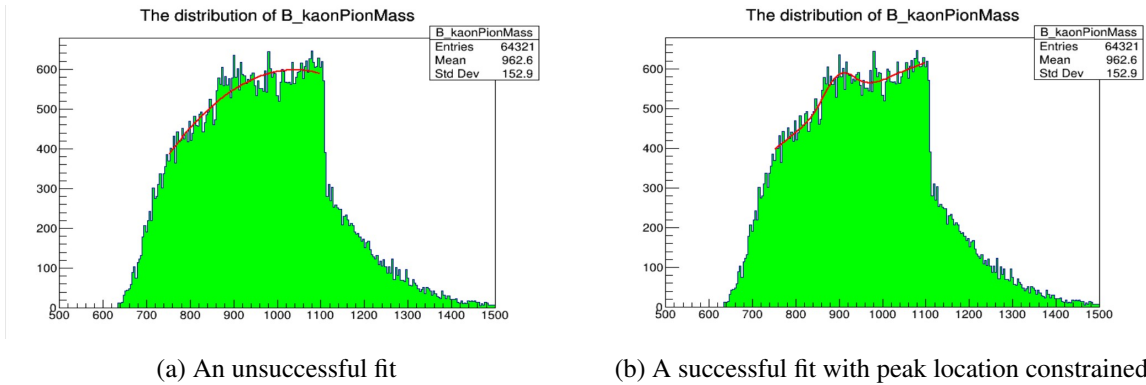
After learning the basic functionalities of the ROOT framework such as histogram fitting, root files reading and processing, a natural practice would be to use real data to perform fitting. By inspecting an n-tuple from a sample derivation of RK\* samples, several histograms regarding masses are fitted, where the fit function is a gaussian peak overlaying a background function of selected forms, such as a second-order polynomial (Fig 4a) or a third-order polynomial (Fig 4b). Cut-offs are seen in the histograms as a result of selection processes, and the fitting range is determined by eye accordingly.



(a) Gaussian peak over second-order polynomial back-ground (b) Gaussian peak over third-order polynomial back-ground

Figure 4: Fitting of a histogram from test n-tuples

In some cases, the peak may not be recognized due to its small amplitude or relatively large background noise. An example of a failed fit is shown in Fig 5a, which only gives a second-order polynomial background fit. In Fig 5b, constraints on the location of the peak are set to a range where the peak seems to be, and the fitted function indeed has a peak plus background form. Note that this may seem to defeat the purpose of using fitting to find the location of the peak and hence determine, for instance, experimental values of masses. However, constrained fits are actually used if the bounds of a value are known, which avoids fitting failure. It also serves as a good exercise to show how parameters constraints can be implemented in code.



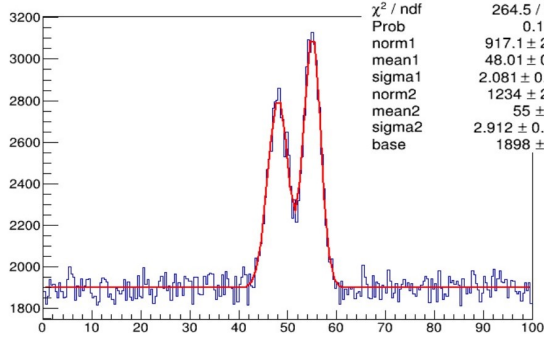
(a) An unsuccessful fit

(b) A successful fit with peak location constrained

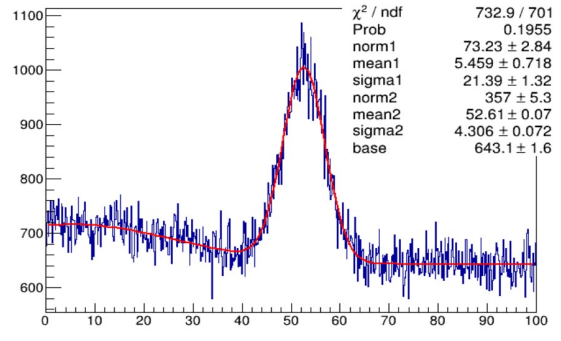
Figure 5: Fitting of a histogram with a smaller peak

### 3.2 Fitting randomly generated two-peak histograms

In real scenarios, there are often two signals with similar variance and size in the same plot. It is crucial to resolve the two peaks, which may sometimes be closely overlapping. In this exercise, sample two-peak histograms are generated with random parameters and fitted with a function of a flat background plus two gaussian peaks. It is noticed that fitting often fails when no constraints are set on the fitting parameters (mean and sigma of the two peaks). As an adjustment, the fitting parameters are constrained to within  $\pm 5\%$  of the true parameters in the following examples shown in Fig 6a and 6b.



(a) An example with two clear peaks



(b) An example with a wide, small amplitude left peak

Figure 6: Fitting two-peak histograms

### 3.3 Pull distributions of fitting

It is crucial to determine not only the accuracy of fit, but also the accuracy of the error given by the fitting algorithm. For each fit, we obtain a pull value for a fitted parameter by

$$\text{pull} = \frac{\text{fitted value} - \text{actual value}}{\text{fitter error}} \quad (4)$$

Using the two-peak histograms, the parameters studied are the means and sigmas of the two peaks, and by doing fits repeatedly on histograms generated with the same parameters, a pull distribution is obtained. Theoretically, the pull distribution should be a standard Gaussian centred at zero with unit width for a truly random sample and an accurate fitter. Shown in Fig 7 is an example of a pull distribution of the left peak mean for a certain set of parameters for histogram generation, which shows an expected result with roughly mean value zero and sigma value one. Also note that the  $\chi^2/\text{ndf}$  ratio is relatively low, indicating a good gaussian fit.

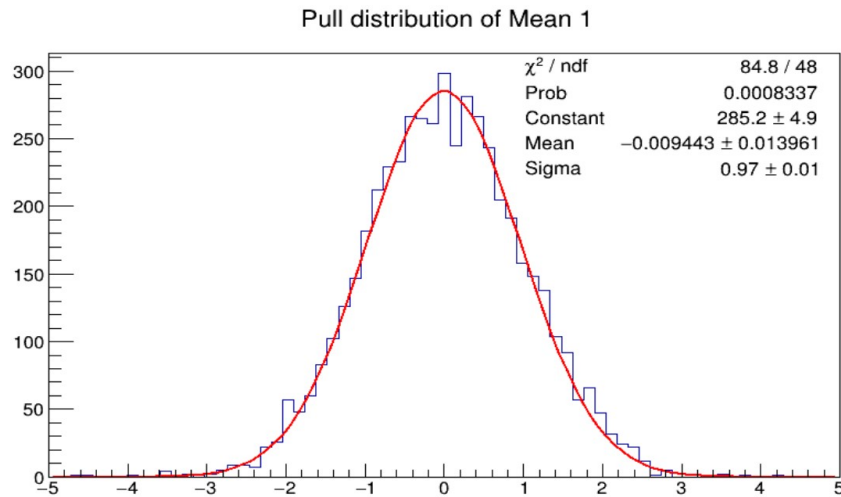


Figure 7: Pull distribution of left peak mean; 500000 entries per histogram, constraints  $\pm 10\%$

As the number of entries in the randomly generated histograms will affect the quality of fit, the pull distributions are affected subsequently. Realistically, particle physics analyses often have a low number of signal events, so it is interesting to study how the fitting accuracies change with data volume. In Fig 7, there are 500,000 entries per histogram, and in Fig 8a and 8b, it is decreased to 10000 and 1000 respectively. The pull values reside near  $\approx \pm 1$ , an indication that fitted values are consistent with the

true values within the uncertainty calculated by the fitter. Here the fitted value is constrained to  $\pm 10\%$  of the true value, and a probable explanation is that due to very low number of entries, the fitter fails to find a good fit within the constraints and hence returns a value at the artificial boundary with similar estimated errors.

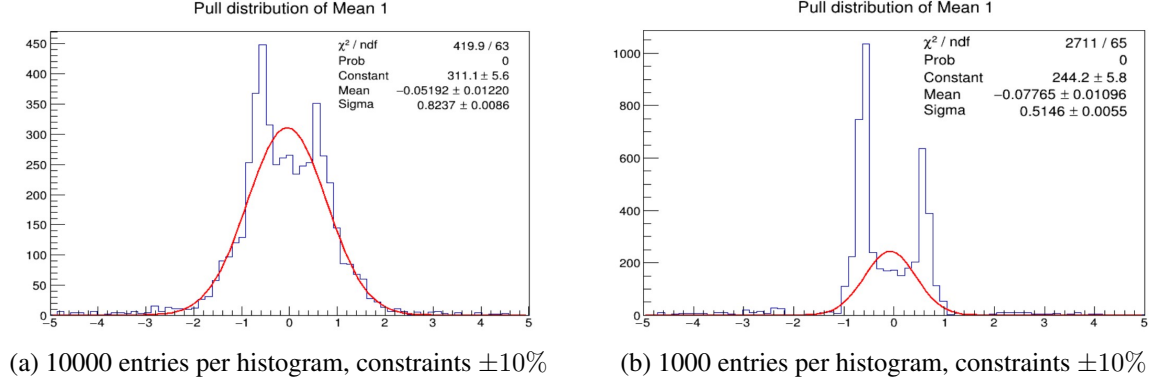


Figure 8: Pull distributions with lower number of entries per histogram

A natural next step is to loosen the constraints. Shown in Fig 9a are constraints set at  $\pm 30\%$ , while in Fig 9b, there are no constraints except the fitted mean values have to be within the histogram x-axis range. Although both distributions are far from gaussian, the no constraints graph peaks nearer to zero and show less asymmetry.

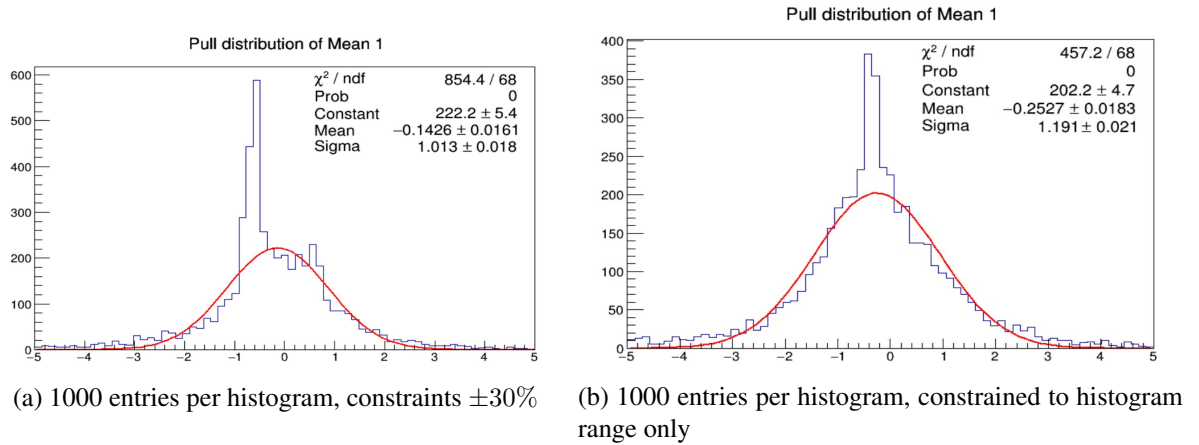


Figure 9: Pull distributions with loosened constraints

## 4 Future Work

Regarding the  $R(K^*)$  analysis overall, the LHC Run 3, which will start in 2022, will provide more data for analysis and hence higher sensitivity, which may help reduce uncertainties in the final measurement. For the muon channel, analysis will be performed once muon data becomes available, with tasks such as deriving Monte Carlo samples to simulate background processes.

## Acknowledgements

I would like to express my sincere gratitude to my supervisor, Prof Luis Roberto FLORES-CASTILLO, for his guidance and advice in the project. I would also like to thank the CERN Summer Student Team for their effort in organizing this meaningful program in this difficult time.

## Bibliography

- [1] Aaij, Roel, et al. "Test of lepton universality in beauty-quark decays." arXiv preprint arXiv:2103.11769 (2021).
- [2] Clarke, P. E. L., et al. "Test of lepton universality with  $B^0 \rightarrow K^{*0} \ell^+ \ell^-$  decays." Journal of High Energy Physics 1708 (2017): 055.