The Search for $\Lambda_b^0 \to J/\psi\Xi^- K^+$ and $\Xi_b^0 \to J/\psi\Xi^-\pi^+$

Noah Luch[1], Luis Miguel Garcia Martin[2], Radoslav Marchevski[2]

[1] *Princeton University, NJ, United States*
[2] *École polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland.*

Abstract

This analysis aimed to measure the branching ratios of $\Lambda_b^0 \to J/\psi\Xi^- K^+$ and $\Xi_b^0 \to J/\psi\Xi^-\pi^+$ with the normalization channels, $\Lambda_b^0 \to J/\psi\Lambda$ and $\Xi_b^- \to J/\psi\Xi^-$ respectively. This is done by comparing the simulated decays to the LHCb data, performing several cuts, training a machine learning algorithm, and finally performing a massfit to find the branching ratios.

Contents

1 Introduction

1.1 Motivation

The measurement of $\Xi_b^0 \to J/\psi \Xi^- \pi^+$ and its branching ratio has not, to our knowledge, been seen yet. This decay would give insights into the $\Xi(1530)$ mixing and pentaquark candidates. Complementing this, $\Lambda_b^0 \to J/\psi \Xi^- K^+$ has been recently measured [1], and we hope to give a more precise value for its branching ratio. Further, this branching ratio would give insight into the $\Lambda(1405)$ resonances and determine parts of the chiral lagrangian.

1.2 Analysis Strategy

By comparing simulated, Monte Carlo (MC) data to data from the LHCb, we hope to remove the majority of the background from the LHCb data to have a clear isolated signal. To do so, we first correct the MC data to ensure that events are weighted properly. Then, we preform an intial, preselection cut, followed by a larger cut, which is determined by a machine learning algorithm. Finally, our analysis is blinded, meaning we mask out the signal region of the mass distribution for the signal channels. We nevertheless fit the mass distribution of our signal channels ($\Lambda_b^0 \to J/\psi \Xi^- K^+$ and $\Xi_b^0 \to J/\psi \Xi^- \pi^+$) under an assumed branching ratio to determine our number of signal events and model validity.

We use data from the LHCb Run II between 2016-2018, under several stripping requirements. For Λ_b^0 we use LLL and LDD tracks. For Ξ_b^0, we use LLL, LDD, and DDD tracks.

We apply several triggers to the stripping requirements. Specifically, L0_Muon, L0_Data_Muon1_pt> 36, HLT1 DiMuonHighMass, DiMuonLowMass, TrackMuonMVA, and HLT2 DiMuonDetachedJpsi.

2 Corrections to the MC

In order to ensure that our simulated data (MC) does not over- or underweight events which have been constructed differently, we must correct our MC using several weights. The corrections we have applied so far are: Trigger corrections, track calibration weights, kinematic weights, and decay tree fitter (DTF) χ^2 weights.

The trigger corrections are calculated using TIS (Trigger Independent of Signal) TOS (Trigger On Signal) weights. These weights are calculated before this analysis, and correct the triggers used in the analysis, as described above. Similarly, track calibration weights are calculated before this analysis, and apply to each particle, essentially weighting how accurate each particle's track is, depending on its momentum and psuedorapidity.

Once these weights have been applied, we calculate the kinematic and DTF weights. Due to various nuanced reasons, the transverse momentum and Decay Tree Fitter χ^2 of the simulation and data differ, and as such, must be corrected. This is done by first isolating the signal from the background in the normalization channels by performing an initial massfit on the normalization channel data after preselection. The process of performing the massfit is the same as described in section 4.1. Following the massfit, we have a certain number of signal events and distribution, which we can then assign each event a probability of signal to, using a package known as `sWeights` with the `COWs` weights [2]. Using this package, we may add weights to the data corresponding to their probability of signal.

From this weighted data, we then compare the transverse momentum of the mother particle and the DTF χ^2. As expected, they do not match, so we take the ratio of the distributions, and fit the result to a sextic polynomial. Finally, evaluating the p_T and DTF χ^2 of each event in the MC on the fitted sextics yields the weights of those events.

3 Event Selection

3.1 Preselection

Following the corrected distributions, we compare the MC with the high-mass sideband data distributions of several variables. As we are confident that the MC represents signal, and the high-mass sideband data, background, we hope to apply cuts on several variables to remove as much of the background as possible, while keeping as much of the signal as possible. If we hope to effectively remove the background, we cannot simply remove an equal number of signal and background, as there is much more background than signal.

Given this, we perform several cuts on the distributions of the particles. If no cut may easily be found the variable is either not cut, or left for the machine learning algorithm.

3.2 Boosted Decision Tree

Following the preselection, we use a machine learning algorithm to further differentiate between the signal (MC) and background (high-mass sideband). We use a Boosted Decision Tree (BDT) algorithm, where future generations of the model learn at a fixed rate, and generations which make better cuts are "boosted" in importance to the next generation.

Examining the distributions of the signal and background data yielded the variables we trained our BDT on. Once we have found a good sample, we also perform an iterative analysis to determine which variables may also be relevant, by re-training our BDT on each of the candidate variables. Finally, we improve our BDT by calibrating the learning rate, testing ratio, and number of estimators to maximize the efficiency of the BDT.

We then give our trained BDT the full preselected data, which will yield the probability of each event being signal. To perform the cut, then, we calculate the Punzi figure of merit,

$$\mathrm{FoM}(t) = \frac{\epsilon_S(t)}{a/2 + \sqrt{B(t)}}, \tag{1}$$

where $\epsilon_S(t)$ is the signal efficiency at each cut location (t), a is the significance (we use $a = 5$), and $B(t)$ is the number of background events, calculated by integrating a sideband massfit to an exponential. In other words, we calculate essentially how much signal and how much background remains after we apply remove all events below a certain probability, t. The cut which maximized this figure of merit (maximizes signal efficiency and minimizes background efficiency) is where we take our cut.

4 Massfit

4.1 Method

Following the BDT cut, we fit the mass distribution of each decay using a two crystal-ball model. This is described by,

$$CB_i(m; \mu, \sigma, \alpha_i, n_i) = N_i \cdot \begin{cases} \exp\left(\frac{(m-\mu)^2}{2\sigma^2}\right), & \text{for } \frac{m-\mu}{\sigma} > -\alpha_i \\ A_i\left(B_i - \frac{m-\mu}{\sigma}\right)^{-n_i} & \text{for } \frac{m-\mu}{\sigma} \leq -\alpha_i \end{cases}, \tag{2}$$

with,

$$\begin{aligned}
A_i &= \left(\frac{n_i}{|\alpha_i|}\right)_i^n \exp\left(-\frac{|\alpha_i|^2}{2}\right), \\
B_i &= \frac{n_i}{|\alpha_i|} - |\alpha_i|, \\
N_i &= \frac{1}{\sigma(C_i + D_i)}, \\
C_i &= \frac{n_i}{|\alpha_i|(n-1)} \exp\left(-\frac{|\alpha_i|^2}{2}\right), \\
D_i &= \sqrt{\frac{\pi}{2}}\left(1 + \mathrm{erf}\left(\frac{|\alpha|}{\sqrt{2}}\right)\right).
\end{aligned} \tag{3}$$

Rather than fitting with the coefficients in front of the crystal ball equation, we allow those coefficients to float, only caring about the ratio of the two crystal ball coefficients.

We fit the MC distribution of each channel with the BDT cut applied to these two crystal balls. In order to ensure that the data is modeled after the MC, we fix the α_i and n_i from these fits. We then fit the data to the two crystal balls, allowing only the mean and sigma to float, and introducing a background term modeled as an exponential.

4.2 Toy Fits

As our analysis is blinded, we do not perform the full massfit on the signal channels, so as to not find out how many signal events we do end up having. Rather, we show that our modeling works by generating toy distributions.

For the normalization channels, we are able to fit the full mass distribution, as they are not blinded. Given this, we generate signal and background event according to a poisson distribution, with the mean given by our original massfit signal and background events. From these, we are able to tell if our model has any kind of bias. If we calculate,

$$\frac{N_{\text{gen}} - N_{\text{sig}}}{\sigma_{\text{gen}}}, \tag{4}$$

we are able to compare the pull of the N_{sig} variable. If the mean is equivalent with zero, and the standard deviation equivalent with one, this shows that our model is not dependent on the number of signal events we have. Indeed for the normalization channels, we find that our mean is zero and sigma is one, which shows that our model has worked well.

Something similar may be done for the signal channel, though we must derive an N_{sig}, as we are not able to fit for it. To do this, we use the formula:

$$\frac{N(A \to B)}{N(C \to D)} = \frac{f_A}{f_B} \frac{\mathcal{B}(A \to B)}{\mathcal{B}(C \to D)} \frac{\epsilon_{A \to B}}{\epsilon_{C \to D}}, \tag{5}$$

where $C \to D$ is our normalization channel, $A \to B$ is our signal channel, $\mathcal{B}(...)$ is the branching ratio, f is the fragmentation fraction, and $N(...)$ is the number of signal events.

To find $N_{\text{sig}}(\Lambda_b^0 \to J/\psi \Xi^- K^+)$, we first find $N_{\text{sig}}(\Lambda_b^0 \to J/\psi \Lambda)$ using the same fitting procedure described above. The ratio of branching ratios $\frac{\mathcal{B}(\Lambda_b^0 \to J/\psi \Xi^- K^+)}{\mathcal{B}(\Lambda_b^0 \to J/\psi \Lambda)}$, may be estimated using the recent measurement [1] of $\frac{\mathcal{B}(\Lambda_b^0 \to J/\psi \Xi^- K^+)}{\mathcal{B}(\Lambda_b^0 \to \psi(2S)\Lambda)} = 0.025 \pm 0.008 \pm 0.009$ and $\frac{\mathcal{B}(\Lambda_b^0 \to \psi(2S)\Lambda)}{\mathcal{B}(\Lambda_b^0 \to J/\psi \Lambda)} = 0.501 \pm 0.033 \pm 0.019$ [3]. Further, we assume that $\frac{\mathcal{B}(\Lambda_b^0 \to J/\psi \Xi^- K^+)}{\mathcal{B}(\Lambda_b^0 \to J/\psi \Lambda)} \approx 6.25 \cdot \frac{\mathcal{B}(\Xi_b^- \to J/\psi \Xi^- \pi^+)}{\mathcal{B}(\Xi_b^- \to J/\psi \Xi^- 1)}$ due to $\mathfrak{su}(3)$. Finally, using the efficiencies described above, the fragmentation fractions cancel out, and a rough estimate of $N(A \to B)$ is calculated.

We then create more toy datasets under the assumption of these signal events, and fit the sidebands of the mass distribution to an exponential for the background events. Generating these, and calculating the pull now of the ratio of branching ratios, we hope still that our mean is close to zero and standard deviation is close to one. While this is very true for Ξ_b^0, we are about three sigmas off in our measurement of the Λ_b^0 sigma. So far, we are not sure why this happens.

5 Systematics

In our systematics, we include the possibility of modeling our distribution using two gaussians instead of two crystal balls, and get an uncertainty of under one percent. We look at how our efficiencies change under the assumption of a higher background category, and we propagate the error of our efficiencies and branching ratios. From these, we get a good estimation of our systematic uncertainties.

6 Conclusions and Further Analysis

Our results thus far show that, for the most part, our models work well to determine the ratio of branching ratios of our signal channels, $\Lambda_b^0 \to J/\psi \Xi^- K^+$ and $\Xi_b^0 \to J/\psi \Xi^- \pi^+$. As such, while we have not yet measured what branching ratio we have found, we are on track to determine it soon, and have begun an analysis note to push the project further into a paper on the first measurement of $\Xi_b^0 \to J/\psi \Xi^- \pi^+$ and its branching ratio. We would also improve the measurement of the branching ratio of $\Lambda_b^0 \to J/\psi \Xi^- K^+$.

We still have yet to check many systematics, and look into our signal massfit for Λ_b^0, but our results look very promising so far, and will hopefully turn into a paper in due time.

Acknowledgements

I would like to thank my wonderful supervisors, Luis Miguel Garcia Martin, and Radoslav Marchevski, who have been so supportive and welcoming in helping me work on this project. I cannot stress enough how much I have learned and how engaging and fun you have made this project for me. Thank you.

References

[1] CMS, CERN Report No., , 2023 (unpublished).

[2] H. Dembinski, M. Kenzie, C. Langenbruch, and M. Schmelling, *Custom Orthogonal Weight functions (COWs) for event classification*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **1040**, 167270 (2022).

[3] ATLAS, G. Aad *et al.*, *Measurement of the branching ratio* $\Gamma(\Lambda_b^0 \to \psi(2S)\Lambda^0)/\Gamma(\Lambda_b^0 \to J/\psi\Lambda^0)$ *with the ATLAS detector*, Phys. Lett. B **751**, 63 (2015), 1507.08202.