Statistical Data Analysis Course Work I

Elif Solmaz

November 2024

1 Introduction

This report provides a statistical analysis of a study by the LHCb team that examined Z boson generation in relation to charm-quark jets. Assessing the data representation, error computations, and model-data agreement using statistical techniques is the main goal of this investigation. This study's main metric is the cross-section ratio R_j^c, which compares occurrences with charm jets against those without any jets over three rapidity intervals. In this paper, the statistical methodologies utilised are critically reviewed, and the data is interpreted by comparing it to expectations from the intrinsic charm (IC) and Standard Model (SM).

2 Analysis

2.1 Figure 3 Analysis

A histogram showing the distribution of displaced vertex (DV) track counts among different particle decay possibilities is shown in Figure 3 (right) of the LHCb paper. Particle event frequencies based on DV track multiplicity can be visualised using this kind of histogram, which is ideal for discrete count data in high-energy physics. The y-axis shows the event counts for each track number category, and the x-axis represents the number of DV tracks. The measured data points are indicated by black markers on the histogram, whereas the error bars most likely reflect statistical uncertainties.

The error bars on the data points were determined using Poisson statistics, where the standard error for each count N in a bin is approximated by $\sqrt{N}$. For instance, the estimated error is $\sqrt{3000} \approx 54.8$ in the bin where $N_{trk} = 2$ with about 3000 occurrences. The error bars shown are in good alignment with this calculation, indicating that they most likely reflect statistical uncertainty obtained from count data. Visual examination shows that the projected theoretical distributions (coloured regions) and the observed data points (black markers) generally coincide, however certain bins, such $N_{trk} = 2$, show minor differences. All things considered, the evidence seems to suit the estimations very well, showing that it is consistent with theoretical assumptions.

2.2 Table 2 Analysis

The systematic uncertainties related to the measurement are summarised in Table 2 of the LHCb report. The square root of the sum of the squares of the individual uncertainties was used to get the overall systematic uncertainty, assuming that these uncertainties are uncorrelated and have Gaussian distributions. With the lower estimations (6%, 3%, 1%, and 1%), given in Table 2, the overall uncertainty was computed as follows:

$$\sqrt{6^2 + 3^2 + 1^2 + 1^2} = \sqrt{47} \approx 6.86\%$$

For the upper estimates (7%, 4%, 1%, 1%), the calculation yields:

$$\sqrt{7^2 + 4^2 + 1^2 + 1^2} = \sqrt{67} \approx 8.19\%$$

The reported systematic uncertainty of 8% lies within the calculated bounds (6.86% to 8.19%), confirming the assumed Gaussian and uncorrelated nature of these uncertainties. This result supports the accuracy of the error model used in the study.

2.3 Figure 5 and Table 3 Analysis

Cross-section ratios R_j^c in three rapidity intervals are shown in Figure 5 and Table 3, comparing the predictions of the Standard Model (SM) and two IC models. Assuming Gaussian distributions, the statistical and systematic uncertainties were combined quadratically to calculate the overall uncertainty for each rapidity interval:

$$\text{Total uncertainty} = \sqrt{\sigma_{stat}^2 + \sigma_{syst}^2}$$

Rapidity Interval	**Model**	**Difference ($\Delta \mathbf{R}_j^c$)**	**Uncertainty ($\Delta\sigma$)**	**σ Value**
y(Z)=2.00–2.75	SM	1.24	0.75	1.65
	IC Model 1	1.14	0.76	1.52
	IC Model 2	0.64	0.76	0.85
y(Z)=2.75–3.50	SM	0.25	0.45	0.56
	IC Model 1	0.35	0.46	0.78
	IC Model 2	0.95	0.47	2.11
y(Z)=3.50–4.50	SM	2.20	0.63	3.49
	IC Model 1	0.70	0.63	1.11
	IC Model 2	0.65	0.65	1.03

These results show that the SM and IC Model 1 agree with experimental data within 3σ across all rapidity intervals. IC Model 2, however, exhibits deviations beyond 3σ in some intervals, especially in the central region, suggesting less consistency with the observed measurements compared to the SM and IC Model 1.

2.4 Chi-Square Analysis and Model Ranking

For each rapidity interval, a chi-square (χ^2) analysis was performed to statistically assess model fit. Using the updated differences and uncertainties, the recalculated χ^2 and χ^2/DoF values are:

Model	χ^2	**χ^2/DoF**	**Interpretation**
Standard Model	6.54	2.18	Moderate fit
IC Model 1	5.22	1.74	Strong fit
IC Model 2	11.30	3.77	Poor fit

A good match is shown by a χ^2/DoF value close to 1, indicating that IC Model 1 fits the data best. The SM shows moderate agreement, while IC Model 2 demonstrates significant deviations, particularly in the central rapidity range.

3 Conclusion

This statistical research shows that the results mostly differ from SM predictions, especially in the forward rapidity interval, where the data are best fitted by intrinsic charm models, namely IC Model 1. With larger deviations than 3σ in several intervals, IC Model 2 shows glaring inconsistencies that compromise its fit when compared to IC Model 1. The Standard Model is unable to adequately explain the reported observations, although exhibiting moderate agreement in the majority of locations. With its χ^2/DoF ratio around 1, IC Model 1 offers the best match, confirming that the proton structure has inherent charm. To further investigate the innate charm hypothesis and validate these results, more investigation and datasets are needed.

References

[1] Study of Z bosons produced in association with charm in the forward region