Data cleaning for K_S^0 decay reconstruction ML model CBM

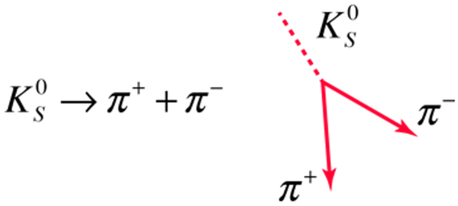
Julian Nowak

VIII 2021

1 K_S^0 (short lived Kaon) reconstruction:

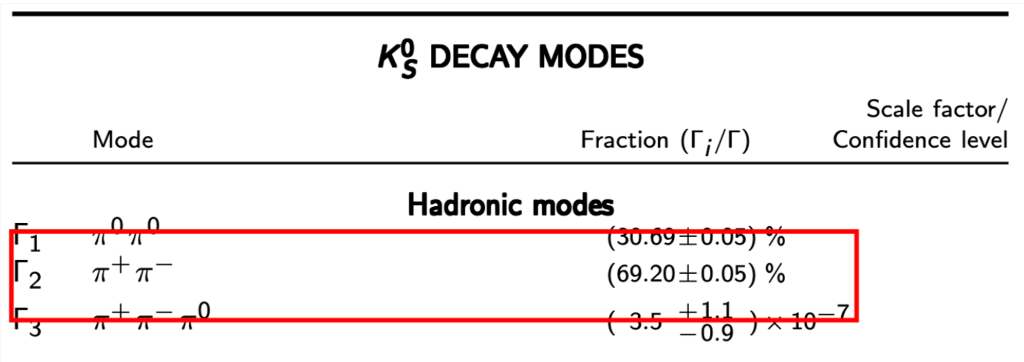
K-short (short lived Kaon) reconstruction:

- Mother particle: K_S^0 (PDG = 310)
- Mass: 497.611±0.013 MeV/c^2
- Mean lifetime: $8.958 \cdot 10^{-11}$ s
- Charge = 0
- Meson, composed of two quarks: $d\ \bar{s}$ or $s\ \bar{d}$
- Strange particle



K_S^0 decay diagram [1]

In the main decay mode:



K_S^0 DECAY MODES

	Mode	Fraction (Γ_i/Γ)	Scale factor/ Confidence level
	Hadronic modes		
Γ_1	$\pi^0\pi^0$	(30.69 ± 0.05) %	
Γ_2	$\pi^+\pi^-$	(69.20 ± 0.05) %	
Γ_3	$\pi^+\pi^-\pi^0$	$(3.5\ ^{+1.1}_{-0.9}) \times 10^{-7}$	

Figure 1: K_S^0 decay modes [2]

K_S^0 particle decays into π^+ (PDG = 211) and π^- (PDG = -211). Mass of each pion equals 139.57039(18) MeV/c^2

2 Objective

Our model will be trained to correctly distinguish between the *signal* - pairs of π^+ and π^- which were produced in the K_S^0 decay, and *background* - pairs of pions which aren't result of the decay. We assume that the majority of the particles of invariant mass in 5σ region around the mass peak $= 0.4981$ GeV/c^2 should be recognized as the K-short signal candidates. The ML model learns, which parameters values should be associated with the signal, and which with the background. As the mathematical ML model has no clue which data values are physically correct, we need to clean the data.

3 Data cleaning

To reject the numeric values of parameters which don't have physical sense, but are present in the data set, we apply some selection criteria before the beginning of the model training. Similarly, we reject some values which might be possible, but are rare enough, so we reject them to reduce the data.

3.1 Invariant mass

As the K-short particle decays into two pions (in the decay mode we're able to reconstruct) its invariant mass cannot be smaller than the mass of the two pions, so:

$$\text{mass} > 0.279 \text{ GeV}/c^2$$

Also, to reduce the amount of data, we only accept the particles with:

$$\text{mass} < 1.5 \text{ GeV}/c^2$$

3.2 Distances and *x, y, z* cordinates

Distance between the primary vertex (the point where the collision of the nuclei happens), and the secondary vertex (the extrapolated point where the two daughter particles should have crossed each other) - l and the distance of closest approach between the two pions - DCA - obviously cannot be smaller than 0. So:

$$DCA,\ l,\ \tfrac{l}{\Delta l} > 0$$

Also, due to the sizes of the tracking system (the largest station has an are of 100 cm^2):

$$DCA < 100 \text{ cm}$$

For the same reason:

$$|x|, |y| < 50 \text{ cm}$$

As the particle has to hit 3 stations of the tracking system, and the last two are placed above 80cm"

$$l < 80\text{cm}$$

For the same reason:

$$|z| < 80 \text{ cm}$$

To reduce the data:

$$\tfrac{l}{\Delta l} < 5000$$

3.3 Momentums

The fixed target geometry of the detector requires that:

$$p_Z > 0 \text{ GeV/c}$$

To reduce the data, we only preserve:

$$p < 20 \text{ Gev/c};\ p_T < 3 \text{ GeV/c}$$

3.4 Chi square

Since χ^2 is a square distance, all the values must be bigger than zero:

$$\chi^2 > 0$$

Following the work of Olga for Λ decays, for now I set the rest of chi square cut values to:

- χ^2 first and second $< 3 \cdot 10^7$
- $\chi^2_{geo} < 1000$
- $\chi^2_{topo} < 100000$

3.5 Pseudorapidity

As pseudorapidity $\eta = -\ln\tan(\frac{\theta}{2})$, and the Silicon Tracking System (STS) cover the polar angles between 2.5° and 25°, we decide to constrain the pseudorapidity values:

$$1.5 < \eta < 3.82$$

References

[1] http://hyperphysics.phy-astr.gsu.edu/hbase/Particles/kaon.html

[2] https://pdg.lbl.gov/2021/listings/contents_listings.html