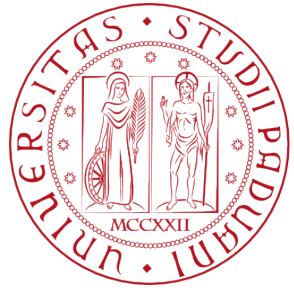
Search for charmless B decay in Belle II experiment

Student: Daniele Dal Santo

Dipartimento di Fisica e Astronomia "Galileo Galilei"
Corso di Laurea in Physics
10th September 2020



Università degli Studi di Padova

The skimming

The higher luminosity that characterizes the data taking of the experiment in the ongoing run gives rise to the necessity of introducing a pre-selection of the data called skimming. Its aim is to discard background events enabling us to save just a small fraction of the increased amount of data.

The skimming

The higher luminosity that characterizes the data taking of the experiment in the ongoing run gives rise to the necessity of introducing a pre-selection of the data called skimming. Its aim is to discard background events enabling us to save just a small fraction of the increased amount of data.

What we expect is that the efficiency of the skimming defined as

$$\frac{\#\ \textit{signal events passing the skimming}}{\#\ \textit{signal events}} \approx 100\%, \qquad (1)$$

but this fact hasn't been checked yet on the channel $B_0 \to \eta'(\eta(\gamma\gamma)\pi\pi)K_S$.

The skimming

The higher luminosity that characterizes the data taking of the experiment in the ongoing run gives rise to the necessity of introducing a pre-selection of the data called skimming. Its aim is to discard background events enabling us to save just a small fraction of the increased amount of data.

What we expect is that the efficiency of the skimming defined as

$$\frac{\#\ \textit{signal events passing the skimming}}{\#\ \textit{signal events}} \approx 100\%, \qquad (1)$$

but this fact hasn't been checked yet on the channel $B_0 \to \eta'(\eta(\gamma\gamma)\pi\pi)K_S$.

The efficiency for the channel $B_0 \to \phi(KK)K_S$ has been recently checked and is $\approx 50\%$.

Signal selection

To be more precise: there could be some signal events which are rejected by the skimming, but we expect them to be included in those rejected by the following selection which is subsequently applied:

Signal selection

To be more precise: there could be some signal events which are rejected by the skimming, but we expect them to be included in those rejected by the following selection which is subsequently applied:

- $E_\gamma > 150\,MeV$
- $0.92 < M_{\eta'} < 1\,GeV/c^2$
- $0.5 < M_\eta < 0.57\,GeV/c^2$
- $0.49 < M_{K_S} < 0.51\,GeV/c^2$
- $cos\theta_{p,v} > 0.99$

Signal selection

To be more precise: there could be some signal events which are rejected by the skimming, but we expect them to be included in those rejected by the following selection which is subsequently applied:

- $E_\gamma > 150\,MeV$
- $0.92 < M_{\eta'} < 1\,GeV/c^2$
- $0.5 < M_\eta < 0.57\,GeV/c^2$
- $0.49 < M_{K_S} < 0.51\,GeV/c^2$
- $cos\theta_{p,v} > 0.99$

If so, then the skimming procedure works well since it has a mild effect on the signal, while still reducing the background.

KS test

The most relevant part of my analysis involved the comparison between a Monte Carlo generated file of 200 000 signal events subject to the skimming procedure and one not subject to it.

KS test

The most relevant part of my analysis involved the comparison between a Monte Carlo generated file of 200 000 signal events subject to the skimming procedure and one not subject to it.

My aim was to check whether there were significant variations in the distribution of the variables in the data frame.

KS test

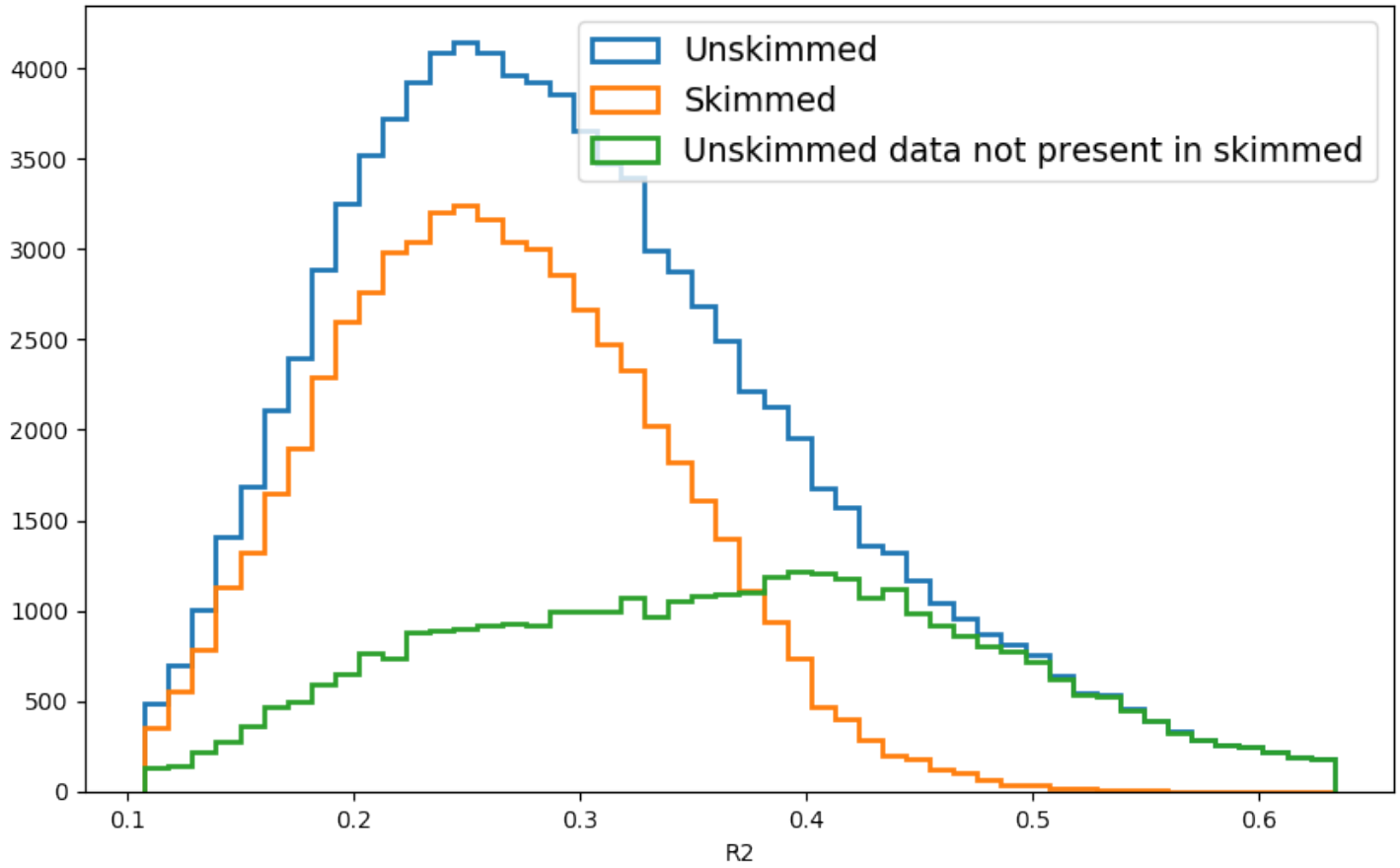
The most relevant part of my analysis involved the comparison between a Monte Carlo generated file of 200 000 signal events subject to the skimming procedure and one not subject to it.

My aim was to check whether there were significant variations in the distribution of the variables in the data frame.

The variables at disposal are $\gtrsim 300$ so a KS test was employed to select variables for which the shape of the distribution changed due to skimming.

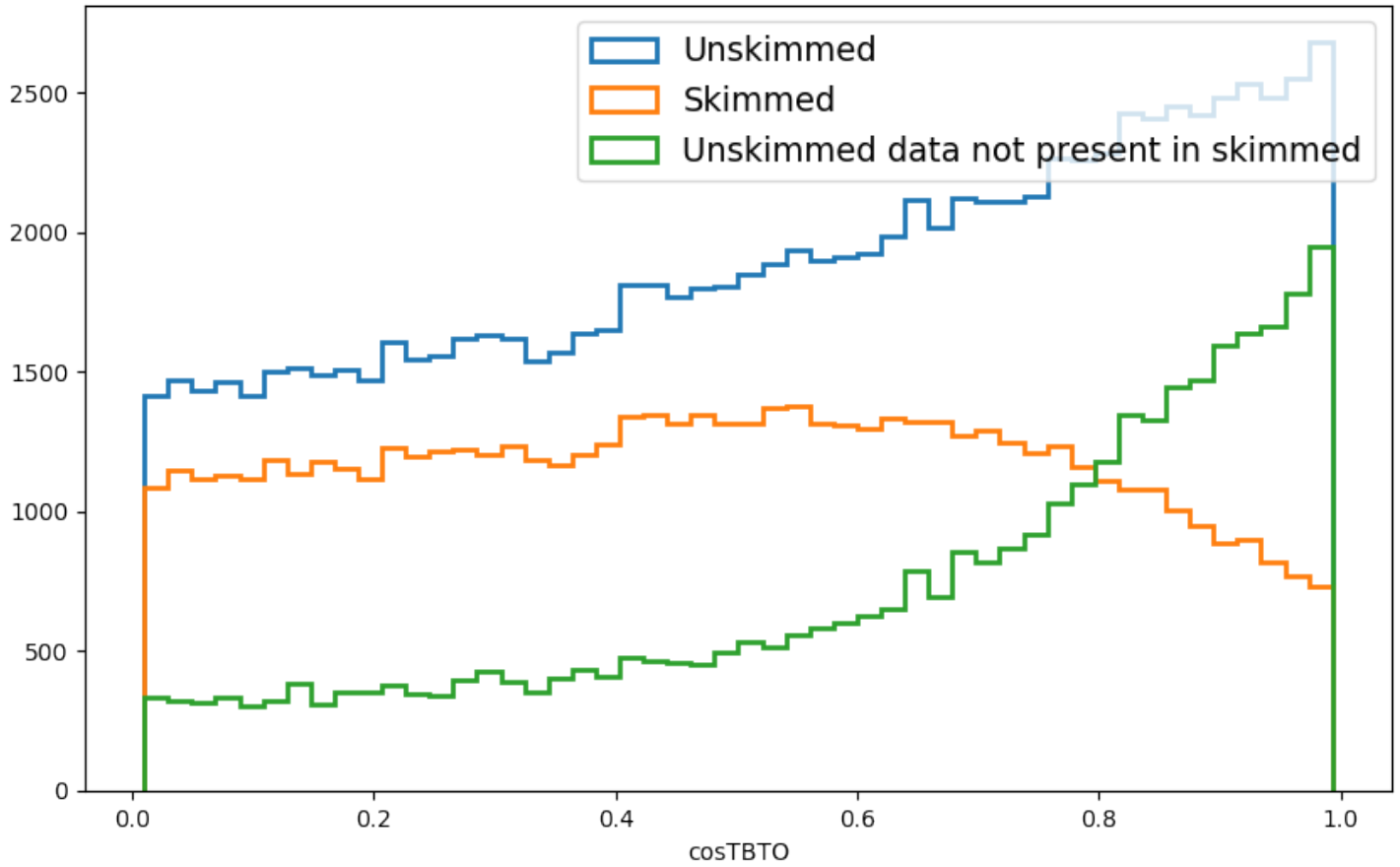
Background related variables

Between the variables highlighted by the KS statistics, many are correlated ones used to distinguish signal and background.



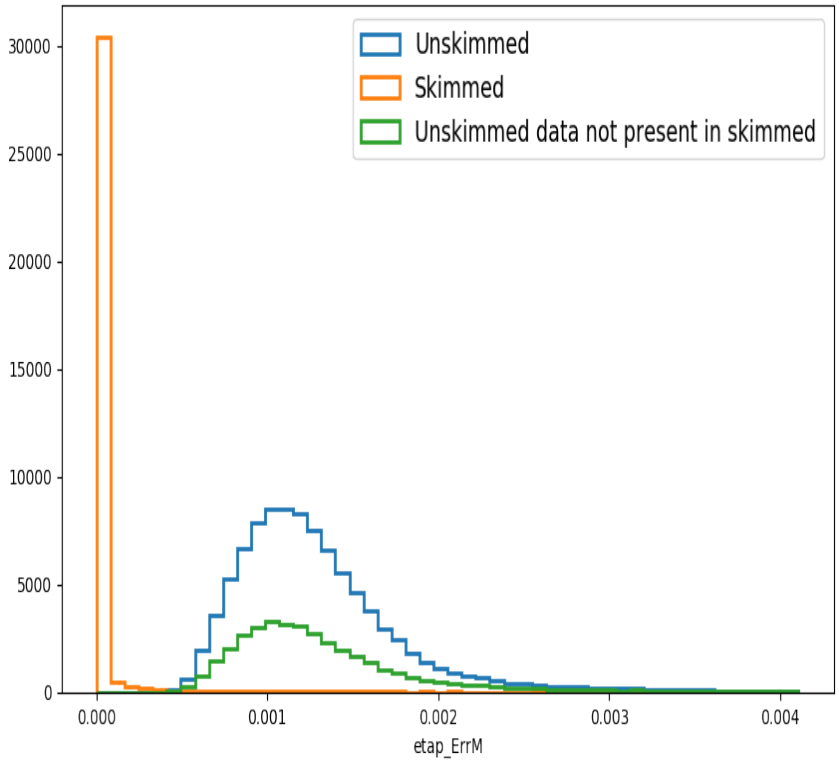
Background related variables

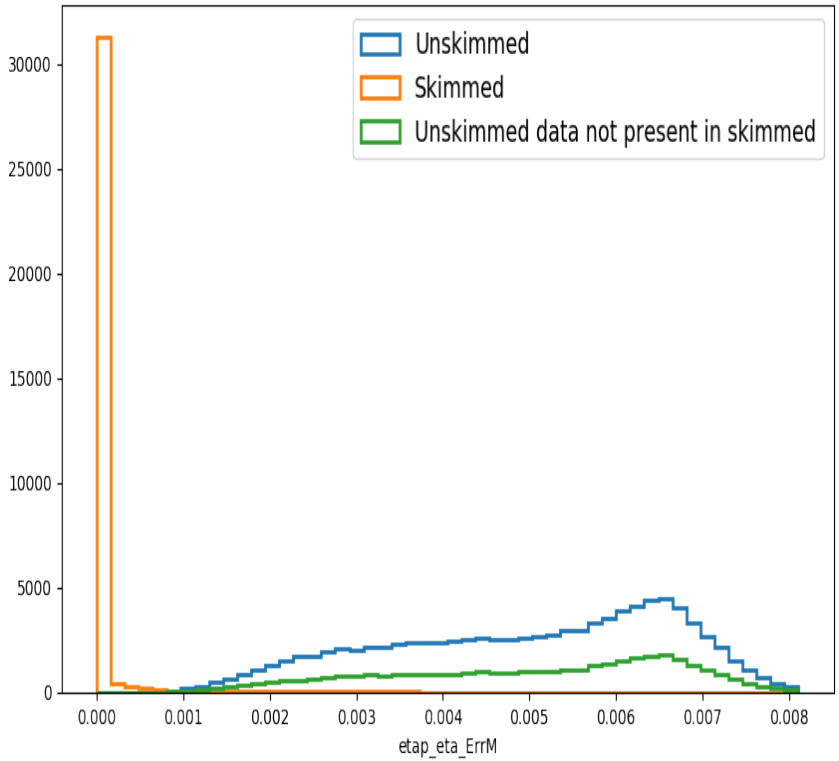
Between the variables highlighted by the KS statistics, many are correlated ones used to distinguish signal and background.



Mass errors

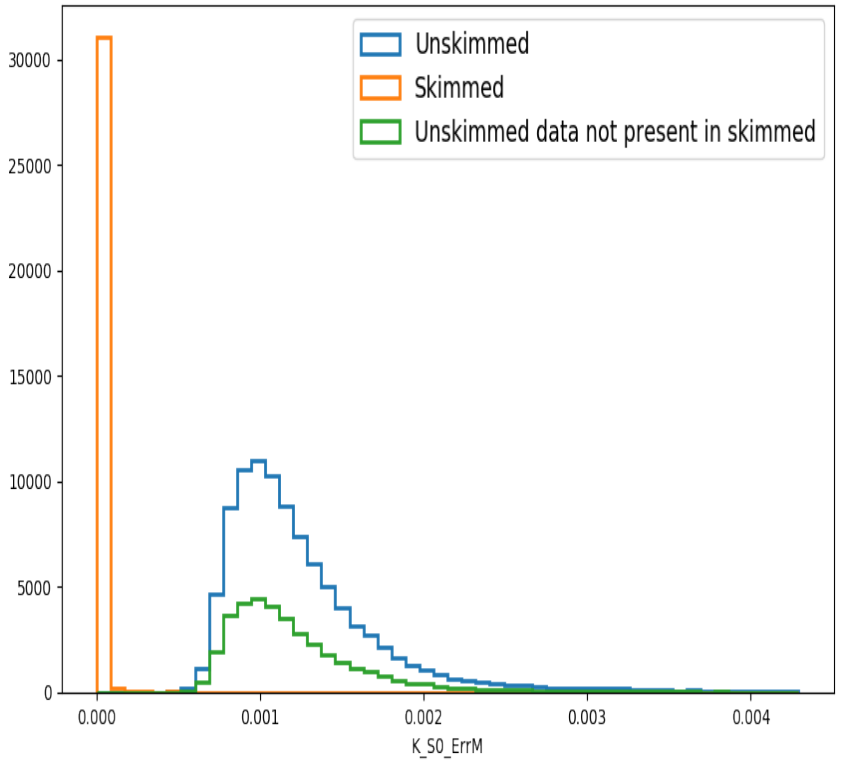
A first relevant observation coming from this qualitative comparison is the fact that errors on masses are suppressed after skimming.





Mass errors

A first relevant observation coming from this qualitative comparison is the fact that errors on masses are suppressed after skimming.



Global efficiency

As a first estimate of the selection efficiency defined in Eq.1, one can compute the ratio between the number of events in the data frame before and after the skim, which gives:

$$\epsilon_1 = \frac{49402}{79267} = 62.3^{+0.3}_{-0.4}\ \%. \tag{2}$$

Global efficiency

As a first estimate of the selection efficiency defined in Eq.1, one can compute the ratio between the number of events in the data frame before and after the skim, which gives:

$$\epsilon_1 = \frac{49402}{79267} = 62.3^{+0.3}_{-0.4}\ \%. \tag{2}$$

Applying also the selections defined before, one obtains a ratio

$$\epsilon_2 = \frac{38971}{62555} = 62.3^{+0.4}_{-0.4}\ \%. \tag{3}$$

Global efficiency

As a first estimate of the selection efficiency defined in Eq.1, one can compute the ratio between the number of events in the data frame before and after the skim, which gives:

$$\epsilon_1 = \frac{49402}{79267} = 62.3^{+0.3}_{-0.4}\ \%. \tag{2}$$

Applying also the selections defined before, one obtains a ratio

$$\epsilon_2 = \frac{38971}{62555} = 62.3^{+0.4}_{-0.4}\ \%. \tag{3}$$

$\rightarrow$ Eq.3 reveals that the selection contained in the skimming affects the data frame in a stronger way than expected.

Background reduction

These same ratios were computed using 2 files (one skimmed, one not) containing $c\bar{c}$ background events. The results are:

$$\epsilon_1^{bkg} = \frac{9705}{74645} \sim 12\%, \tag{4}$$

$$\epsilon_2^{bkg} = \frac{438}{2588} \sim 17\%. \tag{5}$$

Background reduction

These same ratios were computed using 2 files (one skimmed, one not) containing $c\bar{c}$ background events. The results are:

$$\epsilon_1^{bkg} = \frac{9705}{74645} \sim 12\%, \tag{4}$$

$$\epsilon_2^{bkg} = \frac{438}{2588} \sim 17\%. \tag{5}$$

$\rightarrow$ The skimming procedure is indeed able to cut a significant amount of background events.

Efficiency histogram

Then, I considered the variables E_γ, $M_{\eta'}$, M_η, M_{K_S} and $cos\theta_{p,\nu}$ of the skimmed and unskimmed data frames, after having applied the selections defined in the 2^{nd} slide.

Efficiency histogram

Then, I considered the variables E_γ, $M_{\eta'}$, M_η, M_{K_S} and $cos\theta_{p,\nu}$ of the skimmed and unskimmed data frames, after having applied the selections defined in the 2^{nd} slide.

The regions between the quantile 0.05 and 0.95 of the unskimmed distributions were divided in 45 bins and the ratio

$$r_i = \frac{\#\ events\ in\ the\ i^{th}\ bin\ in\ the\ skimmed\ distribution}{\#\ events\ in\ the\ i^{th}\ bin\ in\ the\ unskimmed\ distribution} \quad (6)$$

was computed for all the bins.

Efficiency histogram

Then, I considered the variables E_γ, $M_{\eta'}$, M_η, M_{K_S} and $cos\theta_{p,\nu}$ of the skimmed and unskimmed data frames, after having applied the selections defined in the 2^{nd} slide.

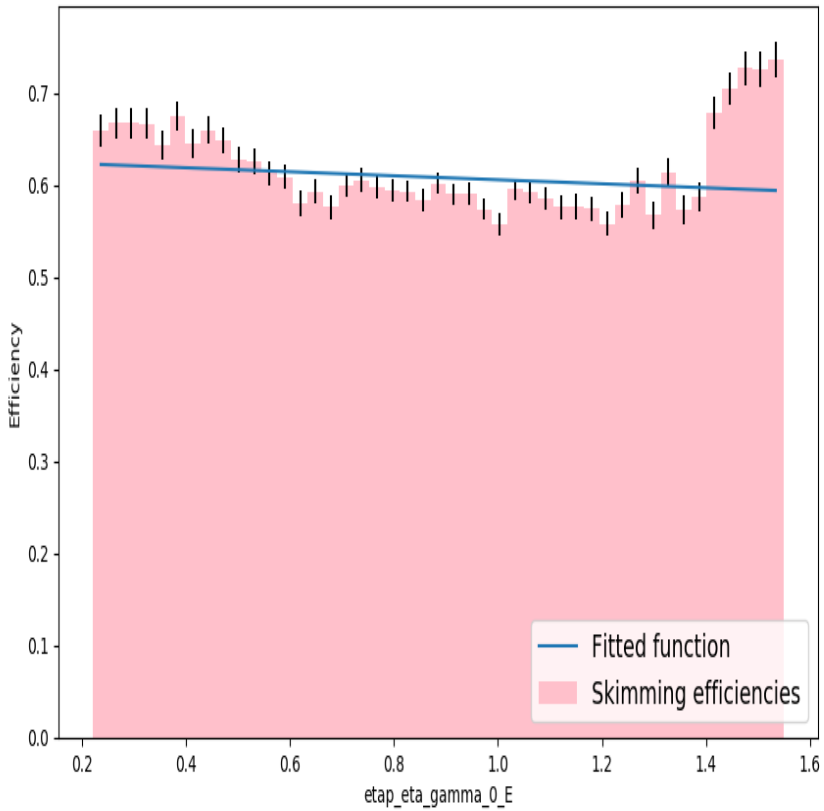
The regions between the quantile 0.05 and 0.95 of the unskimmed distributions were divided in 45 bins and the ratio

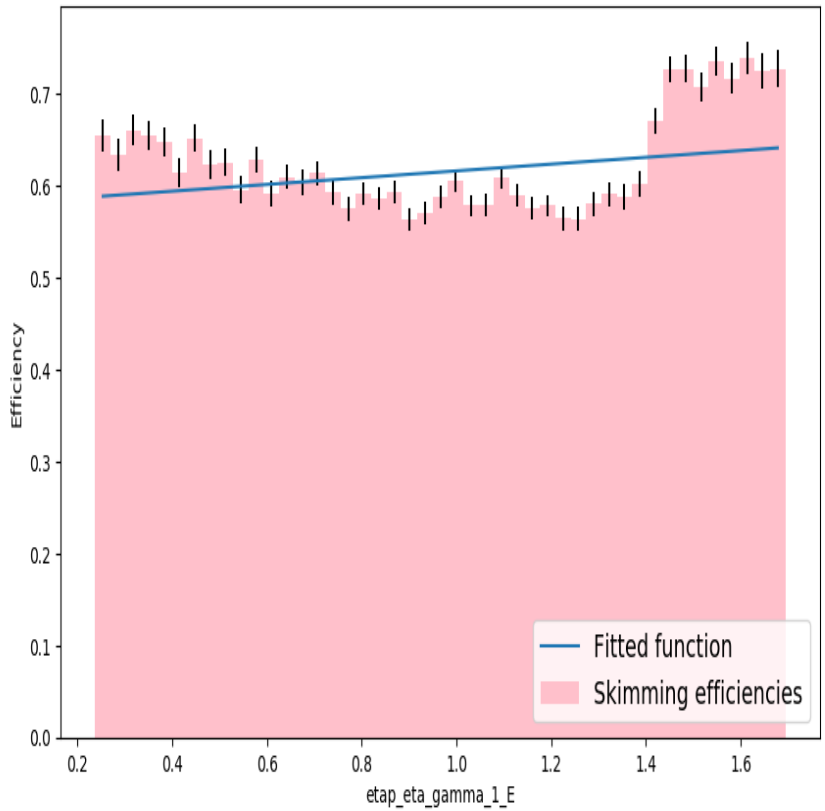
$$r_i = \frac{\#\ events\ in\ the\ i^{th}\ bin\ in\ the\ skimmed\ distribution}{\#\ events\ in\ the\ i^{th}\ bin\ in\ the\ unskimmed\ distribution} \tag{6}$$

was computed for all the bins.

We expected $r_i \approx constant \approx 1$.

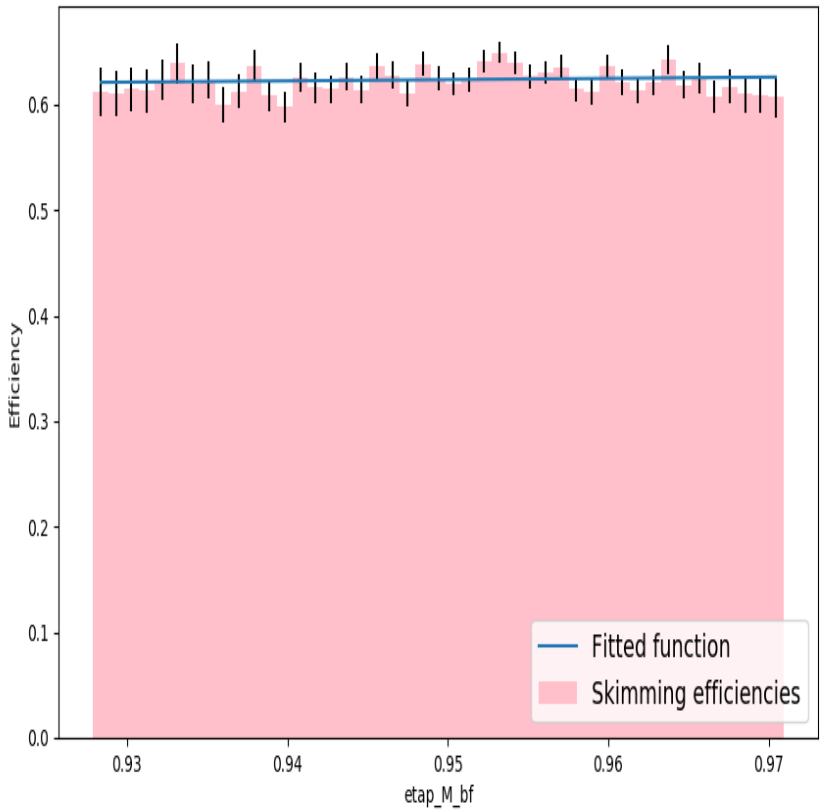
Efficiency histogram

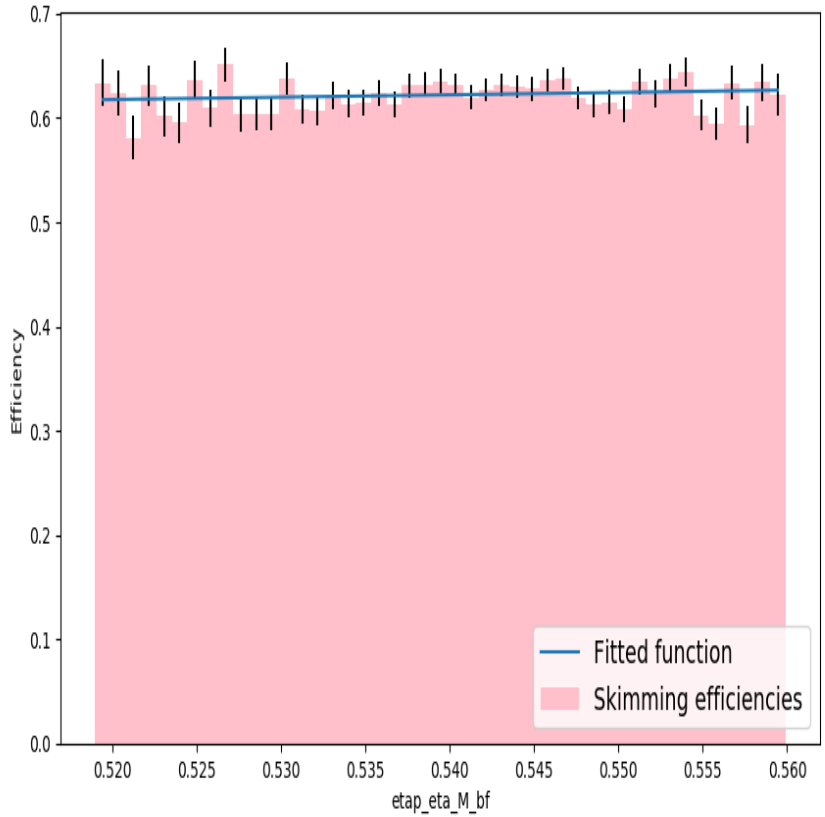




Variable	Intercept	Slope
etap_eta_gamma_0_E	0.609 ± 0.006	-0.02 ± 0.02 GeV^{-1}
etap_eta_gamma_1_E	0.615 ± 0.007	0.04 ± 0.02 GeV^{-1}

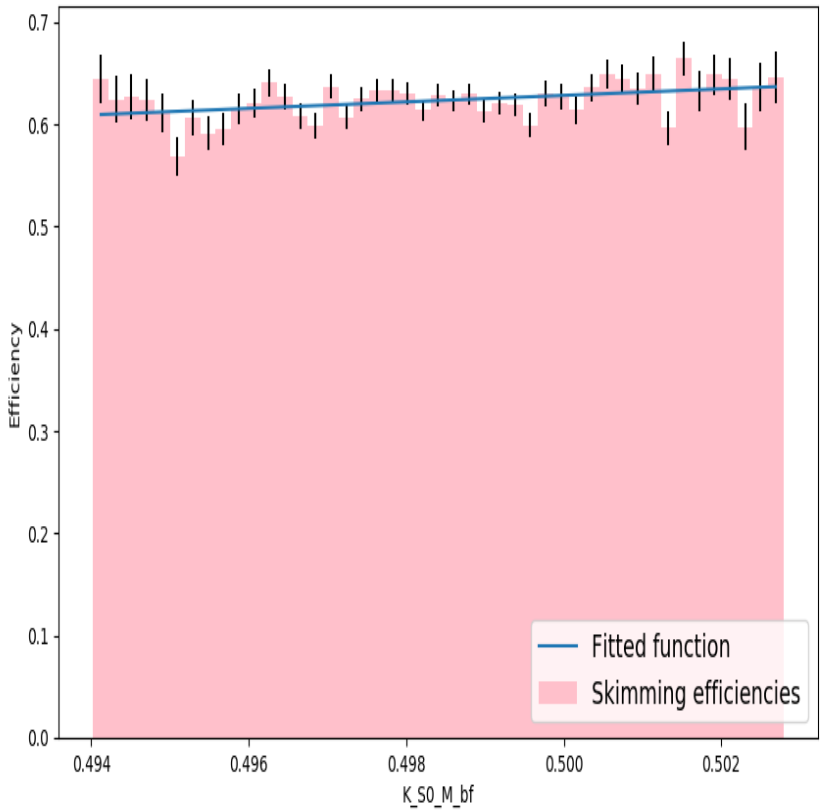
Efficiency histogram

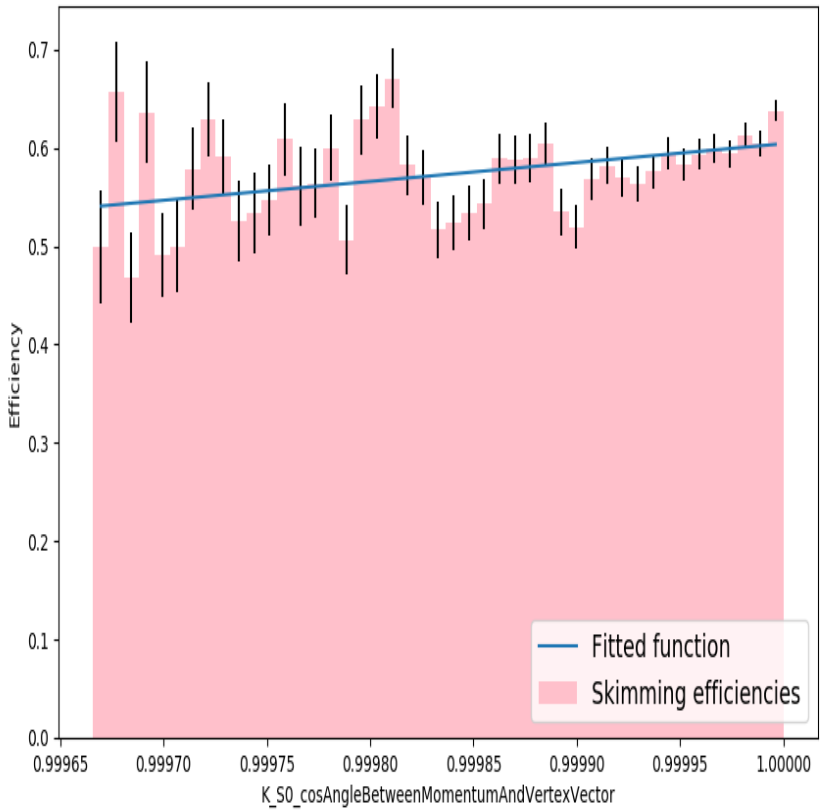




Variable	Intercept	Slope
etap_M_bf	0.624 ± 0.002	0.1 ± 0.2 GeV^{-1}
etap_eta_M_bf	0.622 ± 0.002	0.2 ± 0.2 GeV^{-1}

Efficiency histogram





Variable	Intercept	Slope
K_S0_M_bf	0.623 ± 0.002	3 ± 1 GeV^{-1}
$\cos\theta_{p,v}$	0.572 ± 0.007	190 ± 60

Next step

As it is evident the current skimming procedure determines a (too) strong rejection of signal events with respect to the current analysis procedures.

Next step

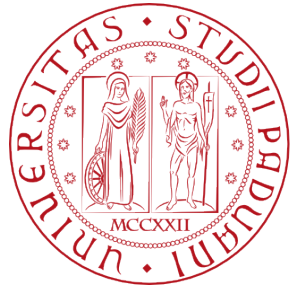
As it is evident the current skimming procedure determines a (too) strong rejection of signal events with respect to the current analysis procedures.

$\rightarrow$ The next step of my work (already started) will be to deeply understand how the skimming procedure works and analyse which is the condition that determines the loss of nearly 40% of the signal.

Search for charmless B decay in Belle II experiment

Student: Daniele Dal Santo

Dipartimento di Fisica e Astronomia "Galileo Galilei"
Corso di Laurea in Physics
10th September 2020



Università degli Studi di Padova