

# LCFIPlus check

Masakazu Kurata

07/25/2018

# Start from here

Ryo's study

1: 50k samples

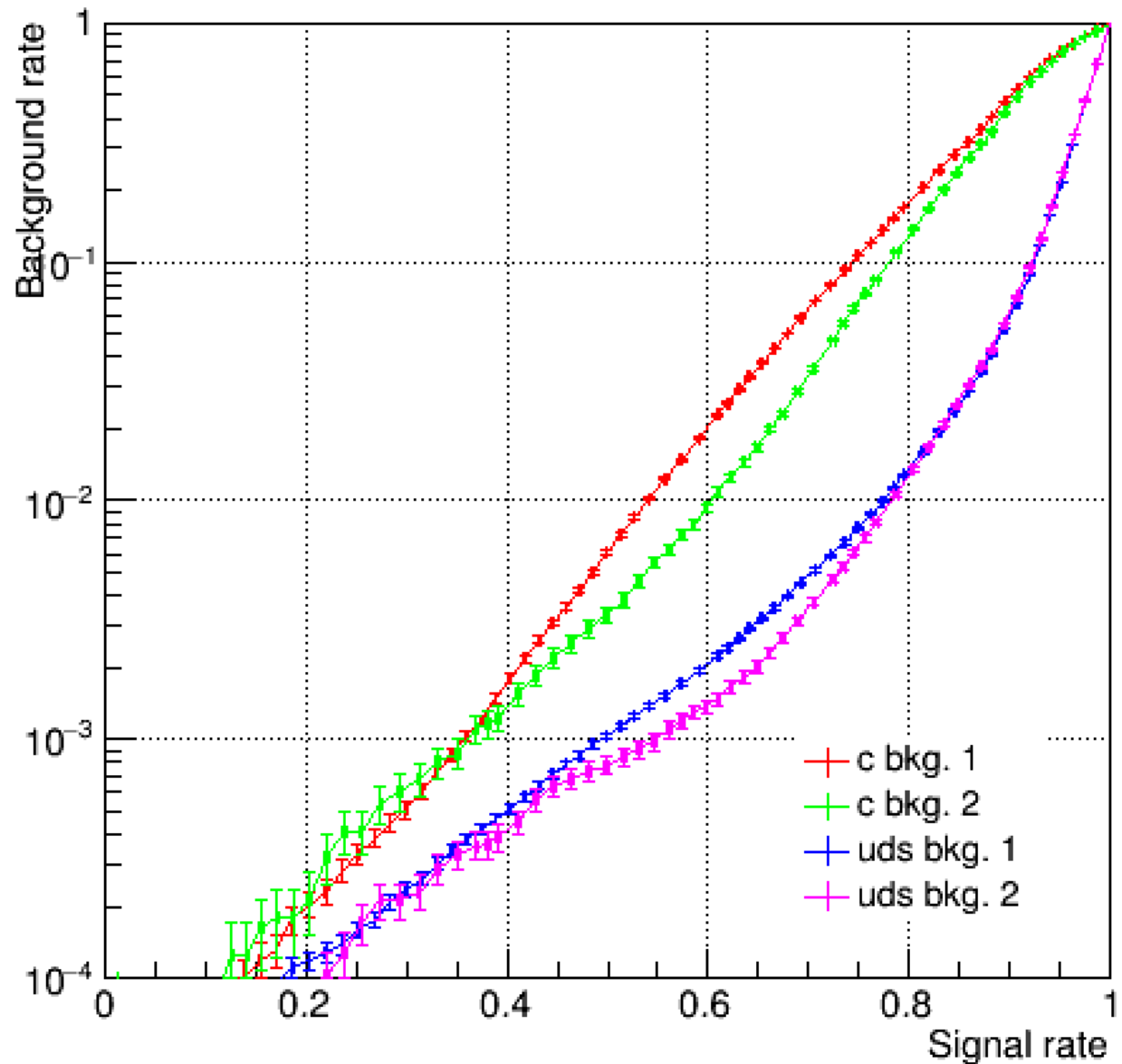
2: 20k samples

Fluctuation very large

And

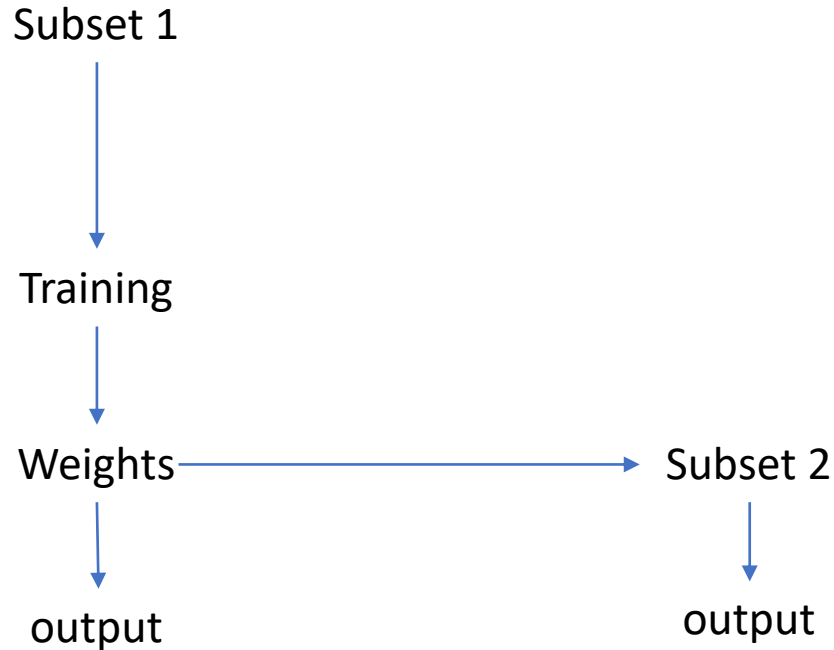
It is expected that worse  
than DBD

Over-training?



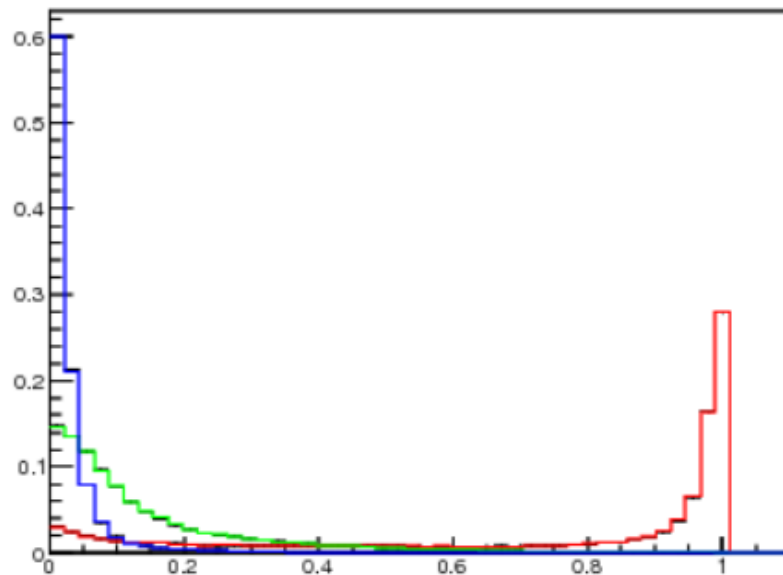
# Weight check

- Apply training weights to different sample

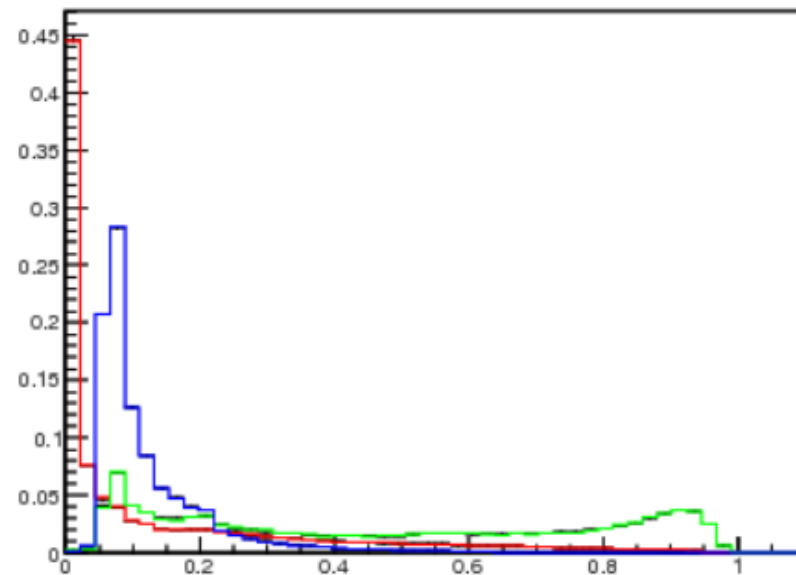


- Weight files are coming from Ryo's study
- Both subsets are independent of each other
  - Both have  $\sim 20k$  events
  - $\sim 50k$  case also checked(do not show)

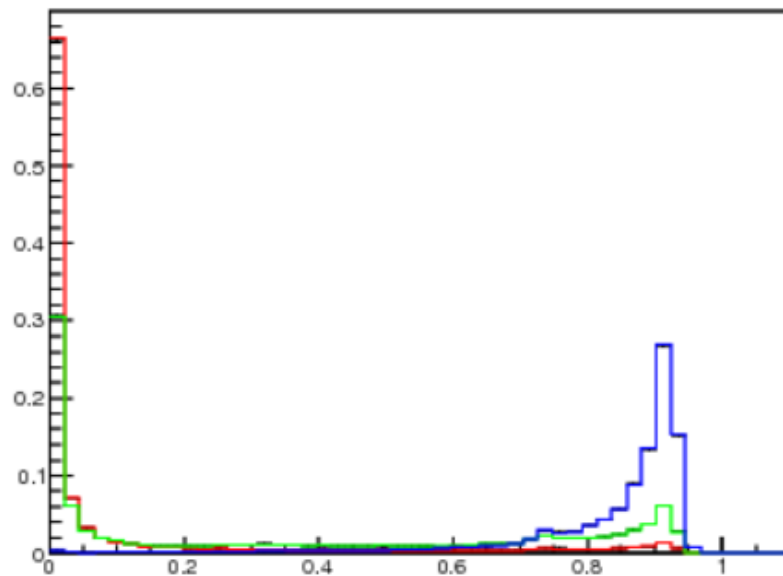
**b likelihood**



**c likelihood**



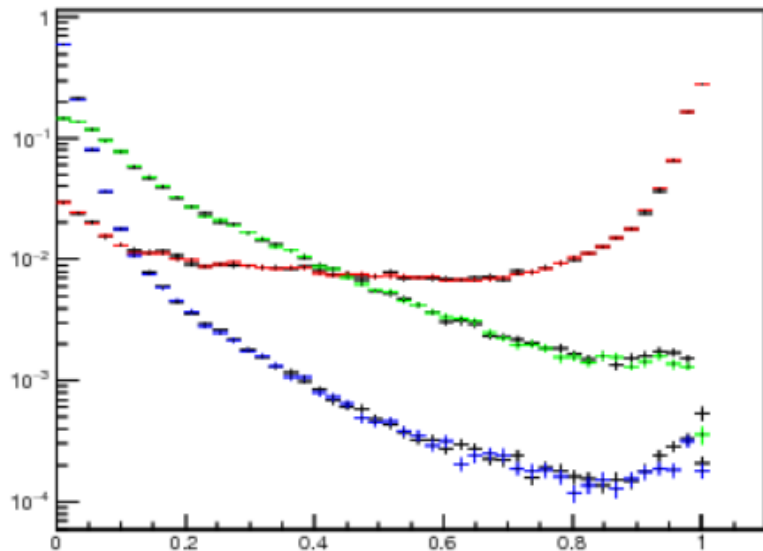
**o likelihood**



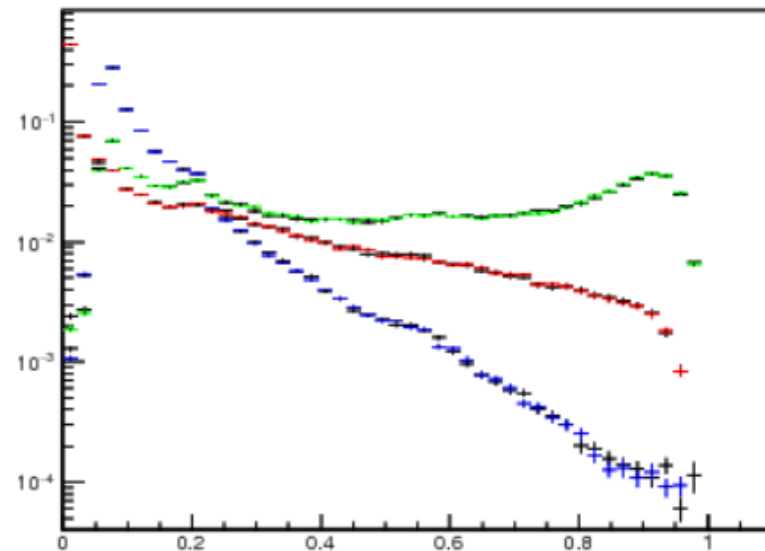
Black dots: training output  
Histograms: testing output  
red: b jets  
green: c jets  
blue: light flavor

Very hard to see difference

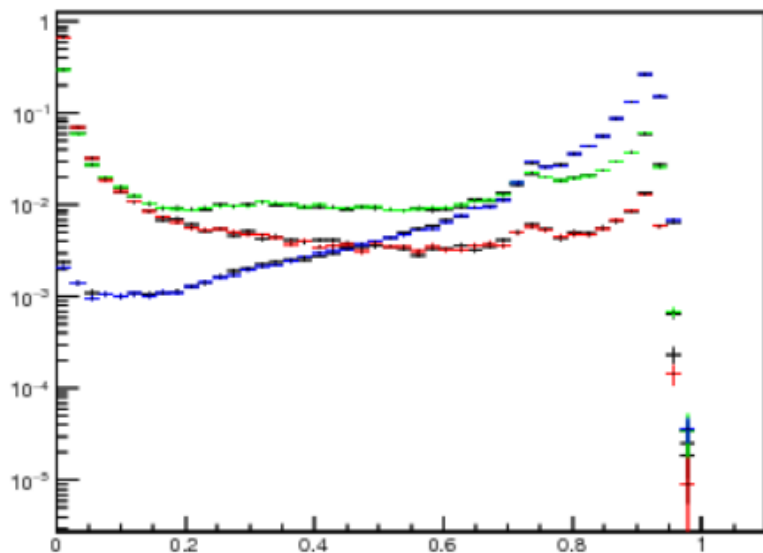
**b likelihood**



**c likelihood**



**o likelihood**



Black dots: training output

Color dots: testing output

red: b jets

green: c jets

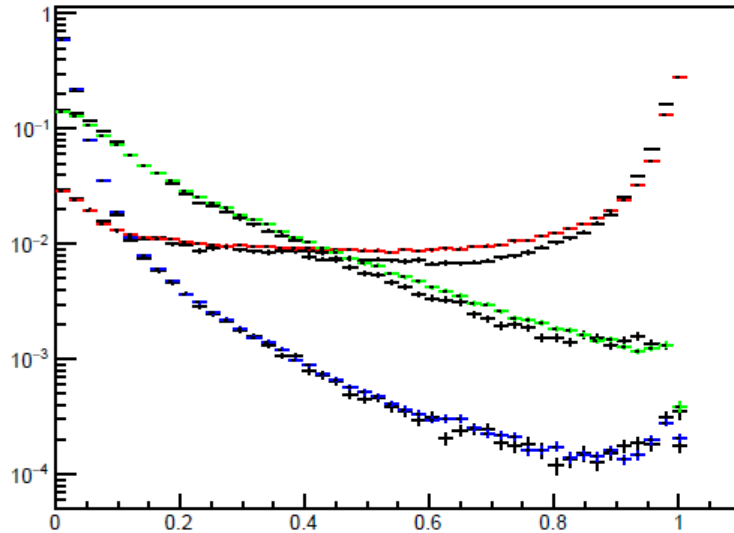
blue: light flavor

Log scale

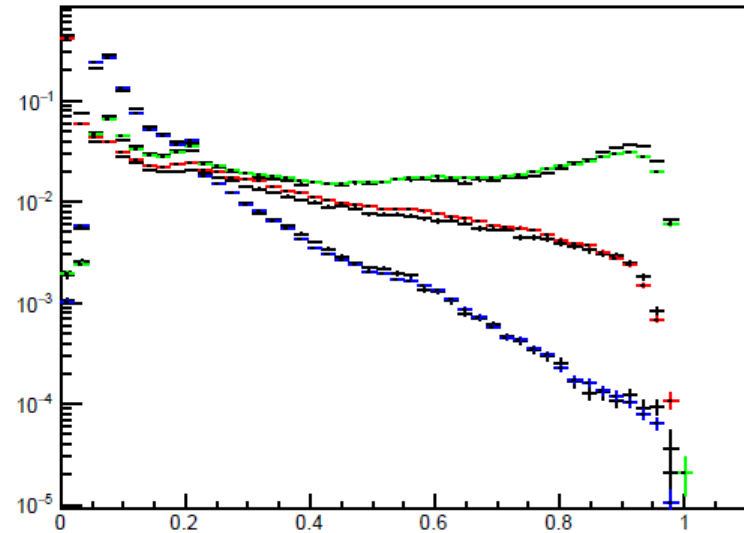
50k case: same result

# Comparison between data amounts

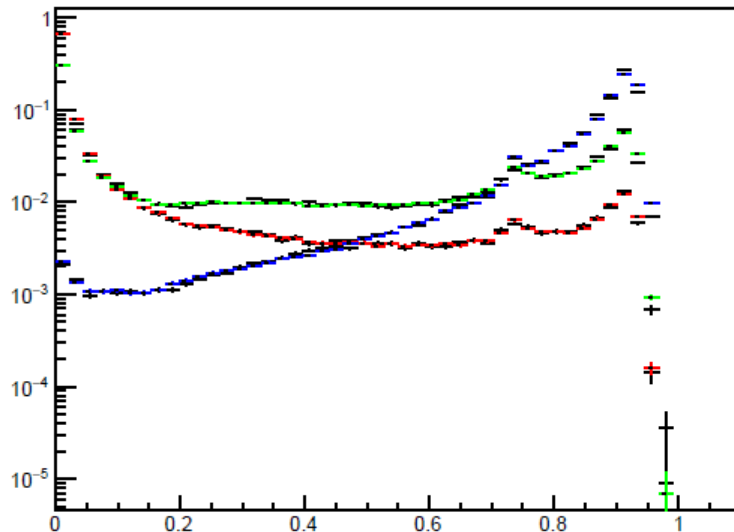
b likelihood



c likelihood



o likelihood



Black dots: 20k training output  
Color dots: 50k training output

red: b jets

green: c jets

blue: light flavor

Log scale

**Different output distribution!**

So,

- Training looks very stable
- But different output dist. between num. of training samples
- We are cleaning up bugs in LCFIPlus
  - Related to primary vertex position smearing
  - But, these bugs do not affect large statistical fluctuation
  - Hyper parameters in BDT?