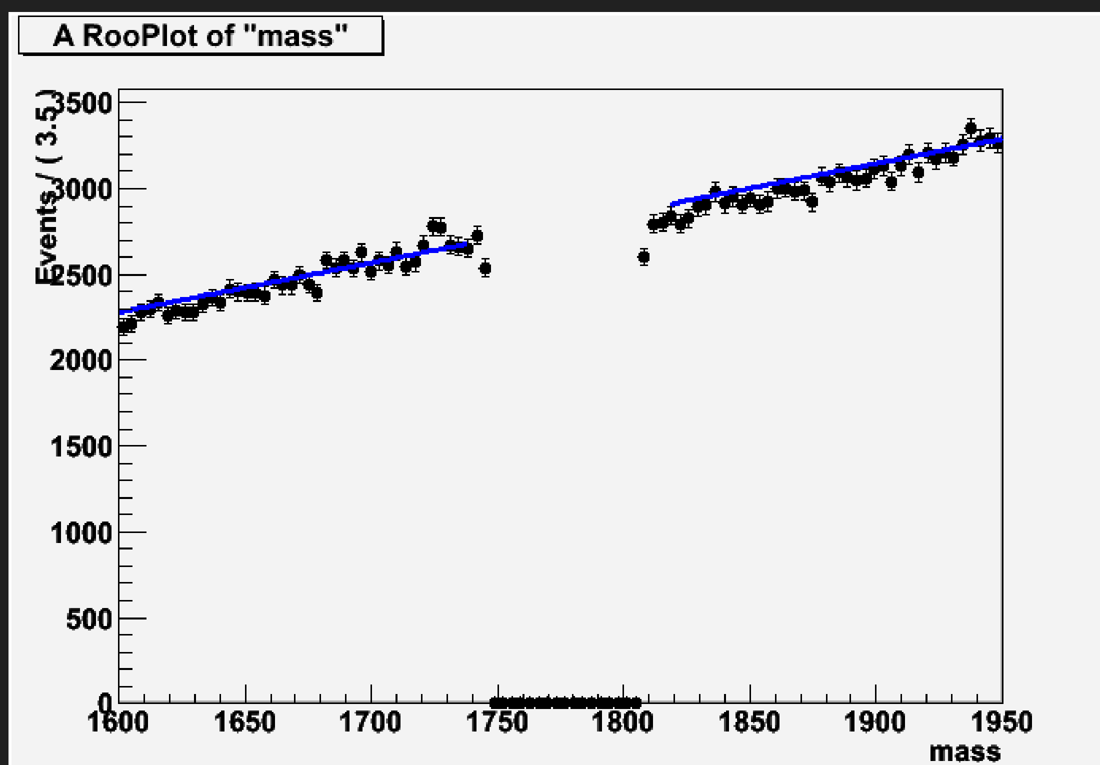
Binning optimization

M.Chrząszcz

Cuts used from now and for always:

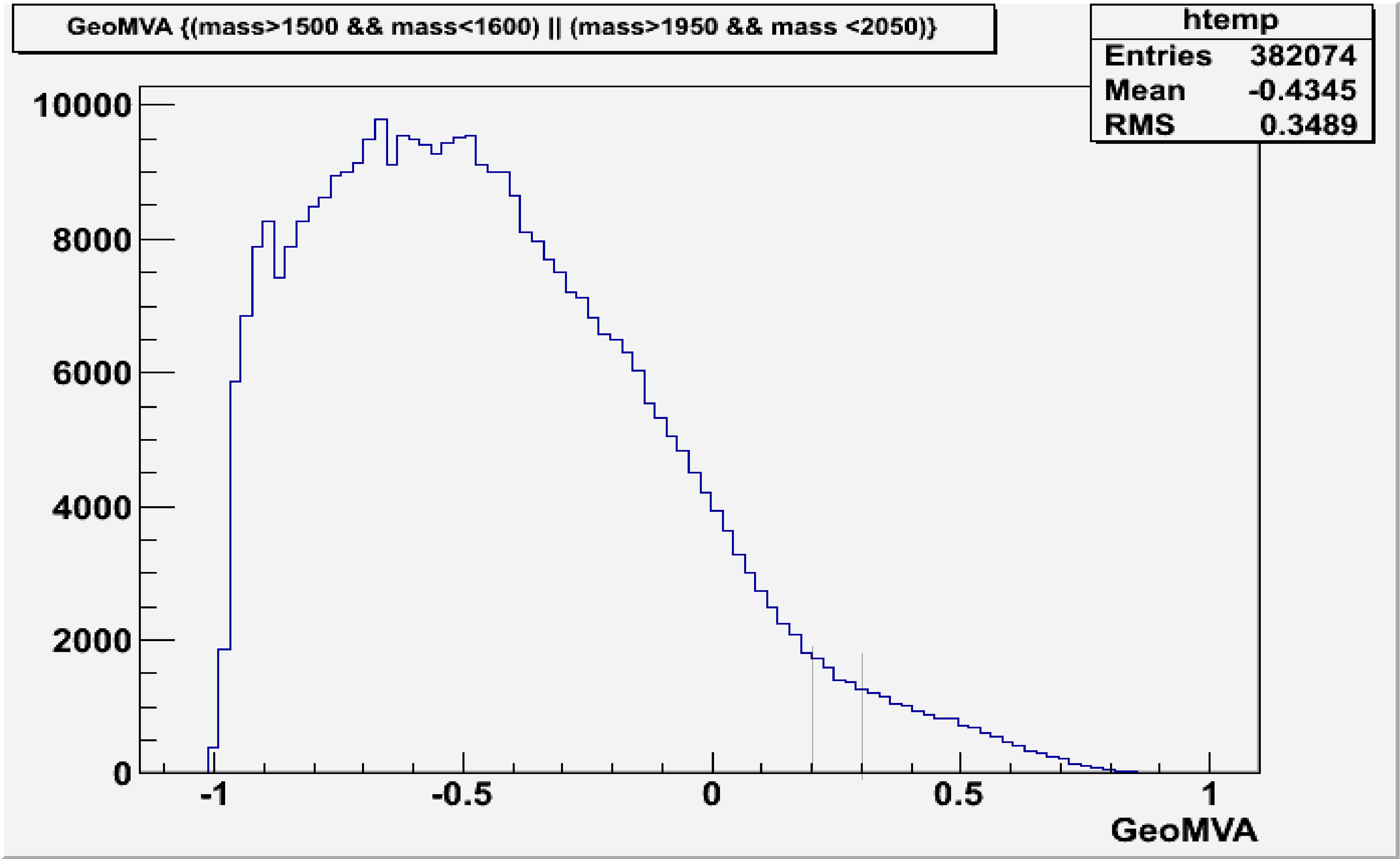
```
 TCut c1 = "L0Dec&&Hlt1Dec&&Hlt2Dec&&cleaningcut";
 TCut c3 = "mass_p0p1>250&&abs(mass_p0p2-1020)>20&&abs(mass_p1p2-1020)>20";
 TCut c31 = "mass_p0p1>550 && mass_p0p2>550 && mass_p1p2>550";
 TCut c32 = "abs(mass_p0p1-782)>20 && abs(mass_p0p1-782)>20 && abs(mass_p0p1-
782)>20";
```

Firstly lets calculate how many events are expected in mass window(1763.4, 1793.4)

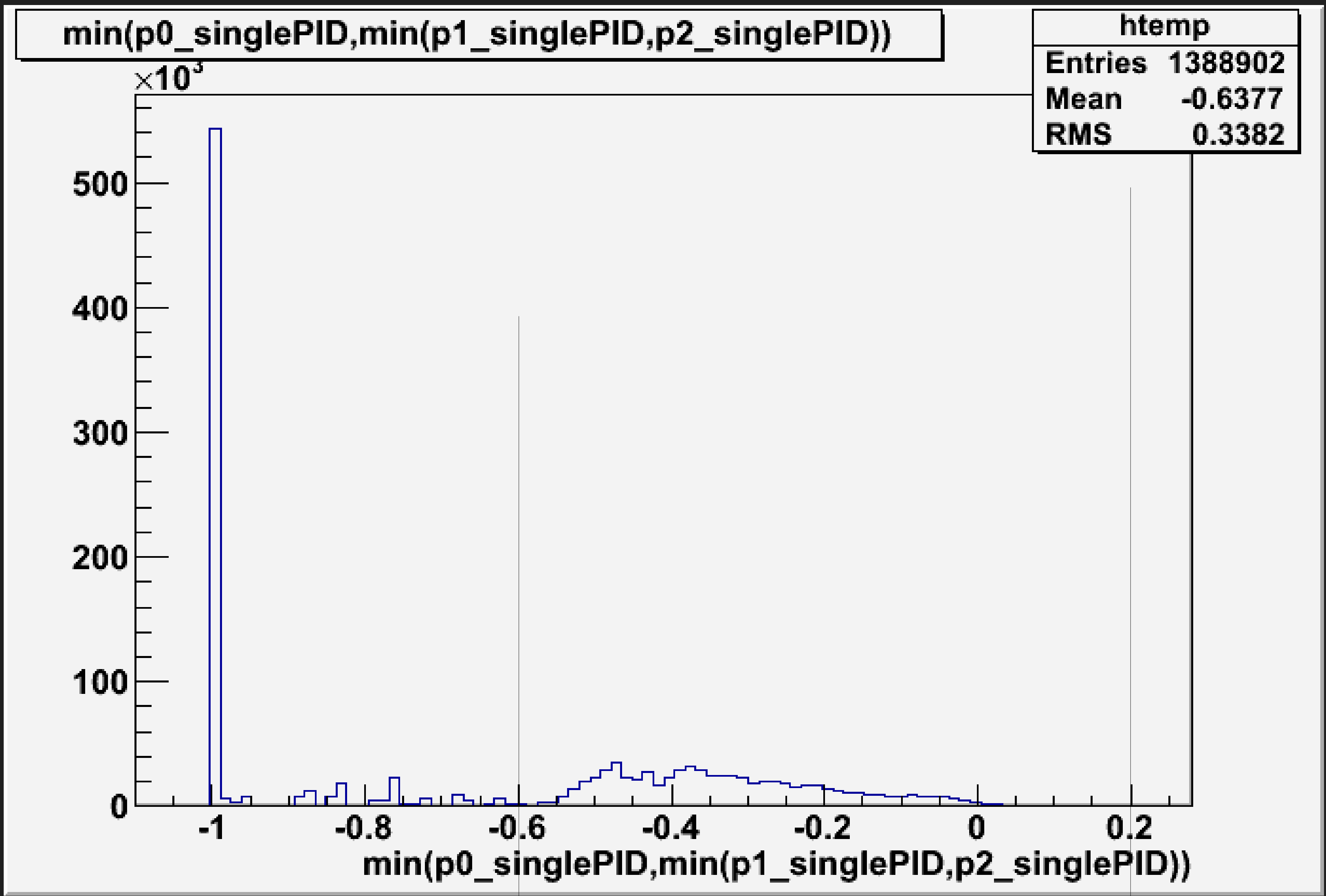


If we know how many events to expect. Than we are interested in the distributions of BDTs:

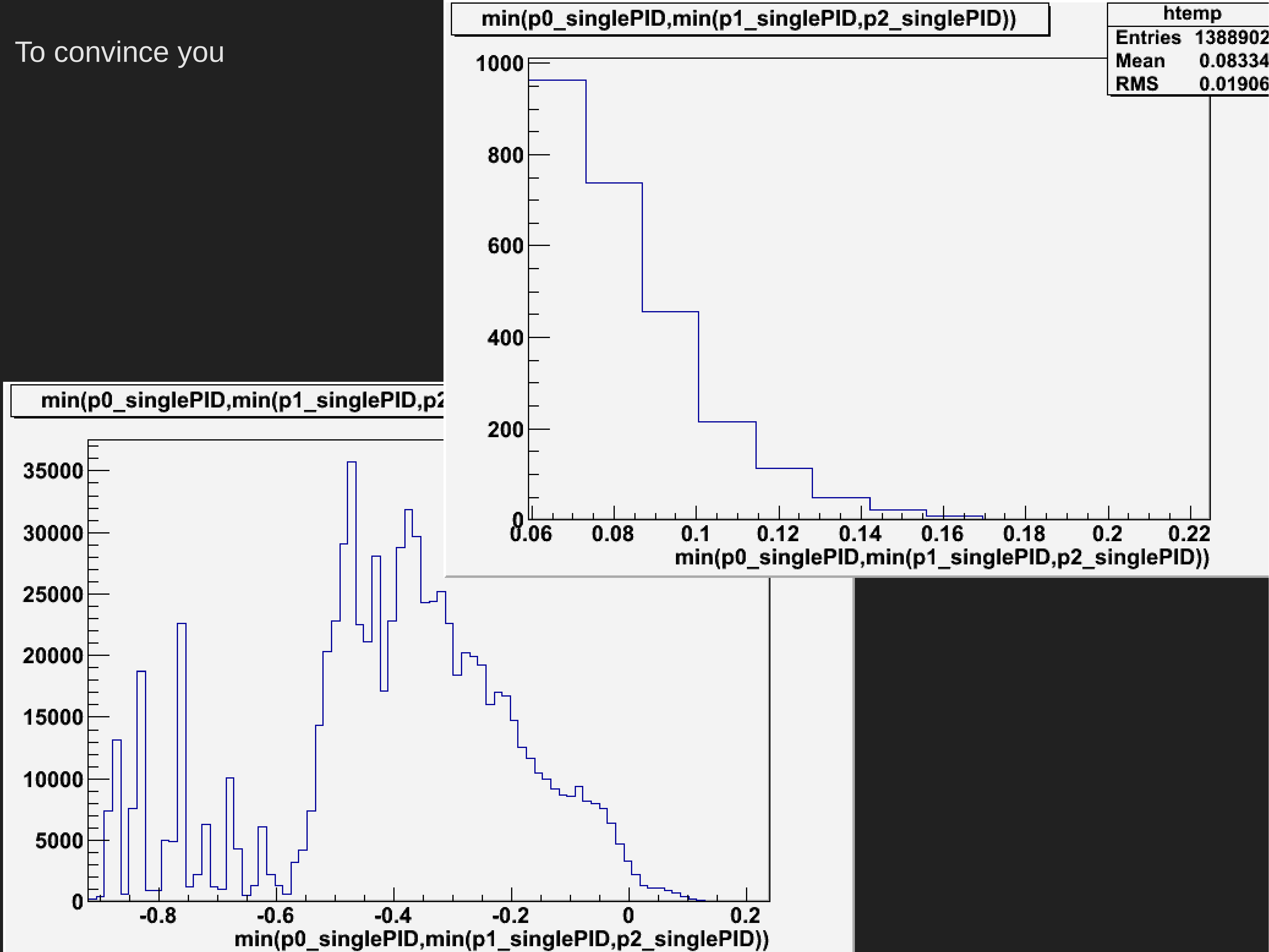
BACKGROUND!!!!!!!!!!!!!!!!!!!!!   d(GeoMVA =0.1) + remember the cuts



We know the bin contente of both BDTs.
Note that for PID I used range from (-0.6, 0.2) They are smaller so I can use d(pid)=0.05



To convince you



For the signal I used MC. Assuming BR is 10^-8. This doesn't have a big influence on results. Using the same method I calculate number off backgrounds in each small bins.

Now lets explain the optimisation on example. I have GEO 20 bins. I use than 21 array of 0 1. For example:

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 - > is equal to binning {0.6,0.7,0.8,0.9,1}
So 1 corresponds to boundary of the bin. Than I calculate the separation:

$$\widetilde{Q_{SB}^{med}} = \prod \mathcal{P}(s_i + b_i, s_i + b_i)/\mathcal{P}(s_i + b_i, b_i) \quad \text{i.e. data=sig+bkg}$$

$$\widetilde{Q_{B}^{med}} = \prod \mathcal{P}(b_i, s_i + b_i)/\mathcal{P}(b_i, b_i.) \quad \text{i.e. data=bkg}$$

$$\Delta LQ = 2ln\widetilde{Q_{SB}^{med}} - 2ln\widetilde{Q_{B}^{med}}.$$

$$f(k,\lambda) = \frac{\lambda^k e^{-\lambda}}{k!},$$

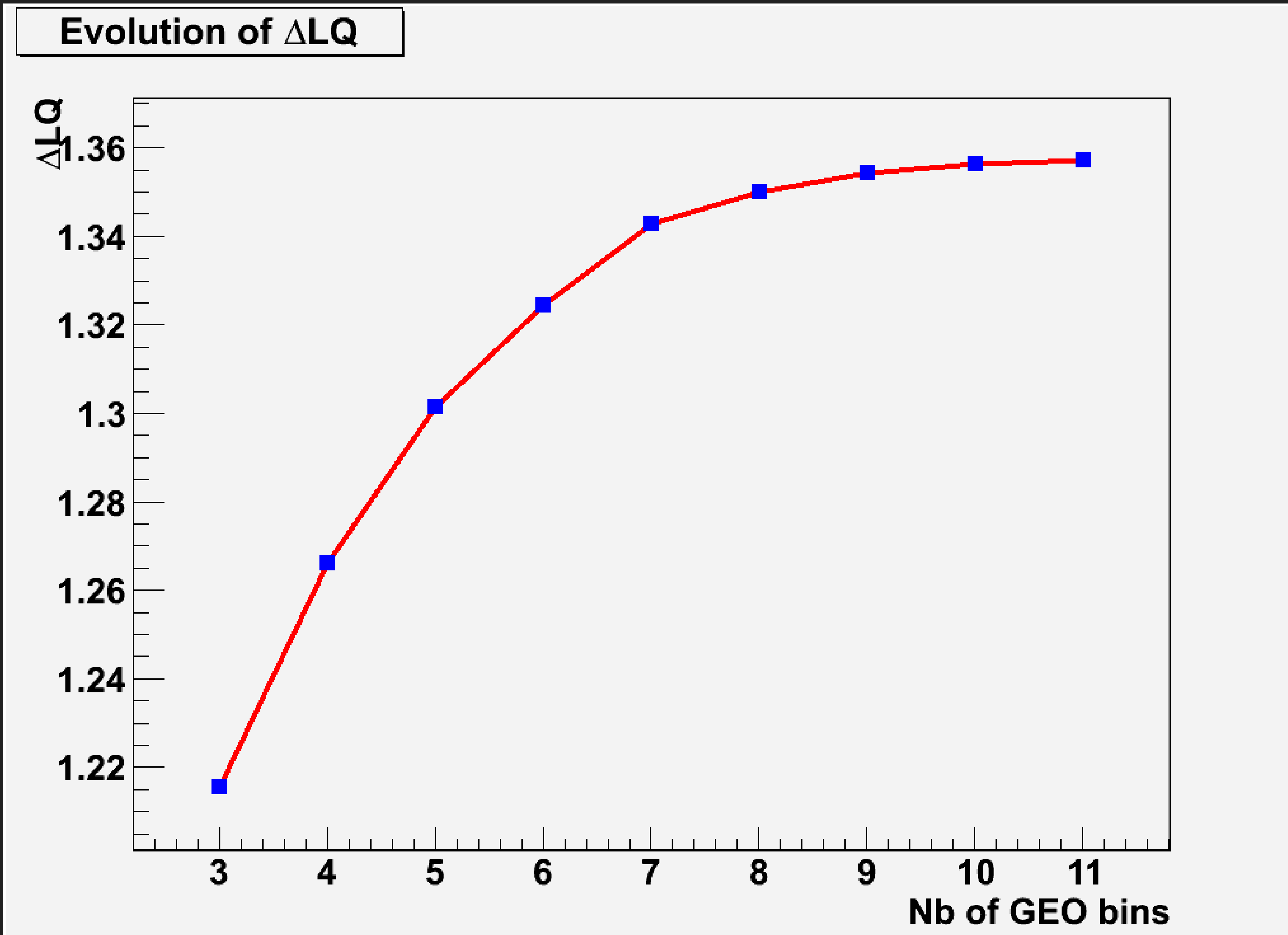
Now I loop over all permutations in:
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1
And get all possible boundaries, example:

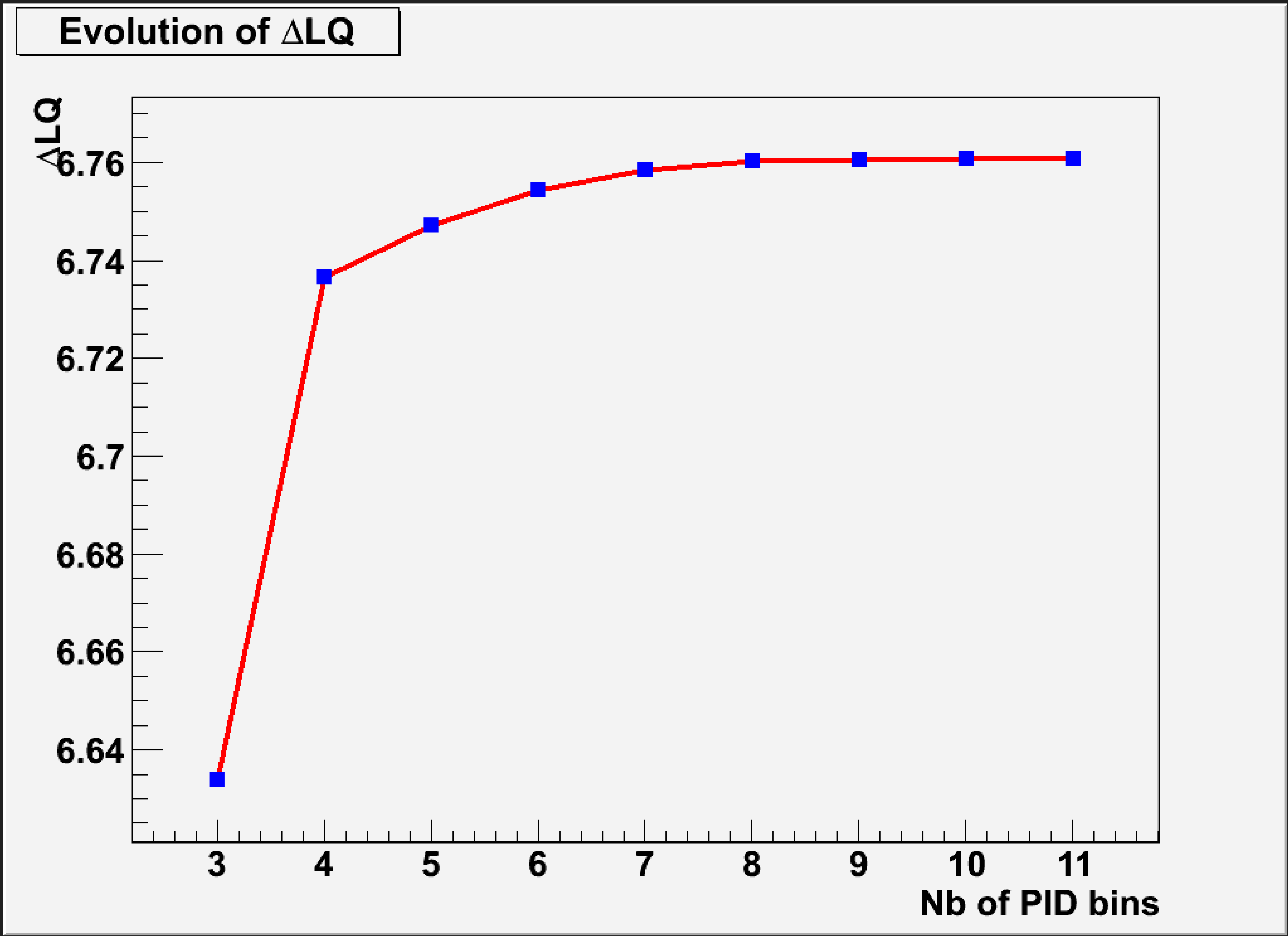
0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 1 0 0 1 1 - > {0.1,0.3,0.6,0.9,1}

Each time LQ is calculated and the best Q with the best binning is stored.
After all permutations we have that in our example we have optimal binning using 4 bins. Than we move to 5 bins.

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 - > OLD
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 - > NEW

And start the procedure again.


Evolution of ΔLQ
ΔLQ
1.36
1.34
1.32
1.3
1.28
1.26
1.24
1.22
3
4
5
6
7
8
9
10
11
Nb of GEO bins


Evolution of ΔLQ
ΔLQ
6.76
6.74
6.72
6.7
6.68
6.66
6.64
3
4
5
6
7
8
9
10
11
Nb of PID bins

Just for sake of clarification and to be 100% sure I am clear about the minus -1 in number of beens. I would say that we are using 6 optimum + 1 trash bins for GEO
And 4 optimum + 1 trash for PID.

Example:
PID 4 optimum from program:{ -0.15, -0.05, 0, 0.1,0.2}
Additional bin{-1, -0.5} and resized last bin to 1.