Have we really observed the final state X ? - I

- We need a criterium to say ok, we have seen the signal or our data are compatible with the background.
- Which statistical uncertainty have we on N_X ?
  - Assume a Poisson statistics to describe N_{cand} negligible uncertainty on ε. We call (using more "popular" symbols):
  - $N = N_{cand}$
  - $B = N_b$
  - $S = N - B = N_x$

$$\left(\frac{\sigma(N_X)}{N_X}\right)^2 = \frac{\sigma^2(N)+\sigma^2(B)}{S^2} = \frac{N+\sigma^2(B)}{S^2}$$

$$\frac{N_X}{\sigma(N_X)} = \frac{S}{\sigma(S)} = \frac{S}{\sqrt{N+\sigma^2(B)}} = \frac{S}{\sqrt{S+B}}$$

Additional assumption: $\sigma^2(B) << N$

$\sigma(S)/S$ is the relative uncertainty on S, its inverse is "how many st.devs. away from 0" → $S/\sqrt{B}$ when low signals on top of large bck

Have we really observed the final state X ? - II

- This quantity is the "significance" of the signal. The higher is $S/\sigma(S) = S/\sqrt{S+B}$, the larger is the number of std.dev. away from 0 of my measurement of S (SCORE FUNCTION)
  - $S/\sqrt{S+B} < 3$ probably I have not osserved any signal (my candidates can be simply a fluctuation of the background)
  - $3 < S/\sqrt{S+B} < 5$ probably I have observed a signal (probability of a background fluctuation very small), however no definite conclusion, more data needed. ➔ ***evidence***
  - $S/\sqrt{S+B} > 5$ observation is accepted. ➔ ***observation***
- NB1: All this is "conventional" it can be discussed
- NB2: $S/\sqrt{S+B}$ is an approximate figure, it relies on some assumptions *(see previous slide)*.

How to optimize a selection ? - I

- The perfect selection is the one with
  - $\varepsilon = 1$
  - $N_b = 0$
- Intermediate situations ? Assume a given ε and a given N_b.

$$N_X = \frac{N_{cand} - N_b}{\varepsilon}$$

- By moving the cut we change each single ingredient. We want to see for which choice of the cut we get the lower statistical error on $\boldsymbol{N_X}$.
  - Again: if we assume a Poisson statistics to describe N_{cand} , negligible uncertainty on ε and on N_b we have to minimize the uncertainty on $S=N_{cand}-N_b$
  - S/sqrt(S+B) ≈ S/sqrt(B) is the good choice: the higher it is the higher is our sensitivity to the final state X. It is the "score function".

Significance as a function of a cut x_{cut} on the x distribution, with x retained if $x<x_{cut}$

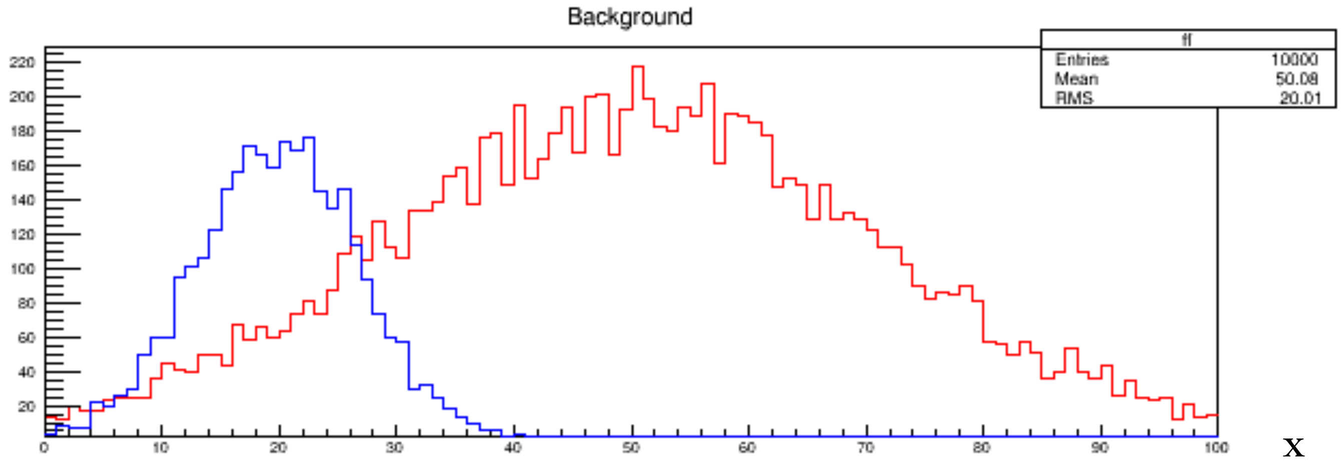
Example - I

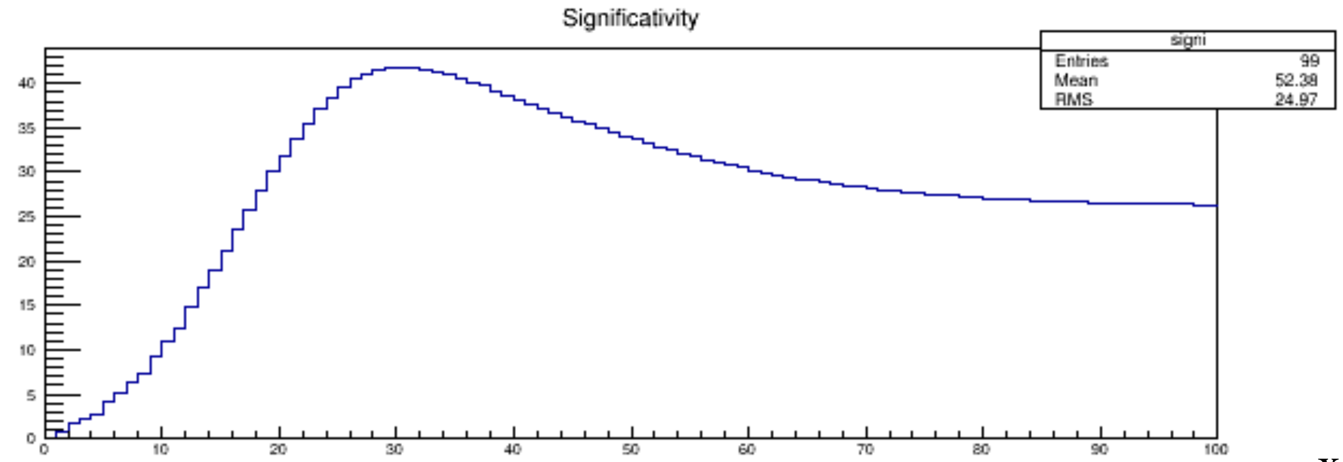
B=10000

σ_x (B) = 15

S=3000

σ_x(S) = 5





Significance as a function of a cut x_{cut} on the x distribution, with x retained if $x<x_{cut}$

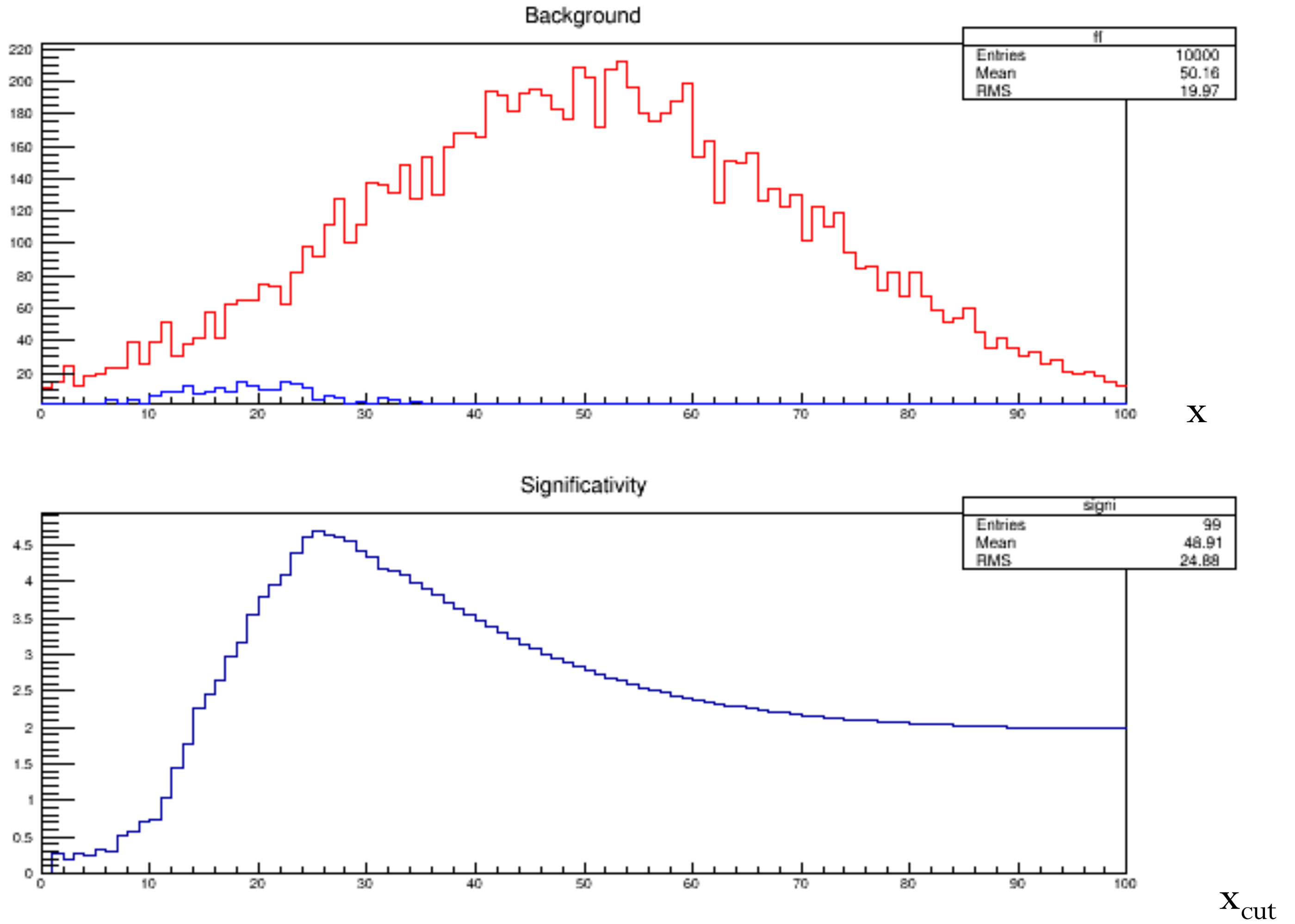
Example - II

$B=10000$

$\sigma_x(B) = 15$

$S=200$

$\sigma_x(S) = 5$



Significance as a function of a cut x_{cut} on the x distribution, with x retained if $x<x_{cut}$

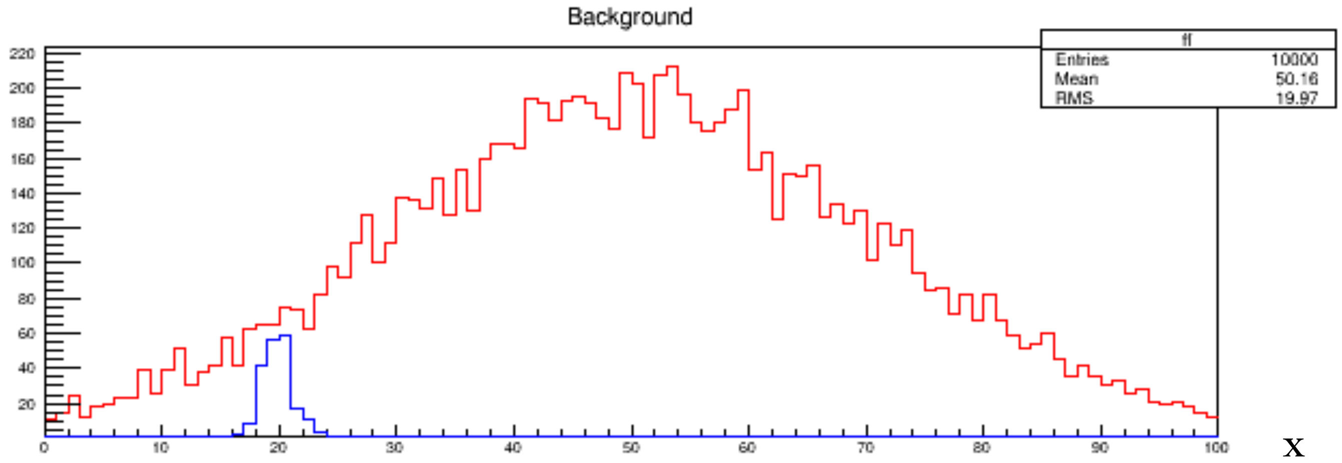
Example - III

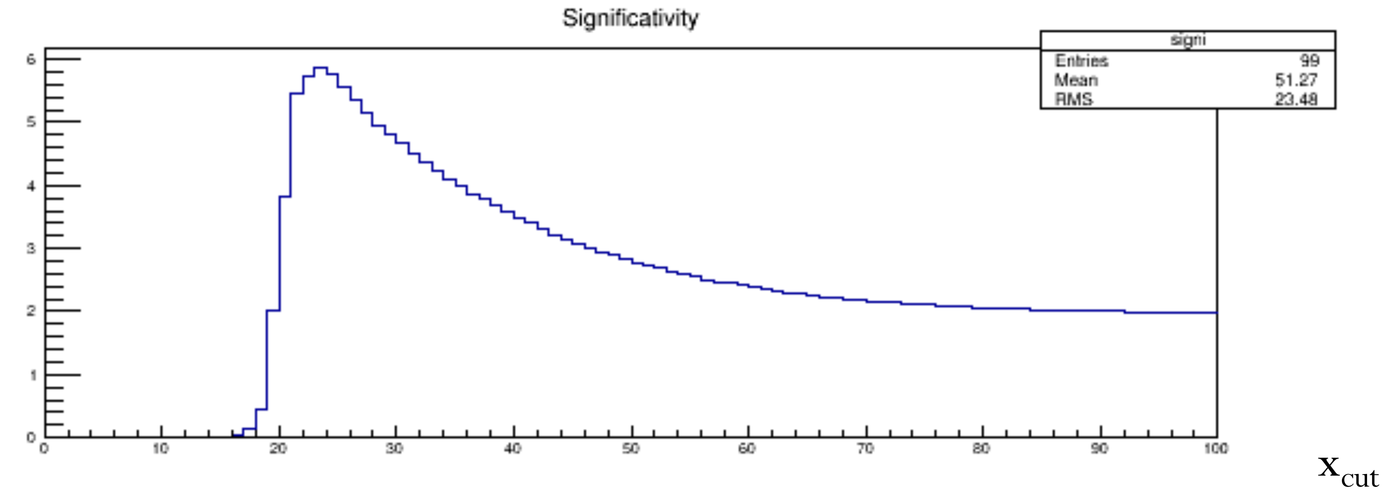
B=10000

σ_x (B) = 15

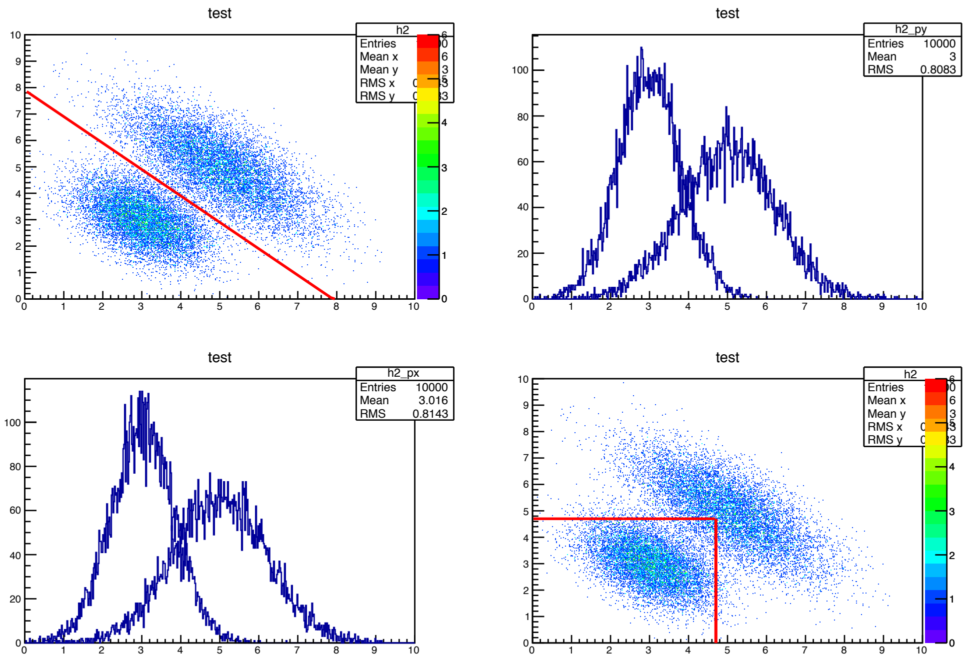
S=200

σ_x(S) = 1





- Cut based analysis
- Multivariate selection e.g. $\alpha x_1 + \beta x_2 < \gamma$



- Discriminant analysis e.g. $t = \sum_{i=1}^{N} \alpha_i x_i < t_{cut}$
  (not only linear combinations -> non linear correlations among variables)
- Multivariate analysis
  e.g. neural network, Boosted decision tree etc..

Multivariate analysis:

N discriminant variables

Training phase on MC signal and MC background samples

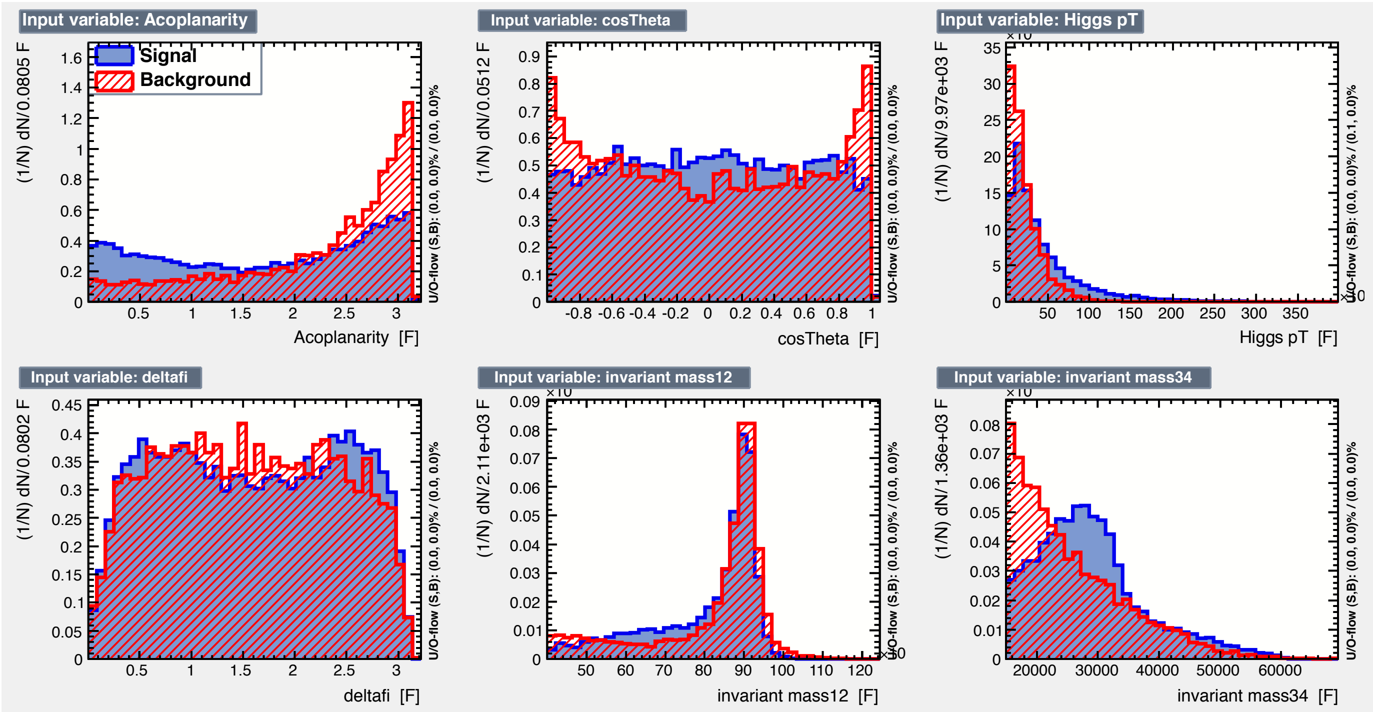
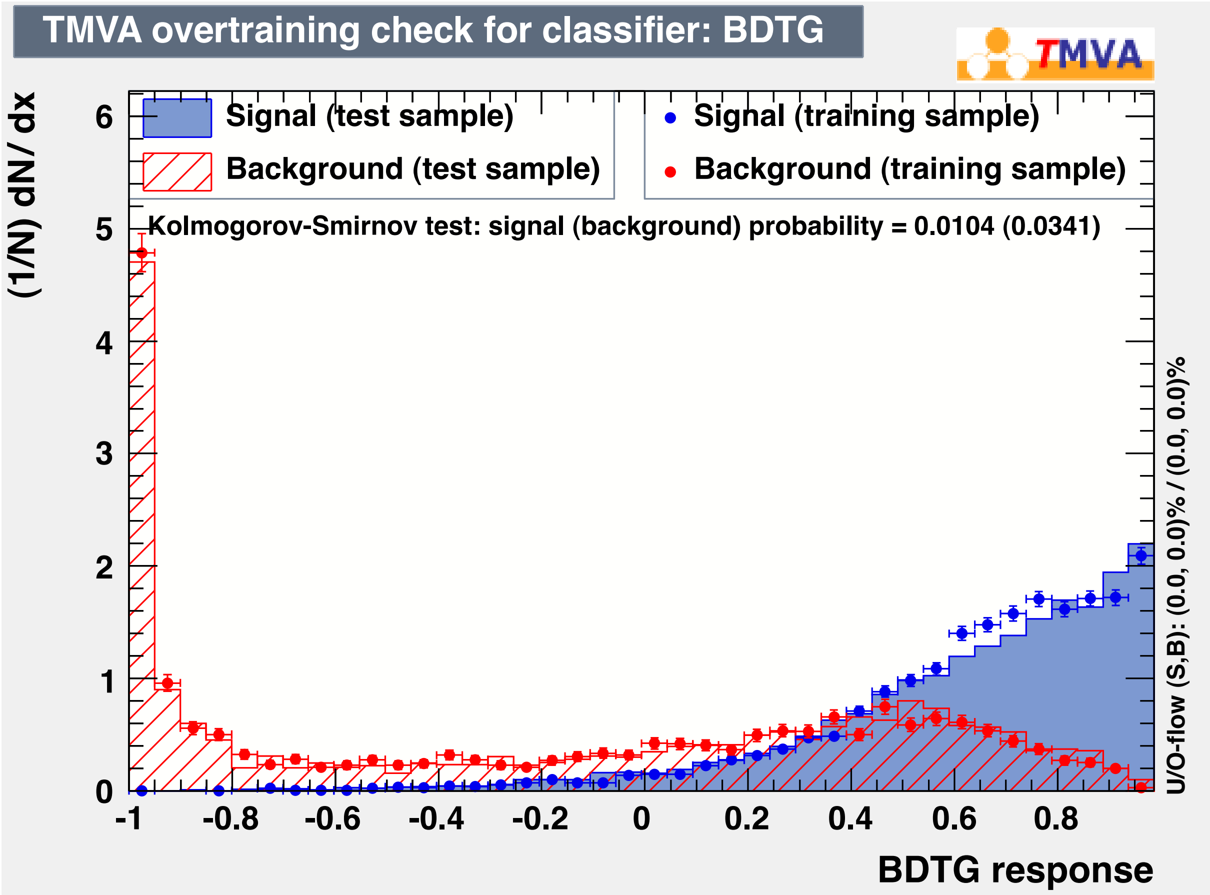


FIGURE 5. Comparison between MC signal (blue) and MC background (red) distributions for the 6 chosen discriminating variables entering in the multivariate analysis (taken from A.Calandri thesis, Sapienza University, A.A. 2011-2012).

Training phase, evaluation of t discriminant variable (e.g. evaluations of coefficients α in linear case)
Test phase, on independent MC samples (t does not depend on specific features of the training sample (overtraining) e.g. a statistical fluactuation)



t

FIGURE 6. Comparison between MC signal (blue) and MC background(red) BDT variable. The points are for the "training" samples, while the histograms correspond to the "test" samples. In the insert the results of compatibility tests between training and test results are given

CUT-BASED SELECTION

The most natural way to proceed is to apply cuts. We find among the physical quantities of each event those that are more "discriminant" and we apply cuts on these variables or on combinations of these variables. The selection procedure is a sequence of cuts, and is typically well described by tables or plots that are called "Cut-Flows".

TABLE 1. Example of cut-flow. The selection of $\eta\pi^0\gamma$ final state with $\eta \to \pi^+\pi^-\pi^0$ from e^+e^- collisions at the ϕ peak ($\sqrt{s}$ = 1019 MeV, is based on the list of cuts given in the first column. The number of surviving events after each cut is shown in the different columns for the MC signal (column 2) and for the main MC backgrounds (other columns). (taken from D. Leone, thesis , Sapienza University A.A. 2000-2001).

Cut	$\eta\pi^0\gamma$	$\omega\pi^0$	$\eta\gamma$	$K_S \to$ neutrals	$K_S \to$ charged
Generated Events	11763	33000	95000	96921	112335
Event Classification	6482	17602	55813	18815	14711
2 tracks + 5 photons	3112	724	110	371	3100
$E_{tot} - \Vert\vec{P}_{tot}\Vert$	2976	539	39	118	1171
Kinematic fit I	2714	236	5	24	66
Combinations	2649	129	1	19	0
Kinematic fit II	2247	2	0	1	0
$E_{rad} > 20$ MeV	2240	1	0	0	0

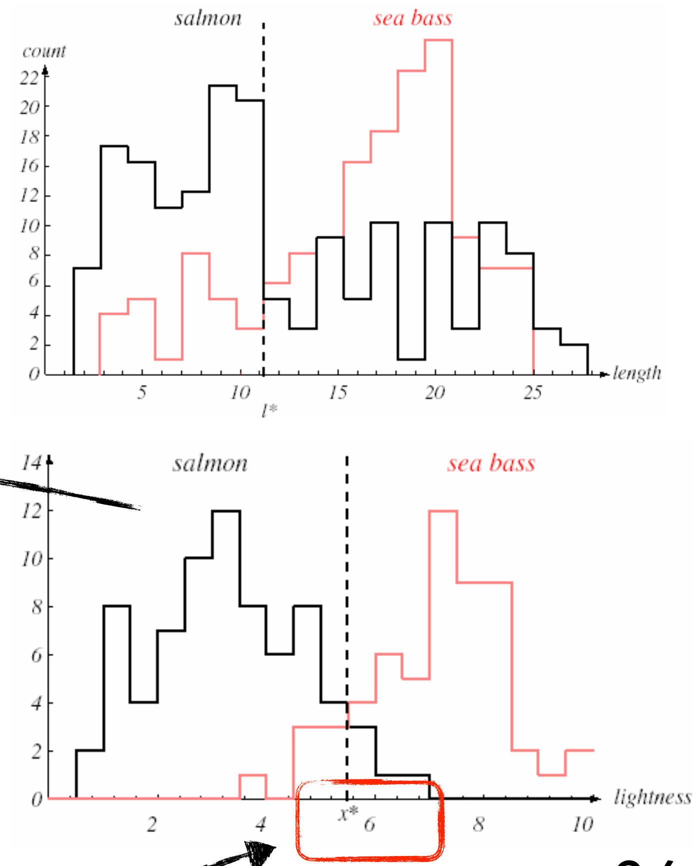
$\varepsilon = 2240/11763 = (19.04 \pm 0.36)$ % (binomial statistics)

R=33000 for $\omega\pi^0$ background.

DECISION BOUNDARIES

- let's assume that we have found that the two best features for our classification task are: length e lightness
- which one we should use for the classification? Which threshold?
- to decide this we make use of the traing set examples

Classification rule:	if $x > x^*$:	object $\in$ class A
	else:	objetc $\in$ class B



decision boundary

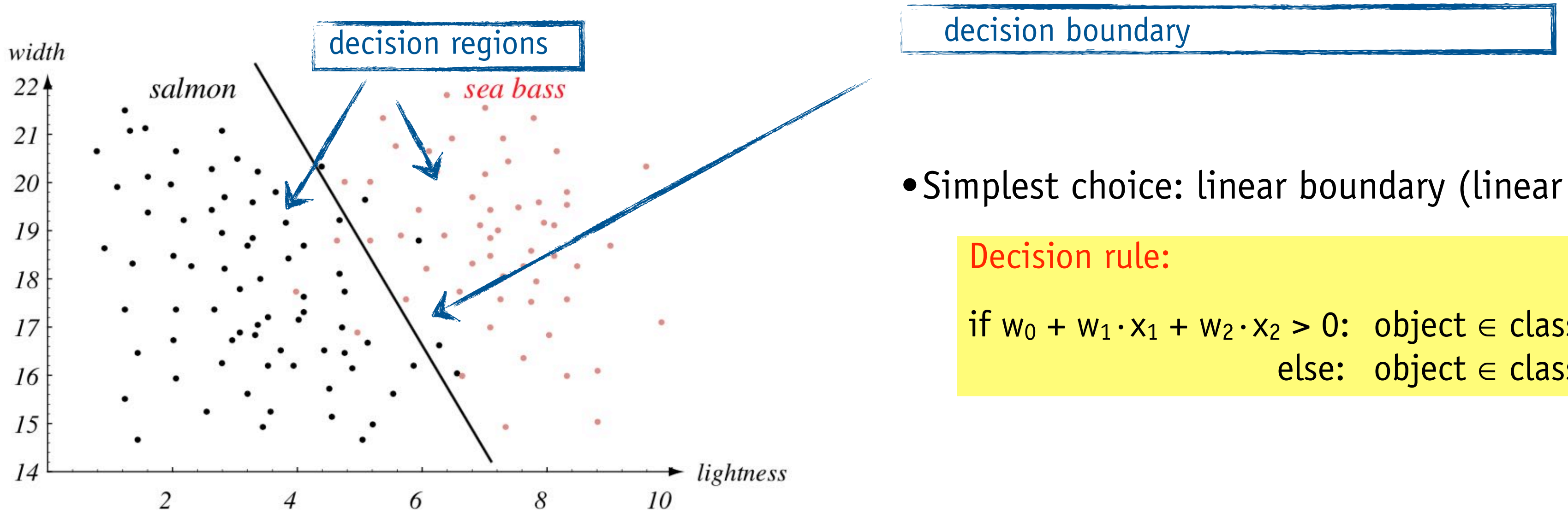
24

the threshold x^* is chosen in order to optimize an appropriate performance measure

example: accuracy, probability of misclassificantion, statistical risk ...

DECISION BOUNDARIES

- to improve P a better strategy woudl be to use more than one feature at the same time
- The classification problem becomes the problem to find the best partition of the feature space, so that the classification error is the smallest one



- Simplest choice: linear boundary (linear classifier)

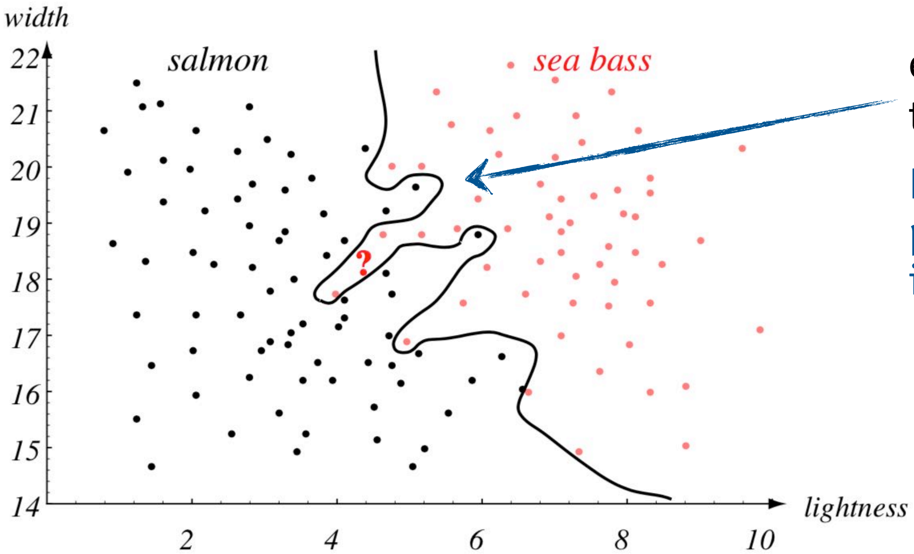
Decision rule:

if $w_0 + w_1 \cdot x_1 + w_2 \cdot x_2 > 0$: object $\in$ class A
else: object $\in$ class B

25

COMPLEX DECISION BOUNDARIES …

- question: it is possible to get rid of all errors with a complex decision boundary?



example: this boundary correctly clasify all the events of the trining set

PROBLEM: this way we are NOT guarantee a good performance of the algorithm when applied to events from independent samples wrt the training set (overfitting)

- The training set has finite dimension and the decision boundary is sensitive to the statistical fluctuation in the training set
- This aspect is called generalisation problem, and sis the crucial aspect in the design and training of any ML algorithm!

26

Optimization of the cut on t => significance as score function

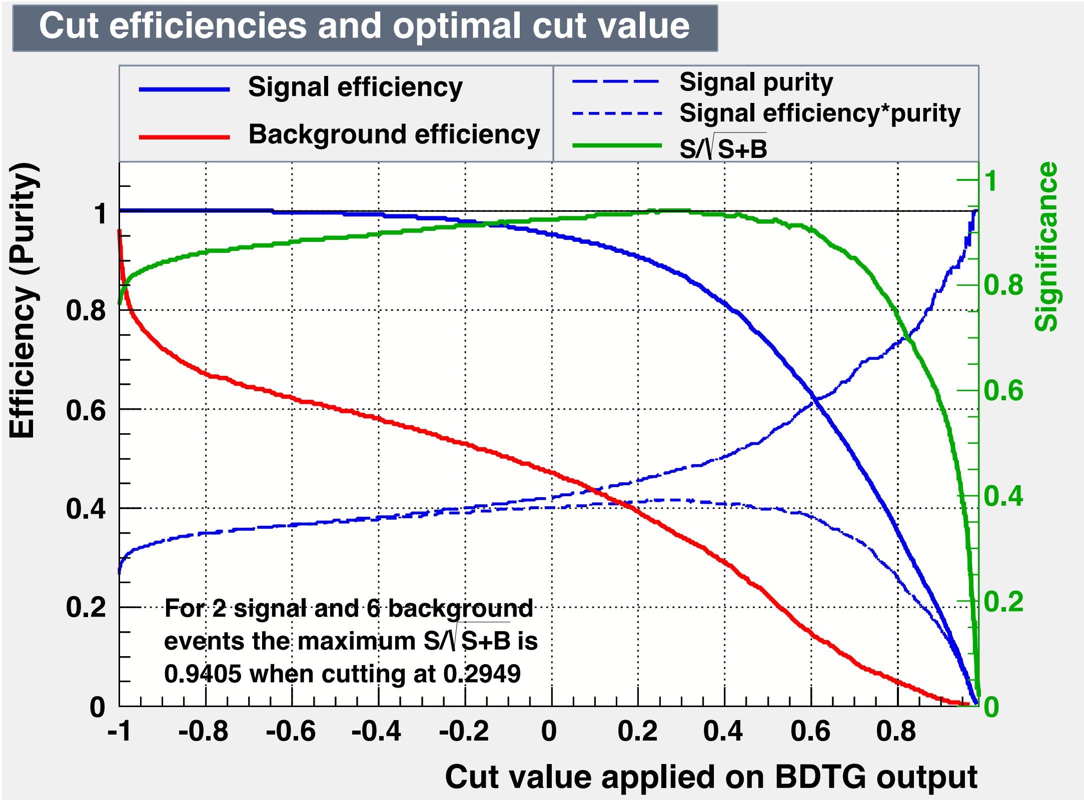


FIGURE 7. Several quantities are shown as a function of the possible value of t_{cut}, the cut on the BDT variable. Blue and red curves show respectively the signal and background efficiency while the green curve is the score function that, in this case, has a maximum around $t_{cut} = 0.25$ although with a very low significance (below 1). (taken from A.Calandri thesis, Sapienza University, A.A. 2011-2012)

Optimization of the cut on t => significance as score function

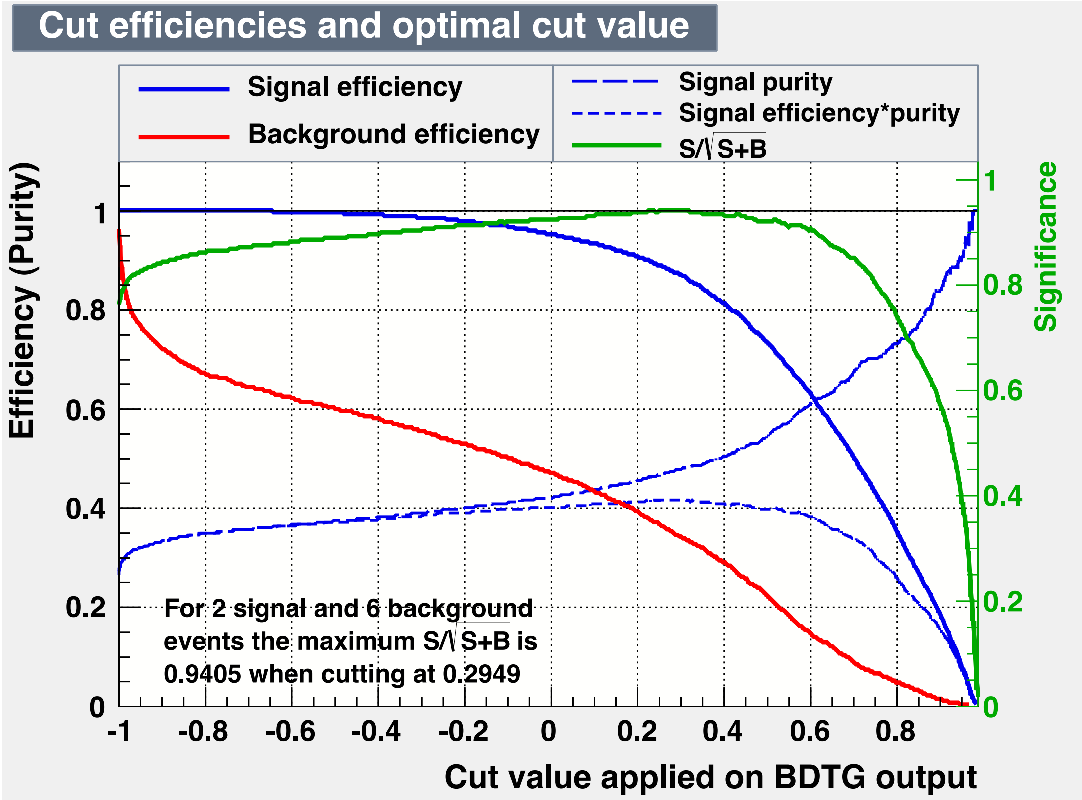
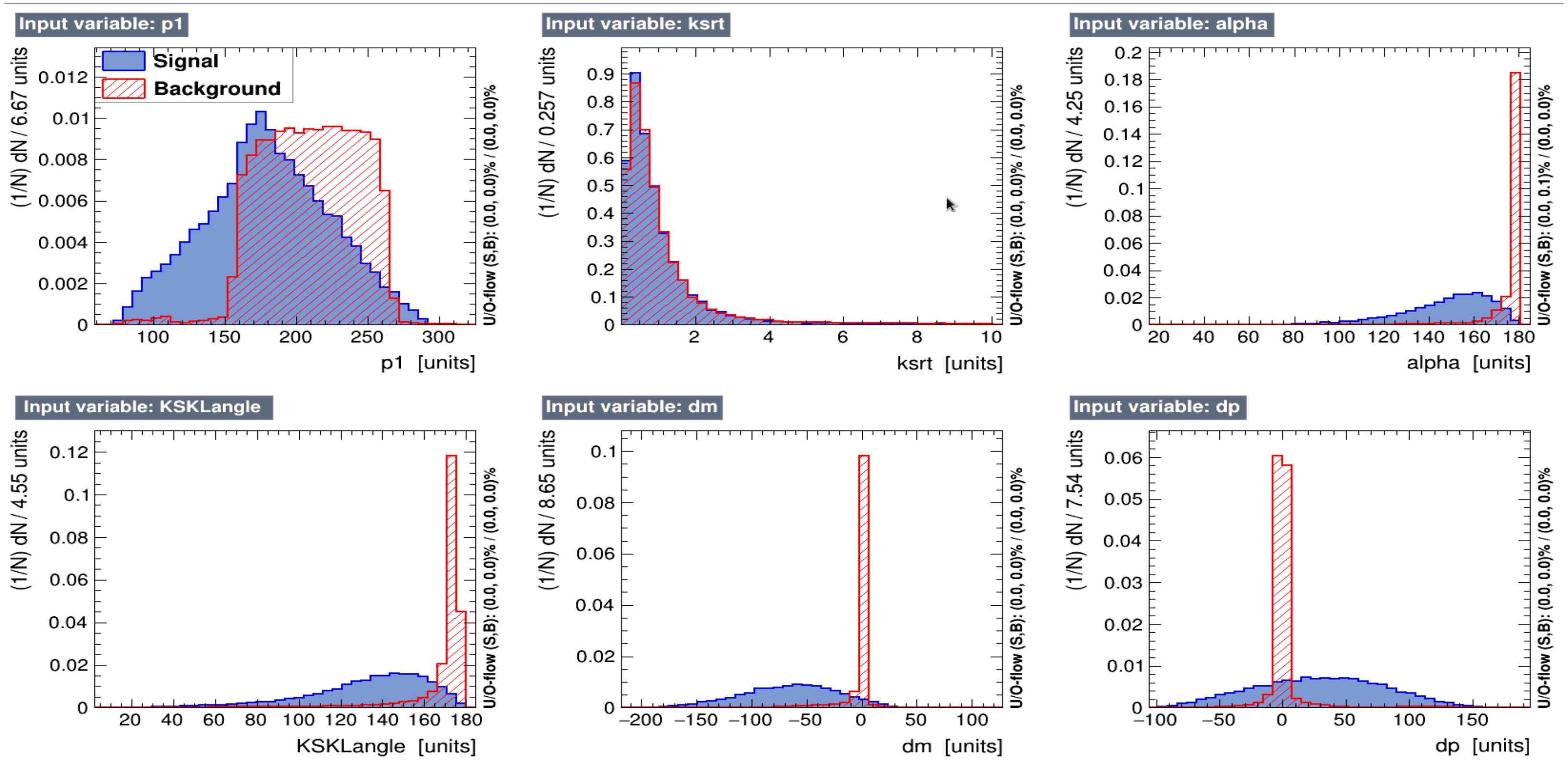


FIGURE 7. Several quantities are shown as a function of the possible value of t_{cut}, the cut on the BDT variable. Blue and red curves show respectively the signal and background efficiency while the green curve is the score function that, in this case, has a maximum around $t_{cut} = 0.25$ although with a very low significance (below 1). (taken from A.Calandri thesis, Sapienza University, A.A. 2011-2012)

Signal observed?

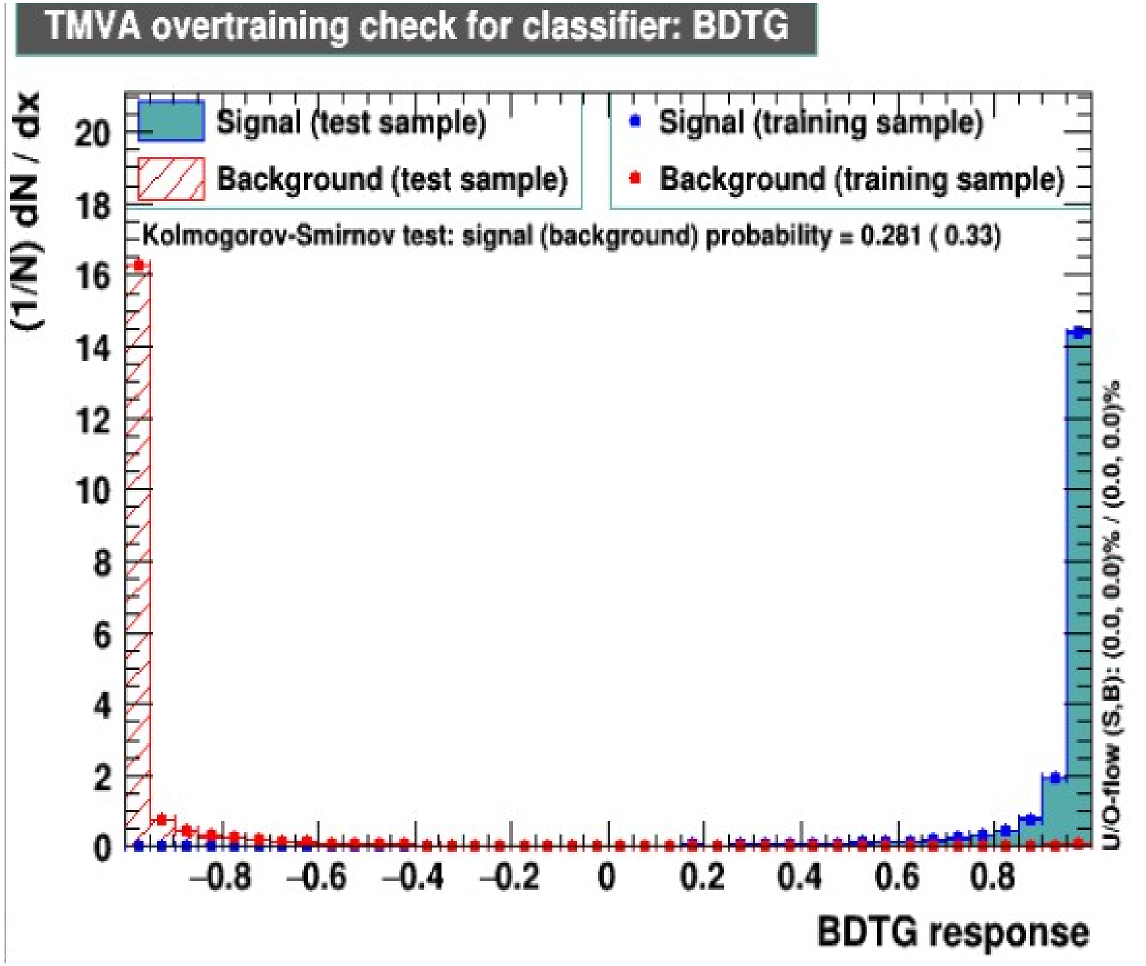
Optimization of the cut on t => significance as score function

Another example



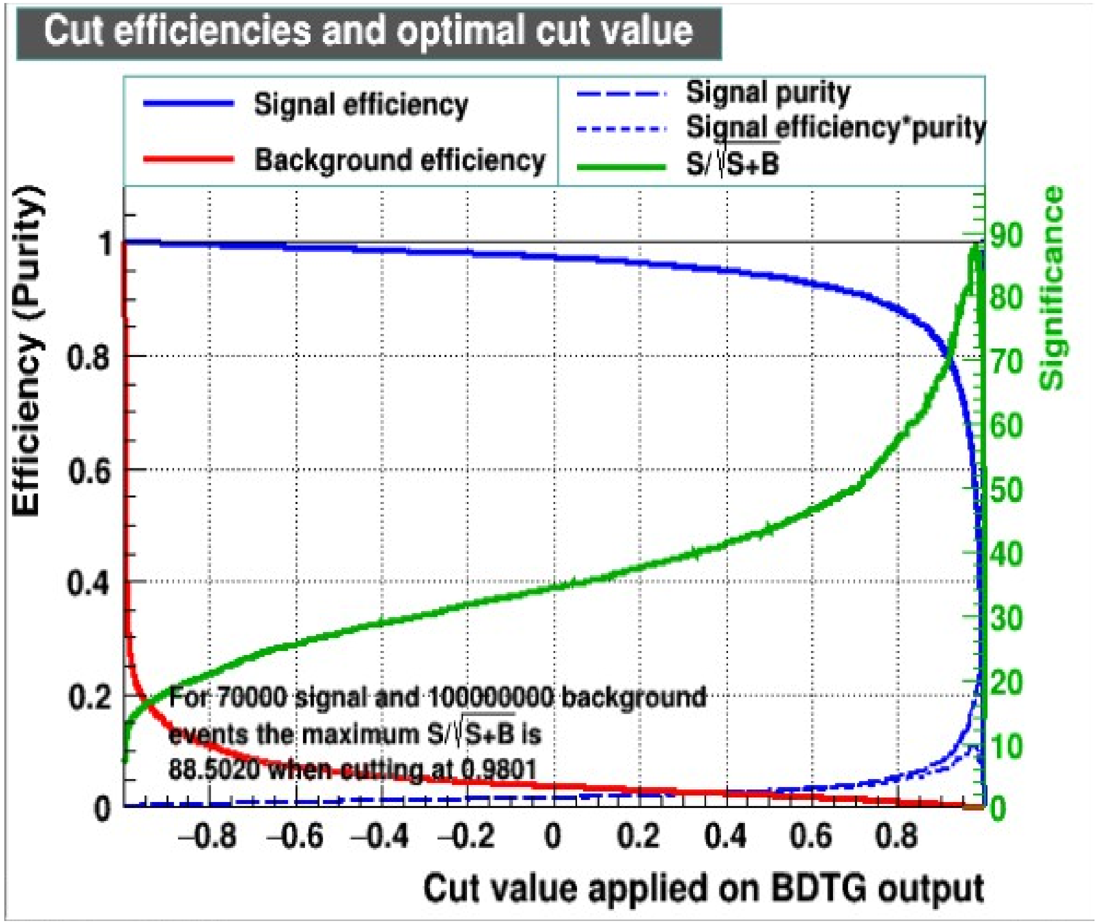
Optimization of the cut on t => significance as score function

Another example



Optimization of the cut on t => significance as score function

Another example



Comments on multivariate methods:

The emphasis is often on controlling systematic uncertainties between the modeled training data and Nature to avoid false discovery.

Although many classifier outputs are "black boxes", a discovery at 5σ significance with a sophisticated (opaque) method will win the competition if backed up by, say, 4σ evidence from a cut-based method.

(see also topical seminar later in the course)

Another score function based on the likelihood ratio test (see later in the course)

$$\sqrt{2(S+B)\ln\left(1+\frac{S}{B}\right)-2S}$$

Efficiency: $\epsilon = \frac{S_f}{S_0}$ Probability that a signal event is identified as signal = ε

Rejection: $R = \frac{B_0}{B_f}$ Probability that a background event is identified as signal = 1/R

Type –I errors:
Efficiency losses, i.e. some signal events discarded

Type-II errors:
Background events contaminate the signal sample

$$P(type - Ierrors) = 1 - \epsilon$$

$$P(type - IIerrors) = \frac{1}{R}$$

Once the selection is performed, CANDIDATE events cannot be distinguished as signal or background on event-by-event basis, only statistically
=> **probability that a given event is a signal event**

In order to evaluate this probability we use the **Bayes theorem**[7]. As usual the Bayes theorem needs two ingredients.

- The so called **likelihood** (we will make use of this word several times in the following). In this context we need essentially on one side the probability that a signal event is identified as signal, and on the other side, the probability that a background event is identified as signal. These two quantities are respectively the efficiency ϵ and the inverse of the rejection power $\beta = 1/R$ defined above.
- The so called **prior** probabilities. In our case they are the expected "cross-sections" of signal and background events respectively.

We call $P(t > t_{cut}/S)$ and $P(t > t_{cut}/B)$ the two likelihood functions we need[8], and π_S and π_B the two prior functions. The Bayes theorem gives:

$$P(S/t > t_{cut}) = \frac{P(t > t_{cut}/S)\pi_S}{P(t > t_{cut}/S)\pi_S + P(t > t_{cut}/B)\pi_B} \tag{60}$$

This probability can be regarded as a **purity** of the sample. It is interesting to write it as follows:

$$purity = P(S/t > t_{cut}) = \frac{1}{1 + \frac{P(t>t_{cut}/B)\pi_B}{P(t>t_{cut}/S)\pi_S}} = \frac{1}{1 + \frac{\pi_B}{R\epsilon\pi_S}} \tag{61}$$

showing that a high purity can be reached only if

Maximize the purity for a given efficiency

$$R\epsilon >> \frac{\pi_B}{\pi_S} \tag{62}$$

If we call r the rate of selected events, the fake rate f is:

$$f = r(1 - purity) \tag{63}$$

In order to evaluate this probability we use the **Bayes theorem**[7]. As usual the Bayes theorem needs two ingredients.

- The so called **likelihood** (we will make use of this word several times in the following). In this context we need essentially on one side the probability that a signal event is identified as signal, and on the other side, the probability that a background event is identified as signal. These two quantities are respectively the efficiency ϵ and the inverse of the rejection power $\beta = 1/R$ defined above.
- The so called **prior** probabilities. In our case they are the expected "cross-sections" of signal and background events respectively.

We call $P(t > t_{cut}/S)$ and $P(t > t_{cut}/B)$ the two likelihood functions we need[8], and π_S and π_B the two prior functions. The Bayes theorem gives:

ε

$$P(S/t > t_{cut}) = \frac{P(t > t_{cut}/S)\pi_S}{P(t > t_{cut}/S)\pi_S + P(t > t_{cut}/B)\pi_B} \tag{60}$$

This probability can be regarded as a **purity** of the sample. It is interesting to write it as follows:

$$purity = P(S/t > t_{cut}) = \frac{1}{1 + \frac{P(t>t_{cut}/B)\pi_B}{P(t>t_{cut}/S)\pi_S}} = \frac{1}{1 + \frac{\pi_B}{R\epsilon\pi_S}} \tag{61}$$

showing that a high purity can be reached only if

Maximize the purity for a given efficiency

$$R\epsilon >> \frac{\pi_B}{\pi_S} \tag{62}$$

If we call r the rate of selected events, the fake rate f is:

$$f = r(1 - purity) \tag{63}$$

In order to evaluate this probability we use the **Bayes theorem**[7]. As usual the Bayes theorem needs two ingredients.

- The so called **likelihood** (we will make use of this word several times in the following). In this context we need essentially on one side the probability that a signal event is identified as signal, and on the other side, the probability that a background event is identified as signal. These two quantities are respectively the efficiency ϵ and the inverse of the rejection power $\beta = 1/R$ defined above.
- The so called **prior** probabilities. In our case they are the expected "cross-sections" of signal and background events respectively.

We call $P(t > t_{cut}/S)$ and $P(t > t_{cut}/B)$ the two likelihood functions we need[8], and π_S and π_B the two prior functions. The Bayes theorem gives:

ε

$$P(S/t > t_{cut}) = \frac{P(t > t_{cut}/S)\pi_S}{P(t > t_{cut}/S)\pi_S + P(t > t_{cut}/B)\pi_B} \tag{60}$$

1/R

This probability can be regarded as a **purity** of the sample. It is interesting to write it as follows:

$$purity = P(S/t > t_{cut}) = \frac{1}{1 + \frac{P(t>t_{cut}/B)\pi_B}{P(t>t_{cut}/S)\pi_S}} = \frac{1}{1 + \frac{\pi_B}{R\epsilon\pi_S}} \tag{61}$$

showing that a high purity can be reached only if

Maximize the purity for a given efficiency

$$R\epsilon >> \frac{\pi_B}{\pi_S} \tag{62}$$

If we call r the rate of selected events, the fake rate f is:

$$f = r(1 - purity) \tag{63}$$

CUT-BASED SELECTION

The most natural way to proceed is to apply cuts. We find among the physical quantities of each event those that are more "discriminant" and we apply cuts on these variables or on combinations of these variables. The selection procedure is a sequence of cuts, and is typically well described by tables or plots that are called "Cut-Flows".

TABLE 1. Example of cut-flow. The selection of $\eta\pi^0\gamma$ final state with $\eta \to \pi^+\pi^-\pi^0$ from e^+e^- collisions at the ϕ peak ($\sqrt{s}$ = 1019 MeV, is based on the list of cuts given in the first column. The number of surviving events after each cut is shown in the different columns for the MC signal (column 2) and for the main MC backgrounds (other columns). (taken from D. Leone, thesis , Sapienza University A.A. 2000-2001).

Cut	$\eta\pi^0\gamma$	$\omega\pi^0$	$\eta\gamma$	$K_S \to$ neutrals	$K_S \to$ charged
Generated Events	11763	33000	95000	96921	112335
Event Classification	6482	17602	55813	18815	14711
2 tracks + 5 photons	3112	724	110	371	3100
$E_{tot} - \lVert\vec{P}_{tot}\rVert$	2976	539	39	118	1171
Kinematic fit I	2714	236	5	24	66
Combinations	2649	129	1	19	0
Kinematic fit II	2247	2	0	1	0
$E_{rad} > 20$ MeV	2240	1	0	0	0

$\varepsilon = 2240/11763 = (19.04 \pm 0.36)$ % (binomial statistics)

R=33000 for $\omega\pi^0$ background.

$R\varepsilon = 6284$; since $\pi_B/\pi_S \sim 100$ => purity ~ 98.4%

Neyman-Pearson Lemma

$$P(type - Ierrors) = 1 - \epsilon = \alpha$$

$$P(type - IIerrors) = \frac{1}{R} = \beta$$

Given the two hypotheses H_s and H_b and given a set of K discriminating variables x_1, x_2,...x_K, we can define the two "likelihoods"

(66) $$L(x_1, ..., x_K/H_s) = P(x_1, ...x_K/H_s)$$

(67) $$L(x_1, ..., x_K/H_b) = P(x_1, ...x_K/H_b)$$

equal to the probabilities to have a given set of values x_i given the two hypotheses, and the **likelihood ratio** defined as

(68) $$\lambda(x_1, ...x_K) = \frac{L(x_1, ..., x_K/H_s)}{L(x_1, ..., x_K/H_b)}$$

Neyman-Pearson Lemma:

For fixed α value, a selection based on the discriminant variable λ has the lowest β value.

=> The "likelihood ratio" is the most powerful quantity to discriminate between hypotheses.

Normalization

- In order to get quantities that can be compared with theory, once we have found a given final state and estimated N_X with its uncertainty we need to normalize to "how many collisions" took place.
- Measurement of:
  - Luminosity (in case of colliding beam experiments);
  - Number of decaying particles (in case I want to study a decay);
  - Projectile rate and target densities (in case of a fixed target experiements).
- Several techniques to do that, all introducing additional uncertainties (discussed later in the course).
- ***Absolute*** vs. ***Relative*** measurements.

The simplest case: rate measurement

- Rate: r = counts / unit time (normally given in Hz). We count N in a time Δt (neglect any possible background) and assume a Poisson process with mean λ

$$r = \frac{\lambda}{\Delta t} = \frac{N}{\Delta t} \pm \frac{\sqrt{N}}{\Delta t}$$

- NB: the higher is N, the larger is the absolute uncertainty on r but the lower the relative uncertainty.

$$\frac{\sigma(r)}{r} = \frac{1}{\sqrt{N}}$$

- Only for large N ($N>20$) it is a 68% probability interval.

Cosmic ray "absolute" flux

- Rate in events/unit surface and time
- My detector has a surface *S*, I take data for a time *Δt* with a detector that has an efficiency ε and I count *N* events (again with no background). The absolute rate *r* is:

$$r = \frac{N}{\varepsilon \Delta t S}$$

- Uncertainty: I combine "in quadrature" all the potential uncertainties.

$$\frac{\sigma(r)}{r} = \sqrt{\frac{1}{N} + \left(\frac{\sigma(\varepsilon)}{\varepsilon}\right)^2 + \left(\frac{\sigma(\Delta t)}{\Delta t}\right)^2 + \left(\frac{\sigma(S)}{S}\right)^2}$$

- Distinction between "***statistical***" and "***systematic***" uncertainty

Combination of uncertainties

- Back to the previous formula.

$$\frac{\sigma(r)}{r} = \sqrt{\frac{1}{N} + \left(\frac{\sigma(\varepsilon)}{\varepsilon}\right)^2 + \left(\frac{\sigma(\Delta t)}{\Delta t}\right)^2 + \left(\frac{\sigma(S)}{S}\right)^2}$$

1. Suppose we have a certain "unreducible" uncertainty on S and/or on ε (the uncertainty on Δt we assume is anyhow negligible..). Is it useful to go on to take data ? Or there is a limit above which it is no more useful to go on ?
2. Suppose that we have a limited amount of time to take data N is fixed: is it useful to improve our knowledge on ε ?

Not only event counting

- Once the candidate sample is obtained many quantities can be measured (particle properties, e.g. particle mass).
- BUT in most cases they are obtained from a **FIT** to a data distribution. So, you divide events in bins and extract the quantity as a ***fit parameter*** ➔ the event counting is still one major source of uncertainty ➔ the uncertainty on the parameter depends on the statistics ≈ $\sqrt{N_i}$.
- Example:
  - Measure the mass of a "imaginary" particle of M=5 GeV.
  - Mass spectrum, gaussian peak over a uniform background
  - FIT in three different cases: 10^3, 10^4 and 10^5 events selected

Mass uncertainty due to statistics

Observations:

→ Poissonian uncertainty on each bin

→ Reduce bin size for higher statistics

→ Fit function = A+B*Gauss(M)

→ Free parameters: A,B,M (fixed width)

→ The fit is good for each statistics

Results

$N=10^3$ events:

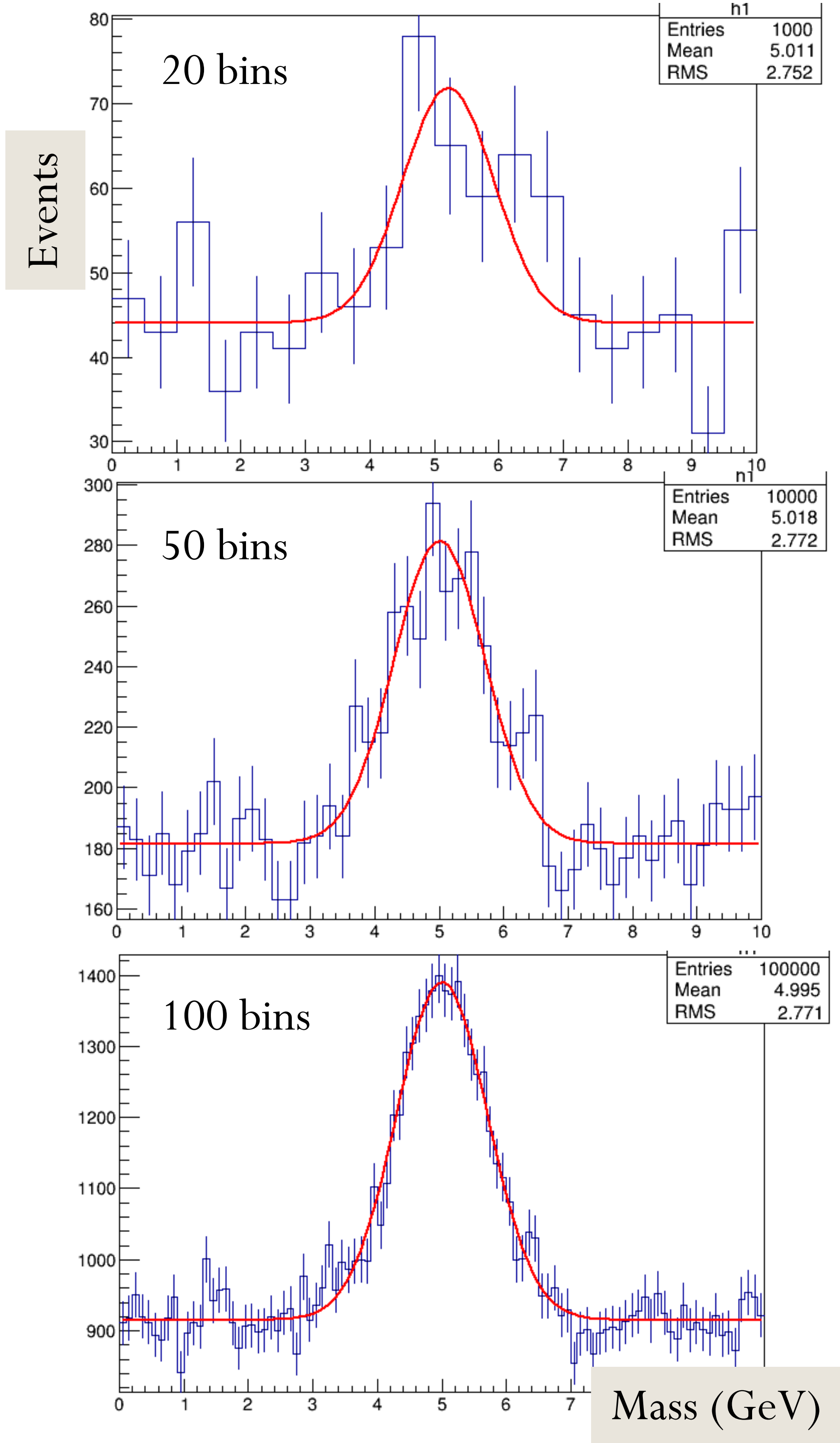
Mass = 5.22±0.22 GeV, $\chi^2 =$ 28 / 18 dof

$N=10^4$ events:

Mass = 5.01±0.06 GeV, $\chi^2 =$ 38 / 48 dof

$N=10^5$ events:

Mass = 5.02±0.02 GeV, $\chi^2 =$ 83 / 98 dof

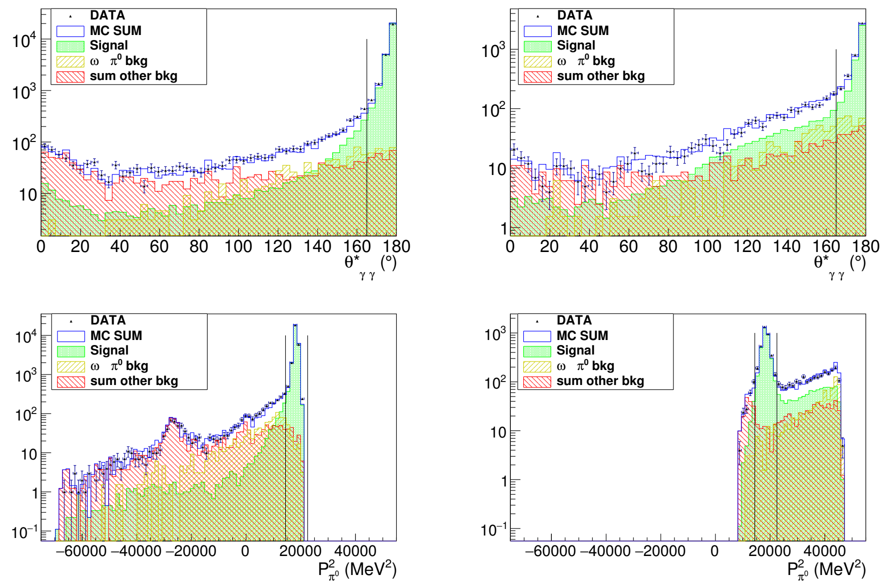


Where could be a systematic uncertainty here ?

- Absolute mass scale: this can be measured using a candle of known mass. Not always it is available. e.g. Z for the Higgs mass at the LHC.
- Mass resolution: in most cases the width of the peak is given by the experimental resolution that sometimes is not perfectly gaussian, giving rise to possible distortion to the curve.
- Physics effects: knowledge of the line-shape, interference with the background…
- In general: M = central value $\pm$ stat.uncert. $\pm$ syst.uncert.

An example:
a recent study of the Dalitz plot of the
$\eta \rightarrow \pi^+\pi^-\pi^0$ decay

$\eta \to \pi^+\pi^-\pi^0$ decay



The light quark masses: study of $\eta \to \pi^+\pi^-\pi^0$decay

$\eta \to \pi\pi\pi$ decay $\Rightarrow$ Isospin violation

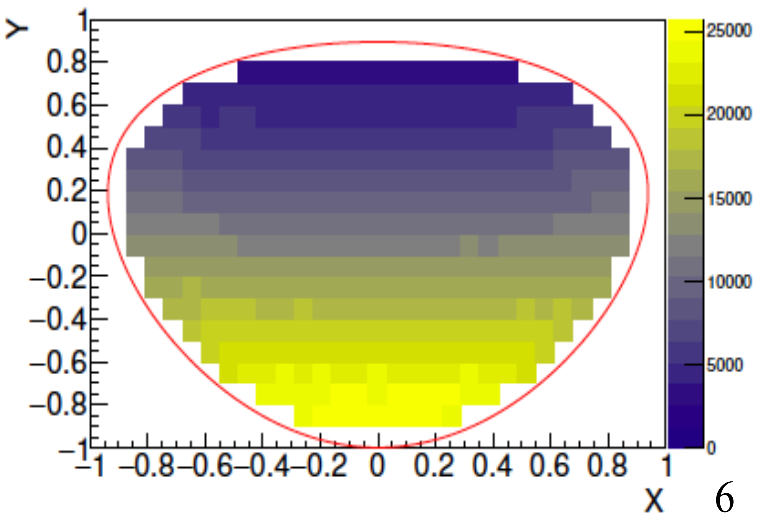
e.m. strongly suppressed, induced dominantly by the strong interaction associated with the u-d quark mass difference

$$X = \sqrt{3}\frac{T_{\pi^+} - T_{\pi^-}}{Q_\eta}$$

$$Y = \frac{3T_{\pi^0}}{Q_\eta} - 1 \qquad Q_\eta = T_{\pi^+} + T_{\pi^-} + T_{\pi^0} = m_\eta - 2m_{\pi^+} - m_{\pi^0}$$

Fit to the Dalitz Plot

$$|A(X,Y)|^2 \propto 1 + aY + bY^2 + cX + dX^2 + eXY + fY^3 + \ldots$$



6

KLOE-2 JHEP 05(2016)019

$\chi^2/\mathrm{dof} = 360/365 \qquad p = 56\%$

$$a = -1.095 \pm 0.003^{+0.003}_{-0.002}$$

$$b = +0.145 \pm 0.003 \pm 0.005$$

$$d = +0.081 \pm 0.003^{+0.006}_{-0.005}$$

$$f = +0.141 \pm 0.007^{+0.007}_{-0.008}$$

$$g = -0.044 \pm 0.009^{+0.012}_{-0.013}$$

c, e param. are C-violating, consistent with zero

syst. error ($\times 10^4$)	Δa	Δb	Δd	Δf	Δg
EGmin	± 6	± 12	± 10	± 5	± 16
BkgSub	± 8	± 7	± 11	± 6	± 38
BIN	± 17	± 13	± 9	± 36	± 44
$\theta_{+\gamma}, \theta_{-\gamma}$ cut	${}^{+0}_{-1}$	${}^{+0}_{-2}$	${}^{+2}_{-2}$	${}^{+3}_{-0}$	${}^{+3}_{-2}$
Δt_e cut	${}^{+6}_{-11}$	${}^{+12}_{-1}$	${}^{+18}_{-1}$	${}^{+3}_{-8}$	${}^{+26}_{-54}$
$\Delta t_e - \Delta t_\pi$ cut	± 0	${}^{+0}_{-1}$	${}^{+3}_{-1}$	± 0	${}^{+2}_{-1}$
$\theta^*_{\gamma\gamma}$ cut	${}^{+14}_{-5}$	${}^{+2}_{-1}$	${}^{+21}_{-12}$	${}^{+5}_{-25}$	${}^{+26}_{-38}$
MM	${}^{+8}_{-10}$	${}^{+46}_{-43}$	${}^{+49}_{-45}$	${}^{+57}_{-62}$	${}^{+100}_{-92}$
ECL	± 0	± 8	± 6	± 9	± 12
TOTAL	${}^{+26}_{-25}$	${}^{+52}_{-48}$	${}^{+59}_{-50}$	${}^{+69}_{-77}$	${}^{+123}_{-129}$

$\eta\to\pi^+\pi^-\pi^0$ decay

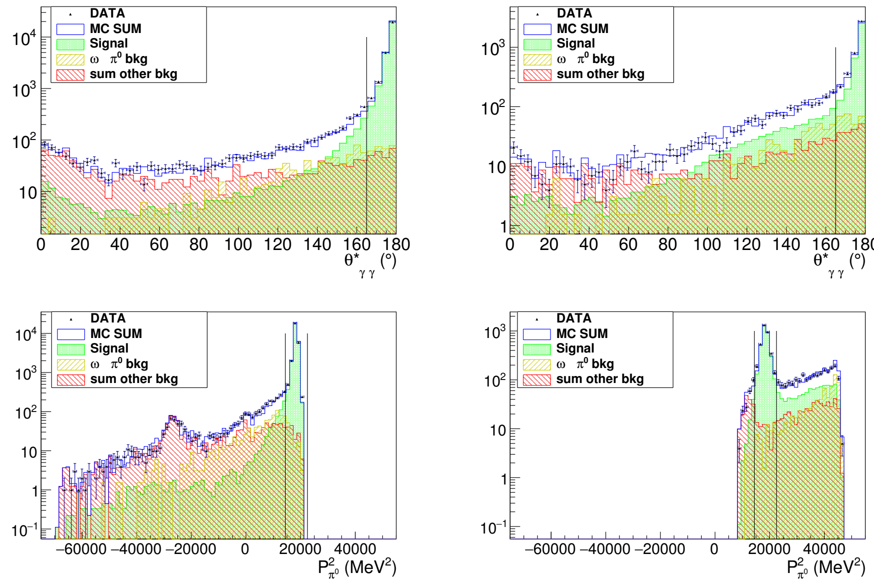
The light quark masses: study of $\eta\to\pi^+\pi^-\pi^0$decay

$\eta\to\pi\pi\pi$ decay $\Rightarrow$ Isospin violation

e.m. strongly suppressed, induced dominantly by the strong interaction associated with the u-d quark mass difference

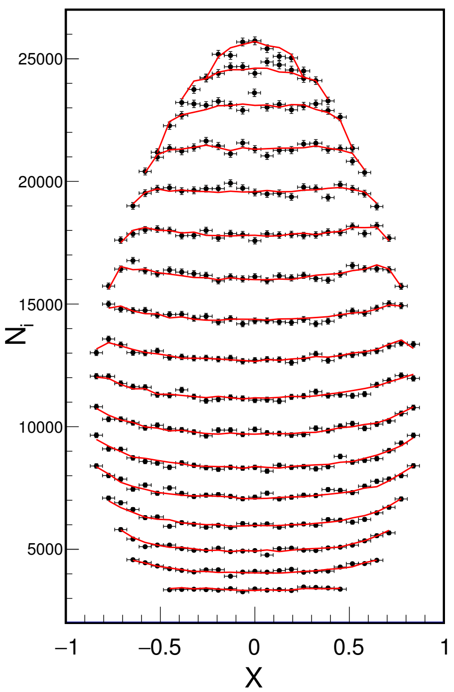
$$X=\sqrt{3}\frac{T_{\pi^+}-T_{\pi^-}}{Q_\eta}$$

$$Y=\frac{3T_{\pi^0}}{Q_\eta}-1 \qquad Q_\eta=T_{\pi^+}+T_{\pi^-}+T_{\pi^0}=m_\eta-2m_{\pi^+}-m_{\pi^0}$$



Fit to the Dalitz Plot

$$|A(X,Y)|^2\propto 1+aY+bY^2+cX+dX^2+eXY+fY^3+\dots$$



$\chi^2/\mathrm{dof}=360/365 \qquad p=56\%$

$$a=-1.095\pm0.003^{+0.003}_{-0.002}$$

$$b=+0.145\pm0.003\pm0.005$$

$$d=+0.081\pm0.003^{+0.006}_{-0.005}$$

$$f=+0.141\pm0.007^{+0.007}_{-0.008}$$

$$g=-0.044\pm0.009^{+0.012}_{-0.013}$$

c, e param. are C-violating, consistent with zero

syst. error ($\times10^4$)	Δa	Δb	Δd	Δf	Δg
EGmin	±6	±12	±10	±5	±16
BkgSub	±8	±7	±11	±6	±38
BIN	±17	±13	±9	±36	±44
$\theta_{+\gamma},\theta_{-\gamma}$ cut	$^{+0}_{-1}$	$^{+0}_{-2}$	$^{+2}_{-2}$	$^{+3}_{-0}$	$^{+3}_{-2}$
Δt_e cut	$^{+6}_{-11}$	$^{+12}_{-1}$	$^{+18}_{-1}$	$^{+3}_{-8}$	$^{+26}_{-54}$
$\Delta t_e-\Delta t_\pi$ cut	±0	$^{+0}_{-1}$	$^{+3}_{-1}$	±0	$^{+2}_{-1}$
$\theta^*_{\gamma\gamma}$ cut	$^{+14}_{-5}$	$^{+2}_{-1}$	$^{+21}_{-12}$	$^{+5}_{-25}$	$^{+26}_{-38}$
MM	$^{+8}_{-10}$	$^{+46}_{-43}$	$^{+49}_{-45}$	$^{+57}_{-62}$	$^{+100}_{-92}$
ECL	±0	±8	±6	±9	±12
TOTAL	$^{+26}_{-25}$	$^{+52}_{-48}$	$^{+59}_{-50}$	$^{+69}_{-77}$	$^{+123}_{-129}$

KLOE-2 JHEP 05(2016)019

$\eta \to \pi^+\pi^-\pi^0$ decay

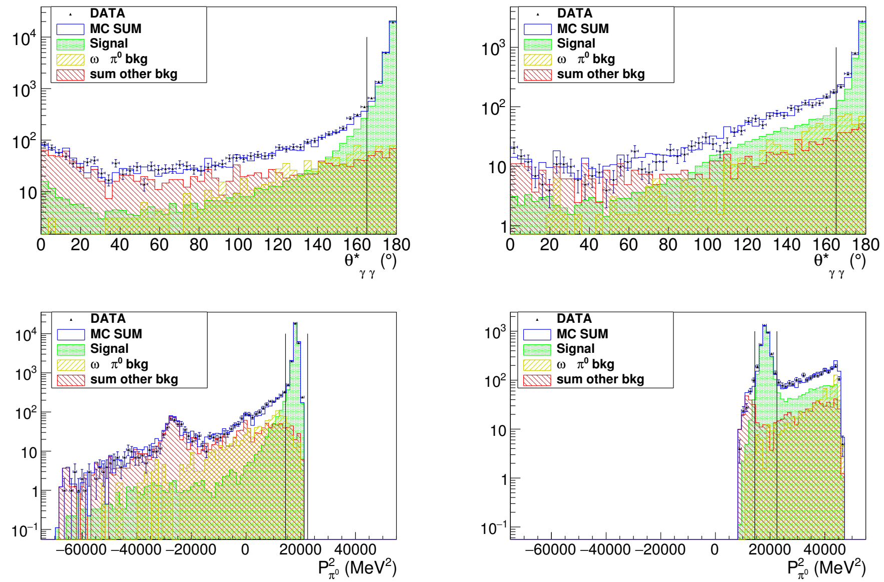
The light quark masses: study of $\eta \to \pi^+\pi^-\pi^0$decay

$\eta \to \pi\pi\pi$ decay $\Rightarrow$ Isospin violation

e.m. strongly suppressed, induced dominantly by the strong interaction associated with the u-d quark mass difference

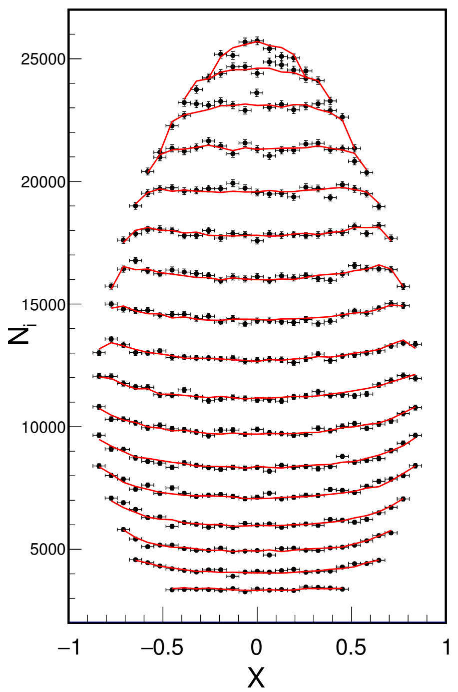
$$X = \sqrt{3}\frac{T_{\pi^+} - T_{\pi^-}}{Q_\eta}$$

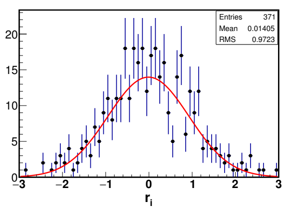
$$Y = \frac{3T_{\pi^0}}{Q_\eta} - 1 \qquad Q_\eta = T_{\pi^+} + T_{\pi^-} + T_{\pi^0} = m_\eta - 2m_{\pi^+} - m_{\pi^0}$$



Fit to the Dalitz Plot

$$|A(X,Y)|^2 \propto 1 + aY + bY^2 + cX + dX^2 + eXY + fY^3 + \ldots$$





$\chi^2/\text{dof} = 360/365 \qquad p = 56\%$

$$a = -1.095 \pm 0.003^{+0.003}_{-0.002}$$

$$b = +0.145 \pm 0.003 \pm 0.005$$

$$d = +0.081 \pm 0.003^{+0.006}_{-0.005}$$

$$f = +0.141 \pm 0.007^{+0.007}_{-0.008}$$

$$g = -0.044 \pm 0.009^{+0.012}_{-0.013}$$

c, e param. are C-violating, consistent with zero

syst. error ($\times 10^4$)	Δa	Δb	Δd	Δf	Δg
EGmin	± 6	± 12	± 10	± 5	± 16
BkgSub	± 8	± 7	± 11	± 6	± 38
BIN	± 17	± 13	± 9	± 36	± 44
$\theta_{+\gamma}, \theta_{-\gamma}$ cut	${}^{+0}_{-1}$	${}^{+0}_{-2}$	${}^{+2}_{-2}$	${}^{+3}_{-0}$	${}^{+3}_{-2}$
Δt_e cut	${}^{+6}_{-11}$	${}^{+12}_{-1}$	${}^{+18}_{-1}$	${}^{+3}_{-8}$	${}^{+26}_{-54}$
$\Delta t_e - \Delta t_\pi$ cut	± 0	${}^{+0}_{-1}$	${}^{+3}_{-1}$	± 0	${}^{+2}_{-1}$
$\theta^*_{\gamma\gamma}$ cut	${}^{+14}_{-5}$	${}^{+2}_{-1}$	${}^{+21}_{-12}$	${}^{+5}_{-25}$	${}^{+26}_{-38}$
MM	${}^{+8}_{-10}$	${}^{+46}_{-43}$	${}^{+49}_{-45}$	${}^{+57}_{-62}$	${}^{+100}_{-92}$
ECL	± 0	± 8	± 6	± 9	± 12
TOTAL	${}^{+26}_{-25}$	${}^{+52}_{-48}$	${}^{+59}_{-50}$	${}^{+69}_{-77}$	${}^{+123}_{-129}$

KLOE-2 JHEP 05(2016)019

Uncertainty combination

central value ± **stat.uncert.** ± **syst.uncert.**

Can we combine stat. and syst. ? If yes how ?

The two uncertainties might have different probability meaning: typically one is a gaussian 68% C.L., the other is a "maximum" uncertainty, so in general it is better to hold them separate.

If needed better to add in quadrature rather than linearly.

Summarizing

- Steps of an PP experiment (assuming the accelerator and the detector are there):
  - Design of a **trigger**
  - Definition of an offline **selection**
  - **Event counting** and **normalization** (including **efficiency** and **background** evaluation)
  - **Fit** of "candidate" distributions
- Uncertainties
  - Statistical due to Poisson fluctuations of the event counting
  - Statistical due to binomial fluctuations in the efficiency measurement
  - Systematic due to non perfect knowledge of detector effects.