Neyman Construction

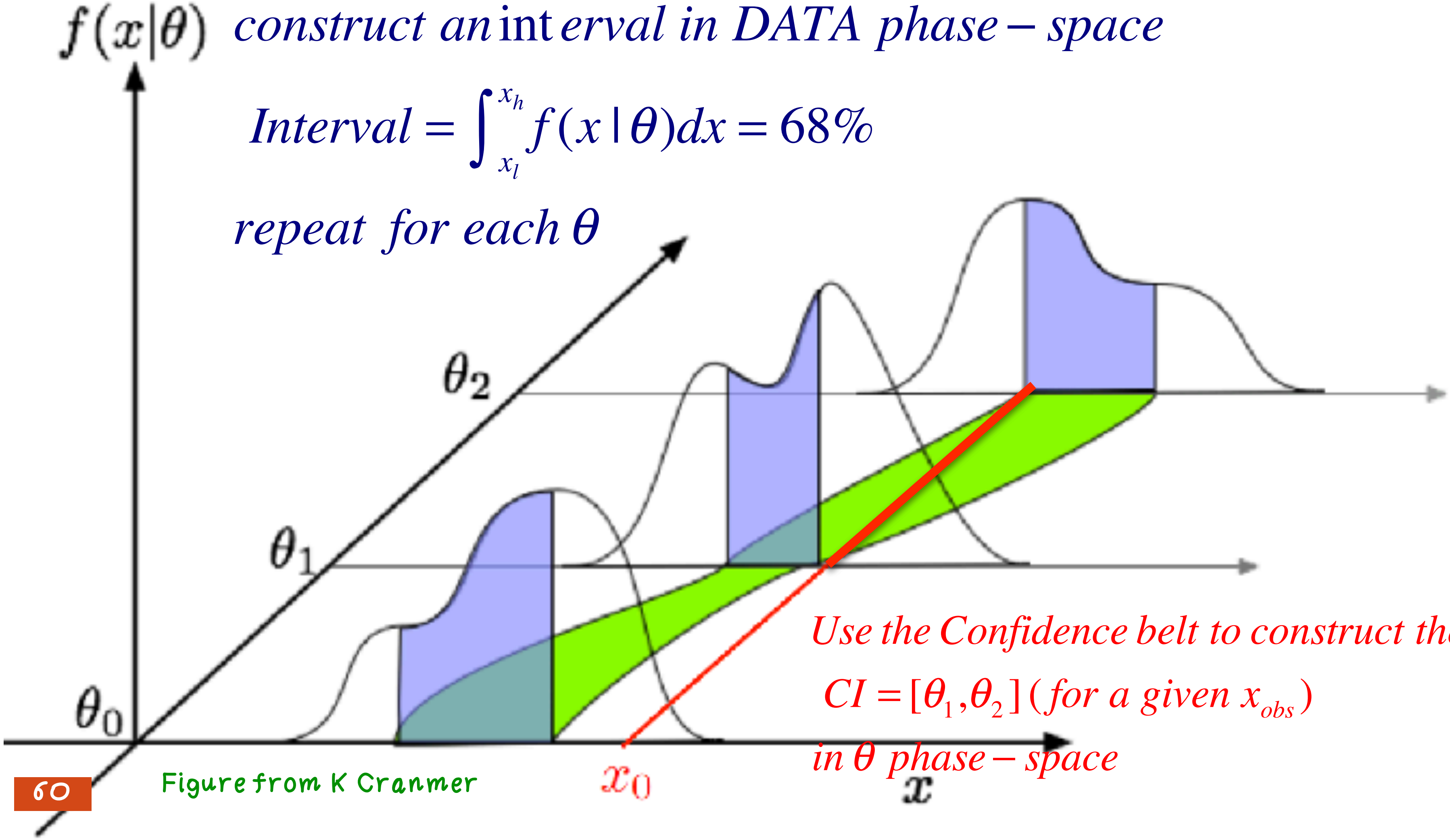
$\theta \equiv s_{true} \quad x \equiv s_{measured} \quad pdf\ f(x|\theta)\ is\ known$

for each prospective θ *generate* x

construct an interval in DATA phase – space

$$Interval = \int_{x_l}^{x_h} f(x|\theta)dx = 68\%$$

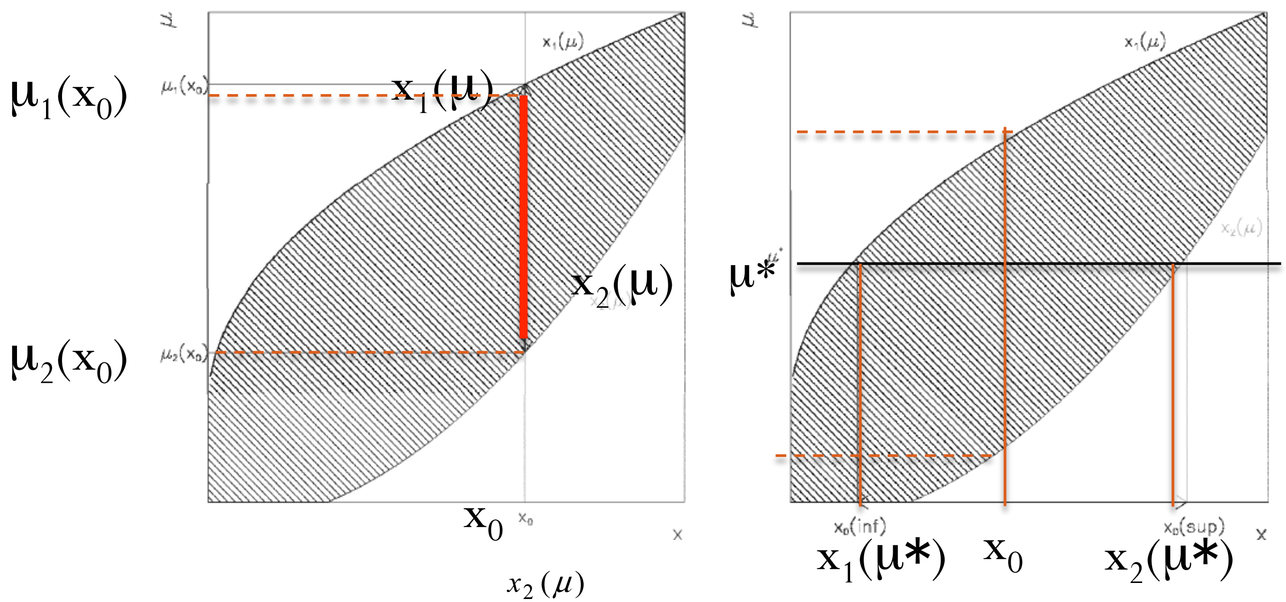
repeat for each θ



Use the Confidence belt to construct the
$CI = [\theta_1, \theta_2]$ *(for a given* x_{obs}*)*
in θ *phase – space*

Figure from K Cranmer

Neyman's construction



$$\int_{x_1(\mu)}^{x_2(\mu)} f(x/\mu)dx = \alpha$$

By construction the probability to measure $x_0' < x_0$ if the true value $\mu = \mu_1(x_0)$ is $(1-\alpha)/2$
$x_0' > x_0$ if the true value $\mu = \mu_2(x_0)$ is $(1-\alpha)/2$

Coverage: suppose μ^* the true value

$$P(x_1(\mu^*) < x_0 < x_2(\mu^*)) = \alpha$$

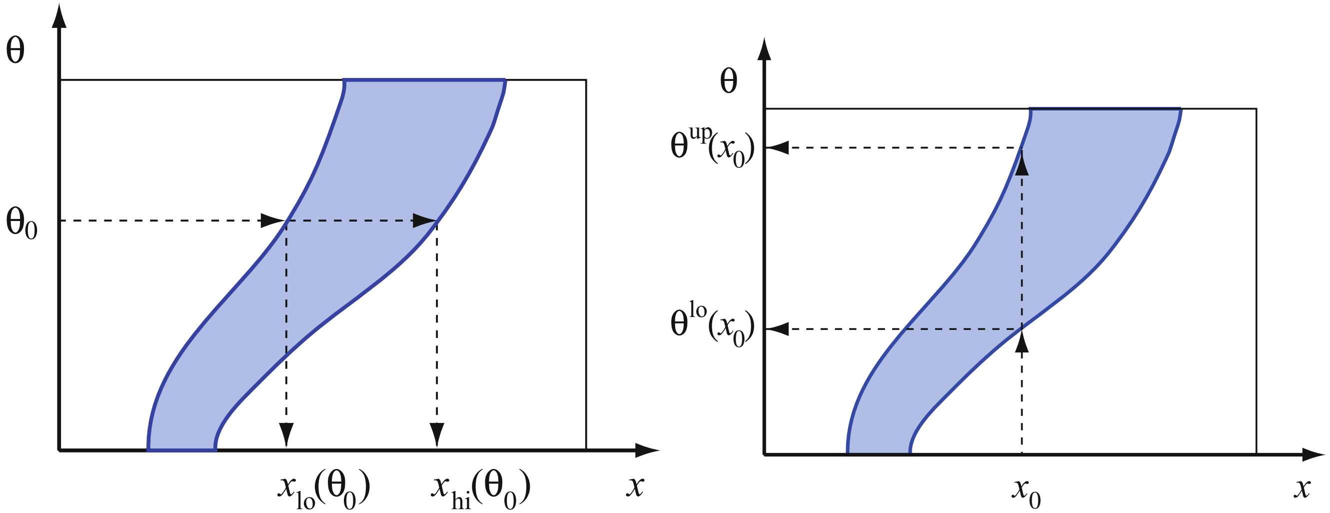


Fig. 7.1 Graphical illustration of Neyman belt construction (*left*) and inversion (*right*)

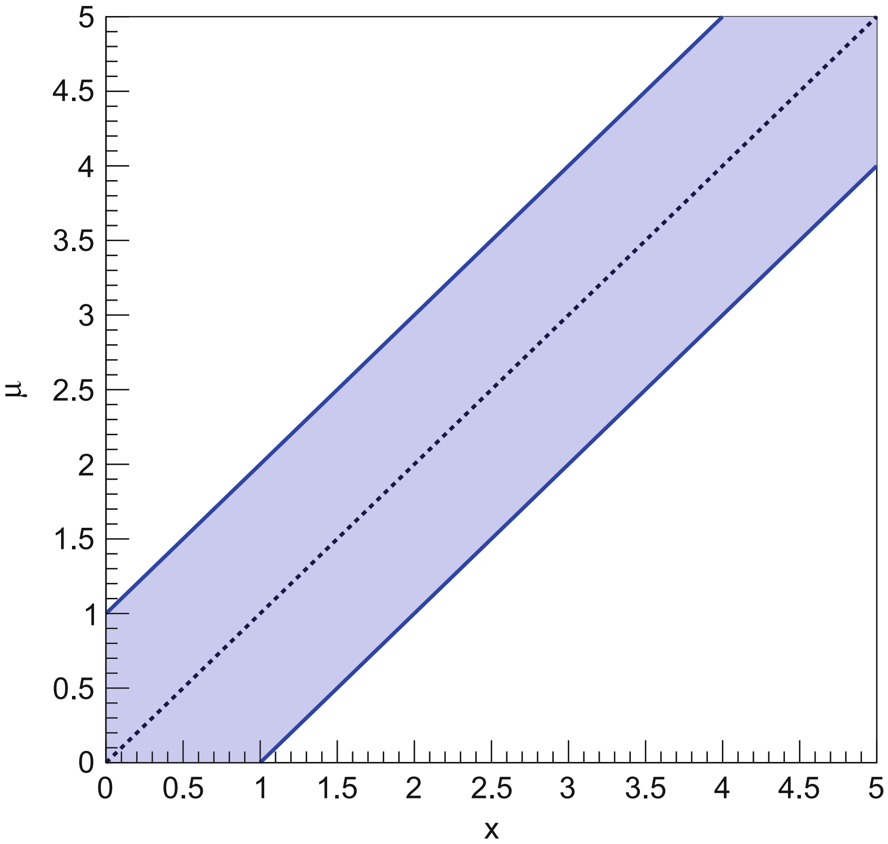


Fig. 7.3 Neyman belt for the parameter μ of a Gaussian with $\sigma = 1$ at the 68.27% confidence level

Suppose Poisson variable and n=0 is measured (no background) Upper limit (lower limit =0)
=> 0±0 (freq) or 1±1 (Beyes) ?

By construction the probability to measure $x_0' < x_0$ if the true value $\mu = \mu_1(x_0)$ is $(1-\alpha)$ (only one limit) or the probability to measure $x_0' > x_0$ if the true value $\mu = \mu_1(x_0)$ is α

$$P(n>0/\lambda) = \sum_{n=1}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} = 1 - e^{-\lambda} = \alpha$$

frequentist

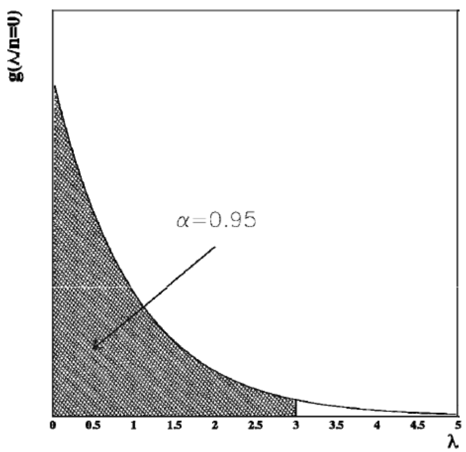
$$\bar{\lambda} = -\ln(1-\alpha)$$

$$g(\lambda/n=0) = \frac{p(n=0/\lambda) f_0(\lambda)}{\int_0^{\infty} p(n=0/\lambda) f_0(\lambda) d\lambda} = \frac{e^{-\lambda}}{\int_0^{\infty} e^{-\lambda} d\lambda} = e^{-\lambda}$$

Bayesian
(uniform prior)

$$p(\lambda < \bar{\lambda}) = \int_0^{\bar{\lambda}} e^{-\lambda} d\lambda = 1 - e^{-\bar{\lambda}} = \alpha$$

	90%	95%	99%
$\bar{\lambda}$	2.3	3.0	4.6



Poisson

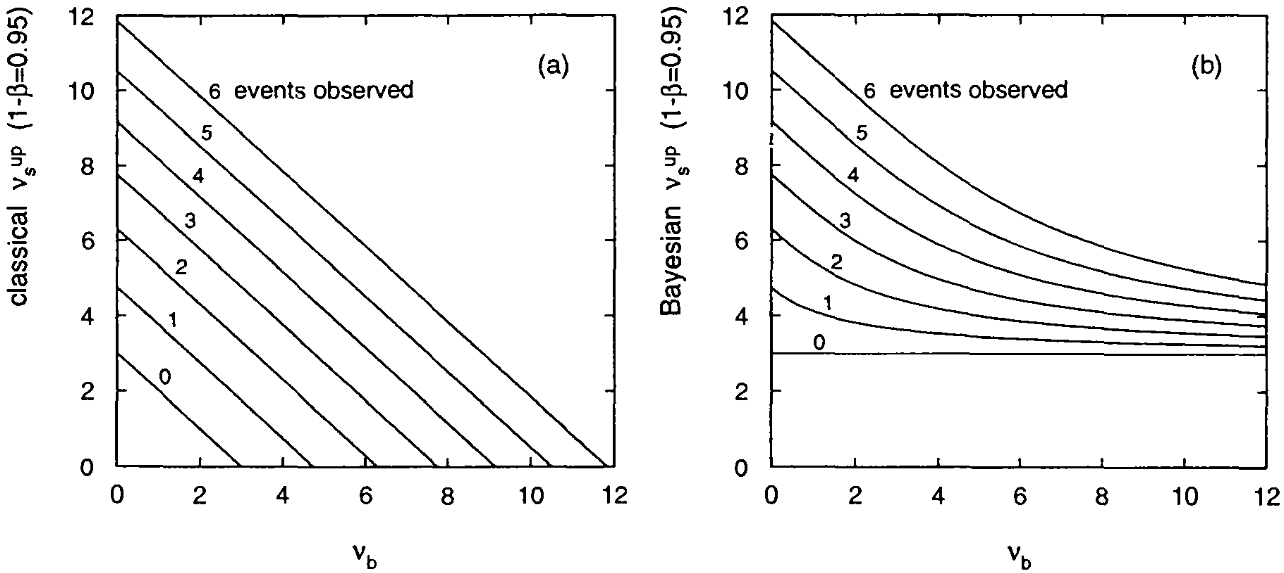


Fig. 9.9 Upper limits ν_s^{up} at a confidence level of $1-\beta = 0.95$ for different numbers of events observed n_{obs} and as a function of the expected number of background events ν_{b}. (a) The classical limit. (b) The Bayesian limit based on a uniform prior density for ν_s.

Combining uncertainties

- Given the uncertainties on $\boldsymbol{N_{cand}}$, $\boldsymbol{\varepsilon}$ and $\boldsymbol{N_b}$, how can we estimate the uncertainty on $\boldsymbol{N_X}$?
- ➔ Uncertainty Propagation. General formulation

$$\left(\frac{\sigma(N_X)}{N_X}\right)^2 = \left(\frac{\sigma(\varepsilon)}{\varepsilon}\right)^2 + \frac{\sigma^2(N_{cand}) + \sigma^2(N_b)}{(N_{cand} - N_b)^2}$$

Assumption: three indipendent contributions

NB: if $\boldsymbol{N_{cand}} \approx \boldsymbol{N_b}$ the relative uncertainty becomes very large (the Formula cannot be applied anymore…)

Can we say we have really observed a signal ???

Or we are simply observing some fluctuation of the background ?

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".

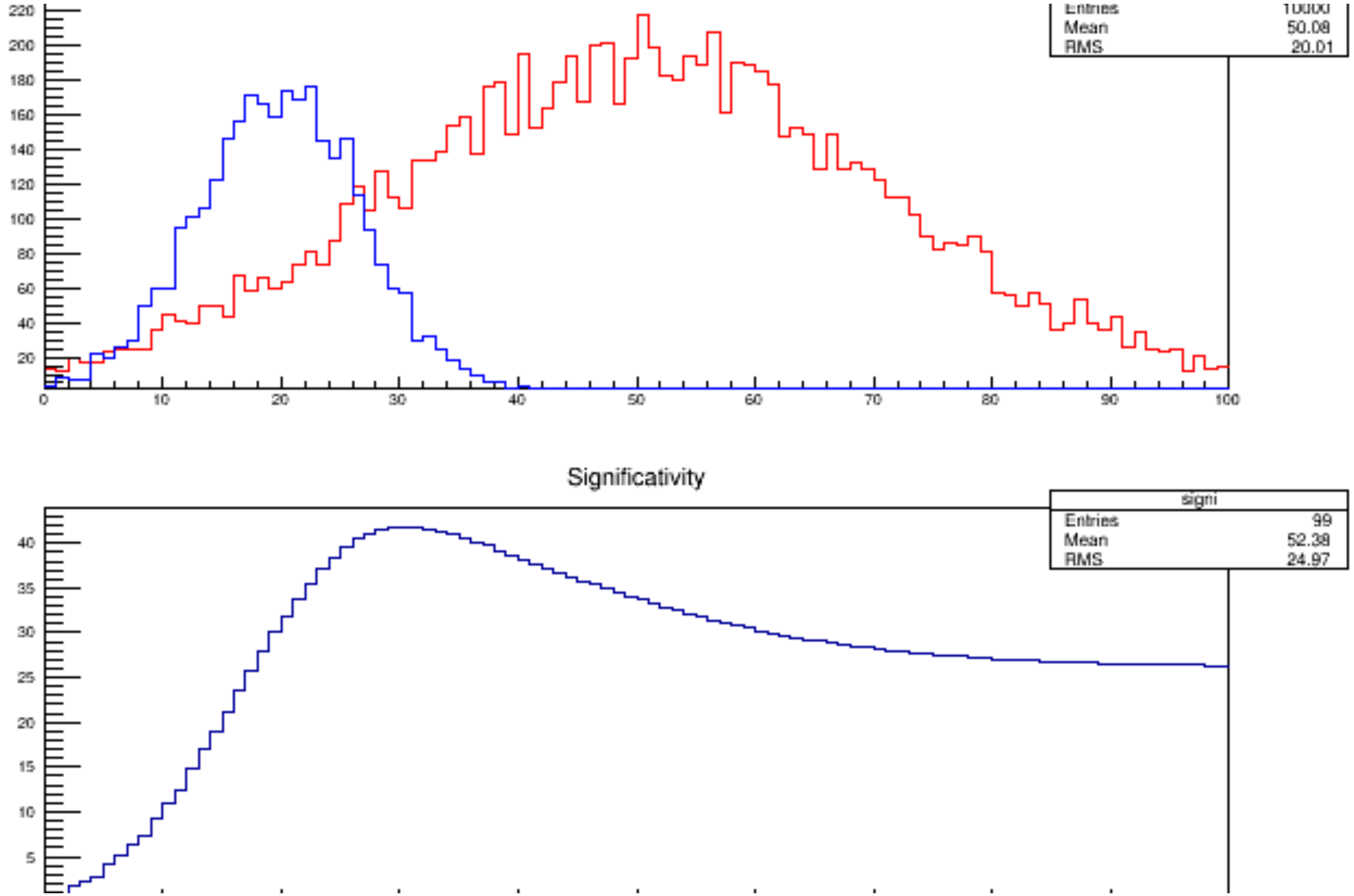
Example - I

B=10000

σ_x (B) = 15

S=3000

σ_x(S) = 5



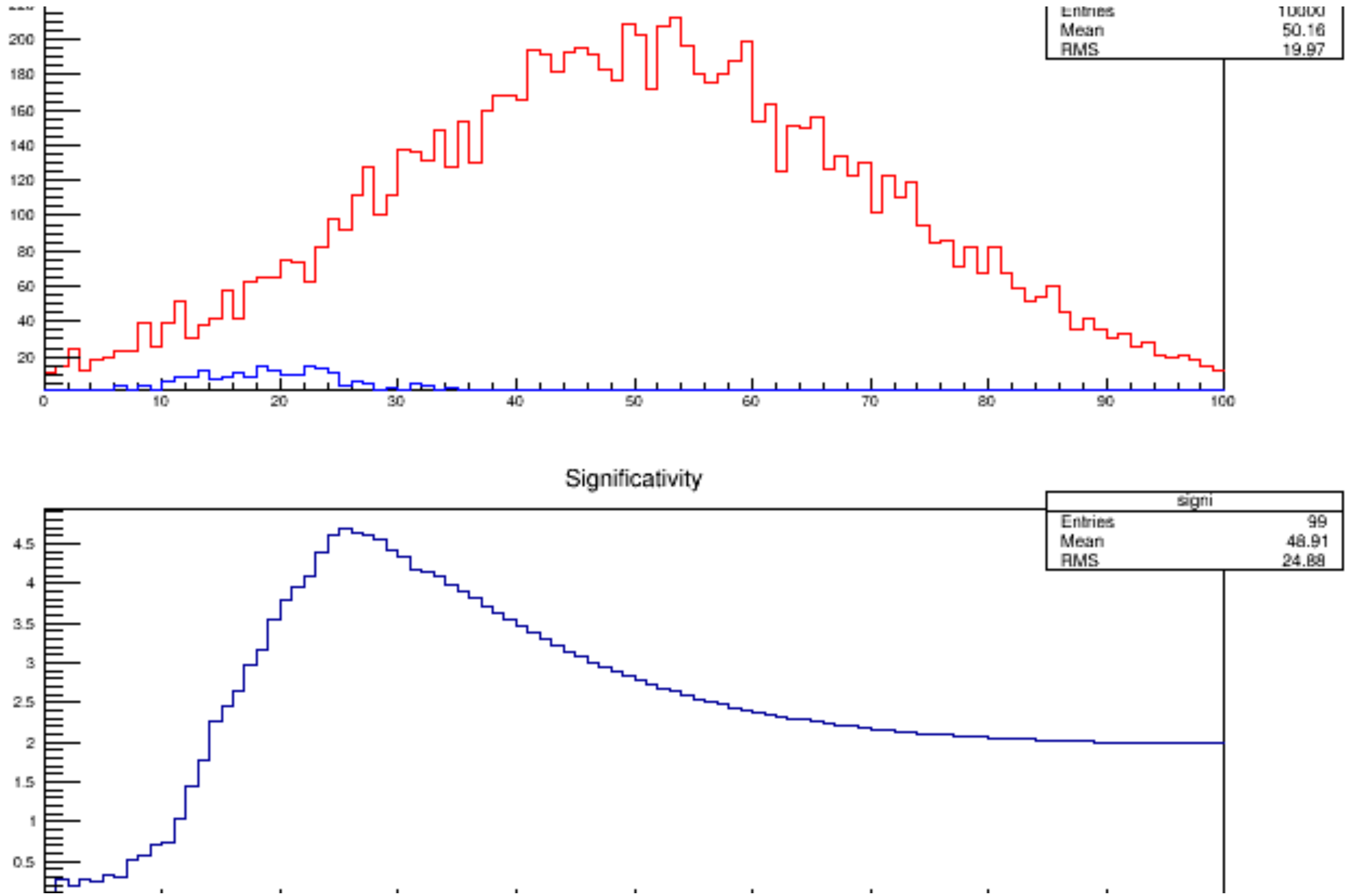
Example - II

B=10000

σ_x (B) = 15

S=200

σ_x(S) = 5



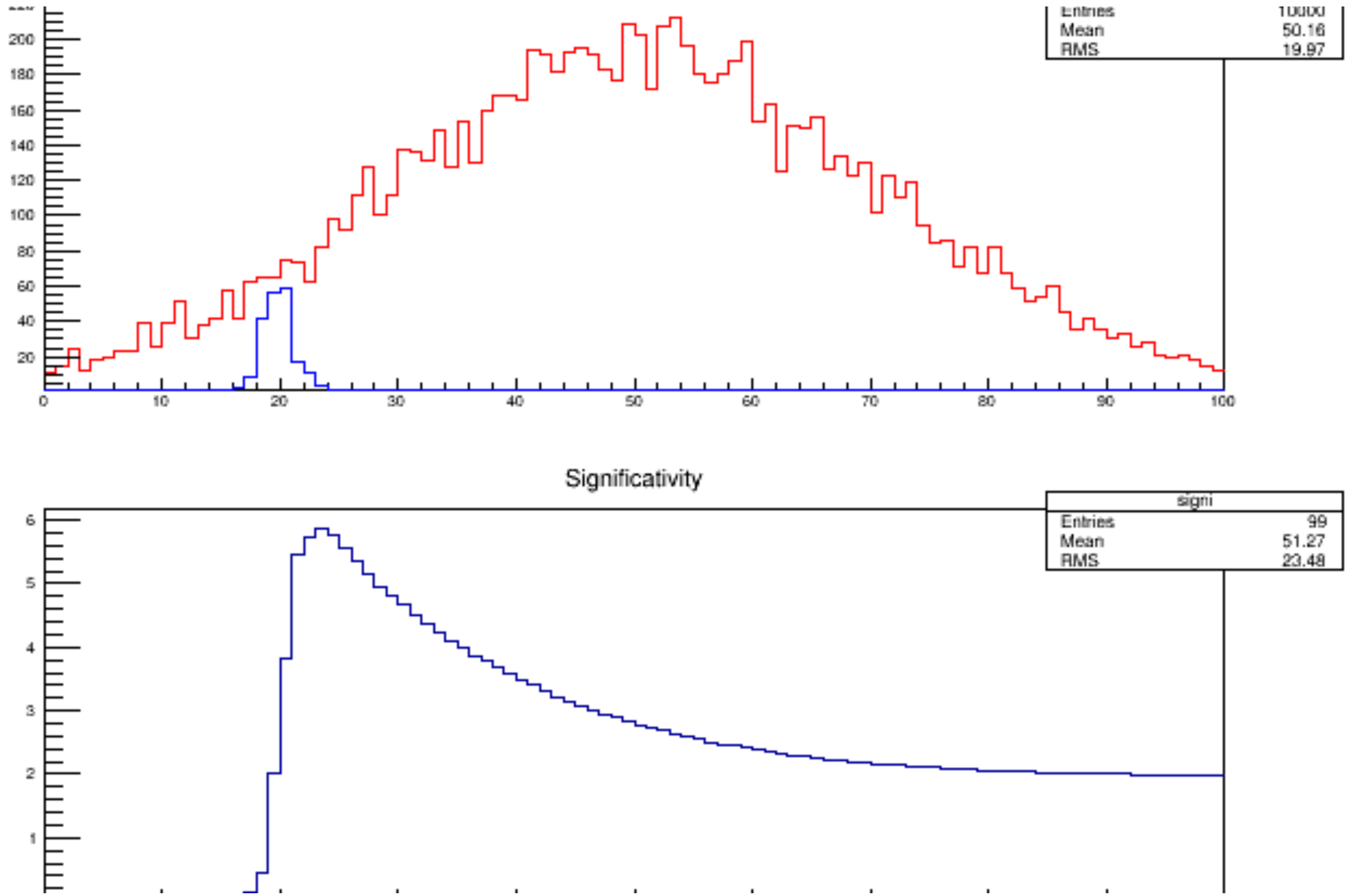
Example - III

B=10000

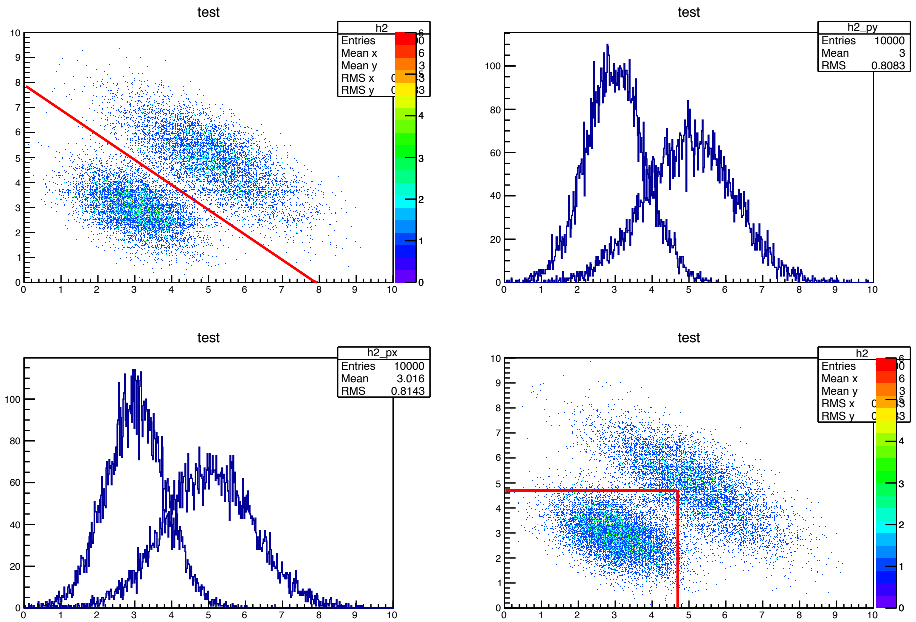
σ_x (B) = 15

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

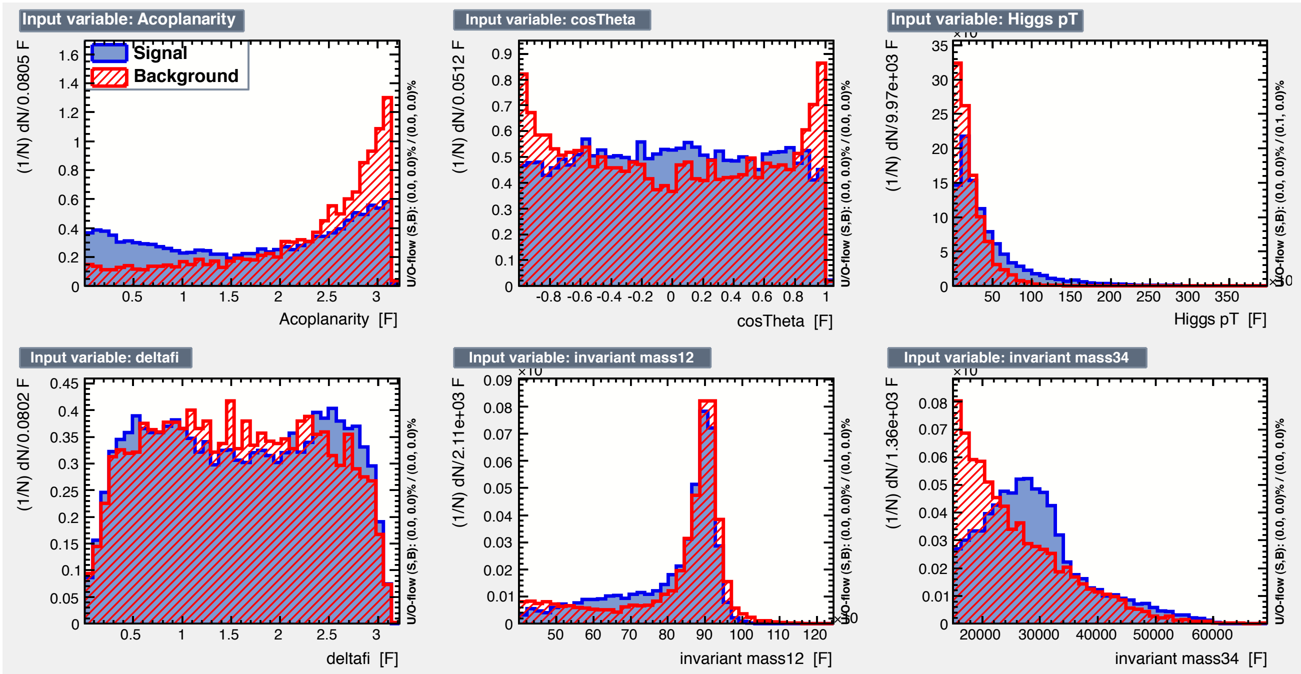
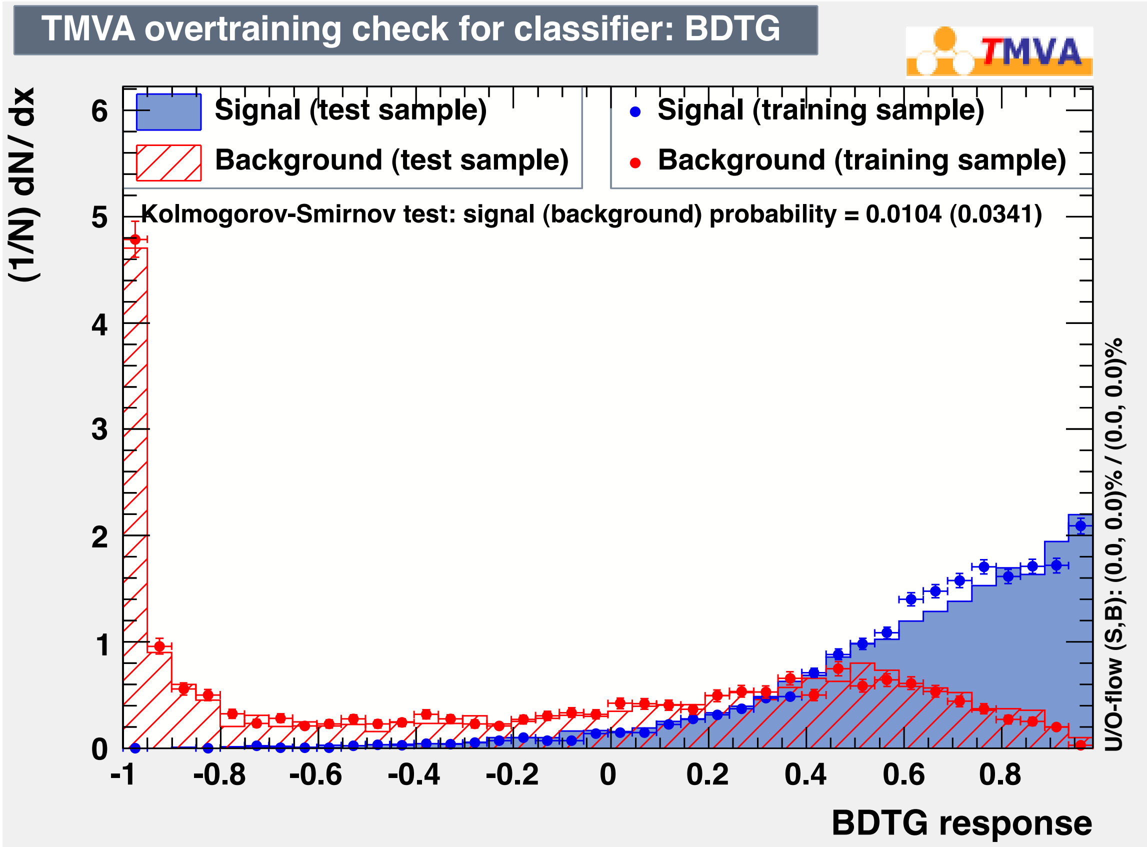


FIGURE 5. Comparison between MC signal (blue) and MC background (red) distributions for the 6 chosen discriminating variables entering in the multivariate analysis (taken from A.Calandri thesis, Sapienza University, A.A. 2011-2012).

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

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

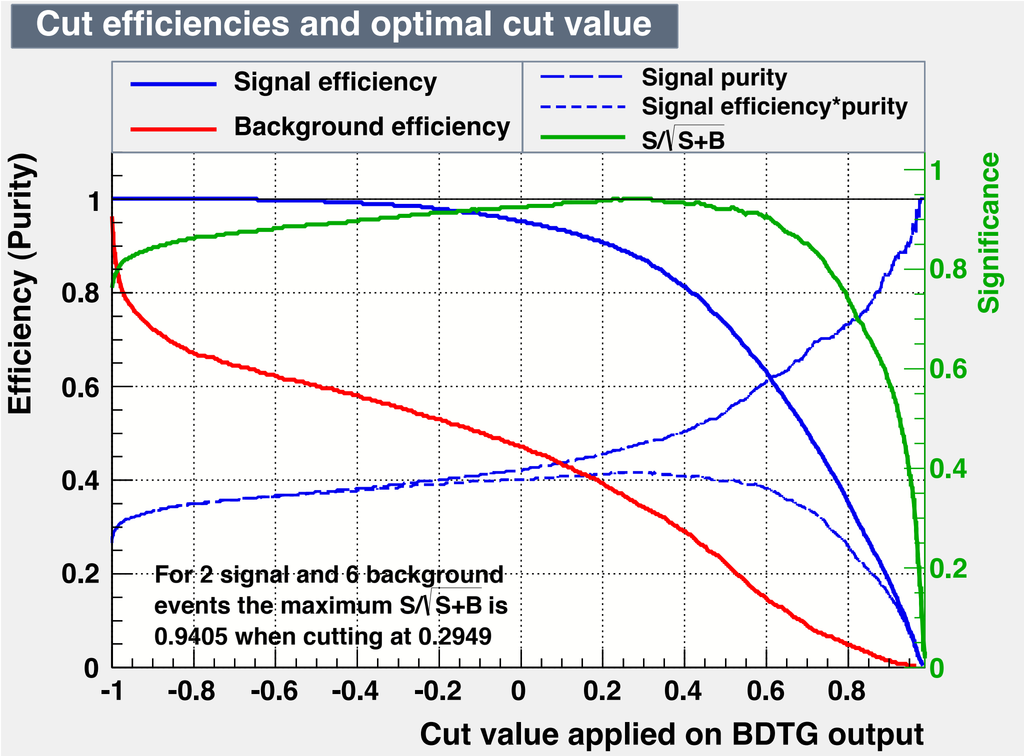


FIGURE 7. Several quantities are shown as a function of the possible value of t_{cut}, the cut on the BDT variable. Blue and red curves show respectively the signal and background efficiency while the green curve is the score function that, in this case, has a maximum around $t_{cut} = 0.25$ although 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

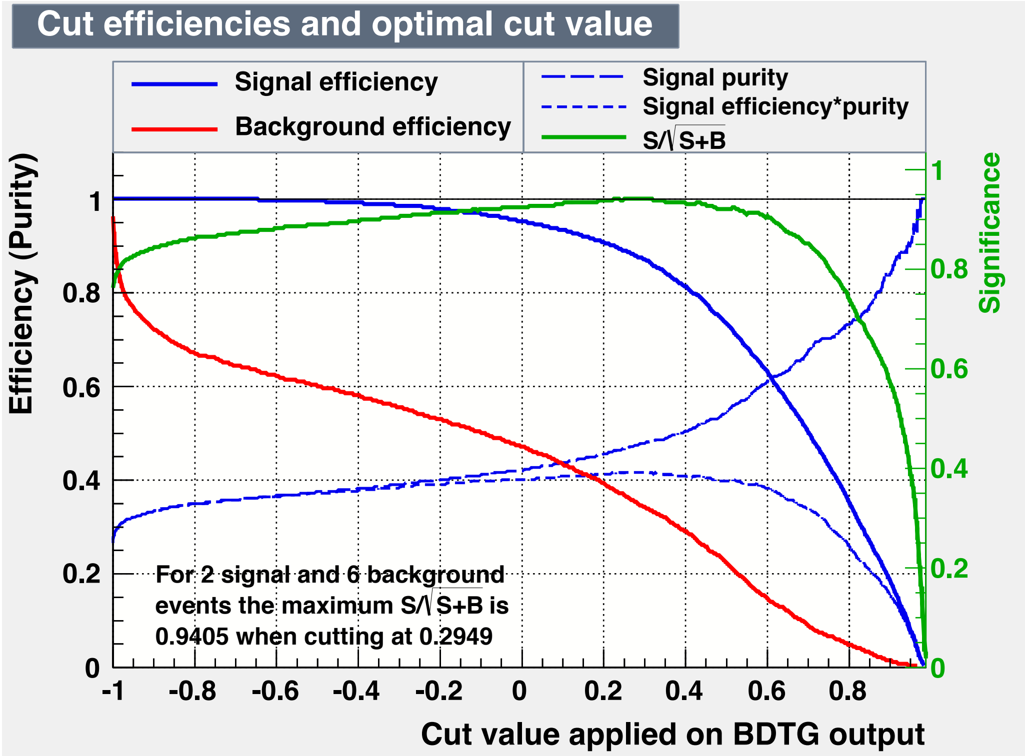


FIGURE 7. Several quantities are shown as a function of the possible value of t_{cut}, the cut on the BDT variable. Blue and red curves show respectively the signal and background efficiency while the green curve is the score function that, in this case, has a maximum around $t_{cut} = 0.25$ although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

Signal observed?

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

$$P(type - I errors) = 1 - \epsilon$$

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

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:

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

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

Maximize the purity for a given efficiency

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

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

3.2. **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". An example of cut-flow is shown in Table 1. The choice of each single cut is motivated by the shape of the MC signal and background distributions in the different variables. From the cut-flow shown in Table 1 we get: $\epsilon = 2240/11763 =$

TABLE 1. Example of cut-flow. The selection of $\eta\pi^0\gamma$ final state with $\eta \rightarrow \pi^+\pi^-\pi^0$ 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	$\eta\pi^0\gamma$	$\omega\pi^0$	$\eta\gamma$	$K_S \rightarrow$ neutrals	$K_S \rightarrow$ charged
Generated Events	11763	33000	95000	96921	112335
Event Classification	6482	17602	55813	18815	14711
2 tracks + 5 photons	3112	724	110	371	3100
$E_{tot} - \|\vec{P}_{tot}\|$	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

$(19.04 \pm 0.36)\%$[3] and $R = 33000$ for $\omega\pi^0$. For the other background channels only a lower limit on R can be given, since in the end no events pass the selection.

Table 1, where we use the $\omega\pi^0$ sample as the only background of the analysis, $R\epsilon = 6284$ and since $\pi_B/\pi_S \sim 10^2$ we have a *purity* of $\sim 98.4\%$.

Neyman-Person 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-Person 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 $\boldsymbol{\varepsilon}$ (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 $\boldsymbol{\varepsilon}$?

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.