Metody eksperymentalne w fizyce wysokich energii

prof. dr hab. A.F.Żarnecki
Zakład Cząstek i Oddziaływań Fundamentalnych IFD

Wykład XV

- Poszukiwanie "nowej fizyki"

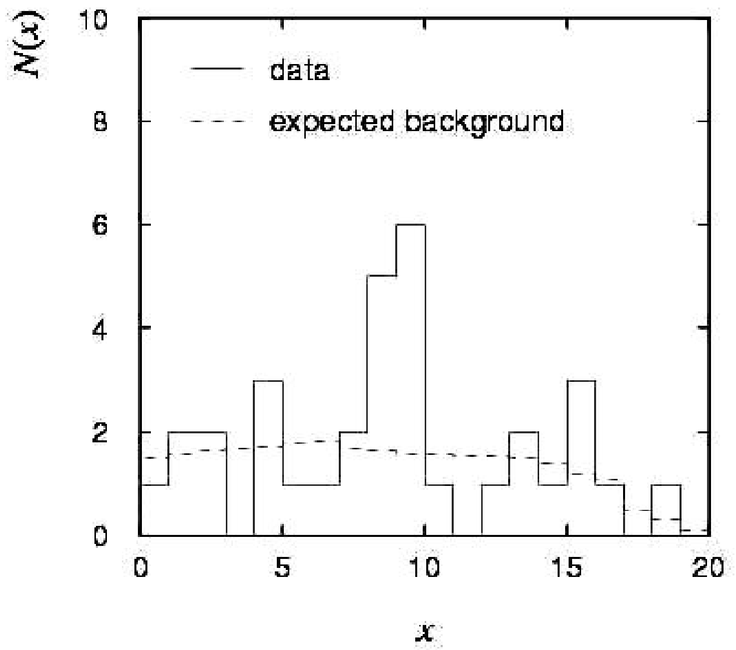
$\Rightarrow$ zliczanie przypadków

$\Rightarrow$ dopasowanie rozkładów

Wprowadzenie

The significance of a peak

Suppose we measure a value x for each event and find:



Each bin (observed) is a Poisson r.v., means are given by dashed lines.

In the two bins with the peak, 11 entries found with $b = 3.2$. The p-value for the $s = 0$ hypothesis is:

$$P(n \geq 11; b = 3.2, s = 0) = 5.0 \times 10^{-4}$$

The significance of a peak (2)

But... did we know where to look for the peak?

→ give $P(n \geq 11)$ in any 2 adjacent bins

Is the observed width consistent with the expected x resolution?

→ take x window several times the expected resolution

How many bins × distributions have we looked at?

→ look at a thousand of them, you'll find a 10^{-3} effect

Did we adjust the cuts to 'enhance' the peak?

→ freeze cuts, repeat analysis with new data

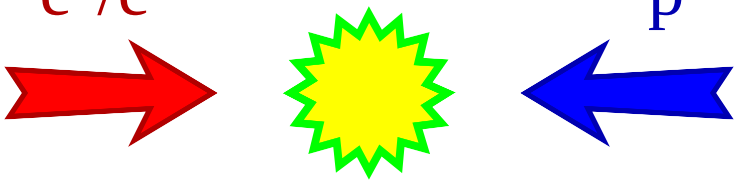
Should we publish????

Introduction

HERA

electron(positron)-proton collider at DESY

e^+/e^- p



27.5 GeV

820 GeV 1994-97

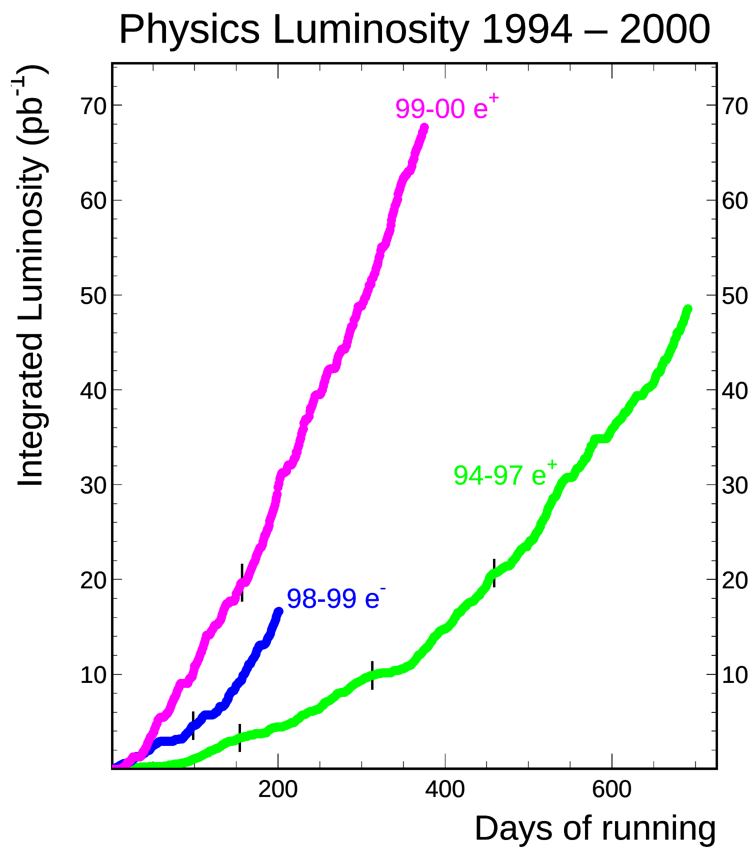
920 GeV 1998-00

Presented results

Year		$\sqrt{s}$	H1	ZEUS
1994-97	e^+p	300 GeV	36 pb^{-1}	48 pb^{-1}
1998-99	e^-p	318 GeV	15 pb^{-1}	16 pb^{-1}
1999-00	e^+p	318 GeV	46 pb^{-1}	64 pb^{-1}

Total of about $\sim$ 220 pb^{-1} of $e^\pm p$ data available.

Very successful HERA operation in 1999-2000:



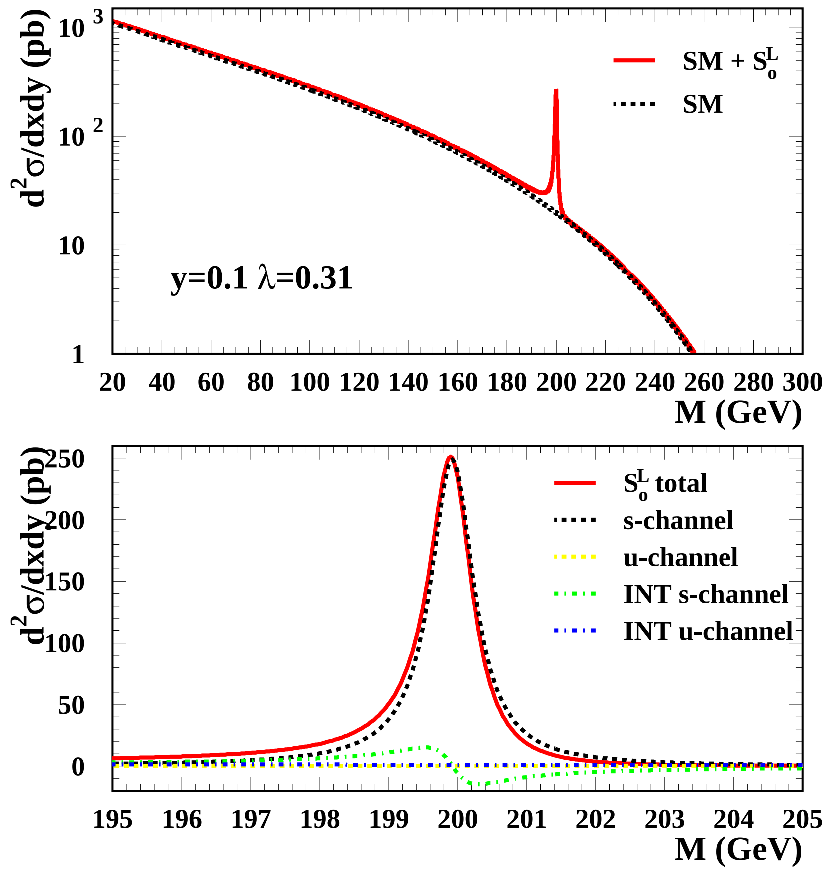
Leptokwarki - model

LEPTOKWARKI:

Cząstki fundamentalne sprzęgające się do leptonów i kwarków, które:

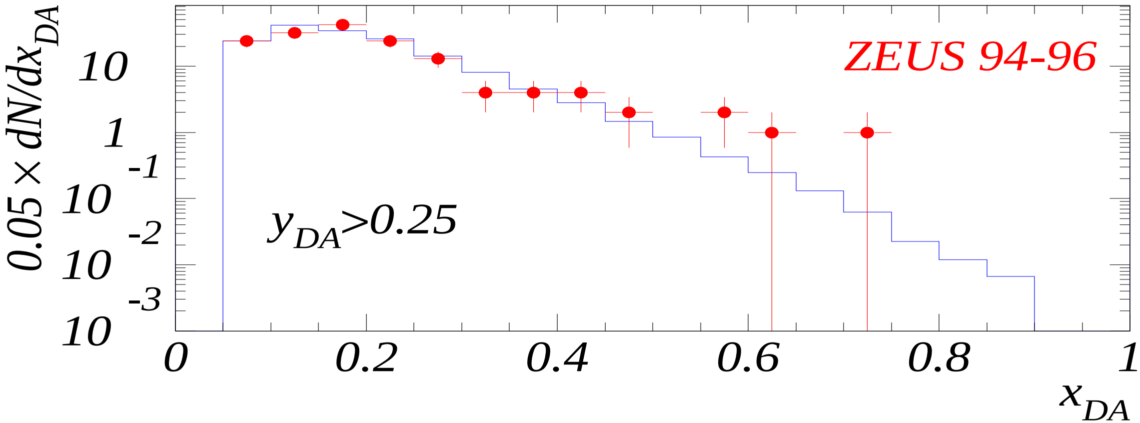
- niosą liczbę barionową, leptonową, kolor i **ułamkowy** ładunek
- w ramach ogólnego modelu Buchmüller'a-Rückl'a-Wyler'a (BRW) dopuszcza się istnienie 14 różnych typów LQ
- mogą być produkowane w parach w zderzeniach e^+e^- i $p\bar{p}$
- możliwa produkcja pojedyńczych LQ w zderzeniach $e^{\pm}p$:

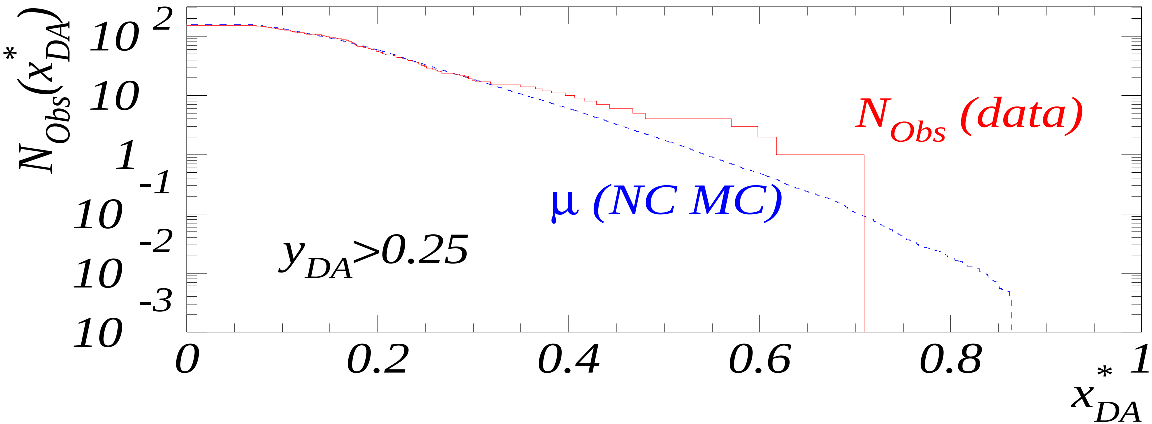
Cząstki takie widzielibyśmy w danych jako rezonanse w rozkładzie masy niezmienniczej $e^{\pm}$-dżet (NC) i ν-dżet (CC).

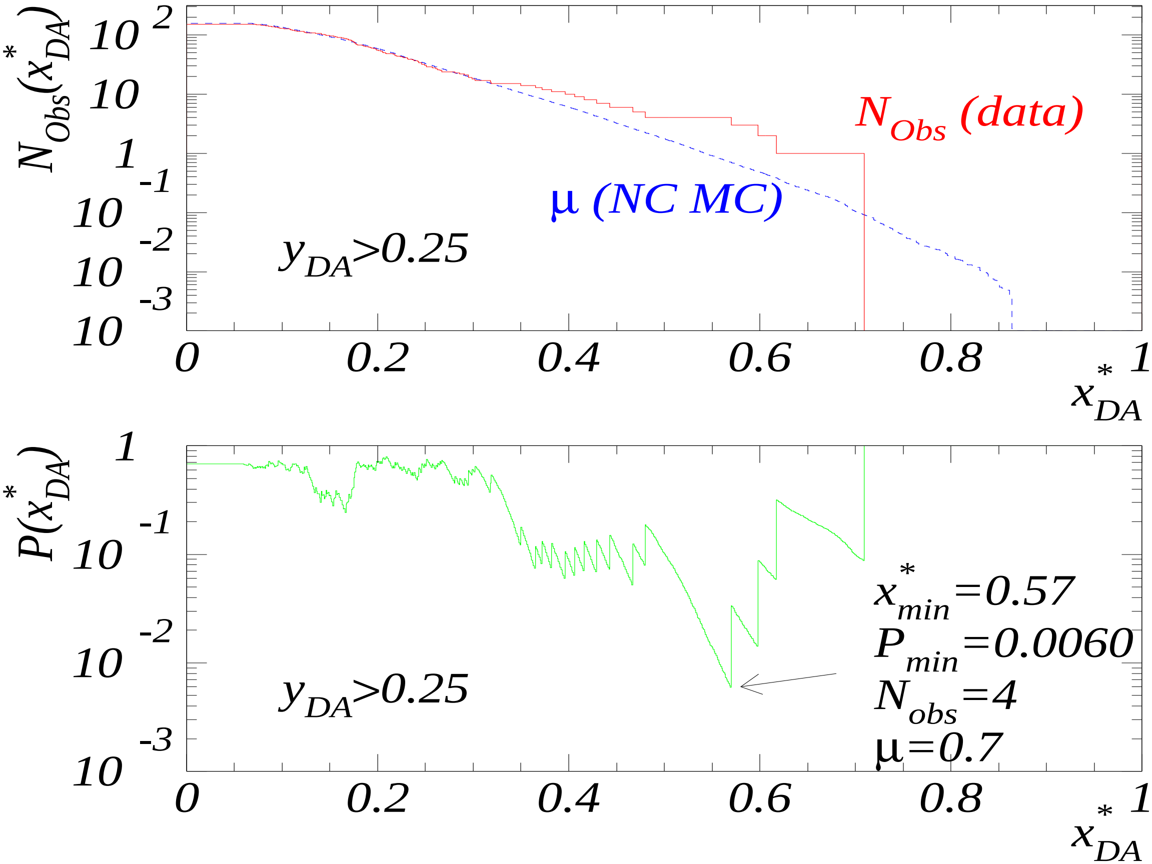


Significance Analysis

Excess in x







$$N_{\rm obs}(x^*_{\rm DA}) \; = \; \int_{x^*_{\rm DA}}^{1} \; dx_{\rm DA} \; dN/dx_{\rm DA}$$

$$\mathcal{P}(x^*_{\rm DA}) \; = \; \sum_{n=N_{obs}}^{\infty} \; e^{-\mu} \, \frac{\mu^n}{n!}$$

Poszukiwanie "nowej fizyki"

Dwa etapy analizy:

* selekcja przypadków

  Podobnie jak w przypadku pomiarów przekroju czynnego
  Naogół optymalizujemy "znaczoność" sygnału
  Optymalna selekcja czuła na poziom sygnału, tło, błędy systematyczne...

* analiza statystyczna

  Na podstawie wybranych przypadków (ich liczby lub rozkładów w odpowiednich zmiennych) musimy zdecydować czy przyjąć czy odrzucić daną hipotezę

Suppose the result of a measurement for an individual event is a collection of numbers $\vec{x} = (x_1, \ldots, x_n)$

x_1 = number of muons,

x_2 = jet p_t of jets,

x_3 = missing energy, ...

$\vec{x}$ follows some n-dimensional joint pdf, which depends on the type of event produced, i.e., was it

$$\mathrm{pp} \to t\bar{t} \;, \qquad \mathrm{pp} \to \tilde{g}\tilde{g} \;, \ldots$$

For each reaction we consider we will have a hypothesis for the pdf of $\vec{x}$, e.g., $f(\vec{x}|H_0)$, $f(\vec{x}|H_1)$, etc.

Often H_0 is the Standard Model, (the background hypothesis), H_1 ... is a signal hypothesis we are searching for

Selecting events

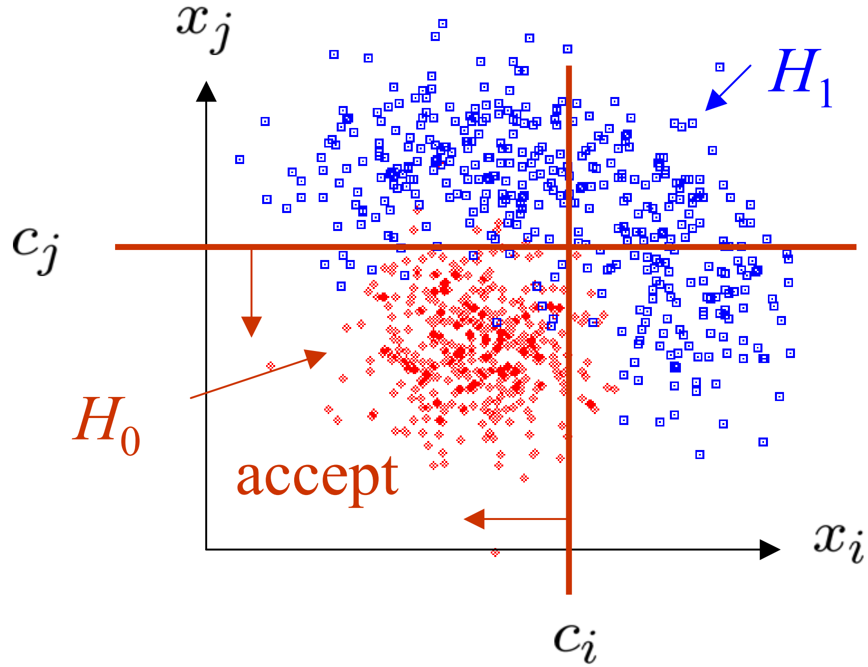
Suppose we have a data sample with two kinds of events, corresponding to hypotheses H_0 and H_1 and we want to select those of type H_0.

Each event is a point in $\vec{x}$ space. What decision boundary should we use to accept/reject events as belonging to event type H_0?

Probably start with cuts:

$$x_i < c_i$$

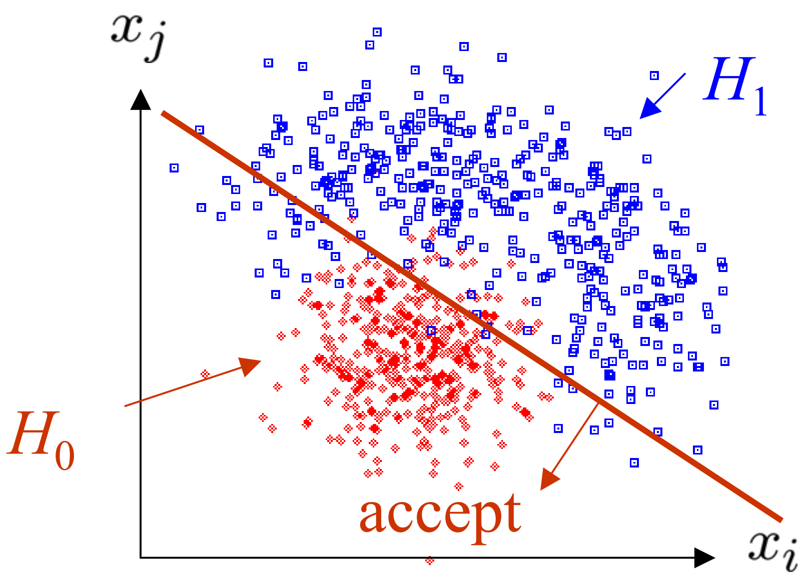
$$x_j < c_j$$



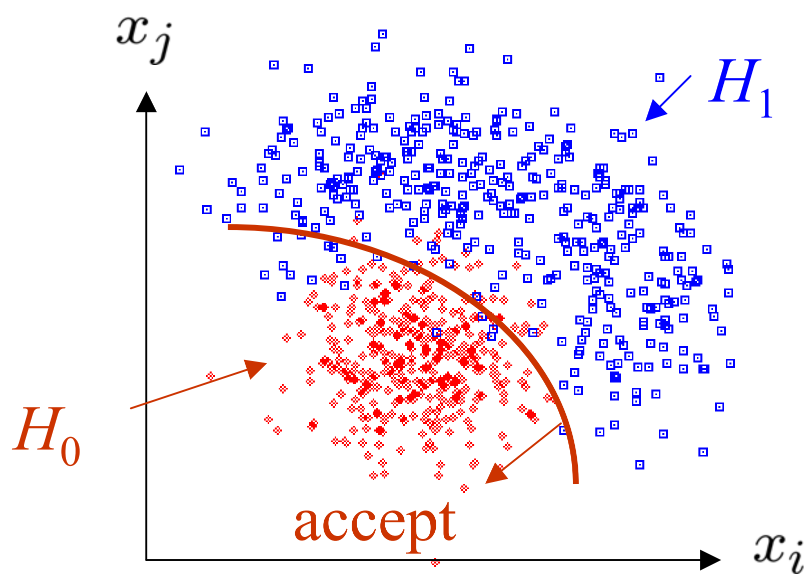
Other ways to select events

Or maybe use some other sort of decision boundary:

linear



or nonlinear



How can we do this in an ‘optimal’ way?

Test statistics

Construct a 'test statistic' of lower dimension (e.g. scalar)

$$t(x_1, \ldots, x_n)$$

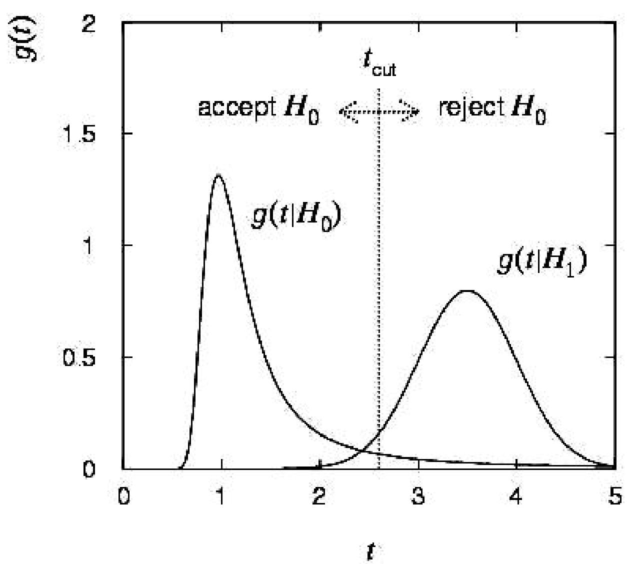
Goal is to compactify data without losing ability to discriminate between hypotheses.

We can work out the pdfs $g(t|H_0),\ g(t|H_1),\ \ldots$

Decision boundary is now a single cut on *t*.

This effectively divides the sample space into two regions where we either:

accept H_0 (acceptance region)
or reject it (critical region).



Constructing a test statistic

How can we select events in an 'optimal way'?

Neyman-Pearson lemma states:

To get the lowest ε_b for a given ε_s (highest power for a given significance level), choose acceptance region such that

$$\frac{f(\vec{x}|\mathsf{s})}{f(\vec{x}|\mathsf{b})} > c$$

where c is a constant which determines ε_s.

Equivalently, optimal scalar test statistic is $\boxed{t(\vec{x}) = \frac{f(\vec{x}|\mathsf{s})}{f(\vec{x}|\mathsf{b})}}$

N.B. any monotonic function of this is just as good.

Significance level and power of a test

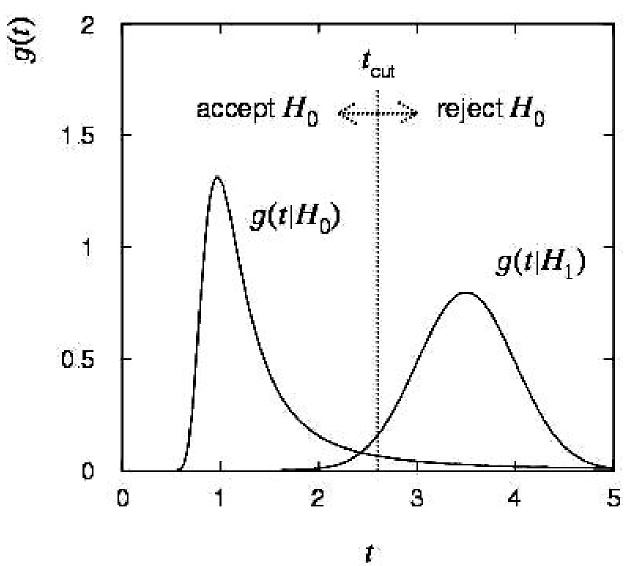
Probability to reject H_0 if it is true (error of the 1st kind):

$$\alpha = P(\text{reject } H_0 \,|\, H_0) = \int_{t_{\text{cut}}}^{\infty} g(t|H_0)\,dt \qquad \text{(significance level)}$$

Probability to accept H_0 if H_1 is true (error of the 2nd kind):

$$\begin{aligned}\beta &= P(\text{accept } H_0 \,|\, H_1) \\ &= \int_{-\infty}^{t_{\text{cut}}} g(t|H_1)\,dt\end{aligned}$$

$(1 - \beta = \text{power})$



Poszukiwanie "nowej fizyki"

Analiza statystyczna:

* gdy widać sygnał

  Staramy się dopasować parametry modelu tak aby uzyskać najlepszy opis danych. Następnie oceniamy na ile znaczący jest to sygnał...

* nie widać sygnału

  Szukamy w przestrzeni parametrów modelu tych obszarów, które powinny były dać widoczny sygnał. Te obszary możemy wykluczyć...

Probability, Probability Density, and Likelihood

- **Poisson *probability* $P(n|\mu) = \mu^n \exp(-\mu)/n!$**
- **Gaussian *probability density function* (pdf) $p(x|\mu,\sigma)$: $p(x|\mu,\sigma)dx$ is differential of probability dP.**
- **In Poisson case, suppose n=3 is observed. Substituting n=3 into $P(n|\mu)$ yields the *likelihood function* $\mathcal{L}(\mu) = \mu^3 \exp(-\mu)/3!$**
  - **Key point is that $\mathcal{L}(\mu)$ is *not* a probability density in μ. (It is not a density!) Area under $\mathcal{L}$ is meaningless. That's why a new word, "likelihood", was invented for this function of the parameter(s), to distinguish from a pdf in the observable(s)!**
  - **Likelihood *Ratios* $\mathcal{L}(\mu_1) / \mathcal{L}(\mu_2)$ are useful and frequently used.**

Likelihood

We have data: x (could be a vector, discrete or continuous) and a probability model: $P(x;\theta)$ (θ could be vector of parameters)

Now evaluate the probability function using the data that we observed and treat it as a function of the parameters. This is the likelihood function:

$$L(\theta) = P(x;\theta) \qquad \text{(here } x \text{ is constant)}$$

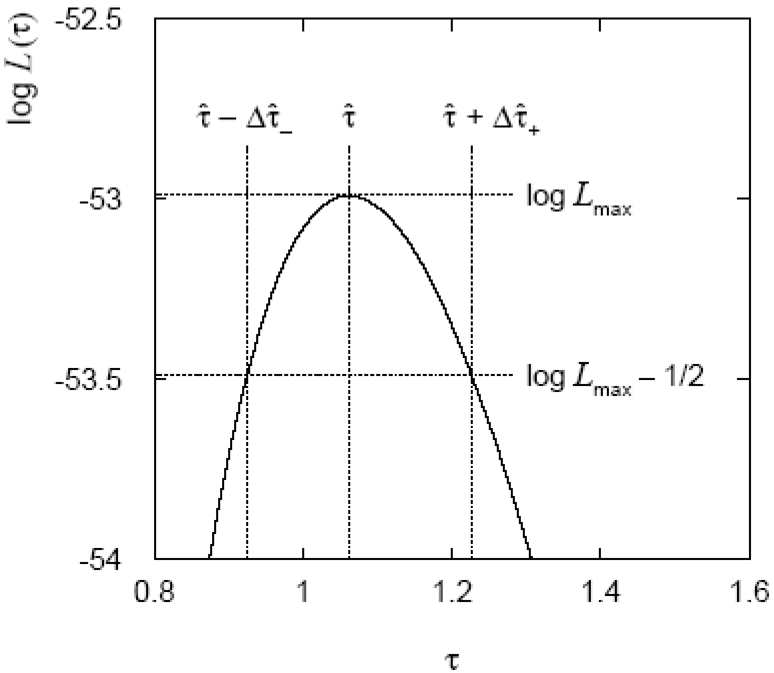
For example, if we have n independent observations of a random variable x, where $x \sim f(x;\theta)$, then

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

Maximum Likelihood

The likelihood function plays an important role in both frequentist and Bayesian statistics.

E.g., to estimate the parameter θ, the method of maximum likelihood (ML) says to take the value that maximizes $L(\theta)$.



ML and other parameter estimation methods would be a large part of a longer course on statistics — for now need to move on...

Definition of “Probability”

- **Abstract mathematical probability P can be defined in terms of sets and axioms that P obeys. If the axioms are true for P, then P obeys Bayes’ Theorem (see next slide)**

  P(B|A) = P(A|B) P(B) / P(A).

- **Two established* incarnations of P are:**

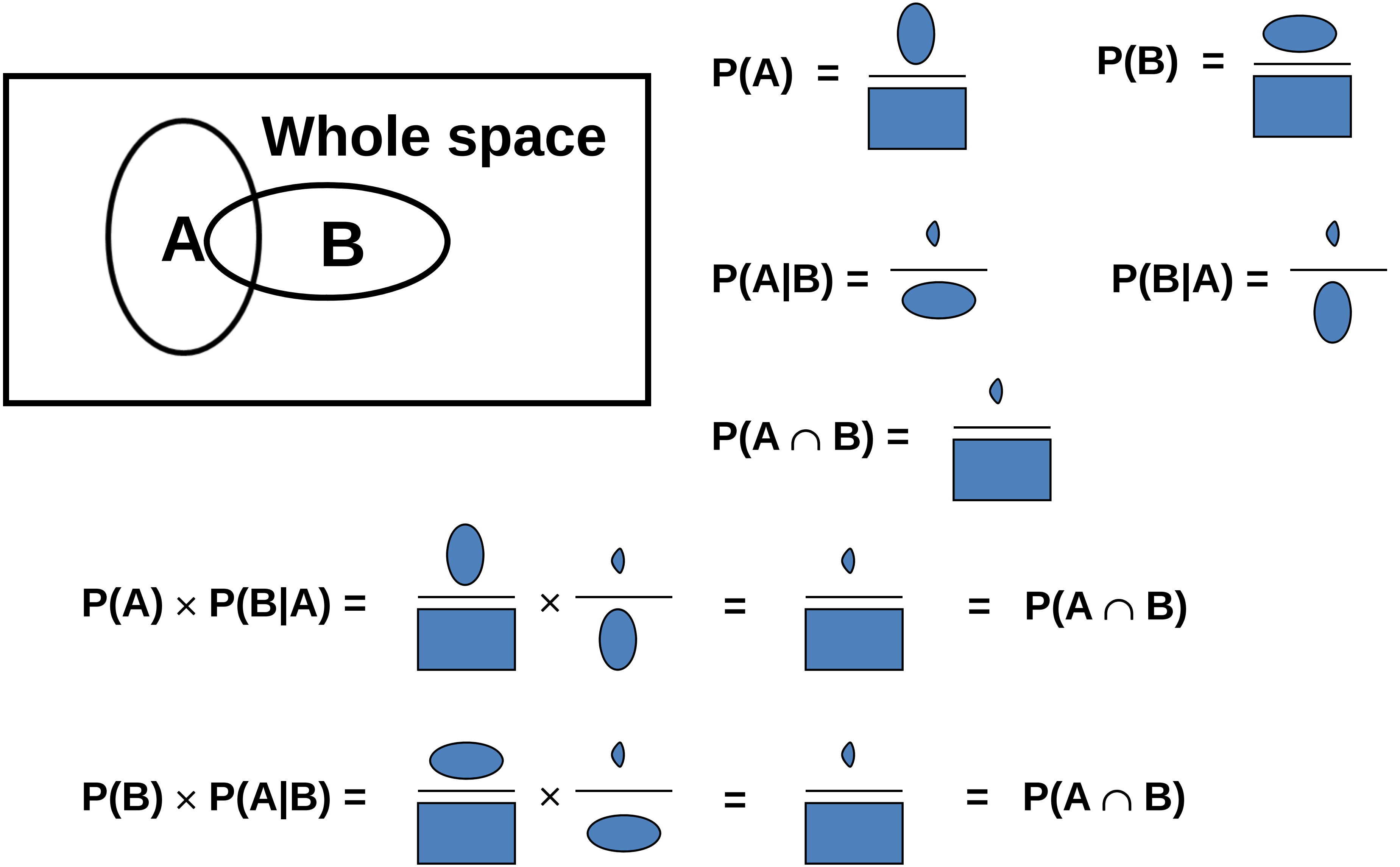
  ***1) Frequentist P*: limiting frequency in ensemble of imagined repeated samples (as usually taught in Q.M.).**
  P(constant of nature) and P(SUSY is true) do not exist (in a useful way) for this definition of P (at least in one universe).

  ***2) (Subjective) Bayesian P*: subjective (personalistic) *degree of belief.* (de Finetti, Savage)**
  P(constant of nature) and P(SUSY is true) exist for You.
  Shown to be basis for coherent personal decision-making.

- ***It is important to be able to work with either definition of P, and to know which one you are using!***

***Of course they are still argued about, but to less practical effect, I think.**

P, Conditional P, and Derivation of Bayes' Theorem in Pictures



⇒ P(B|A) = P(A|B) × P(B) / P(A)

Frequentist Statistics − general philosophy

In frequentist statistics, probabilities are associated only with the data, i.e., outcomes of repeatable observations.

Probability = limiting frequency

Probabilities such as

P (Higgs boson exists),
$P\,(0.117 < \alpha_s < 0.121)$,

etc. are either 0 or 1, but we don't know which.

The tools of frequentist statistics tell us what to expect, under the assumption of certain probabilities, about hypothetical repeated observations.

The preferred theories (models, hypotheses, ...) are those for which our observations would be considered 'usual'.

Bayesian Statistics – general philosophy

In Bayesian statistics, interpretation of probability extended to degree of belief (subjective probability). Use this for hypotheses:

probability of the data assuming hypothesis *H* (the likelihood)

prior probability, i.e., before seeing the data

$$P(H|\vec{x}) = \frac{P(\vec{x}|H)\pi(H)}{\int P(\vec{x}|H)\pi(H)\,dH}$$

posterior probability, i.e., after seeing the data

normalization involves sum over all possible hypotheses

Bayesian methods can provide more natural treatment of non-repeatable phenomena:

systematic uncertainties, probability that Higgs boson exists,...

No golden rule for priors ("if-then" character of Bayes' thm.)

Example: Particles entering a threshold Cerenkov can be e, π or K,

$$P(e) = 1\% \qquad P(\pi) = 70\% \qquad P(K) = 29\%$$

The probabilities that the detector fires (*efficiencies*) are

$$P(C \mid e) = 99\% \qquad P(C \mid \pi) = 2\% \qquad P(C \mid K) = 1\%$$

If a particle fired the detector, what's the probability that it's an e?

$$P(e \mid C) = \frac{P(C \mid e)P(e)}{P(C|e)P(e) + P(C|\pi)P(\pi) + P(C|K)P(K)}$$

$$= \frac{0.99 \times 0.01}{0.99 \times 0.01 \; + \; 0.02 \times 0.70 \; + \; 0.01 \times 0.29} = 37\%$$

Notice that is is a rather selective detector,

yet 63% of signals will be background (π and K).

- To invert probabilities, $P(A\,|\,B) \to P(B\,|\,A)$, need $P(B)$

  $P(C\,|\,e) \to P(e\,|\,C)$, need $P(e)$

- $P(A\,|\,B) \neq P(B\,|\,A)$

  $P(C\,|\,e) \neq P(e\,|\,C)$

Or, with a real life example:

A = female or male	$P(\text{pregnant}\,|\,\text{female}) \approx 0.5\%$
B = pregnant or non-pregnant	$P(\text{female}\,|\,\text{pregnant}) \ggg 1\%$

Example of Bayes' Theorem Using Bayesian P

In a background-free experiment, a theorist uses a "model" to predict a signal with Poisson mean of 3 events. From Poisson formula we know

$P(0 \text{ events} \mid \text{model true}) = 3^0 e^{-3}/0! = 0.05$

$P(0 \text{ events} \mid \text{model false}) = 1.0$

$P(>0 \text{ events} \mid \text{model true}) = 0.95$

$P(>0 \text{ events} \mid \text{model false}) = 0.0$

The experiment is performed and zero events are observed.

Question: Given the result of the expt, what is the probability that the model is true? I.e., What is P(model true | 0 events) ?

Example of Bayes' Theorem Using Bayesian P

In a background-free experiment, a theorist uses a "model" to predict a signal with Poisson mean of 3 events. From Poisson formula we know

P(0 events | model true) = $3^0 e^{-3}/0!$ = 0.05

P(0 events | model false) = 1.0

P(>0 events | model true) = 0.95

P(>0 events | model false) = 0.0

The experiment is performed and zero events are observed.

Question: Given the result of the expt, what is the probability that the model is true? I.e., What is P(model true | 0 events) ?

Answer: *Cannot be determined from the given information!* Need in addition: P(model true), the *degree of belief* in the model *prior* to the experiment. Then Bayes' Thm inverts the conditionality:

P(model true | 0 events) $\propto$ P(0 events | model true) P(model true)

If "model" is S.M., then still very high degree of belief after experiment! (Compare with news releases that would say "there is 5% chance the S.M. is true.")

If "model" is large extra dimensions, then low prior belief becomes even lower.

N.B. Of course this example is over-simplified; it gets more interesting when there is more than one model which predicts the signal-type events. Also when an event is seen; and when normalization factor is included.

Dwa podejścia do wyznaczania limitów:

Bayesowskie

próbujemy "zrekonstruować" rozkład prawdopodobieństwa dla parametru modelu. Traktujemy ten parametr jak zmienną losową.

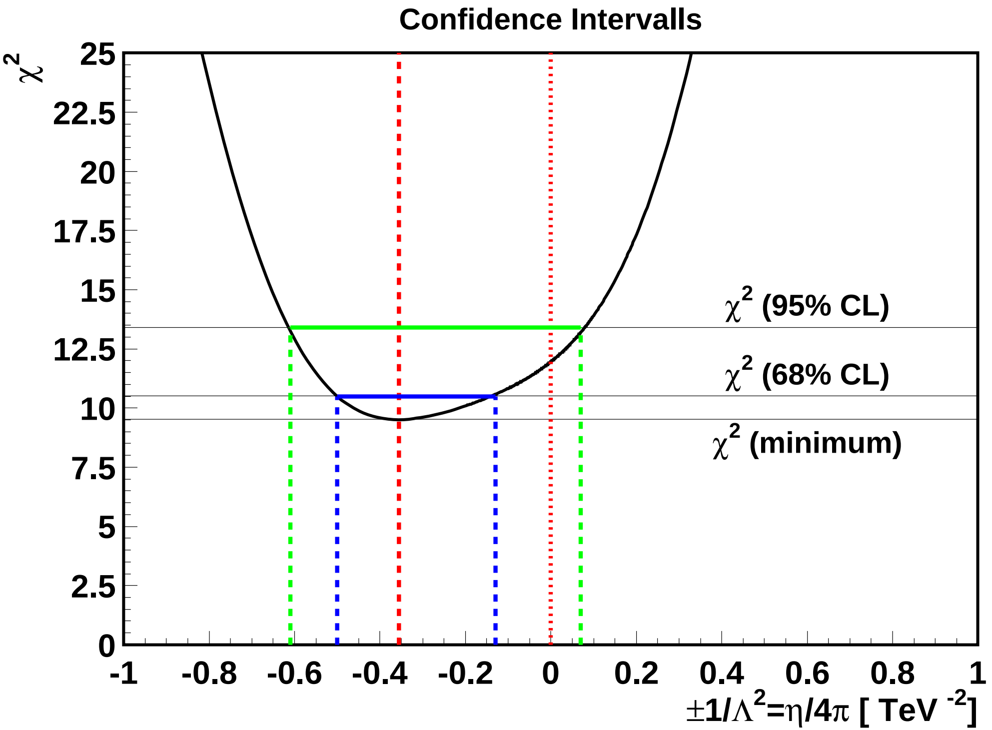
Z tego rozkładu liczymi limit tak jak dla zwykłago rozkładu prawdopodobieństwa zmiennej losowej (np. rozkładu Gaussa).
Np. wykluczone są wartości parametru x większe niż X_{lim} jeśli

$$P(x>X_{lim}) = 1 - CL$$

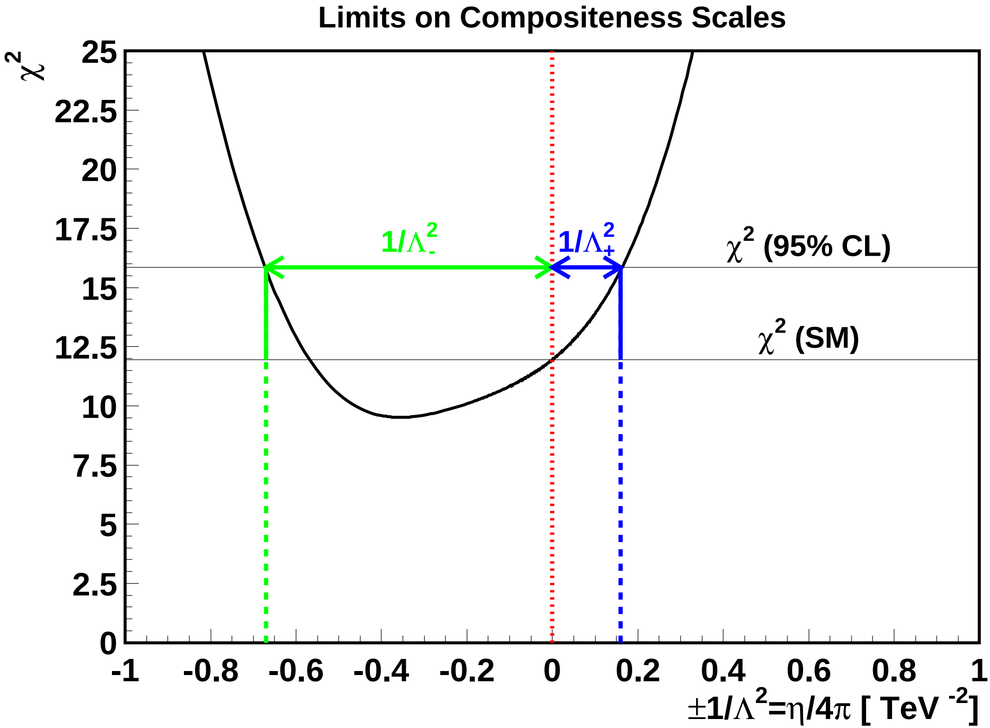
CL - confidence level, naogół CL=0.95

- \+ proste i intuicyjne
- \+ nie wymaga czasochłonnych obliczeń
- \- zależy od wyboru 'prior distribution'
- \- zależy od wyboru parametru modelu

Extraction of Limits

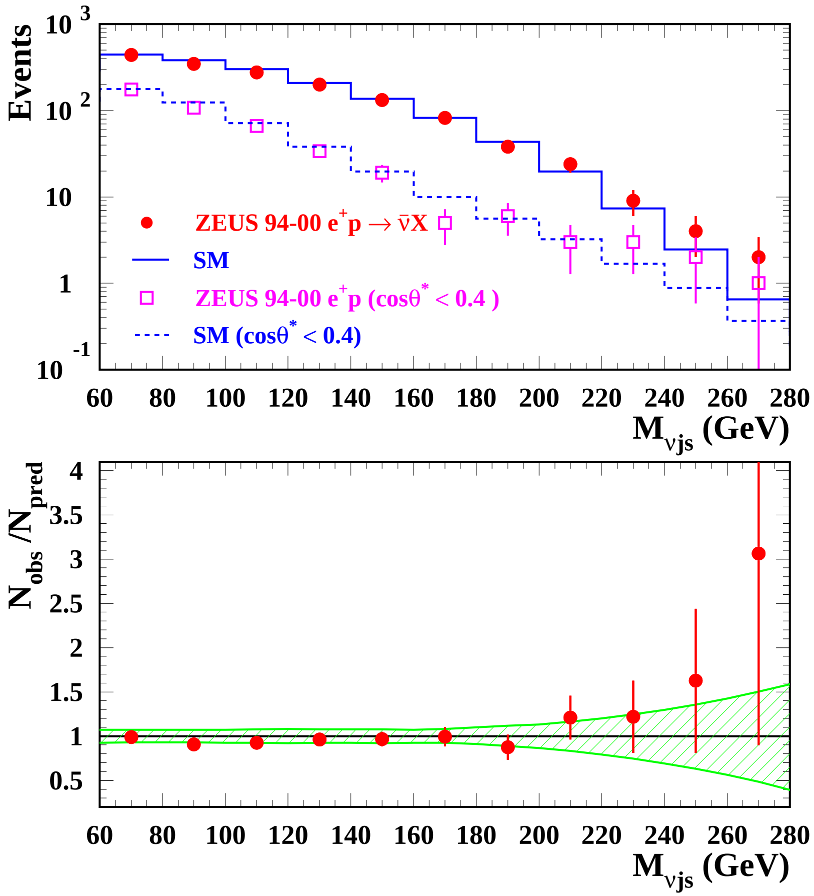


Likelihood ratio

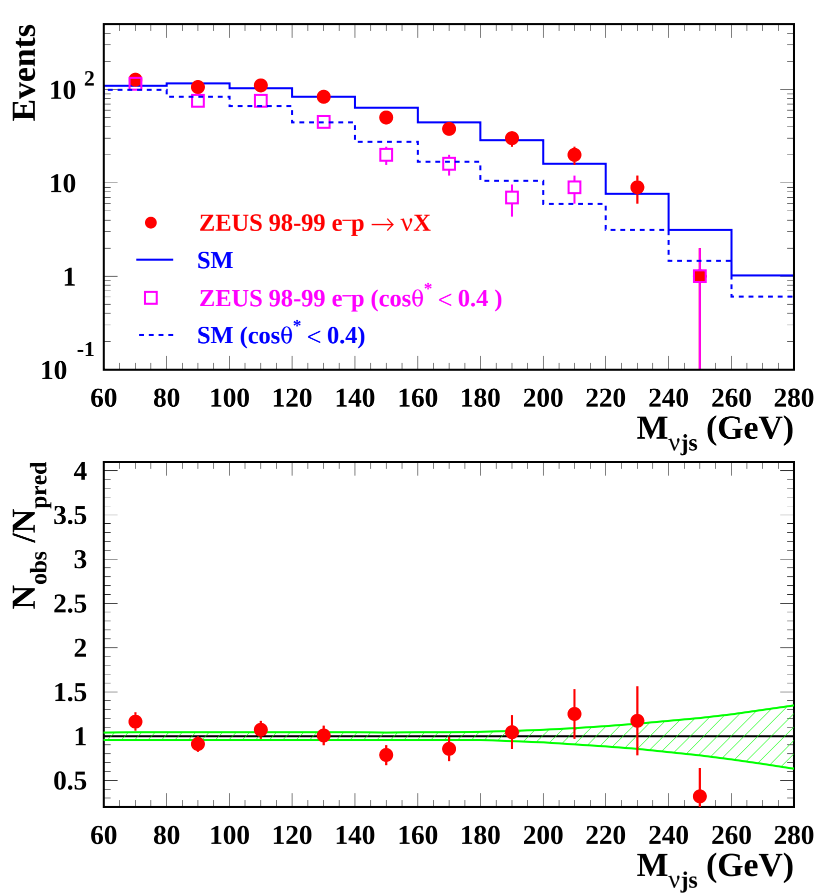


Rozkład masy niezmienniczej νq i porównanie z SM

Dane e^+p:



Dane e^-p:



Dobra zgodność danych z przewidywaniami Modelu Standardowego → brak sygnału LQ

Using shape of a distribution in a search

Suppose we want to search for a specific model (i.e. beyond the Standard Model); contains parameter θ.

Select candidate events; for each event measure some quantity x and make histogram: $\vec{n} = (n_1, \ldots, n_M)$

Expected number of entries in ith bin: $E[n_i] = \underbrace{s_i(\theta)}_{\text{signal}} + \underbrace{b_i}_{\text{background}}$

Suppose the 'no signal' hypothesis is $\theta = \theta_0$, i.e., $s(\theta_0) = 0$.

Probability is product of Poisson probabilities:

$$P(\vec{n}|\theta) = \prod_{i=1}^{M} \frac{(s_i(\theta) + b_i)^{n_i}}{n_i!} e^{-(s_i(\theta)+b_i)}$$

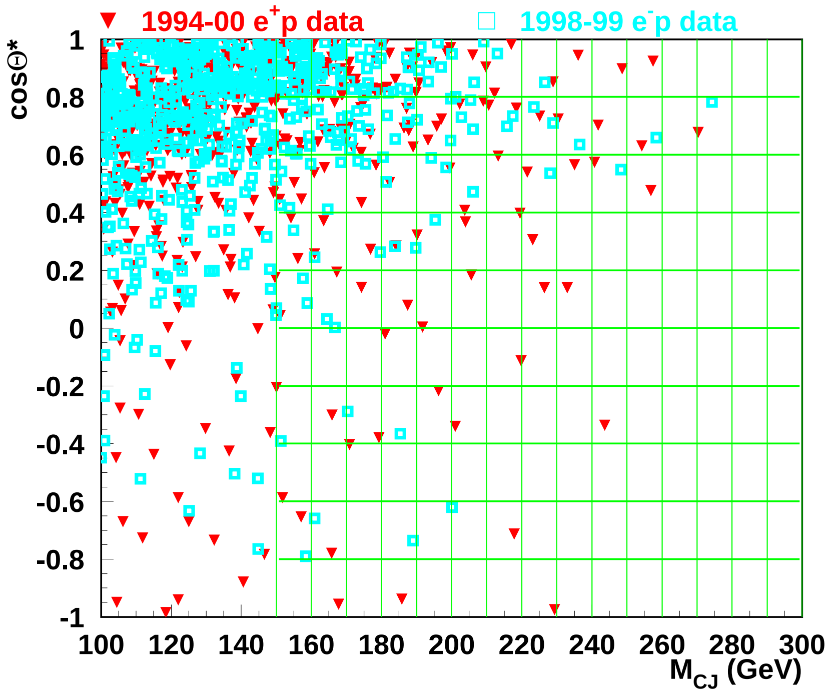


Figure 5.6: Distribution of selected NC DIS type events in the M_{ejs} - $\cos\theta^{\star}_{ejs}$ plane, for the e^-p and the e^+p data. The grid indicates bins used in the likelihood analysis.

L_i is the function of N_i and μ_i, thus also M_{LQ} and λ_{LQ}. The two dimensional likelihood L is the product of Poisson probabilities over all considered $\cos\theta^*$–M_{ljs} bins:

$$L(M_{LQ}, \lambda_{LQ}) = \prod_i L_i = \prod_i e^{(-\mu_i)} \frac{{\mu_i}^{N_i}}{N_i!}. \tag{5.6}$$

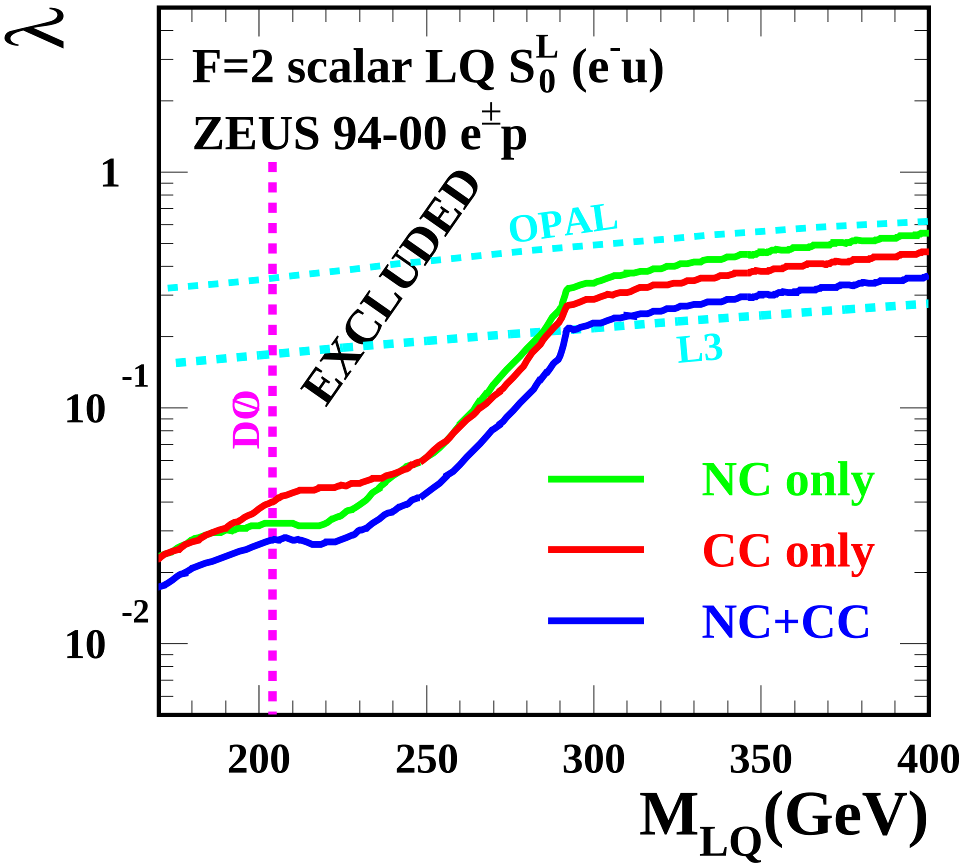
In this analysis we adopted the Bayesian approach and the upper limit on the coupling strength as a function of M_{LQ}, $\lambda_{limit}(M_{LQ})$, was obtained by solving the equation[1]

$$\int_0^{\lambda^2_{\rm limit}} d\lambda^2 L(M_{LQ}, \lambda) = 0.95 \int_0^{\infty} d\lambda^2 L(M_{LQ}, \lambda). \tag{5.7}$$

The confidence level of the limit calculated with this method is not exactly equal, but is expected to be close to 95%. This assumption was verified using the so called Monte Carlo Experiments method. More details can be found in Appendix E.

Leptokwarki - wyniki

Brak widocznego sygnału LQ $\Rightarrow$ Granice na sprzężenie Yukawy λ w funkcji M_{LQ}:



Łączna analiza NC $+$ CC

$\Rightarrow$ silniejsze granice

Dla $\lambda = \sqrt{4\pi\alpha} \approx 0.3$

górne granice na M_{LQ}

wynoszą od 274 do 400 GeV

Dla $M_{LQ} \gg \sqrt{s}$ dolne granice

na stosunek M_{LQ}/λ_{LQ}

wynosi od 0.27 TeV do 1.26 TeV

Podejście klasyczne

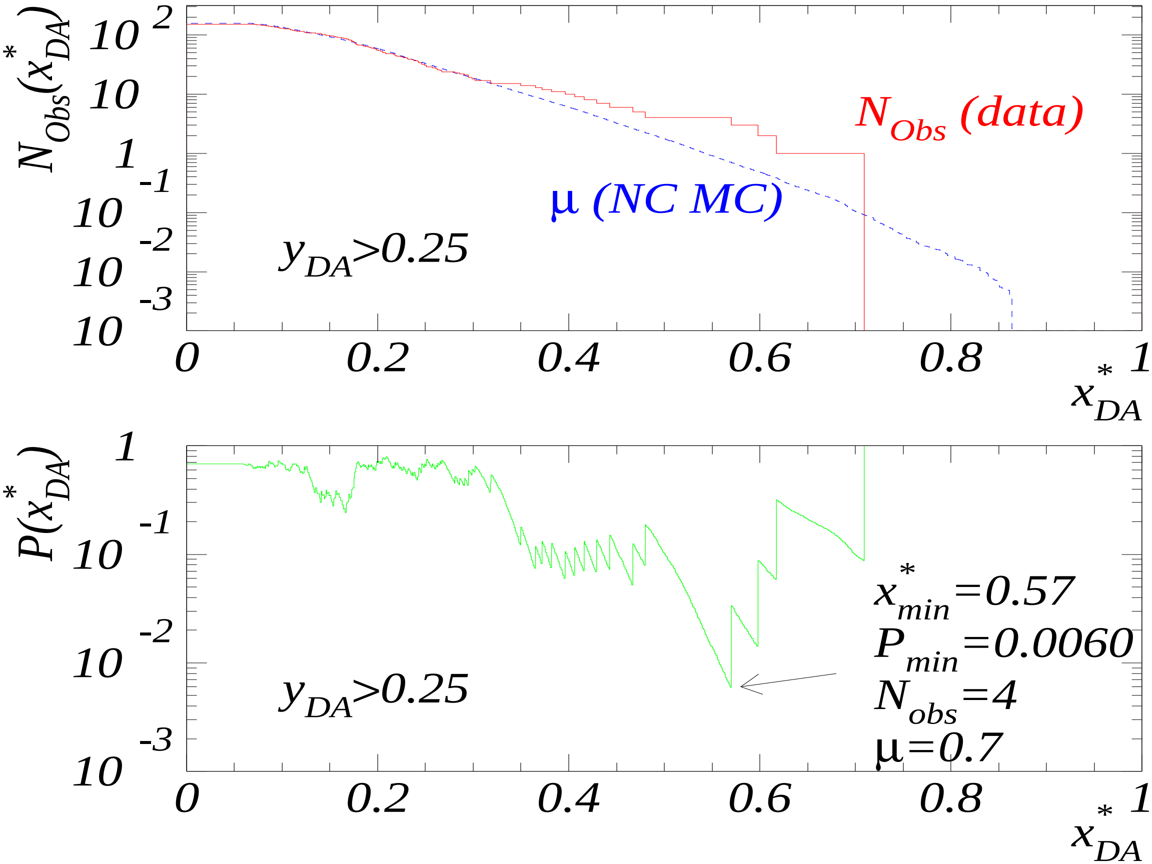
Parametr modelu nie jest zmienną losową!

O przyjęciu (lub odrzuceniu) danego modelu "nowej fizyki" decyduje prawdopodobieństwo, że (przy powtórzeniu pomiaru) dałby on wynik y lepiej (gorzej) zgodny z przewidywaniami SM niż jest to obserwowane w rzeczywiście zebranych danych - y_{data}.

Naogół wykluczamy modele, które dają gorszą niż obserwowana zgodność z SM w 95% przypadków:

$$P(y > y_{data} \mid X_{lim}) = 1 - CL$$

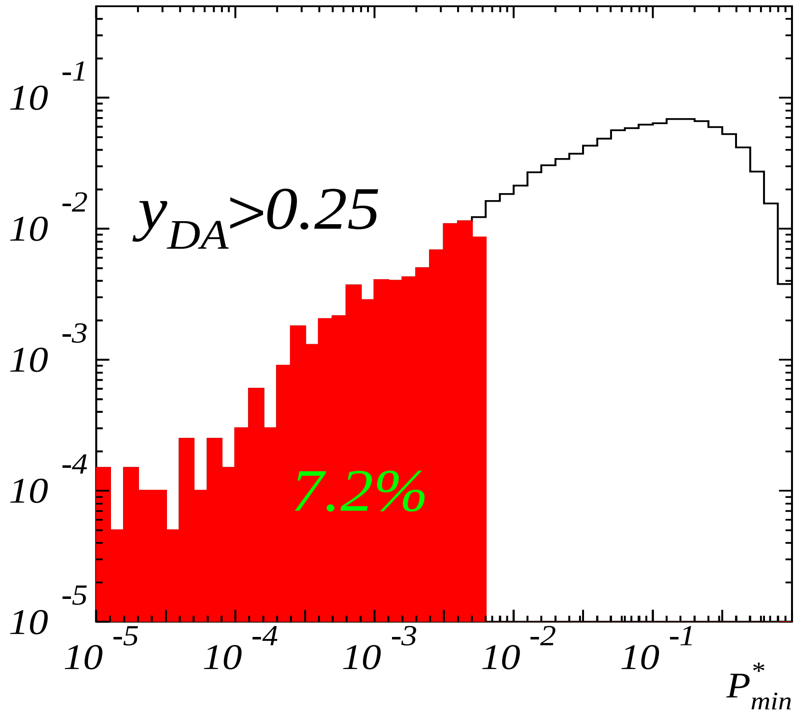
+ jednoznaczna definicja, nie potrzebujemy żadnych 'prior'
+ ścisła interpretacja probabilistyczna
+ nie zależy od wyboru parametru modelu
- wymaga wyboru sposobu oceny zgodności - y (niejednoznaczność)
- wymaga czasochłonnych obliczeń
  (symulacji MC wielu powtórzeń eksperymentu)



$$N_{\rm obs}(x^*_{\rm DA}) \; = \; \int\limits_{x^*_{\rm DA}}^{1} \; dx_{\rm DA} \; dN/dx_{\rm DA}$$

$$\mathcal{P}(x^*_{\rm DA}) \; = \; \sum_{n=N_{obs}}^{\infty} \; e^{-\mu} \, \frac{\mu^n}{n!}$$

Excess in x — continued



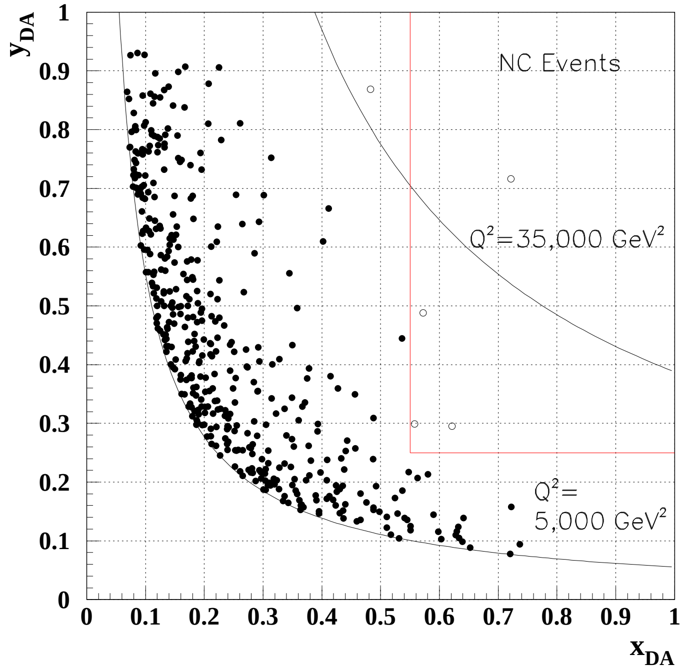
Minimal Poisson probabilities of the x_{DA} distributions for different y_{DA} cuts

y_{DA} range	$\mathcal{P}_{min}(x^*_{\mathrm{DA}})$ [%]	x^*_{DA}	N_{obs}	μ	P_{SM} [%]
$y_{\mathrm{DA}} > 0.05$	1.61	0.708	4	0.95	16.0
$y_{\mathrm{DA}} > 0.15$	2.57	0.708	2	0.25	23.0
$y_{\mathrm{DA}} > 0.25$	0.60	0.569	4	0.71	7.2

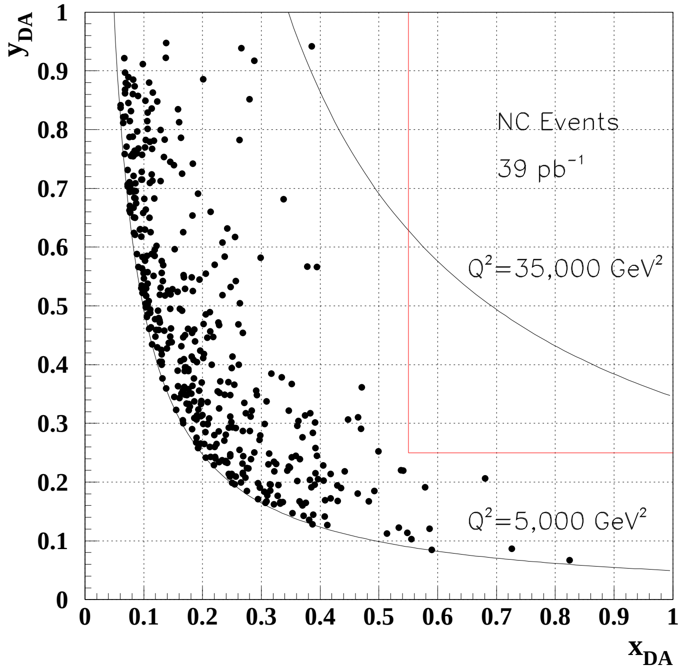
High-Q^2 DIS

High-Q^2 / high-x excess ?

ZEUS 1994-1997 e^+ preliminary



ZEUS 1999-2000 e^+ preliminary 39 pb^{-1}



Excess observed in old data not confirmed

ZEUS

N_{obs} (N_{exp})	1994-97	1998-99	1999-00(part)
$Q^2 > 35\,000$ GeV^2	2 (0.34)	2 (1.02)	1 (0.53)
$x > 0.55$ & $y > 0.25$	4 (1.9)	1 (1.3)	0 (1.6)

1994-96
2 (0.15)
4 (0.91)

Przedział ufności (Neyman, 1937)

Recepta:

- zmienna losowa x o rozkładzie $f_x(x|\theta)$,
- trzy **zadane** liczby: $0 < \alpha < 1$ oraz $0 < \beta < \gamma < 1$, takie że:

$$\gamma - \beta = 1 - \alpha,$$

- przy ustalonej wartości θ, wynik x_0 pomiaru ustanawiamy kwantylami x_β i x_γ rzędu β i rzędu γ rozkładu $f(x|\theta)$:

$$P(x \le x_0 \mid \theta) = \beta, \quad P(x \le x_0 \mid \theta) = \gamma$$

(kwantyl x_p rzędu p to taka wartość x_p zmiennej x, że $P(x \le x_p) = p$)

- rozwiązania obu równań względem θ, wyznaczaja **przedział ufności** $[\theta_{\min}, \theta_{\max}]$ na poziomie ufności $1 - \alpha$

$$1 - \alpha = P\left(x_\beta \le x \le x_\gamma \mid \theta\right) = P\left(\theta_{\min} \le \theta \le \theta_{\max} \mid x_0\right) \ (???)$$

Konstrukcja Neymana

Przy ustalonej wartości θ, wyznaczamy kwantyle x_β i x_γ:

$$P\left(x \le x_\beta \mid \theta\right) = \beta, \quad P\left(x \le x_\gamma \mid \theta\right) = \gamma$$

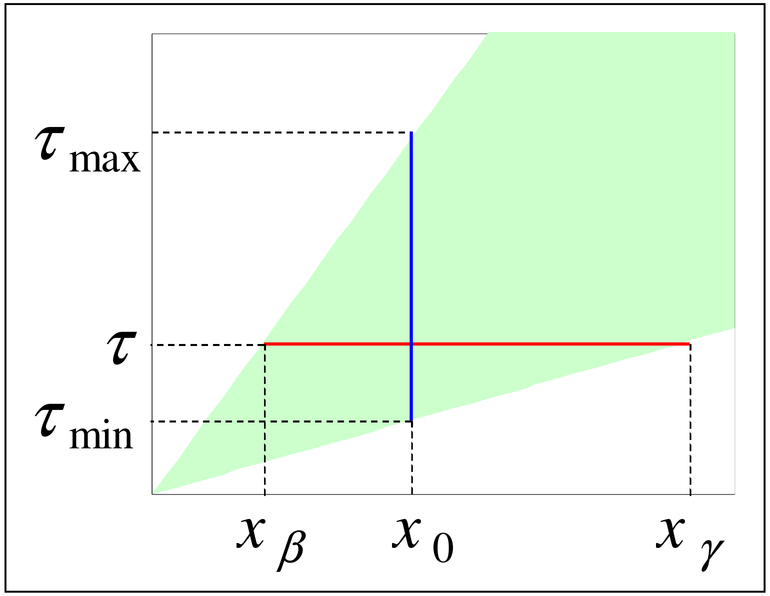
Przykład – rozkład wykładniczy:

$$P\left(x \le x_\beta \mid \tau\right) = \frac{1}{\tau}\int_0^{x_\beta} \exp\left(-\frac{x}{\tau}\right) dx = \beta$$

$$x_\beta = -\tau \ln\left(1-\beta\right)$$

$$P\left(x \le x_\gamma \mid \tau\right) = \frac{1}{\tau}\int_0^{x_\gamma} \exp\left(-\frac{x}{\tau}\right) dx = \gamma$$

$$x_\gamma = -\tau \ln\left(1-\gamma\right)$$



Wykres konstruujemy „w poziomie”, przedziały ufności odczytujemy „w pionie”.

Consider a p.d.f. $f(x;\theta)$ where x represents the outcome of the experiment and θ is the unknown parameter for which we want to construct a confidence interval. The variable x could (and often does) represent an estimator for θ. Using $f(x;\theta)$ we can find for a pre-specified probability $1-\alpha$ and for every value of θ a set of values $x_1(\theta,\alpha)$ and $x_2(\theta,\alpha)$ such that

$$P(x_1 < x < x_2;\theta) = 1-\alpha = \int_{x_1}^{x_2} f(x;\theta)\,dx\ . \tag{32.39}$$

This is illustrated in Fig. 32.3: a horizontal line segment $[x_1(\theta,\alpha), x_2(\theta,\alpha)]$ is drawn for representative values of θ. The union of such intervals for all values of θ, designated in the figure as $D(\alpha)$, is known as the *confidence belt*. Typically the curves $x_1(\theta,\alpha)$ and $x_2(\theta,\alpha)$ are monotonic functions of θ, which we assume for this discussion.

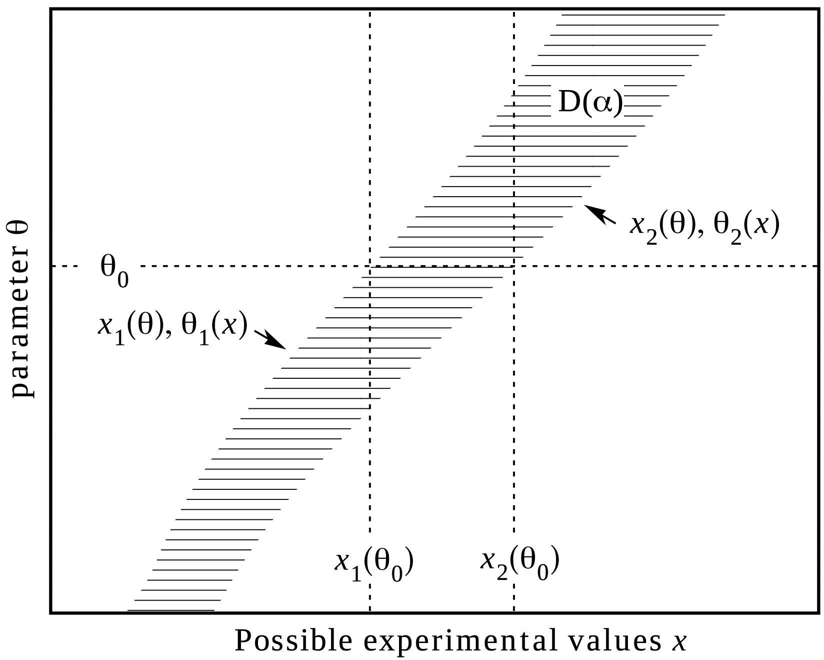
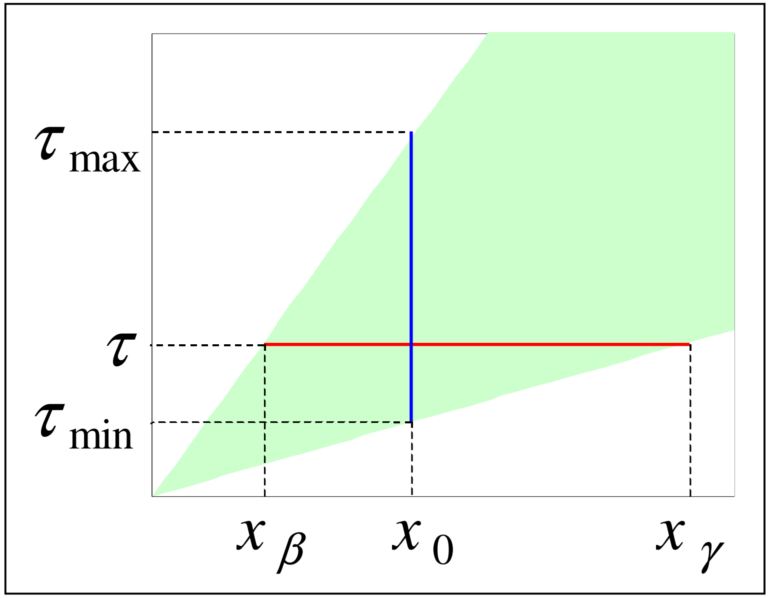


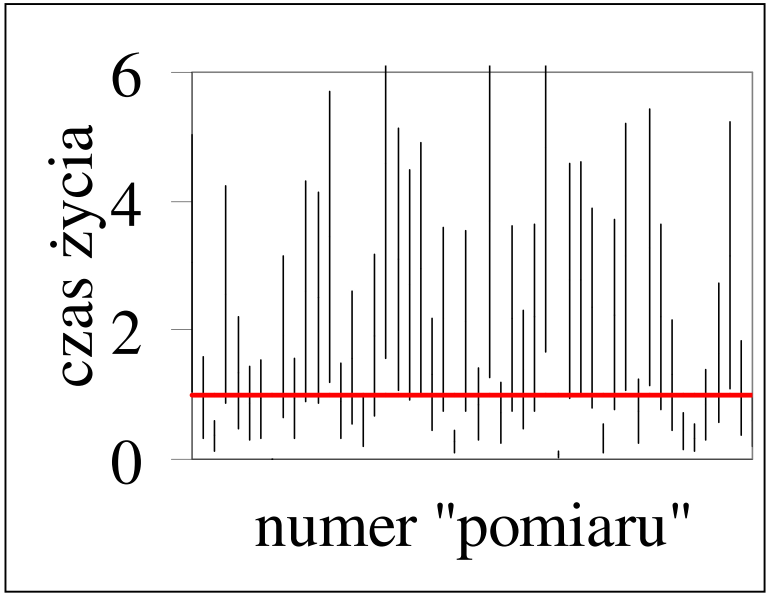
Figure 32.3: Construction of the confidence belt (see text).

Przedział ufności – interpretacja

Zapis konwencjonalny:

$$P\left(x_\beta \le x \le x_\gamma \mid \theta\right) = P\left(\theta_{\min} \le \theta \le \theta_{\max} \mid x_0\right) = 1-\alpha$$





Liczba 1 – α określa prawdopodobieństwo:

- *że przedział* [θ_{min}, θ_{max}], *zawiera* (***pokrywa***) *parametr* θ,

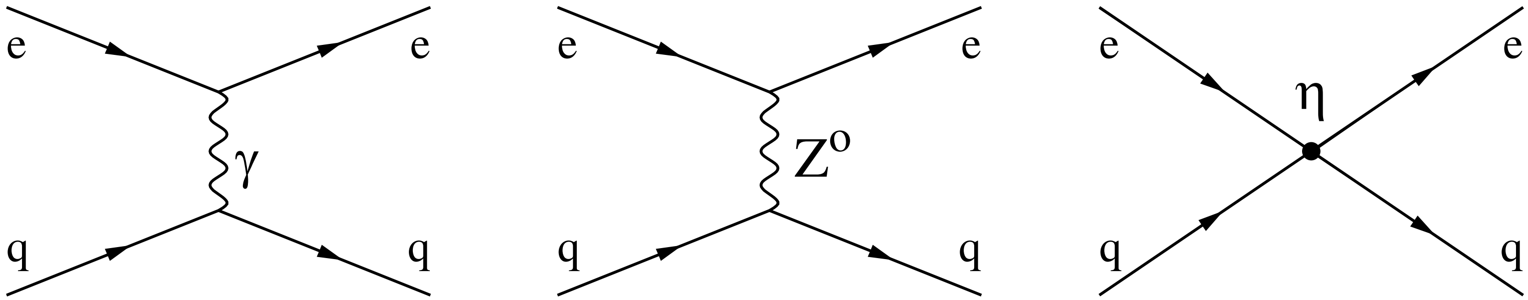
  a nie

- *prawdopodobieństwo znalezienia parametru w tym przedziale!!!*

Models

Contact Interactions

Contact Interactions modify tree level $eq \rightarrow eq$ scattering amplitudes $M^{eq}_{\alpha\beta}$:



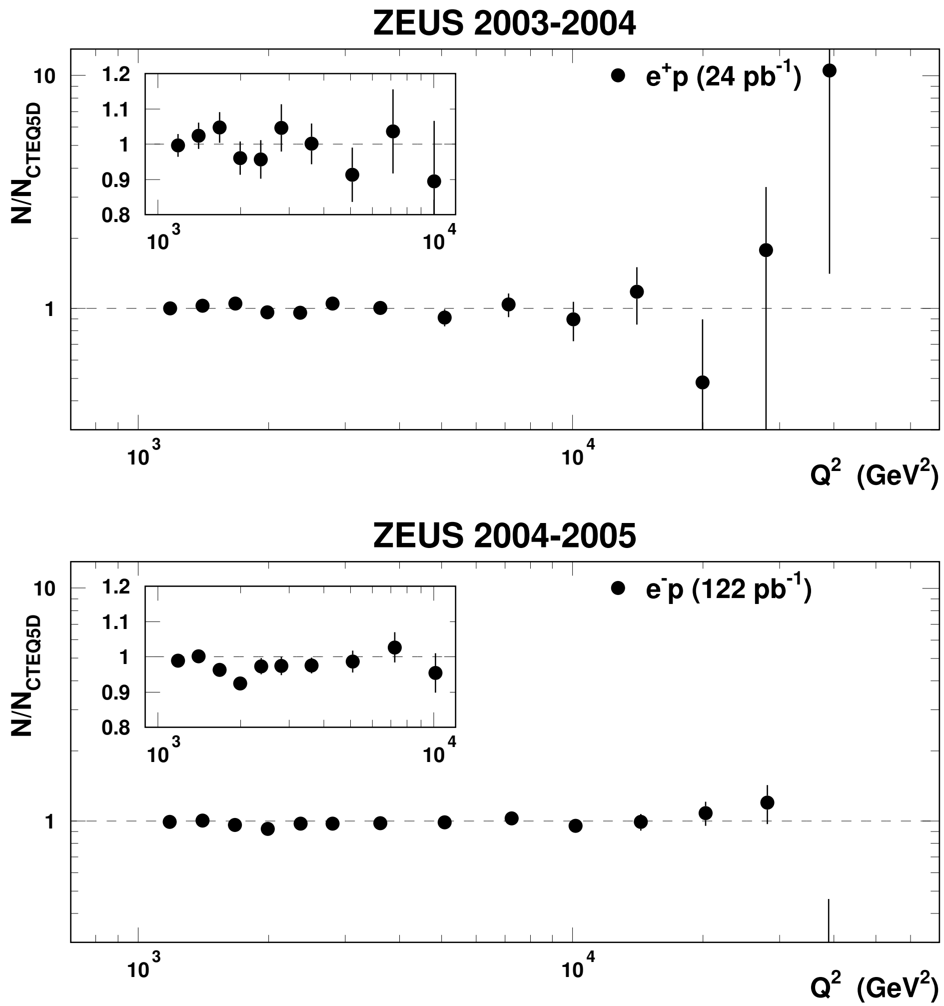
$$M^{eq}_{\alpha\beta}(Q^2) = \underbrace{\frac{e^2 e_q}{Q^2}}_{\gamma} - \frac{e^2}{sin^2\theta_W \cdot cos^2\theta_W} \cdot \underbrace{\frac{g^e_\alpha g^q_\beta}{Q^2 + m^2_Z}}_{Z^\circ} + \underbrace{\eta^{eq}_{\alpha\beta}}_{?}$$

$\eta^{eq}_{\alpha\beta}$ - 4 possible couplings for every flavor q

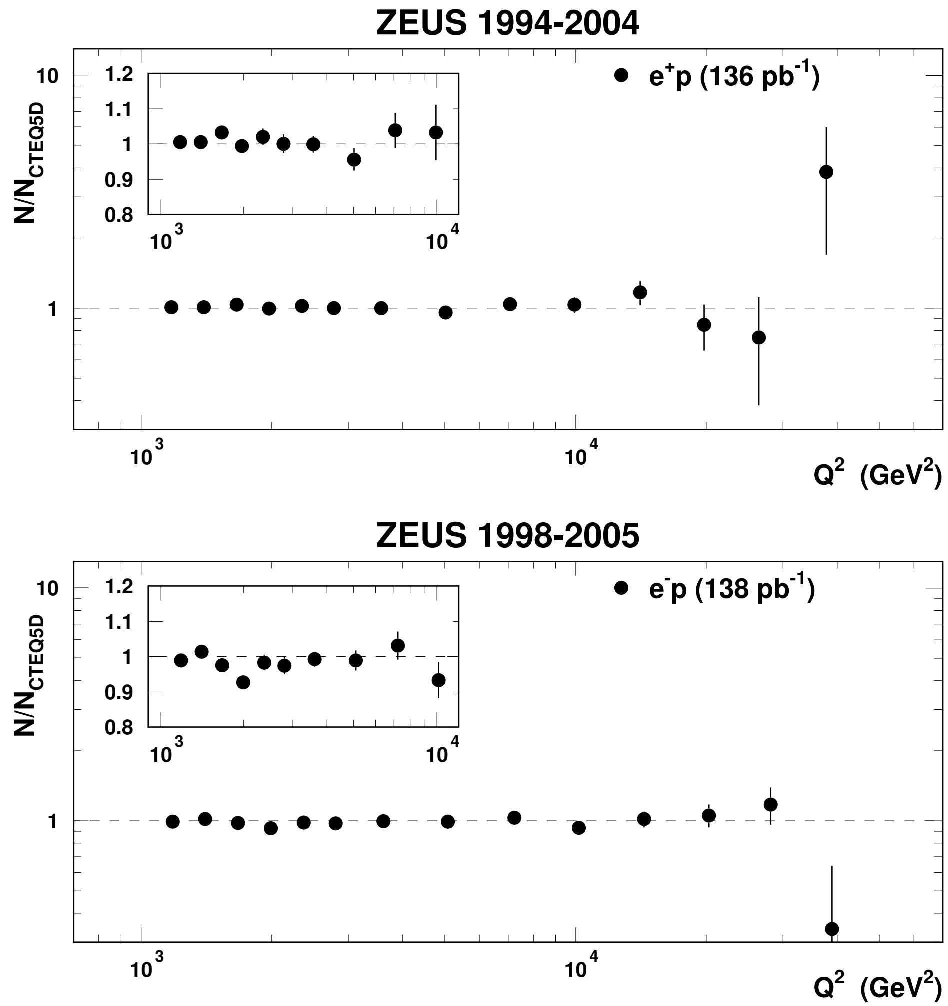
Different models assume different helicity structure of new interactions

Data and analysis

HERA-II data



All HERA data



Likelihood function

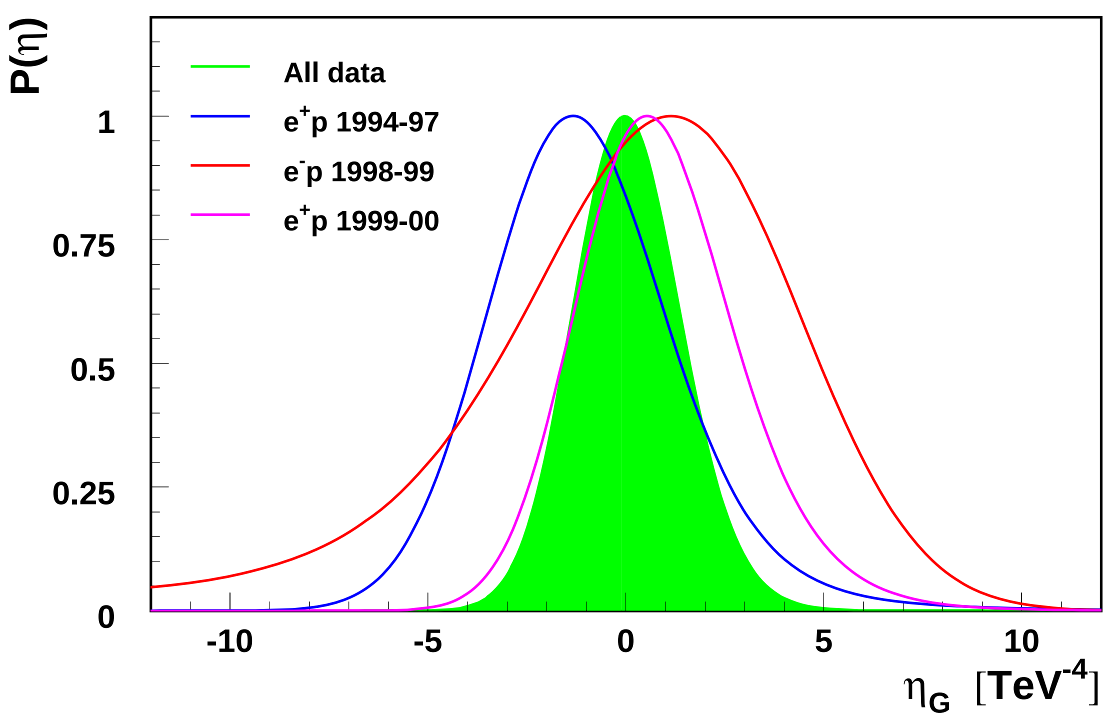
Observed numbers of events in Q^2 bins n_i are compared with the CI model expectations $\mu_i(\eta_G)$ using the probability function:

$$P(\eta_G) \sim \prod_i \frac{\mu_i(\eta_G)^{n_i} \cdot e^{-\mu_i(\eta_G)}}{n_i!}$$

where i runs over 14 Q^2 bins $\times$ 3 data taking periods.

$$\eta_G \equiv \pm \frac{\lambda}{M_S^4}$$

Resulting likelihood function for the nominal data:



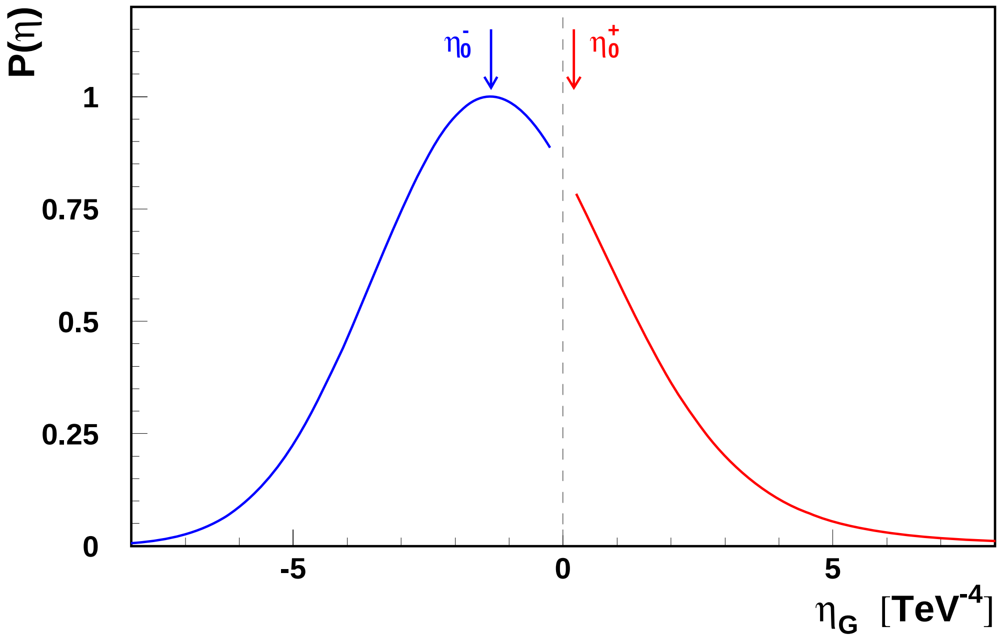
normalized to $\max_\eta P(\eta_G) = 1$

Analysis

Limit setting (1)

- Find coupling values giving best description of the data, separately for negative and positive couplings:

example



For ED model $P(\eta_G)$ has always only one maximum: either $\eta_\circ^+$ or $\eta_\circ^-$ is zero.

In general case (other CI models) two maxima can be found.

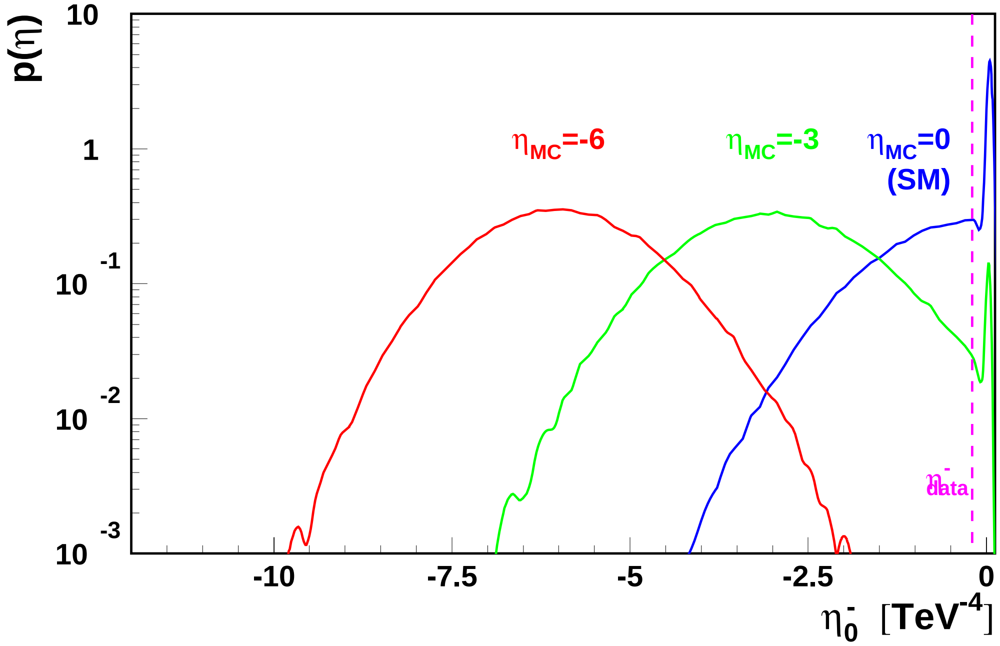
ED model, ZEUS 1994-2000 data:

$$\eta_\circ^- = -0.02\ \text{TeV}^{-4}$$
$$\eta_\circ^+ = 0$$

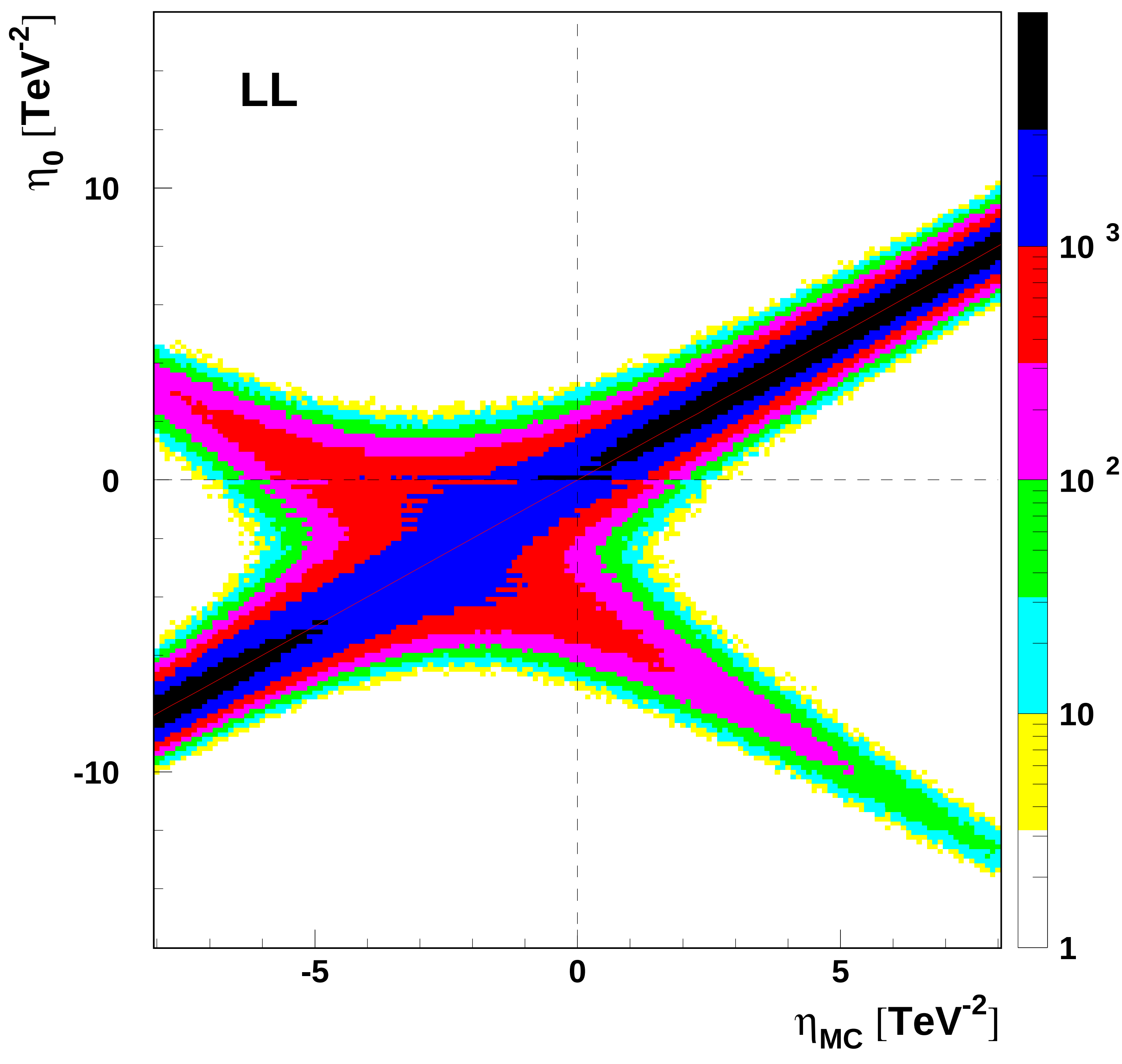
very good agreement with the Standard Model

Limit setting (2)

- Perform "MC experiments" (MCE) to find the expected distribution of $\eta_\circ^+$ and $\eta_\circ^-$ for Standard Model and for ED model with arbitrary coupling value η_{MC}



95% CL limit on η_G (for $\eta_G < 0$) is defined as η_{MC} value for which 95% of Monte Carlo experiments result in η_0^- value lower than the value η_{data}^- found for nominal data.



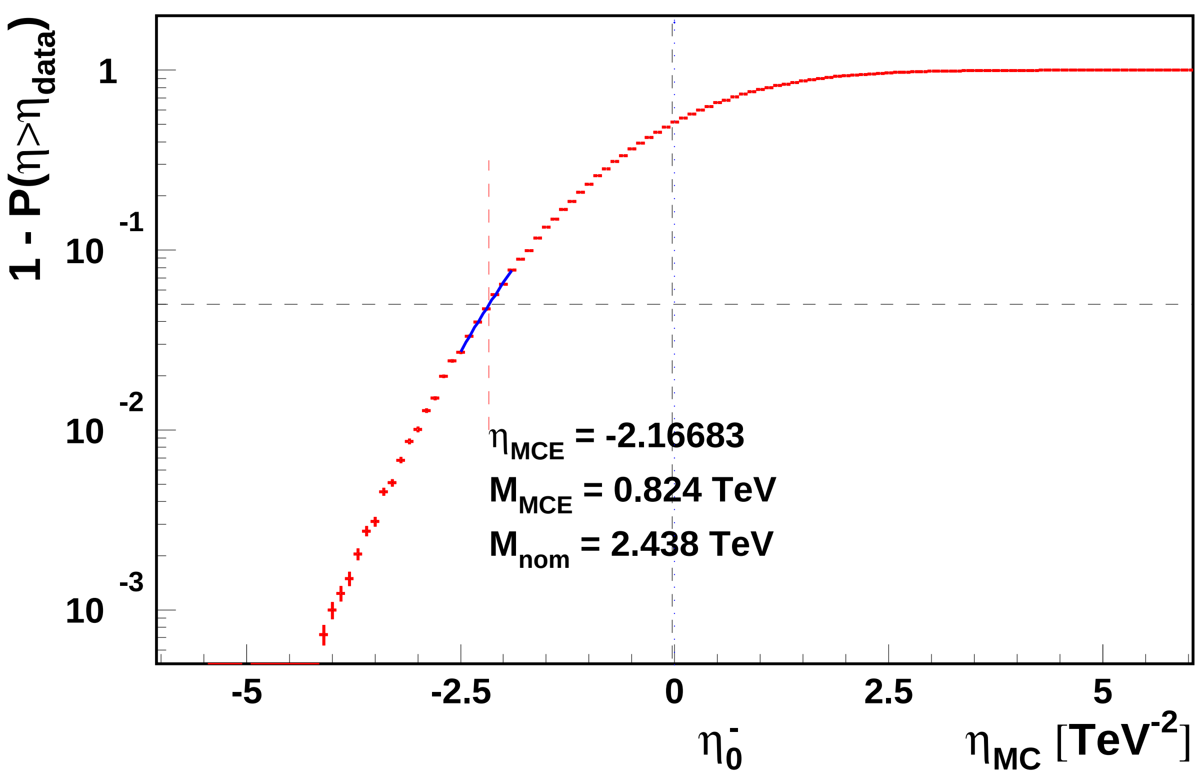
LL
η0 [TeV-2]
10
0
-10
-5
0
5
ηMC [TeV-2]
1
10
10 2
10 3

Limit setting (3)

Method used to find 95% CL coupling limits with high (statistical) precision:

- calculate probability $P(|\eta_\circ^\pm| > |\eta_{data}|)$ for selected η_{MC} values (grid).
- interpolate between grid points using polynomial fit to $\ln(P)$.

nominal data, no systematics



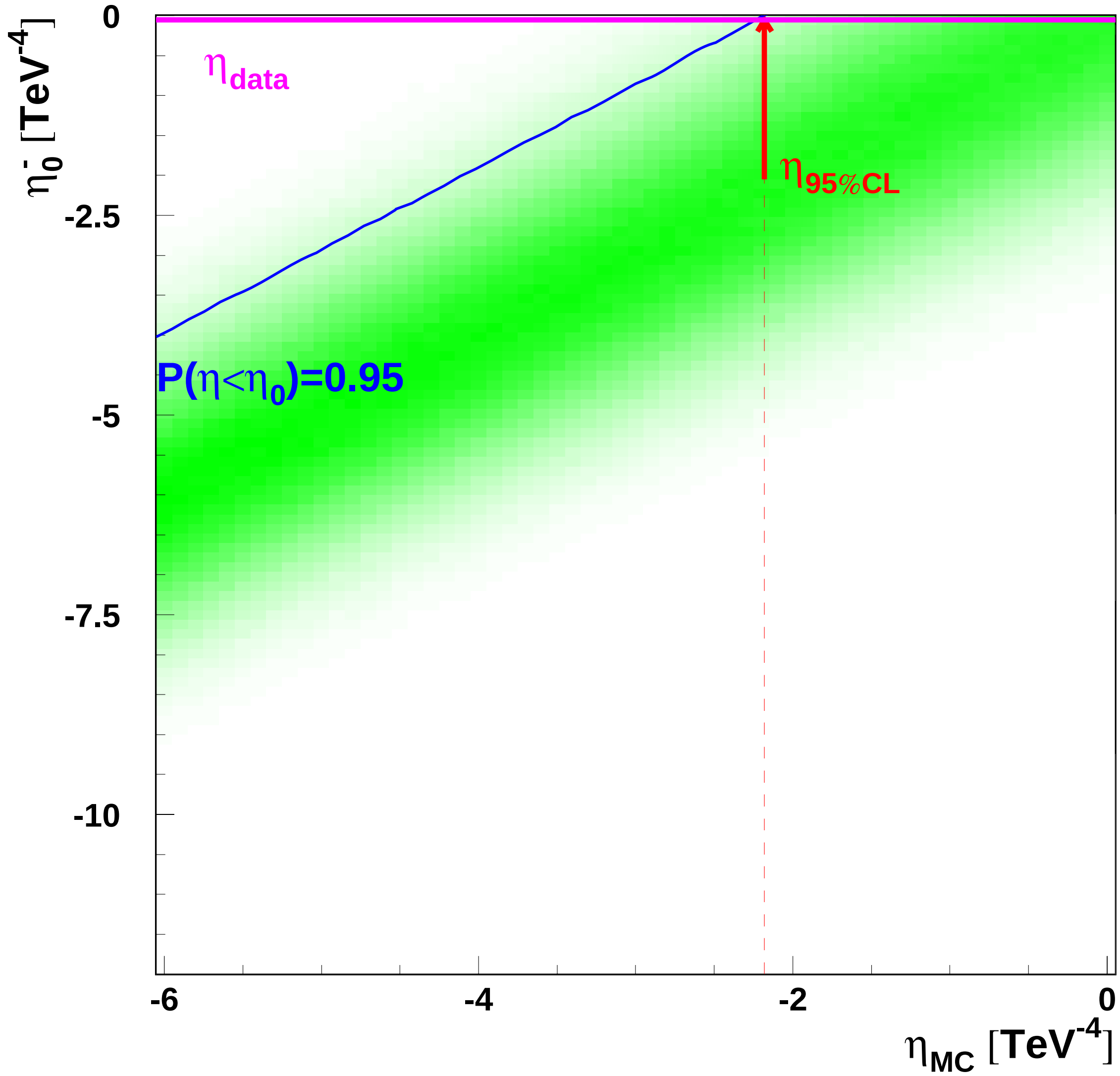
$\eta_G < -2.167$ TeV^{-4} on 95% CL

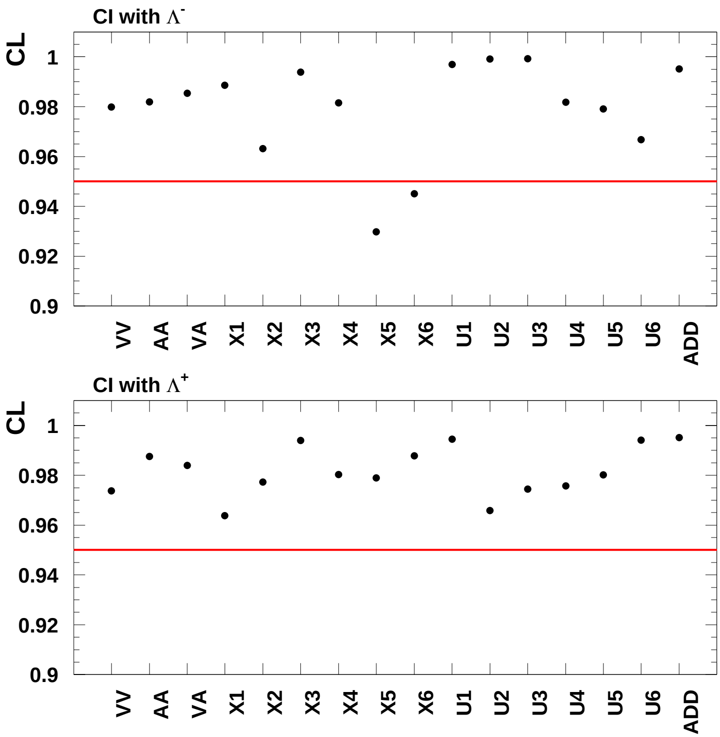
Analysis

Limit setting

2-D probability distribution for η_0^- as a function of η_{MC}

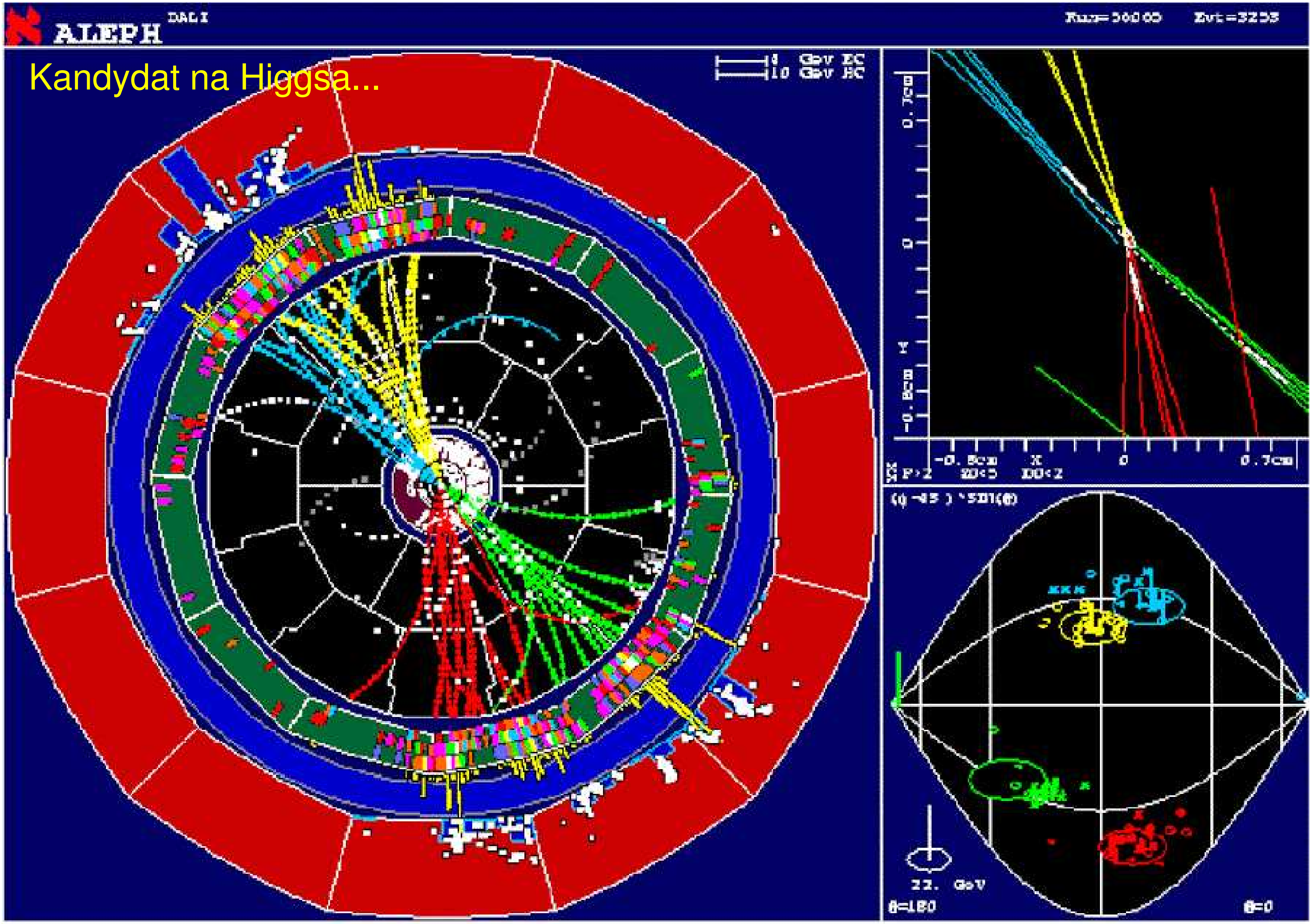
$$P(\eta_\circ^-;\eta_{MC})$$





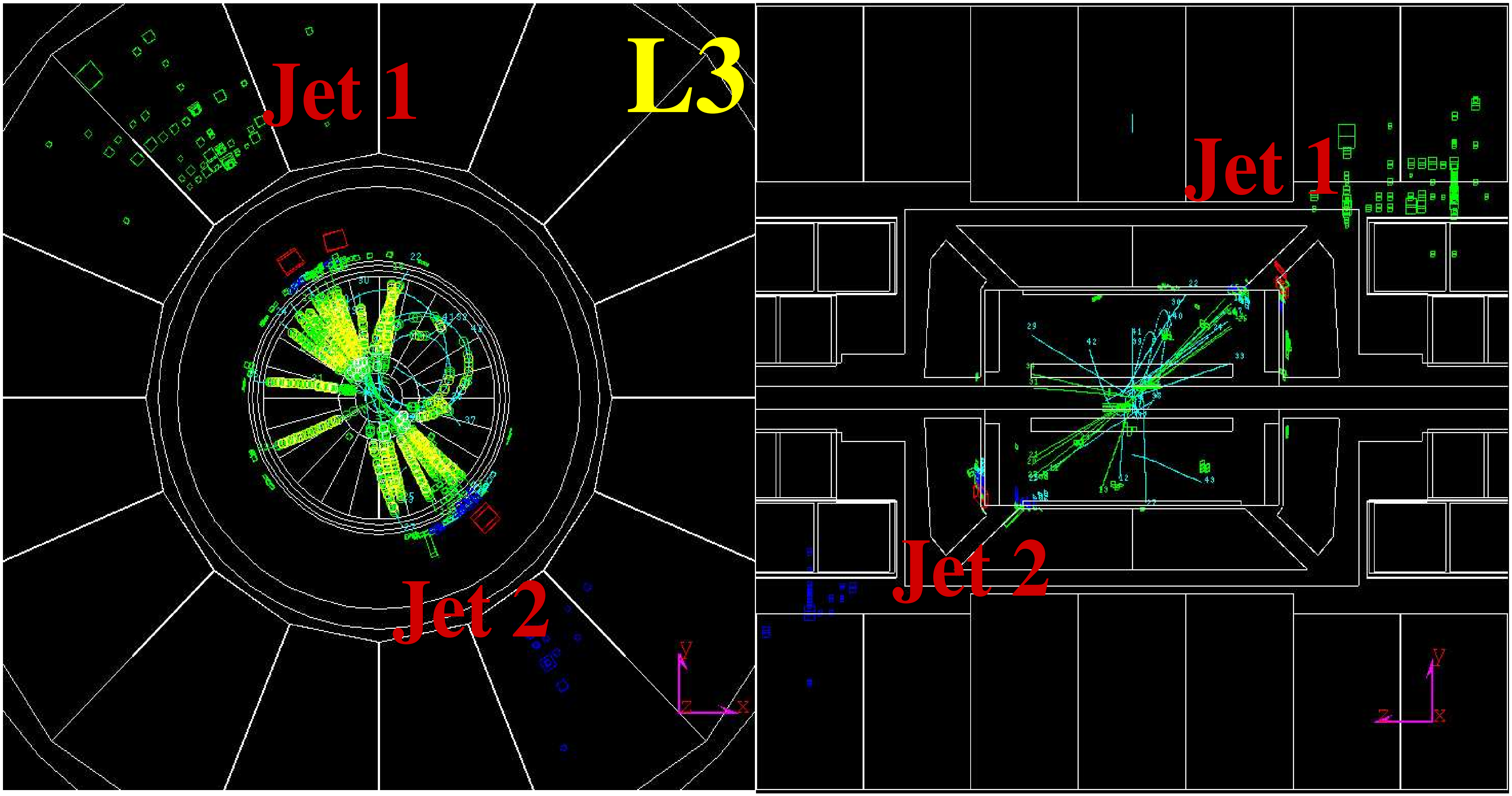
Limits obtained with Bayesian approach

Figure E.5: Confidence Levels for mass scale limits Λ^- and Λ^+, for different contact interaction models considered in this analysis.


ALEPH
DALI
Run=56065 Evt=3253
Kandydat na Higgsa...
22. GeV
θ=180
θ=0

candidate for

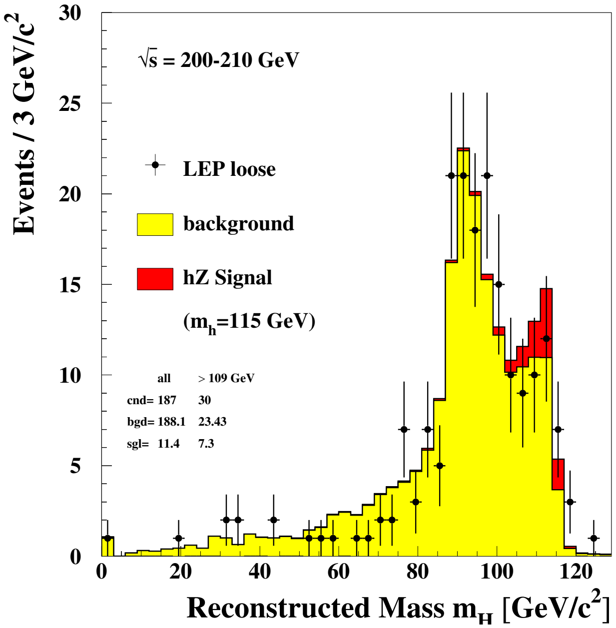
$e^+e^- \rightarrow H\nu\bar{\nu} \rightarrow$ 2 jets + missing energy



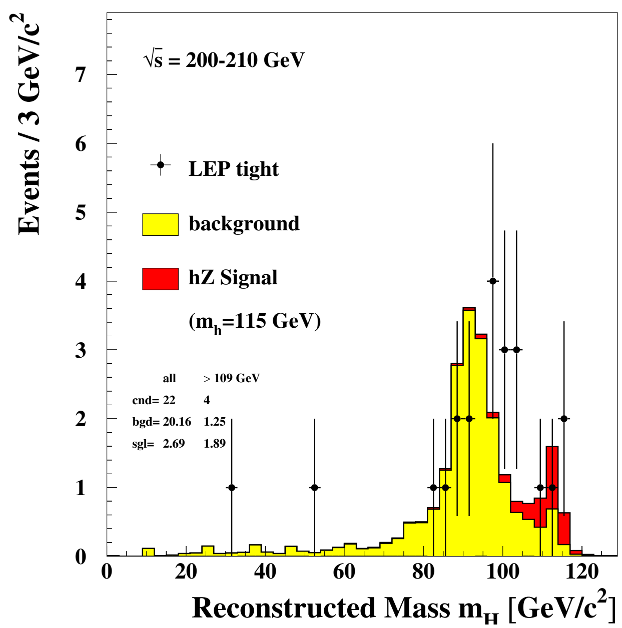
Poszukiwanie Higgsa

Rozkład masy

łagodna selekcja



ostra selekcja



W obszarze $m_h \sim$ 115 GeV widać niewielki nadmiar przypadków, który może pochodzić od produkcji Higgsa

Niestety, jest to efekt na poziomie $\sim 2\sigma$

LEP wyłączono zanim zdołał wyjaśnić ten efekt...

Statistics ... Definitions

TASK ... to combine "channels" from four experiments

Data sets @ different E_{cm} and Luminosities

Different decay-channels

$$e^+e^- \longrightarrow \text{Z H} \longrightarrow b\bar{b},\ \tau^+\tau^-$$
$$\hookrightarrow\ q\bar{q}, \nu\bar{\nu}, e^+e^-, \mu^+\mu^-, \tau^+\tau^-$$

(1) **INPUTS** ... for each "channel" ... binned in two discriminating variables (both contribute to the search sensitivity)

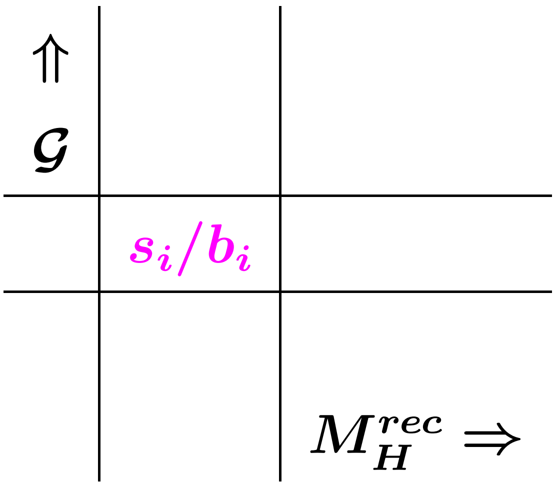
- Reconstructed Higgs mass M_H^{rec}
- Global variable $\mathcal{G}$ ... containing

  b-tag, kinematics, jet-properties ...

In each bin i ...

- Bkgd. (MC) b_i
- Signal (MC) $s_i(m_H)$

  for "test-mass" m_H
- Nbr of candidates N_i



MC estimates of $s_i(m_H)$ and b_i take into account the exp'tal details (e.g. E_{cm}, lumi, signal eff., mass-resol., bkgds ...)

For "test-mass" m_H ...

(2) LIKELIHOOD TEST ... "sig + bkgd" $\Longleftrightarrow$ "bkgd"

$$-2\ln Q(m_H) = 2s_{tot} - 2\sum N_i \ln[1 + s_i(m_H)/b_i]$$

$$Q(m_H) = \mathcal{L}(s+b)/\mathcal{L}(b)$$ **"test-statistic"**

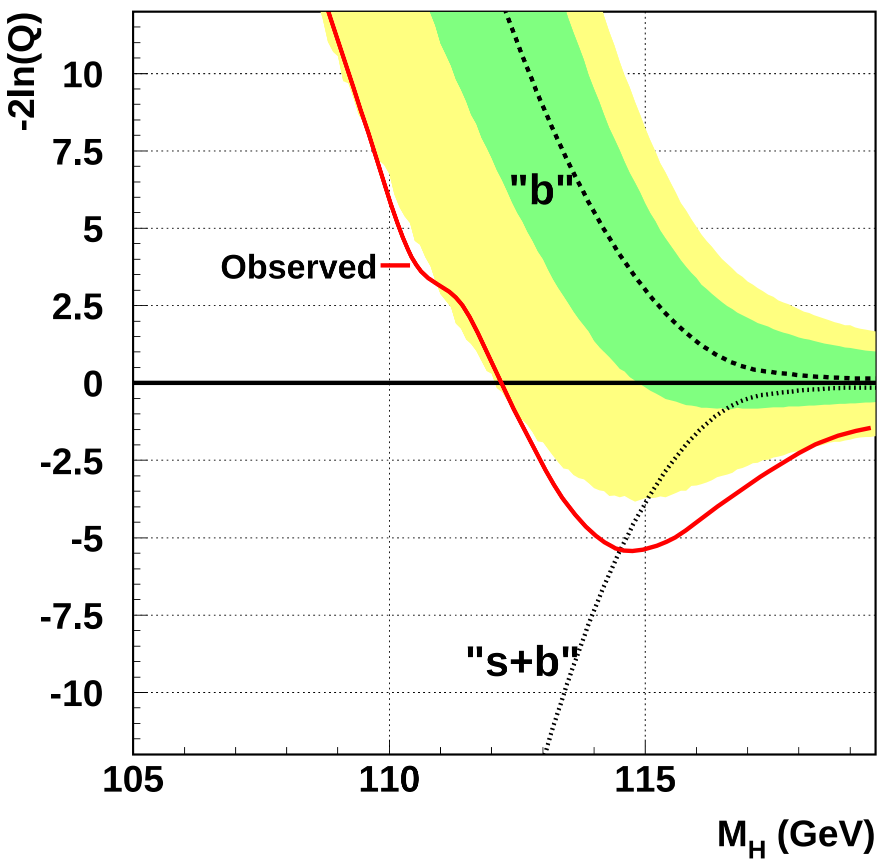
to rank the observed event configuration

between "$s+b$" and "b" hypotheses

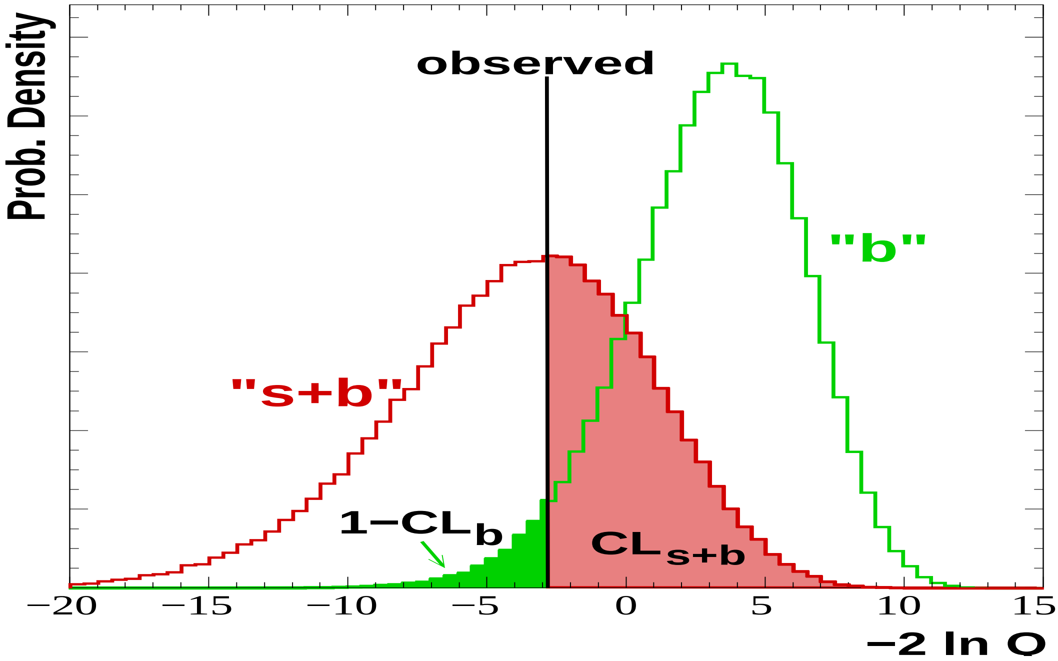
For arbitrary test-mass m_H ... and replacing the data set by

ficticious MC sets of "$s+b$" and "b" configurations

$\Rightarrow$ **expected curves ... and statistical spread**



(3) CONFIDENCE LEVELS ...



$\Leftarrow$ "$s+b$" "b" $\Rightarrow$

- $1 - CL_b$ ... **a measure of incompatibility with "b"**

  Given an ensemble of "b" experiments ...

  probability to obtain an event configuration less bkgd-like

  than the observed event configuration

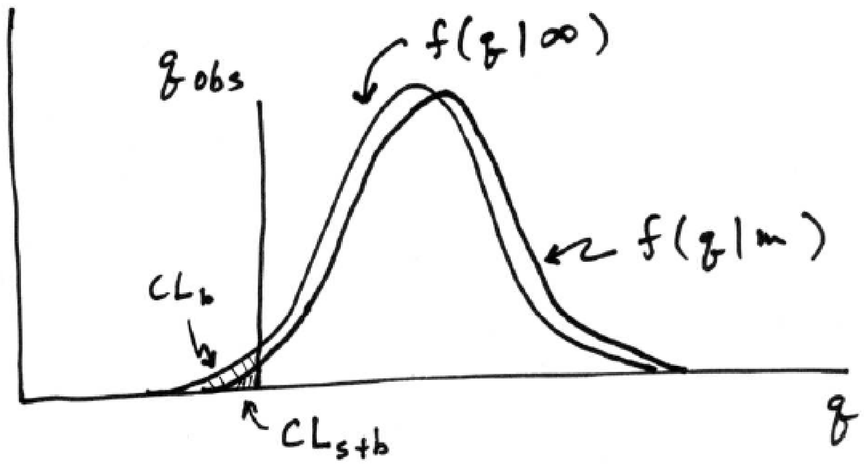
$1 - CL_b$	0.32	0.046	2.7×10^{-3}	6.3×10^{-5}	5.7×10^{-7}
	1 σ	2 σ	3 σ	4 σ	5 σ

- CL_{s+b} ... **a measure of incompatibility with "$s+b$"**

  $CL_s = CL_{s+b}/CL_b \Rightarrow$ **lower bound on Higgs mass**

LEP-style analysis: CL_s

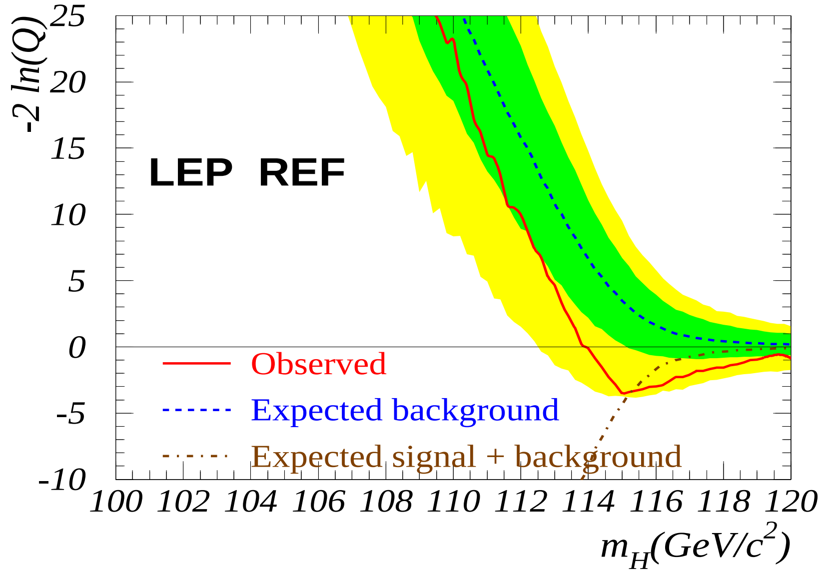
The problem with the CL_{s+b} method is that for high *m*, the distribution of *q* approaches that of the background-only hypothesis:

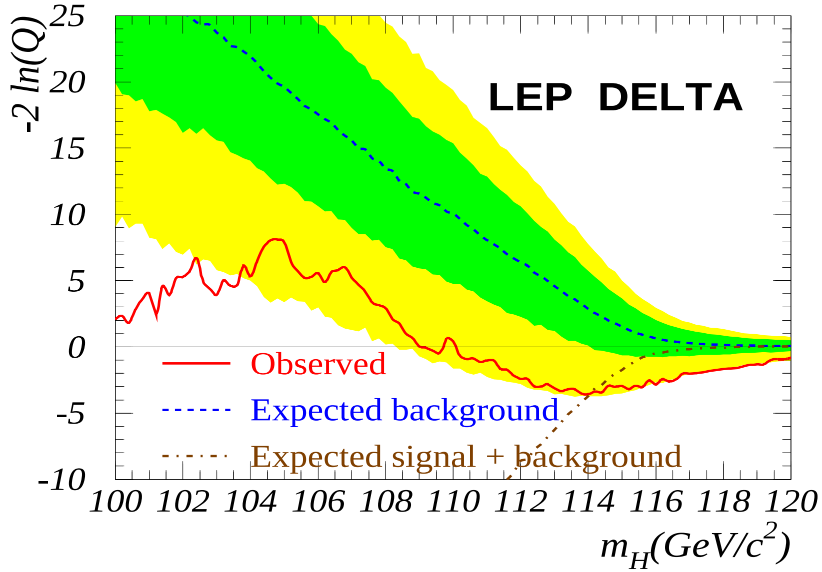


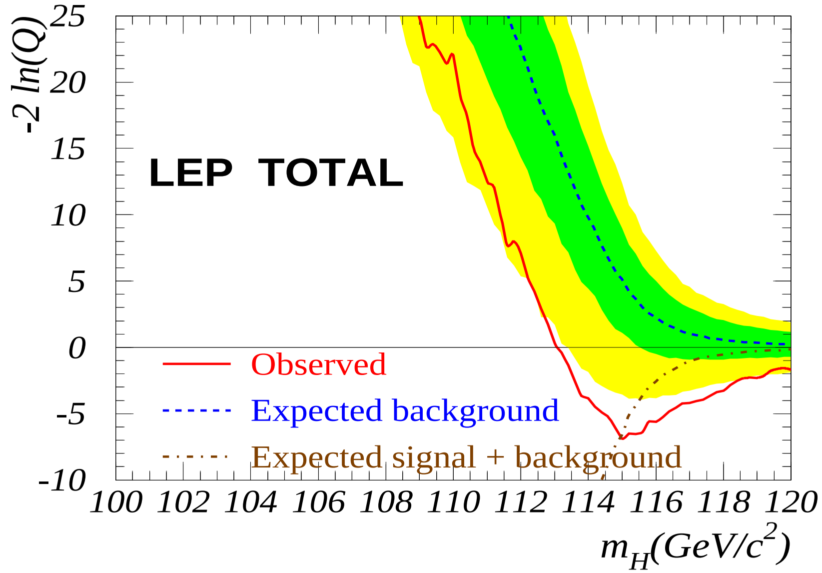
So a low fluctuation in the number of background events can give $CL_{s+b} < \alpha$

This rejects a high *m* value even though we are not sensitive to Higgs production with that mass; the reason was a low fluctuation in the background.

$-2\ln(Q)$ ... REF, DELTA, TOTAL







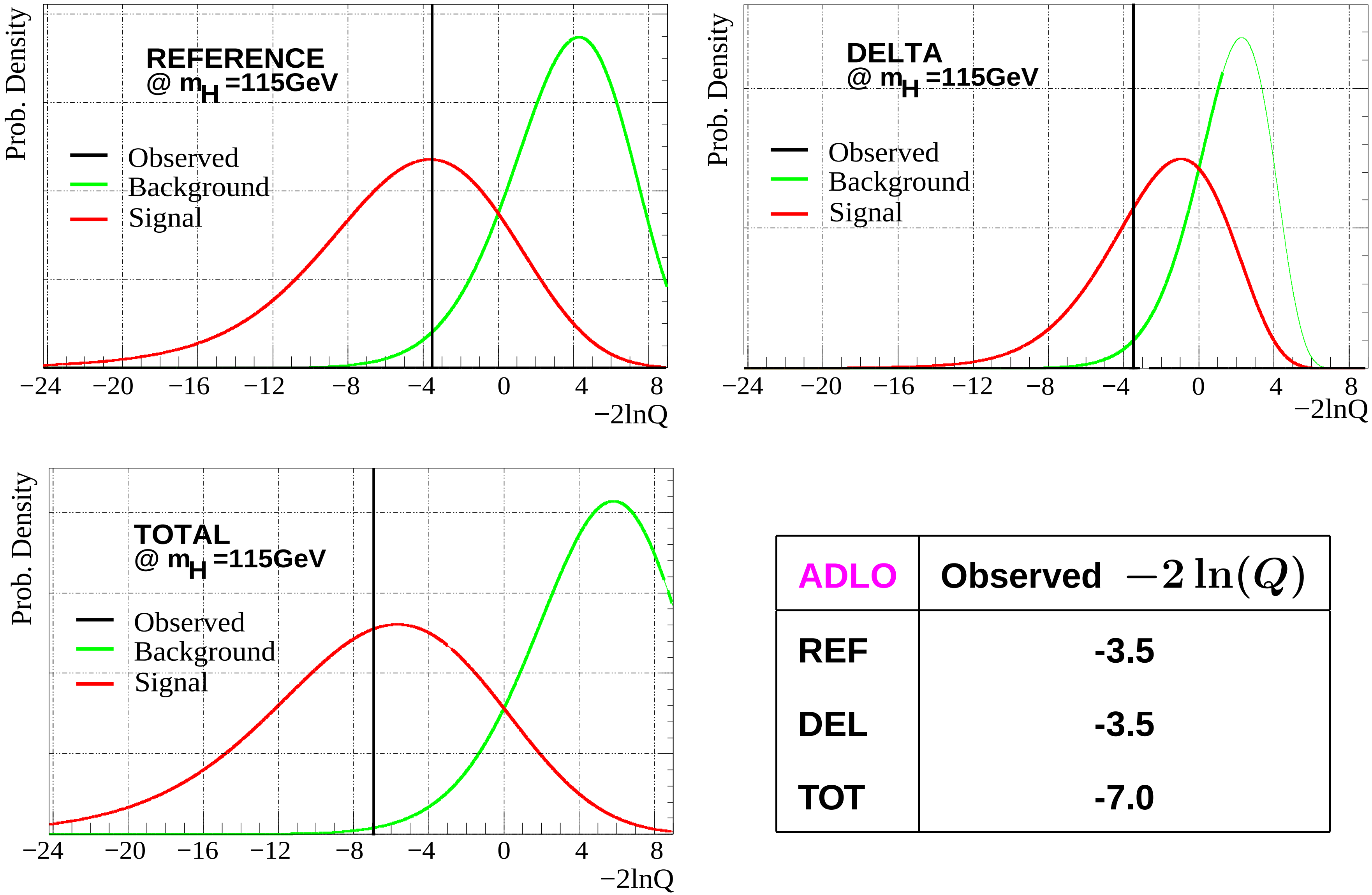
$\Leftarrow$

Minimum @$m_H \approx 115$ GeV

Agreement with SM Higgs cross-sect. for

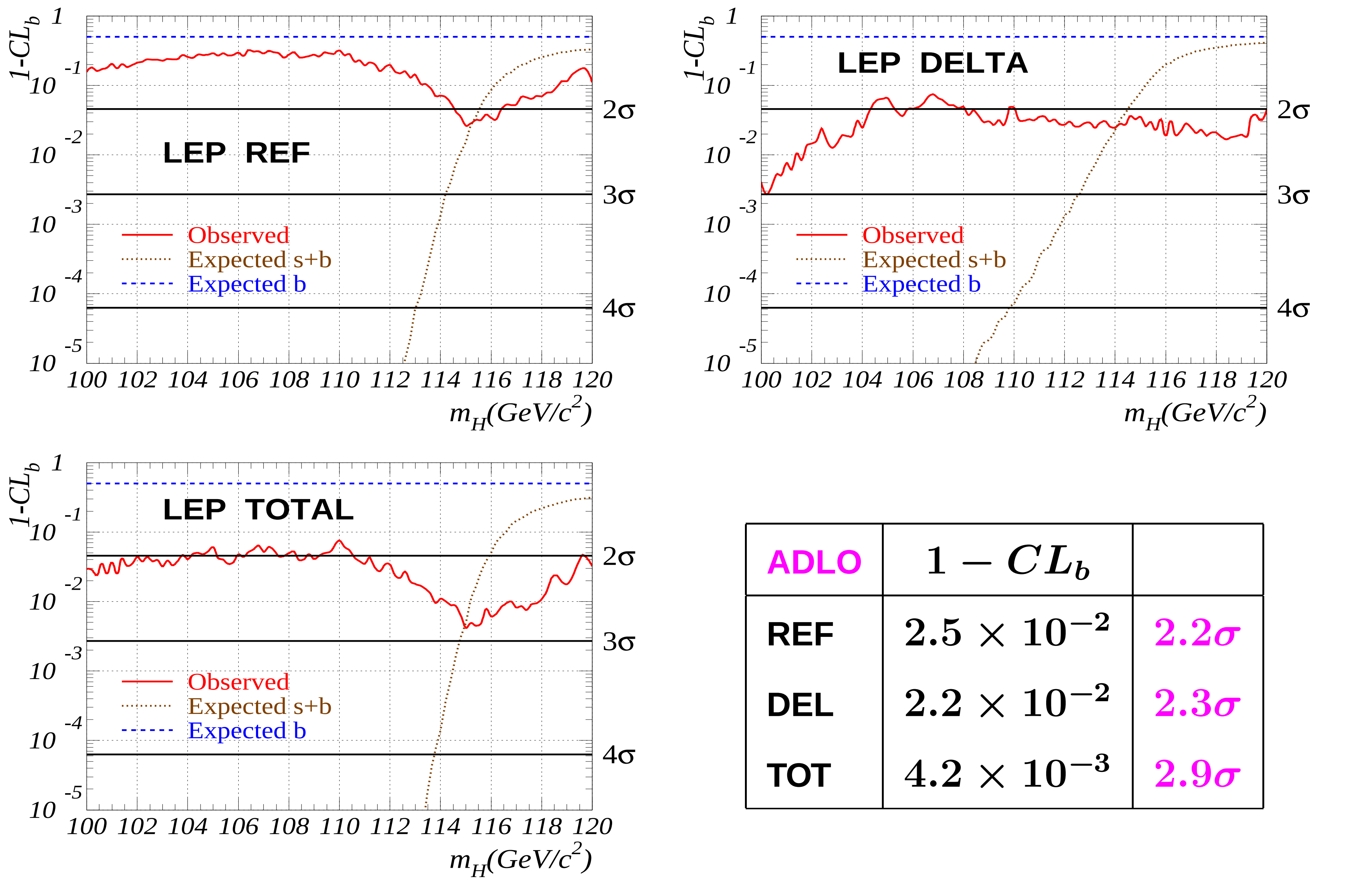
$$m_H = 115.0^{+1.3}_{-0.9}\ \text{GeV}$$

Prob. density @$m_H = 115$ GeV ... REF, DELTA, TOTAL



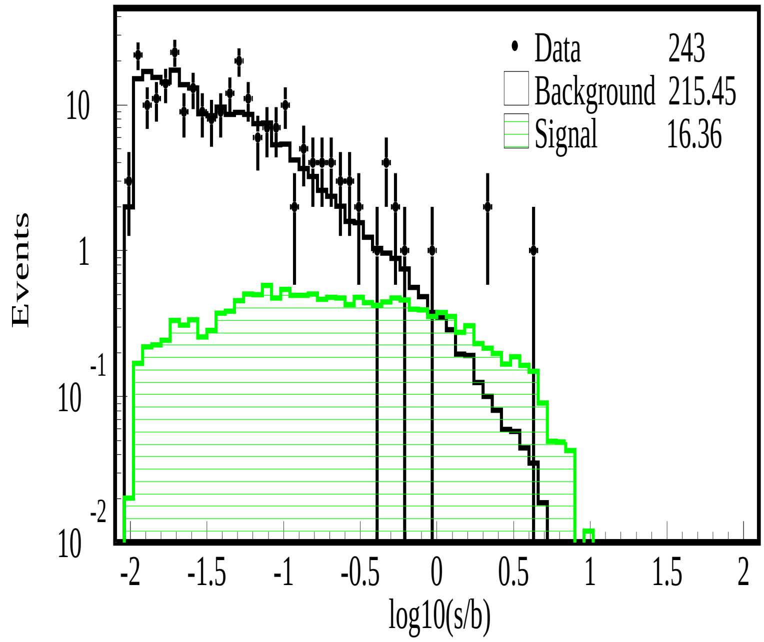
ADLO	Observed $-2\ln(Q)$
REF	-3.5
DEL	-3.5
TOT	-7.0

$1 - CL_b$ ... REF, DELTA, TOTAL



ADLO	$1 - CL_b$	
REF	2.5×10^{-2}	2.2σ
DEL	2.2×10^{-2}	2.3σ
TOT	4.2×10^{-3}	2.9σ

Expected rates @$m_H = 115$ GeV ..... TOTAL



Integrating bkgd, signal and data ...

for $s/b \gtrsim 1$

		Backgd	Signal	Candidates
ADLO	4-jet	0.93	1.60	3
	E-miss	0.30	0.46	1
	Lept	0.35	0.68	0
	Taus	0.14	0.29	0
ADLO	All chan.	1.72	3.03	4

U.L. in Poisson Process, n=3 observed: 3 ways

1. **Bayesian upper limit at 90% credibility:**
   find μ_u such that posterior probability $p(\mu > \mu_u) = 0.1$.
2. **Likelihood ratio method for approximate 90% C.L. U.L.:**
   find μ_u such that $\mathcal{L}(\mu_u) / \mathcal{L}(3)$ has prescribed value.
3. **Frequentist one-sided 90% C.L. upper limit:**
   find μ_u such that $P(n \leq 3 \mid \mu_u) = 0.1$.

Deep foundational issues

- **Only #3 has guaranteed ensemble properties (though issues arise with systematics.) Good ?!?**
- **Only #3 uses $P(n|\mu)$ for $n \neq$ observed value. Bad?!? (See below re likelihood principle)**

These issues will not be resolved: aim to have software for reporting all 3 answers, and sensitivity to prior.

68% intervals by various methods for Poisson process with n=3 observed

Method	Prior	Interval	Length	Coverage?
rms deviation $n \pm \sqrt{n}$	–	(1.27, 4.73)	3.46	no
Bayesian central	1	(2.09, **5.92**)	3.83	no
Bayesian shortest	1	(1.55, 5.15)	3.60	no
Bayesian central	$1/\mu$	(**1.37**, 4.64)	3.27	no
Bayesian shortest	$1/\mu$	(0.86, 3.85)	2.99	no
Likelihood ratio	–	(1.58, 5.08)	3.50	no
Frequentist central	–	(**1.37, 5.92**)	4.55	yes
Frequentist shortest	–	(1.29, 5.25)	3.96	yes
Frequentist LR ordering	–	(1.10, 5.30)	4.20	yes

For the Jeffreys prior ($1/\sqrt{\mu}$), Bayesian central interval is (1.72, 5.27).

Frequentist intervals over-cover due to discreteness of n.

Adapted from Cousins05 and
R. Cousins, Am. J. Phys. 63 398 (1995)

Metoda Bayesa przy pracy

Rozkład Poissona: $\mathcal{P}(n|\mu+\lambda) = \mathcal{L}(n|\mu,\lambda)$ z tłem, $z_\mu(\mu)$ = const

$$f_\mu(\mu\,|\,n,\lambda) = \frac{\mathcal{P}(n\,|\,\mu+\lambda)}{\sum_{k=0}^{N}\mathcal{P}(k\,|\,\lambda)}$$

$$1-\alpha = \int_0^\mu f_\mu(\mu'\,|\,n,\lambda)\,d\mu' \quad \Rightarrow \quad \alpha = \sum_{n=0}^{N} f_\mu(\mu\,|\,n,\lambda)$$

$N = 0,\ \alpha = 0{,}1$					
metoda	$\lambda =$	0	1	2	3
pre FC	$\mu_{\max} =$	2,30	1,30	0,30	–0,70
zunifikowana	$\mu_{\max} =$	2,44	1,61	1,26	1,08
bayesowska	$\mu_{\max} =$	2,30	2,30	2,30	2,30

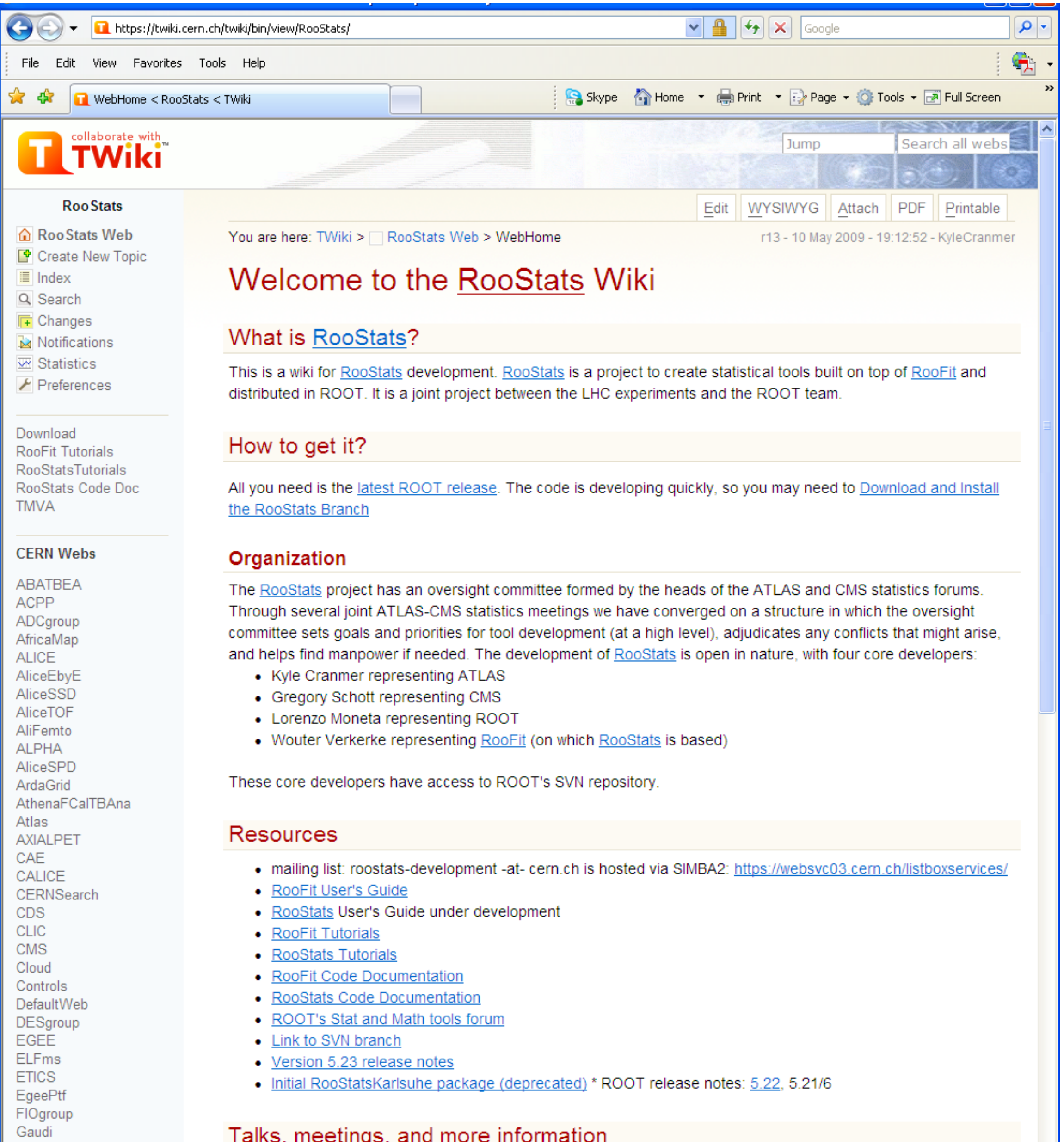
Summary of Three Ways to Make Intervals

	Bayesian Credible	Frequentist Confidence	Likelihood Ratio
Requires prior pdf?	Yes	No	No
Obeys likelihood principle?	Yes (exception re Jeffreys prior)	No	Yes
Random variable in "$P(\mu_t \in [\mu_1, \mu_2])$":	μ_t	μ_1, μ_2	μ_1, μ_2
Coverage guaranteed?	No	Yes (but over-coverage…)	No
Provides P(parameter|data)?	Yes	No	No

Goal for the LHC a Few Years Ago

- **Have in place tools to allow computation of results using a variety of recipes, for problems up to intermediate complexity:**
  - **Bayesian with analysis of sensitivity to prior**
  - **Frequentist construction with approximate treatment of nuisance parameters**
  - **Profile likelihood ratio (Minuit MINOS)**
  - **Other "favorites" such as LEP's CL_S (which is an HEP invention)**
- **The community can then demand that a result shown with one's preferred method also be shown with the other methods, *and sampling properties studied*.**
- **When the methods all agree, we are in asymptopic nirvana.**
- **When the methods disagree, we learn something!**
  - **The results are answers to different questions.**
  - **Bayesian methods can have poor frequentist properties**
  - **Frequentist methods can badly violate likelihood principle**

ATLAS/CMS/ROOT Project: RooStats built on RooFit



https://twiki.cern.ch/twiki/bin/view/RooStats/

File Edit View Favorites Tools Help

WebHome < RooStats < TWiki

RooStats

RooStats Web
Create New Topic
Index
Search
Changes
Notifications
Statistics
Preferences

Download
RooFit Tutorials
RooStatsTutorials
RooStats Code Doc
TMVA

CERN Webs

ABATBEA
ACPP
ADCgroup
AfricaMap
ALICE
AliceEbyE
AliceSSD
AliceTOF
AliFemto
ALPHA
AliceSPD
ArdaGrid
AthenaFCalTBAna
Atlas
AXIALPET
CAE
CALICE
CERNSearch
CDS
CLIC
CMS
Cloud
Controls
DefaultWeb
DESgroup
EGEE
ELFms
ETICS
EgeePtf
FIOgroup
Gaudi

Edit WYSIWYG Attach PDF Printable

You are here: TWiki > RooStats Web > WebHome

r13 - 10 May 2009 - 19:12:52 - KyleCranmer

Welcome to the RooStats Wiki

What is RooStats?

This is a wiki for RooStats development. RooStats is a project to create statistical tools built on top of RooFit and distributed in ROOT. It is a joint project between the LHC experiments and the ROOT team.

How to get it?

All you need is the latest ROOT release. The code is developing quickly, so you may need to Download and Install the RooStats Branch

Organization

The RooStats project has an oversight committee formed by the heads of the ATLAS and CMS statistics forums. Through several joint ATLAS-CMS statistics meetings we have converged on a structure in which the oversight committee sets goals and priorities for tool development (at a high level), adjudicates any conflicts that might arise, and helps find manpower if needed. The development of RooStats is open in nature, with four core developers:

- Kyle Cranmer representing ATLAS
- Gregory Schott representing CMS
- Lorenzo Moneta representing ROOT
- Wouter Verkerke representing RooFit (on which RooStats is based)

These core developers have access to ROOT's SVN repository.

Resources

- mailing list: roostats-development -at- cern.ch is hosted via SIMBA2: https://websvc03.cern.ch/listboxservices/
- RooFit User's Guide
- RooStats User's Guide under development
- RooFit Tutorials
- RooStats Tutorials
- RooFit Code Documentation
- RooStats Code Documentation
- ROOT's Stat and Math tools forum
- Link to SVN branch
- Version 5.23 release notes
- Initial RooStatsKarlsuhe package (deprecated) * ROOT release notes: 5.22, 5.21/6

Talks, meetings, and more information

Core developers:
K. Cranmer (ATLAS)
Gregory Schott (CMS)
Wouter Verkerke (RooFit)
Lorenzo Moneta (ROOT)

Open project, all welcome to contribute.

Included in ROOT production releases since v5.22, more soon to come

Example macros in $ROOTSYS/tutorials/roostats

RooFit *extensively* documented, RooStats manual catching up, code doc in ROOT.

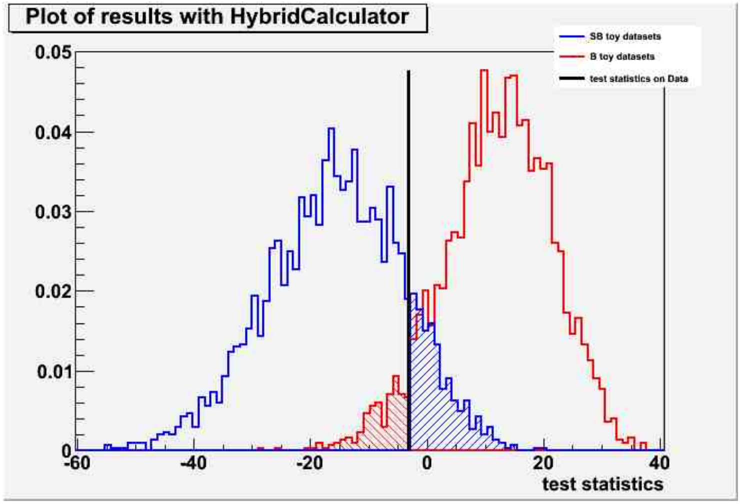
https://twiki.cern.ch/twiki/bin/view/RooStats/WebHome

Running RooStats in Latest (dev) ROOT 5.23.04

CERN lxplus machines: examples run with a few lines, e.g.:

setenv ROOTSYS /afs/cern.ch/sw/lcg/app/releases/ROOT/5.23.04/slc4_ia32_gcc34/root
setenv LD_LIBRARY_PATH ${ROOTSYS}/lib
setenv PATH ${PATH}:${ROOTSYS}/bin
root -x $ROOTSYS/tutorials/roostats/rs201_hybridcalculator.C

Problem considered: Is histogram all background or is there a signal superimposed?
H_0 is uniform background “B” with Poisson mean estimated to be 40±10 events
H_1 is has Gaussian signal in addition to background (“SB”), Poisson mean 20 events.
Calculates test statistic in frequentist way for signal mean, but treats 25% uncertainty on background in Bayesian way. Produces distributions of test statistics under H_0 and H_1 on, from which tail probabilities are calculated and printed.



Code in this example is well-commented but to fully understand it requires consulting the code doc as well, some knowledge of Roofit and RooStats.

Confidence Interval with Nuisance parameters

Estimation of confidence interval for a physics parameter of interest when there are uncertainties in quantities such as acceptance, luminosity, background or selection efficiencies. These are called

- In Statistics: nuisance parameters
- In Particle Physics: sources of systematic uncertainty

Probability model for the data depends on parameters of interest $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_k)$ and nuisance parameters $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_j)$

Have n independent observations $\boldsymbol{X} = (X_1, .., X_n)$ with pdf $f(x|\boldsymbol{\mu}, \boldsymbol{\theta})$

The full likelihood function is given by $\mathscr{L}(\boldsymbol{\mu}, \boldsymbol{\theta} \,|\, \boldsymbol{X}) = \prod_{i=1}^{n} f\,(X_i \,|\, \boldsymbol{\mu}, \boldsymbol{\theta})$

Fully Bayesian Treatment

Requires (joint) prior for the (correlated) nuisance parameters $\pi(\boldsymbol{\theta})$

The posterior is
$$p(\mu) = \frac{\int \mathscr{L}(\mu, \boldsymbol{\theta})\pi(\mu)\pi(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}}{\int\int \mathscr{L}(\mu, \boldsymbol{\theta})\pi(\mu)\pi(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}\,\mathrm{d}\mu}$$

Mathematically equivalent to eliminate nuisance param. from $\mathscr{L}(\mu, \boldsymbol{\theta})$

$$\mathscr{L}(\mu) = \int \mathscr{L}(\mu, \boldsymbol{\theta})\pi(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}$$

and then get the posterior

$$p(\mu) = \frac{\mathscr{L}(\mu)\pi(\mu)}{\int \mathscr{L}(\mu)\pi(\mu)\,\mathrm{d}\mu}$$

The posterior is integrated to define an interval or limit

Podsumowanie

„If your experiment needs statistics,

you ought to have done a **better** experiment."

Ernest Rutherford