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

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:

$$L(\theta) = P(x;\theta)$$ (here x is constant)

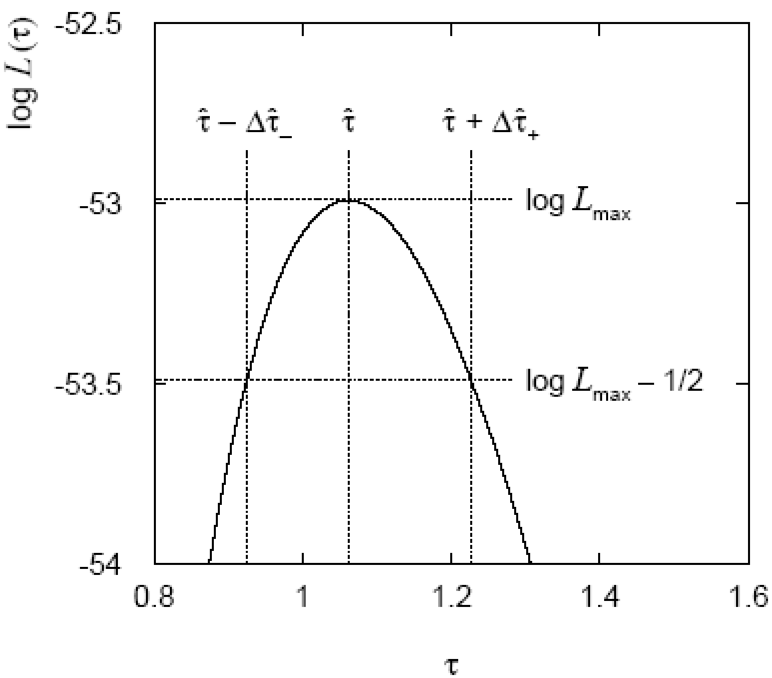
For example, if we have n independent observations of a random variable x, where $x \sim f(x;\theta)$, then

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

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...

Statistical tests (in a particle physics context)

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

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

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

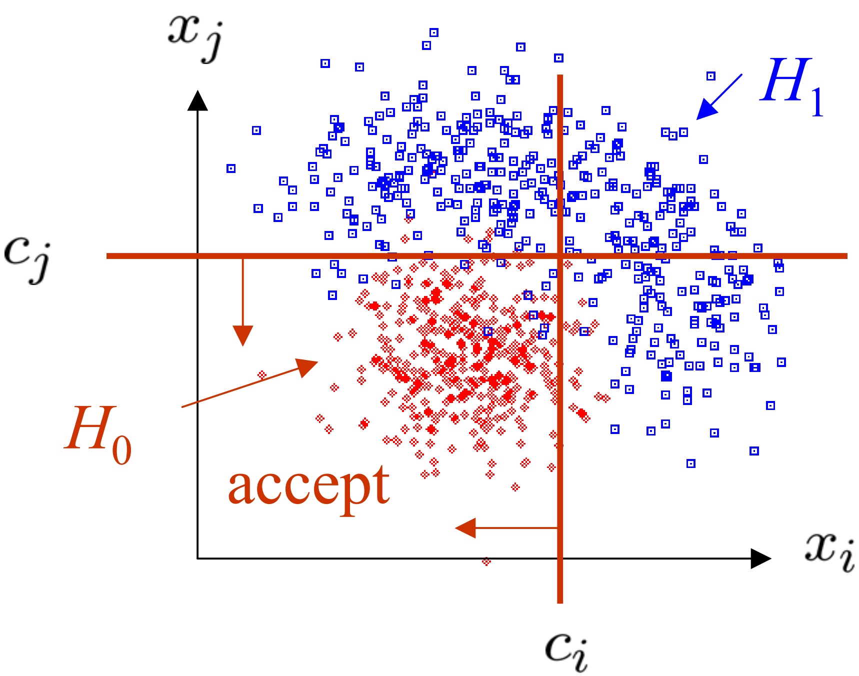
Suppose we have a data sample with two kinds of events, corresponding to hypotheses H_0 and H_1 and we want to select those of type H_0.

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:

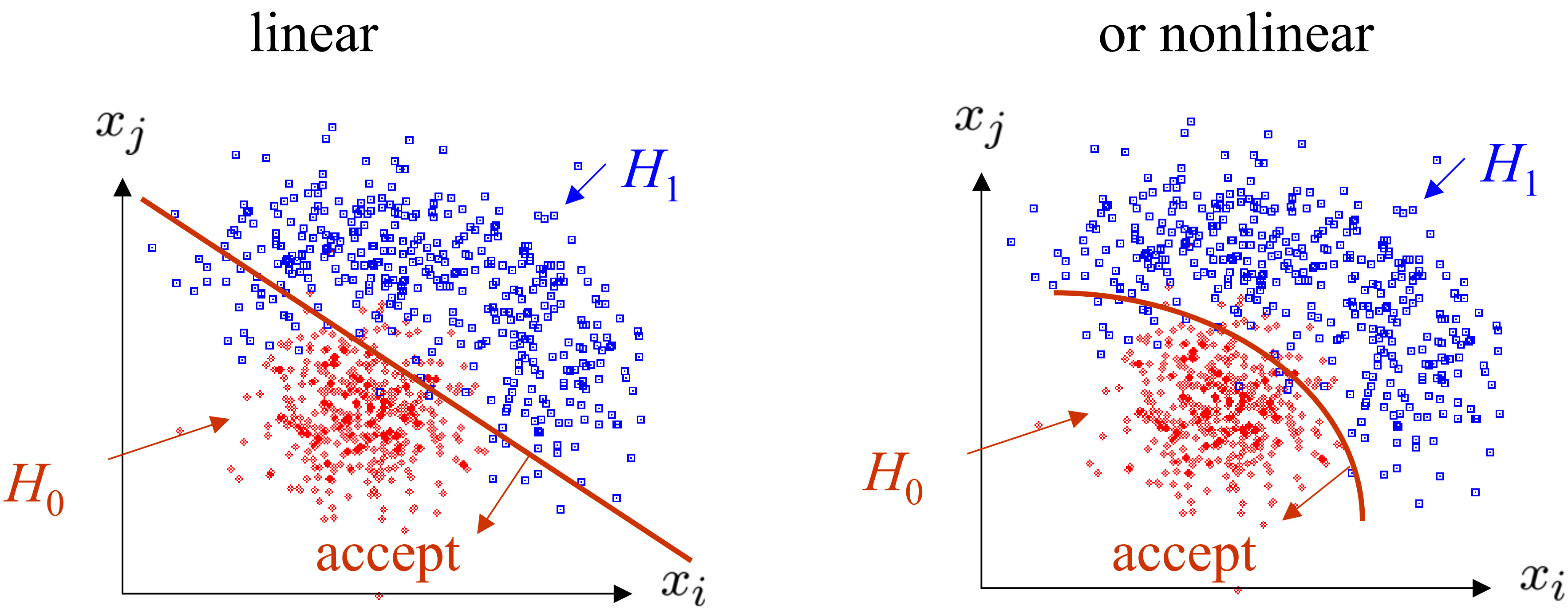
$$x_i < c_i$$

$$x_j < c_j$$



Other ways to select events

Or maybe use some other sort of decision boundary:



How can we do this in an 'optimal' way?

Test statistics

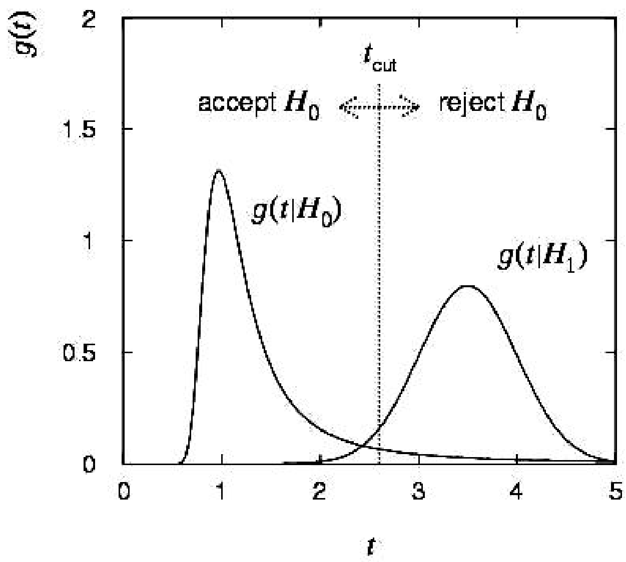
Construct a 'test statistic' of lower dimension (e.g. scalar)

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

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).

Significance level and power of a test

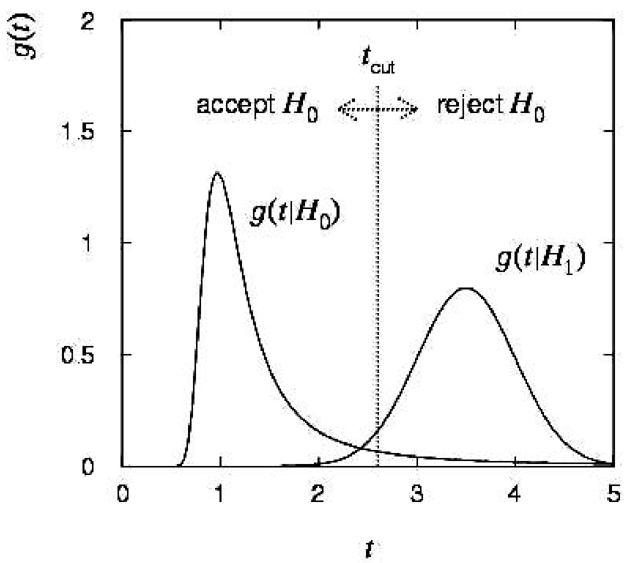
Probability to reject H_0 if it is true (error of the 1st kind):

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

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

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

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



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

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

where c is a constant which determines ε_s.

Equivalently, optimal scalar test statistic is

$$t(\vec{x}) = \frac{f(\vec{x}|\mathrm{s})}{f(\vec{x}|\mathrm{b})}$$

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

Why Neyman-Pearson doesn't always help

The problem is that we usually don't have explicit formulae for the pdfs $f(\vec{x}|\mathsf{s}),\ f(\vec{x}|\mathsf{b})$.

Instead we may have Monte Carlo models for signal and background processes, so we can produce simulated data, and enter each event into an n-dimensional histogram.

Use e.g. M bins for each of the n dimensions, total of M^n cells.

But n is potentially large, → prohibitively large number of cells to populate with Monte Carlo data.

Compromise: make Ansatz for form of test statistic $t(\vec{x})$ with fewer parameters; determine them (e.g. using MC) to give best discrimination between signal and background.

Product of one-dimensional pdfs

First rotate to uncorrelated variables, i.e., find matrix *A* such that

for $\vec{x}' = A\vec{x}$ we have $\mathrm{cov}[x'_i, x'_j] = \delta_{ij}\sigma_i^2$.

Estimate the *d*-dimensional joint pdf as the product of 1-d pdfs,

$$\hat{f}(\vec{x}) \approx \prod_{i=1}^{d} \hat{f}_i(x_i)$$ (here ***x*** decorrelated)

This does not exploit non-linear features of the joint pdf, but simple and may be a good approximation in practical examples.

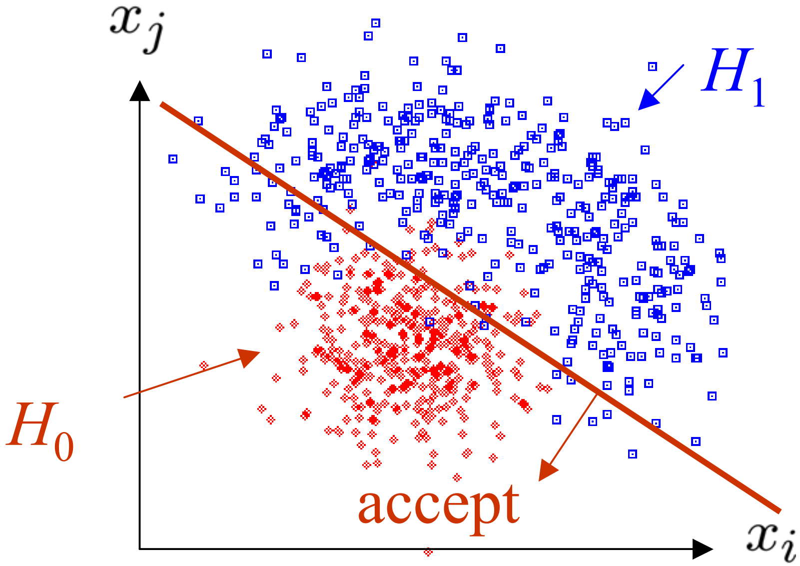
Fisher discriminant

Assume linear test statistic,

$$t(\vec{x}) = a_0 + \sum_{i=1}^{n} a_i x_i$$

and maximize 'separation' between the two classes:

Corresponds to a linear decision boundary.



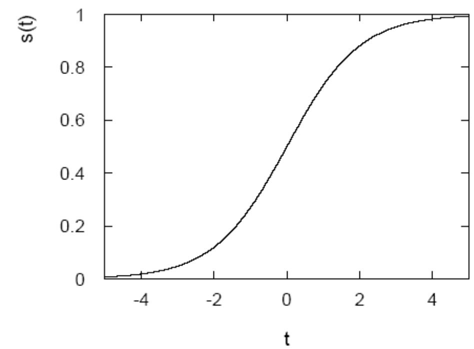
Equivalent to Neyman-Pearson if the signal and background pdfs are multivariate Gaussian with equal covariances; otherwise not optimal, but still often a simple, practical solution.

Sometimes first transform data to better approximate Gaussians.

Neural networks: the multi-layer perceptron

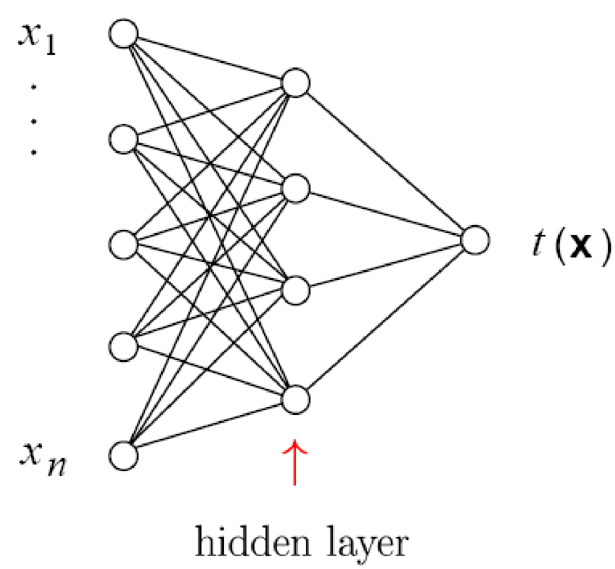
Use e.g. logistic sigmoid activation function,

$$s(u) = \frac{1}{1 + e^{-u}}$$



Define values for 'hidden nodes'

$$h_i(\vec{x}) = s\left(w_{i0} + \sum_{j=1}^{n} w_{ij} x_j\right)$$



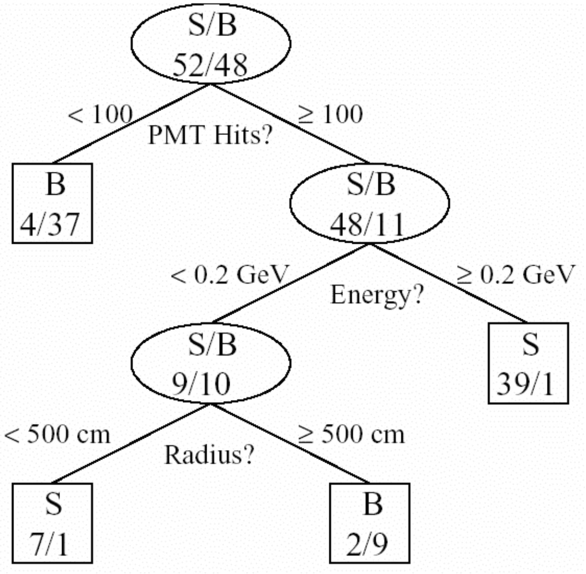
The network output is given by

$$t(\vec{x}) = s\left(a_0 + \sum_{i=1}^{n} a_i h_i(\vec{x})\right) .$$

Decision trees

A training sample of signal and background data is repeatedly split by successive cuts on its input variables.

Order in which variables used based on best separation between signal and background.



Iterate until stop criterion reached, based e.g. on purity, minimum number of events in a node.

Resulting set of cuts is a 'decision tree'.

Tends to be sensitive to fluctuations in training sample.

Example by Mini-Boone, B. Roe et al., NIM A **543** (2005) 577

Boosted decision trees

Boosting combines a number classifiers into a stronger one; improves stability with respect to fluctuations in input data.

To use with decision trees, increase the weights of misclassified events and reconstruct the tree.

Iterate → forest of trees (perhaps > 1000). For the *m*th tree,

$$T_m(\vec{x}) = \begin{cases} 1 & \vec{x} \text{ in signal acceptance region} \\ -1 & \text{otherwise} \end{cases}$$

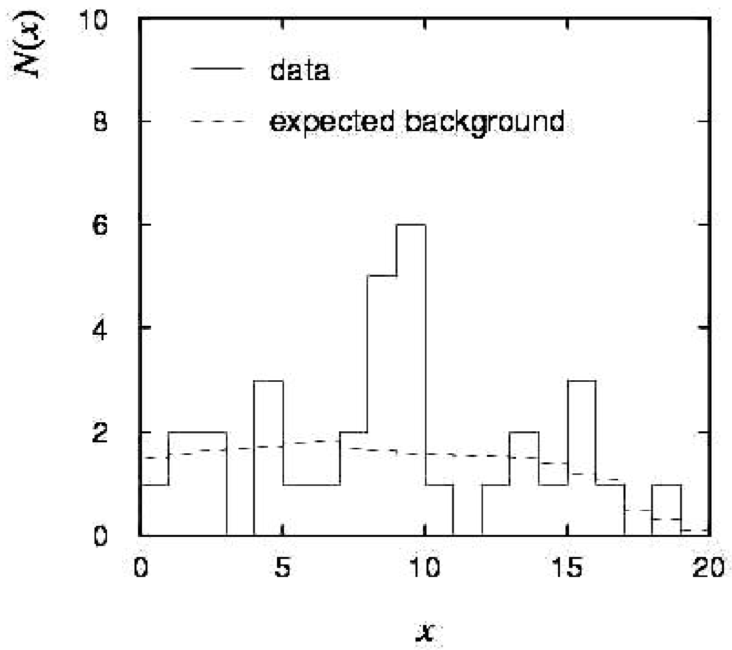
Define a score α_m based on error rate of *m*th tree.

Boosted tree = weighted sum of the trees: $T(\vec{x}) = \sum_m \alpha_m T_m(\vec{x})$

Algorithms: AdaBoost (Freund & Schapire), ε-boost (Friedman).

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?

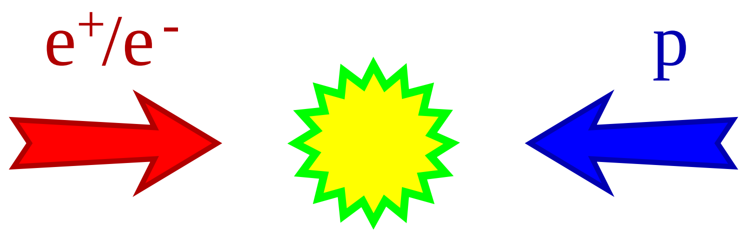
→ freeze cuts, repeat analysis with new data

Should we publish????

Introduction

HERA

electron(positron)-proton collider at DESY



27.5 GeV

820 GeV 1994-97

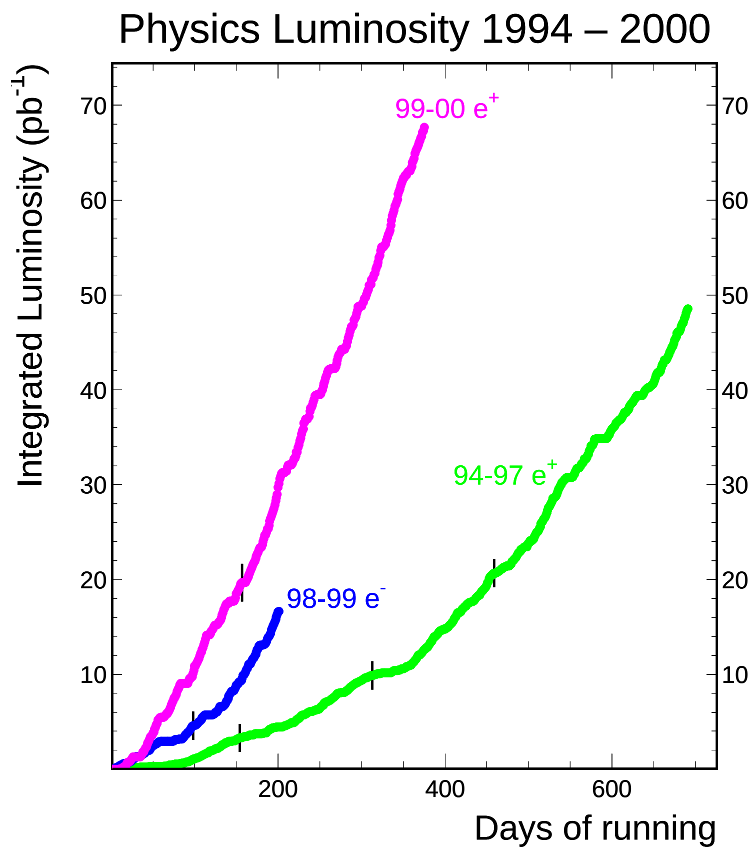
920 GeV 1998-00

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}

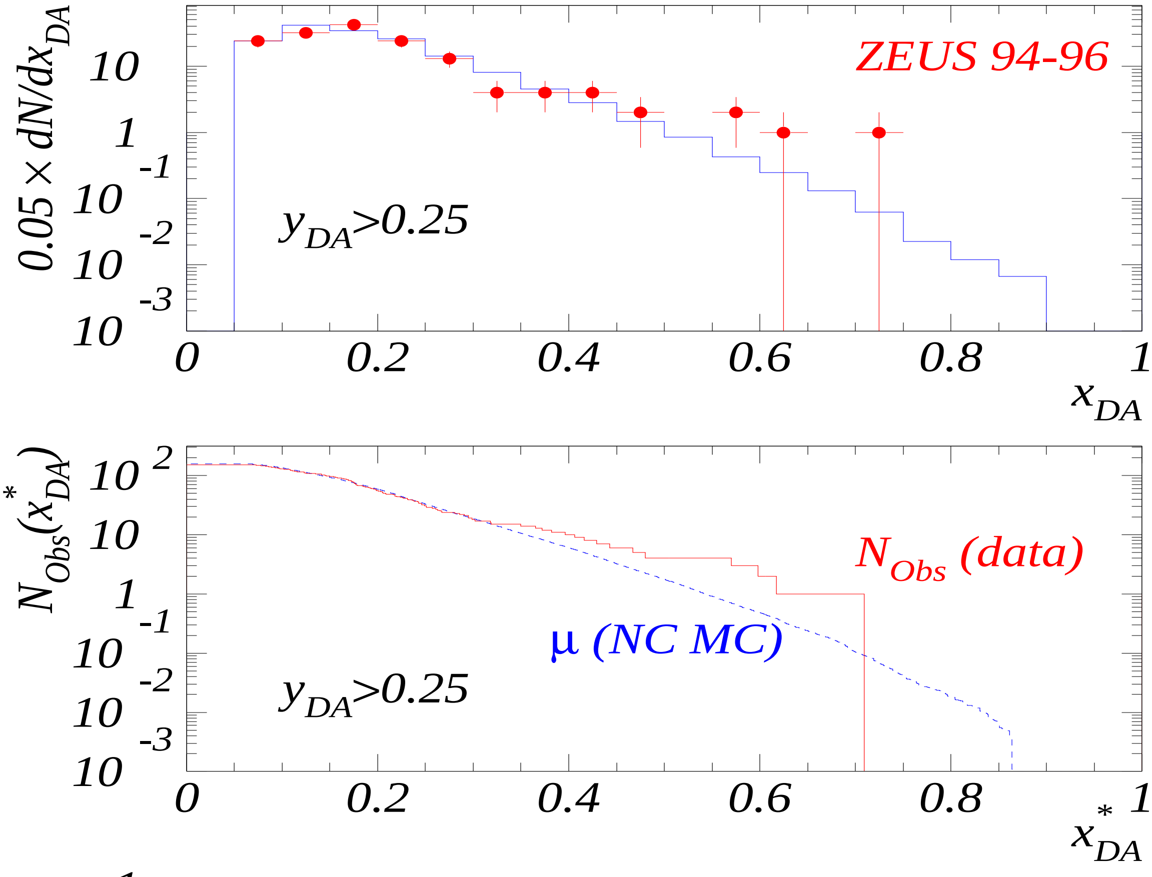
Total of about $\sim$ 220 pb^{-1} of $e^{\pm}p$ data available.

Very successful HERA operation in 1999-2000:



Significance Analysis

Excess in x



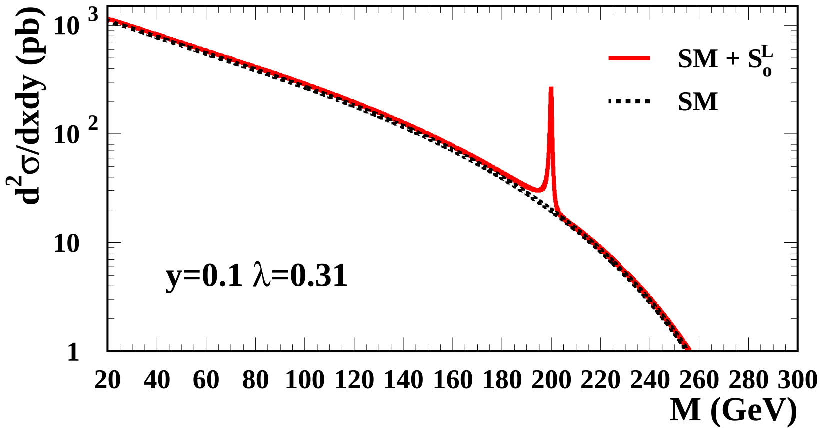
Leptokwarki - model

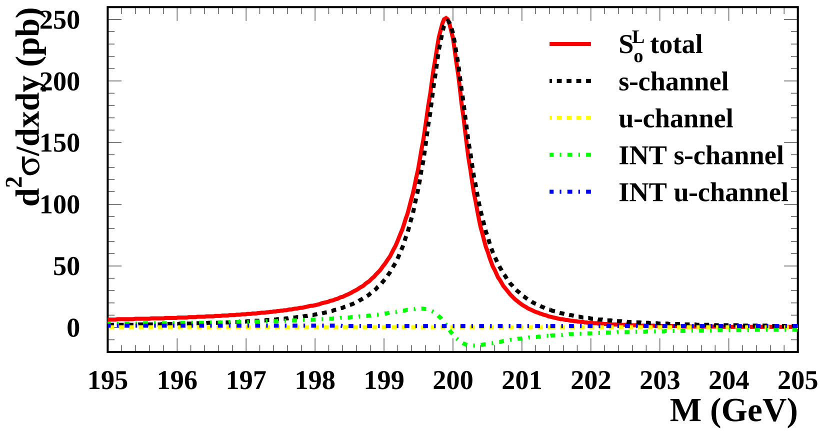
LEPTOKWARKI:

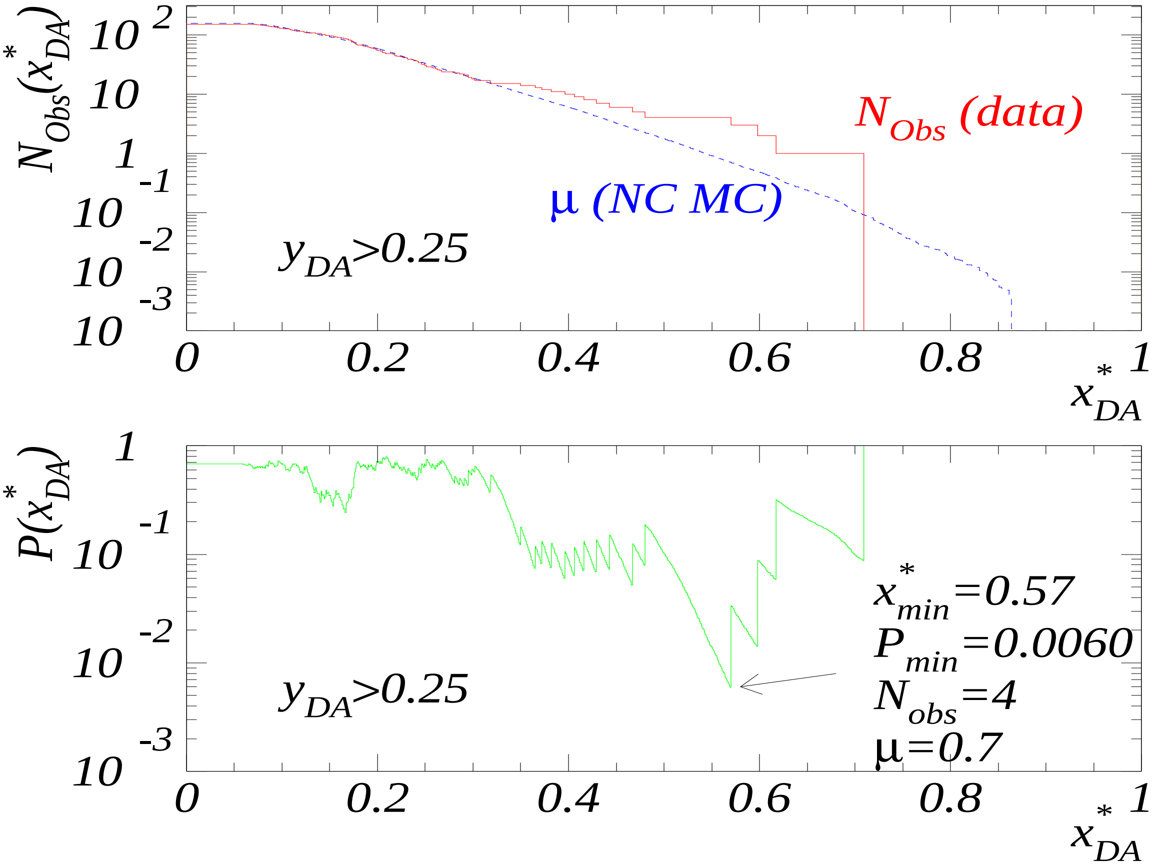
Cząstki fundamentalne sprzęgające się do leptonów i kwarków, które:

- 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).



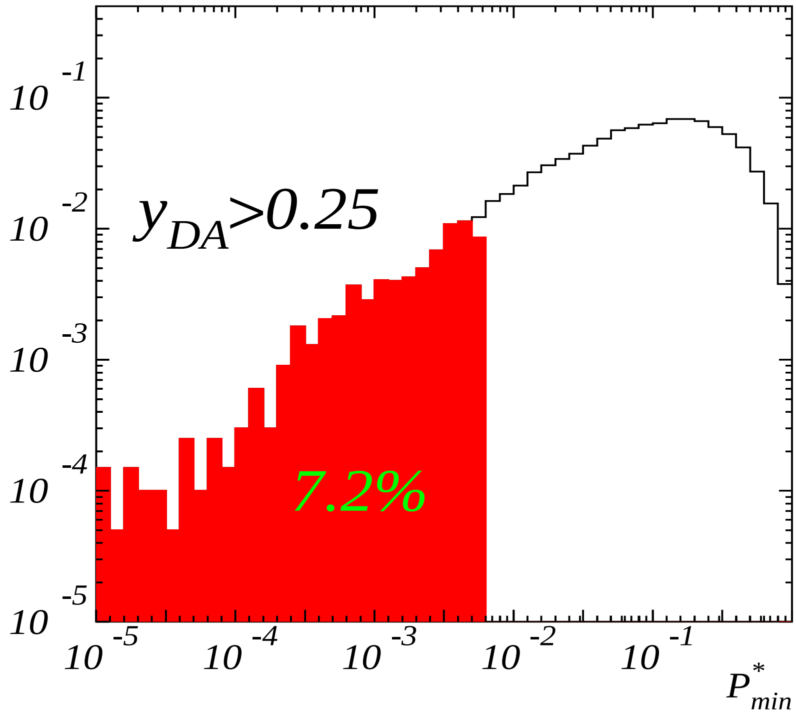




$$N_{\rm obs}(x^*_{\rm DA}) \;=\; \int\limits_{x^*_{\rm DA}}^{1} \; dx_{\rm DA}\; dN/dx_{\rm DA}$$

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

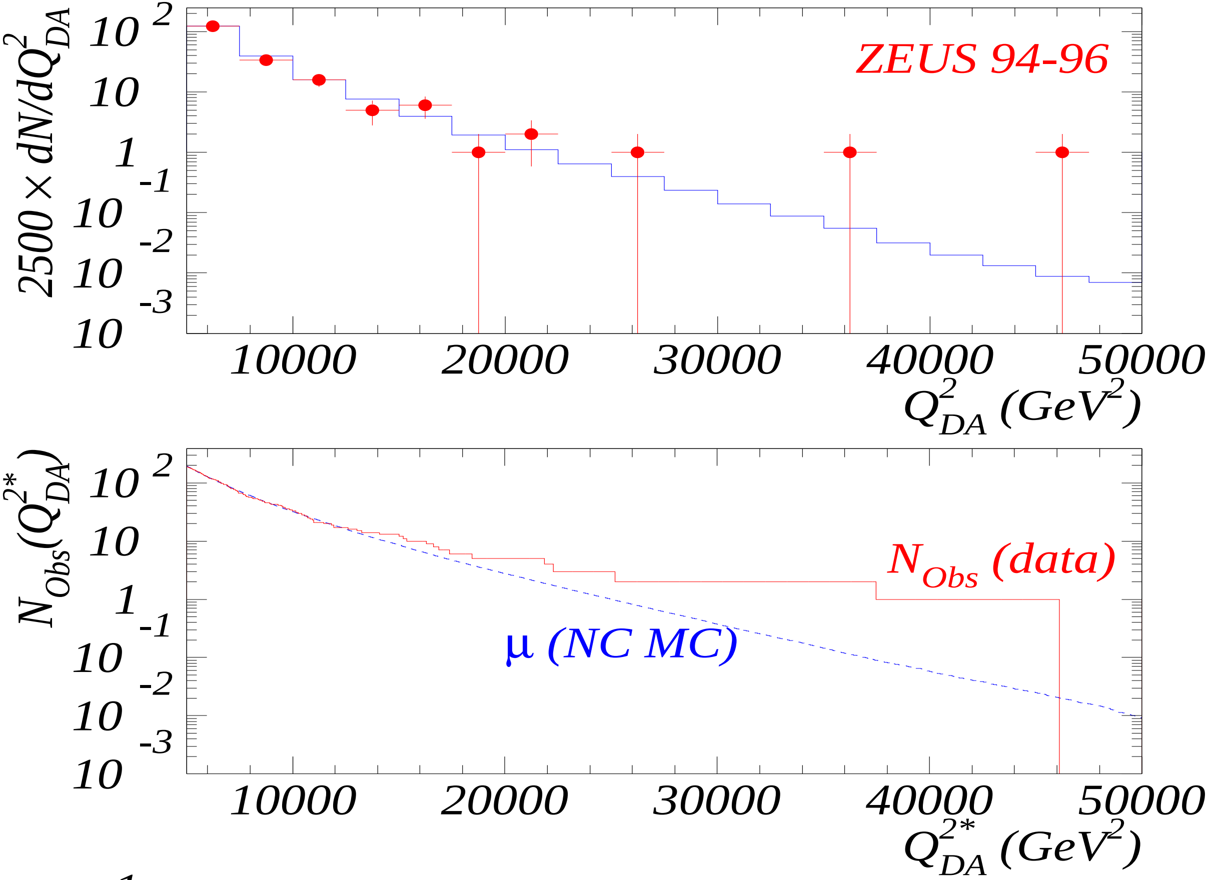
Excess in x — continued

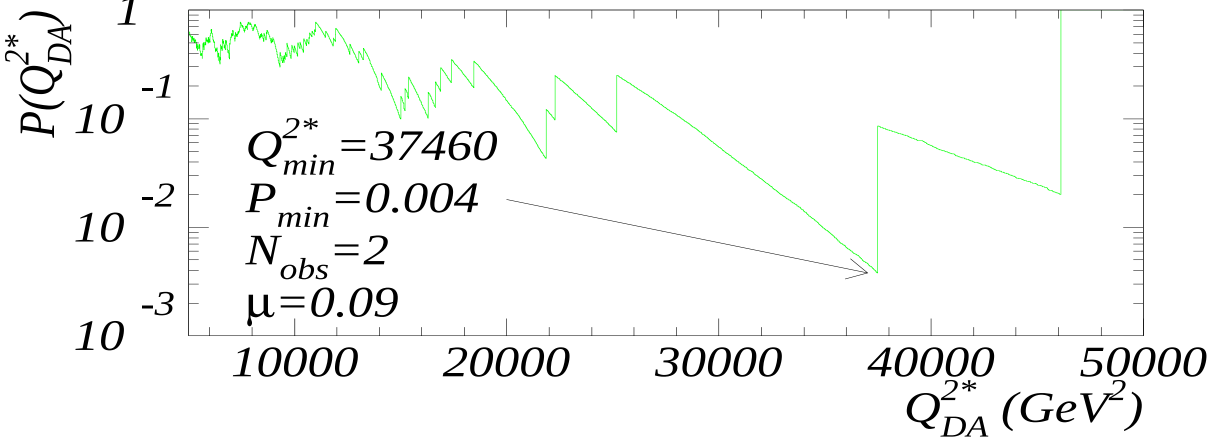


Minimal Poisson probabilities of the x_{DA} distributions for different y_{DA} cuts

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

Excess in Q^2

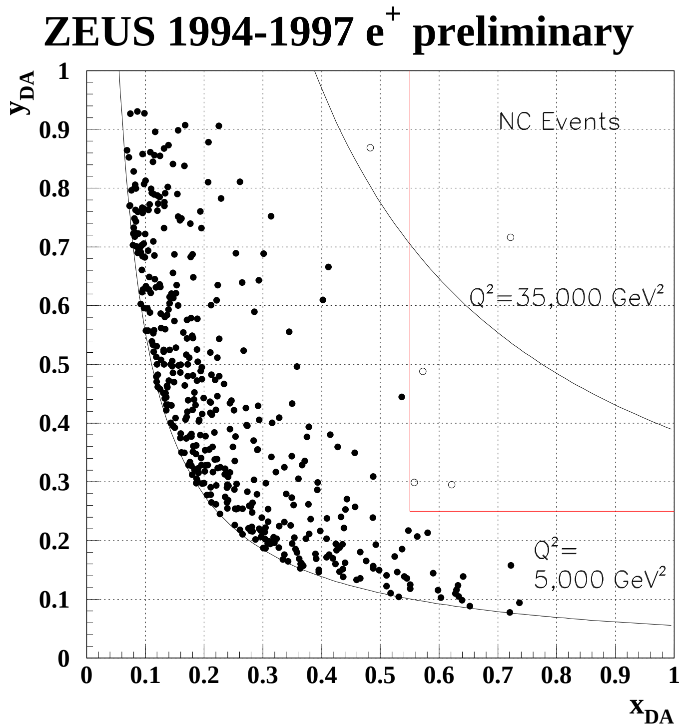

ZEUS 94-96
2500 × dN/dQ²_DA
Q²_DA (GeV²)
N_Obs(Q²*_DA)
N_Obs (data)
μ (NC MC)
Q²*_DA (GeV²)
10000
20000
30000
40000
50000

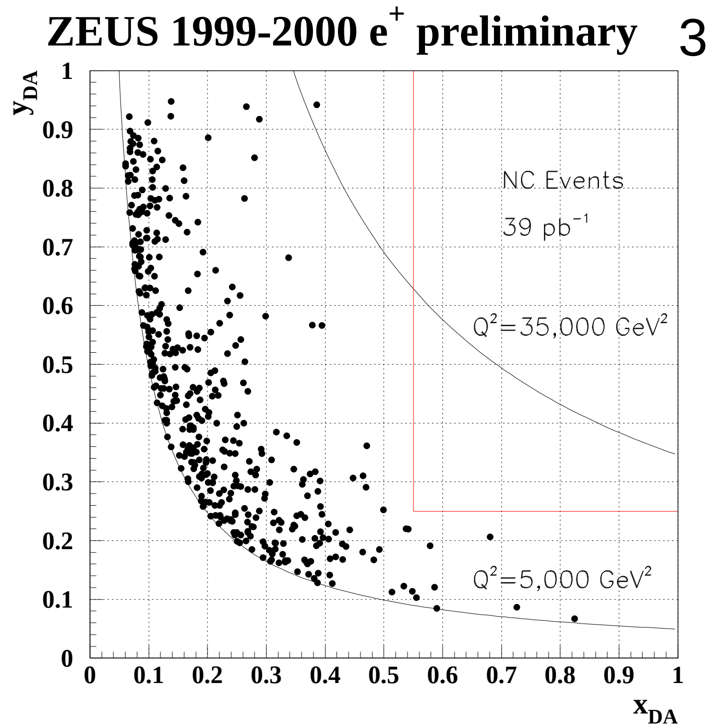


The probability for a simulated experiment to obtain $\mathcal{P}_{min}((Q^2_{\rm DA})^*) < 0.004$ is 6.0%.

High-Q^2 DIS

High-Q^2 / high-x excess ?





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)

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: $E[n_i] = s_i(\theta) + b_i$

signal background

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

Probability is product of Poisson probabilities:

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

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 $P(e)$

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

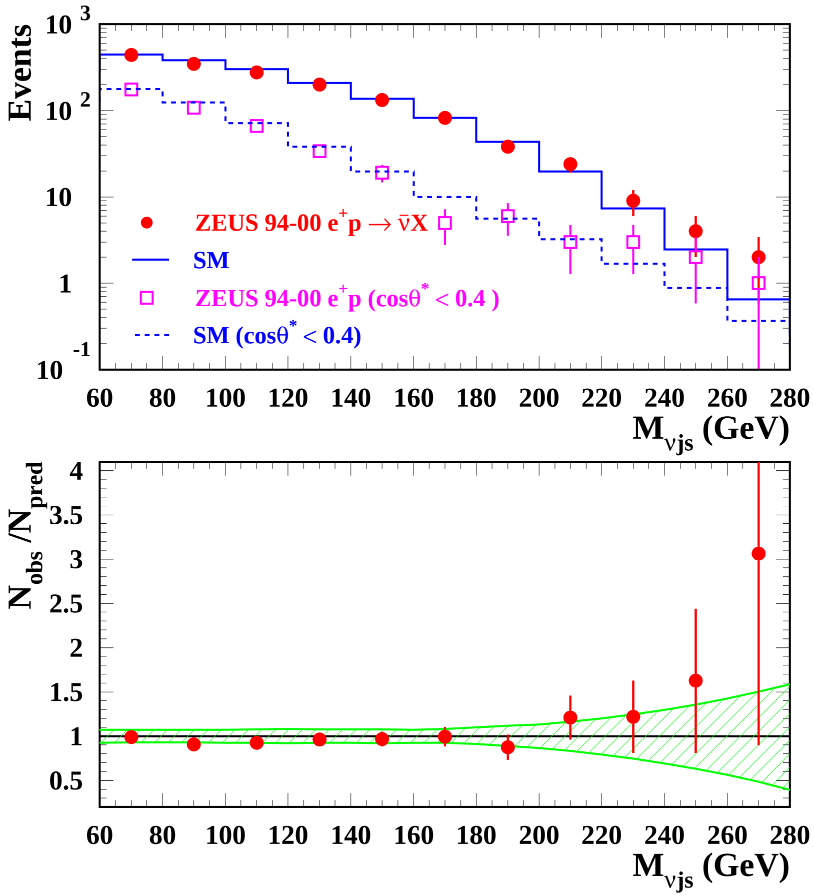
  $P(C \,|\, e) \neq P(e \,|\, C)$

Or, with a real life example:

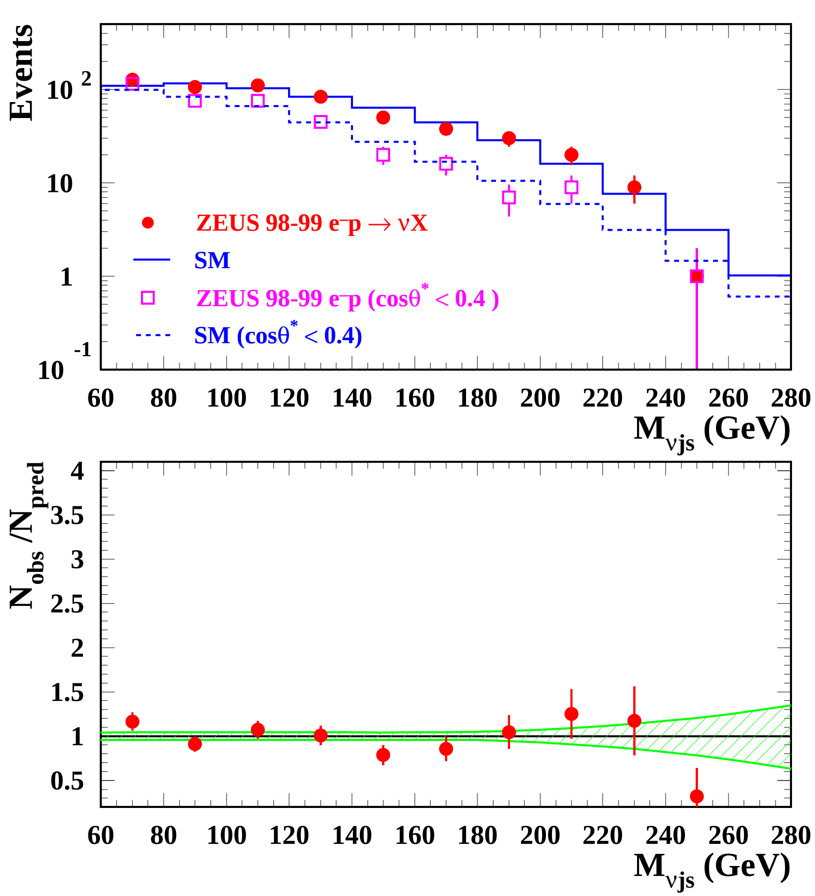
A = female or male	$P(\text{pregnant} \,|\, \text{female}) \approx 0.5\%$
B = pregnant or non-pregnant	$P(\text{female} \,|\, \text{pregnant}) \ggg 1\%$

Rozkład masy niezmienniczej νq i porównanie z SM

Dane e^+p:



Dane e^-p:



Dobra zgodność danych z przewidywaniami Modelu Standardowego → brak sygnału LQ

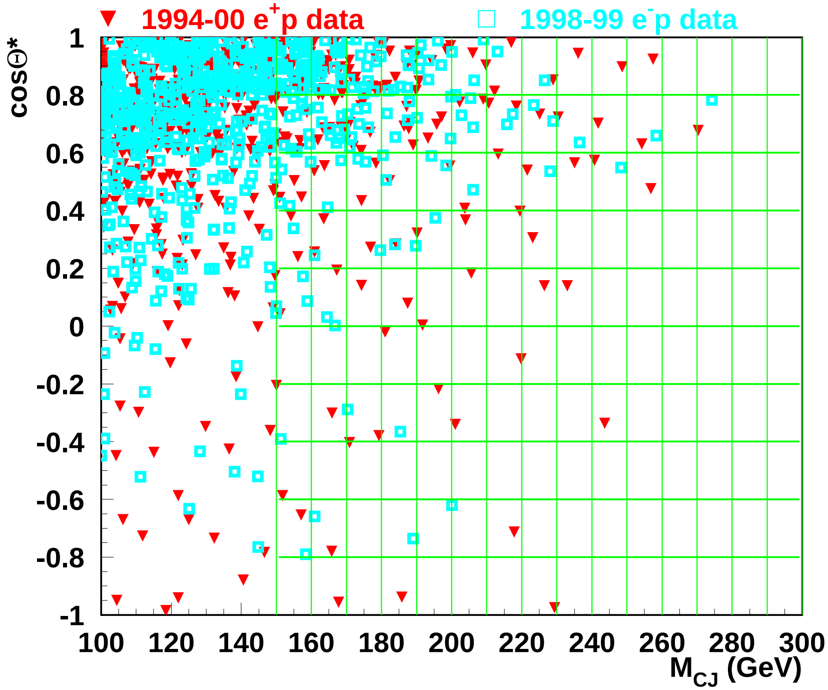


Figure 5.6: Distribution of selected NC DIS type events in the M_{ejs} - $\cos\theta^{\star}_{ejs}$ plane, for the $e^{-}p$ and the $e^{+}p$ data. The grid indicates bins used in the likelihood analysis.

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:

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

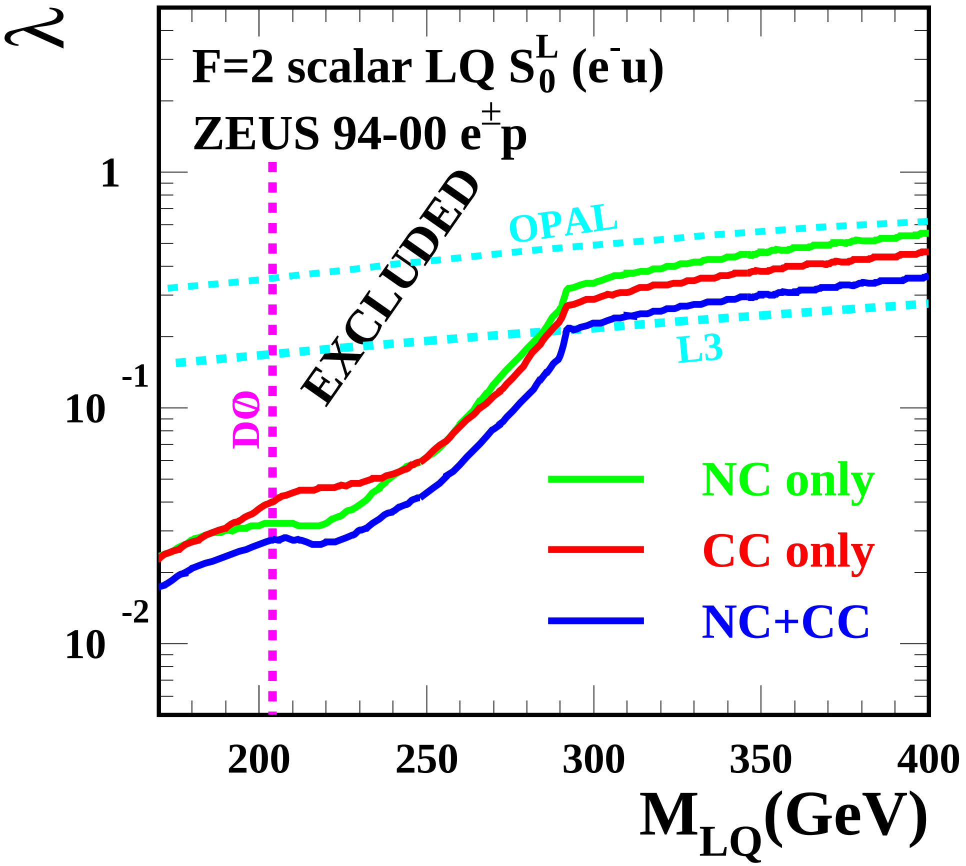
In this analysis we adopted the Bayesian approach and the upper limit on the coupling strength as a function of M_{LQ}, $\lambda_{limit}(M_{LQ})$, was obtained by solving the equation[1]

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

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

Dwa podejścia do wyznaczania limitów:

Bayesowskie

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

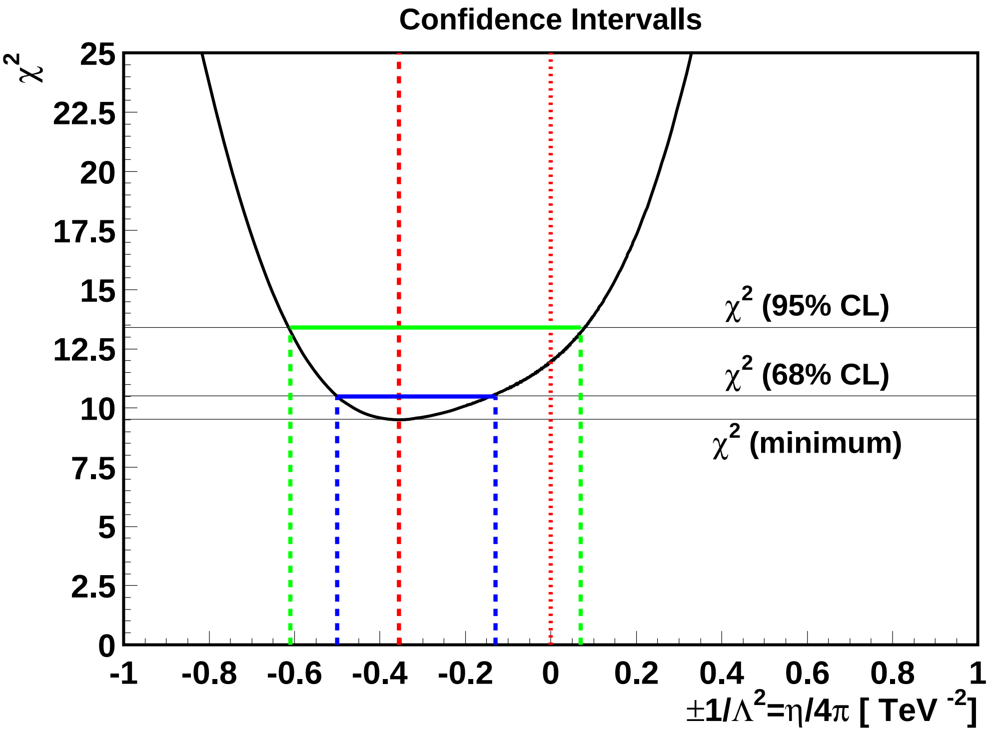
Z tego rozkładu liczymi limit tak jak dla zwykłago rozkładu prawdopodobieństwa zmiennej losowej (np. rozkładu Gaussa). Np. wykluczone są wartości parametru x większe niż X_{lim} jeśli

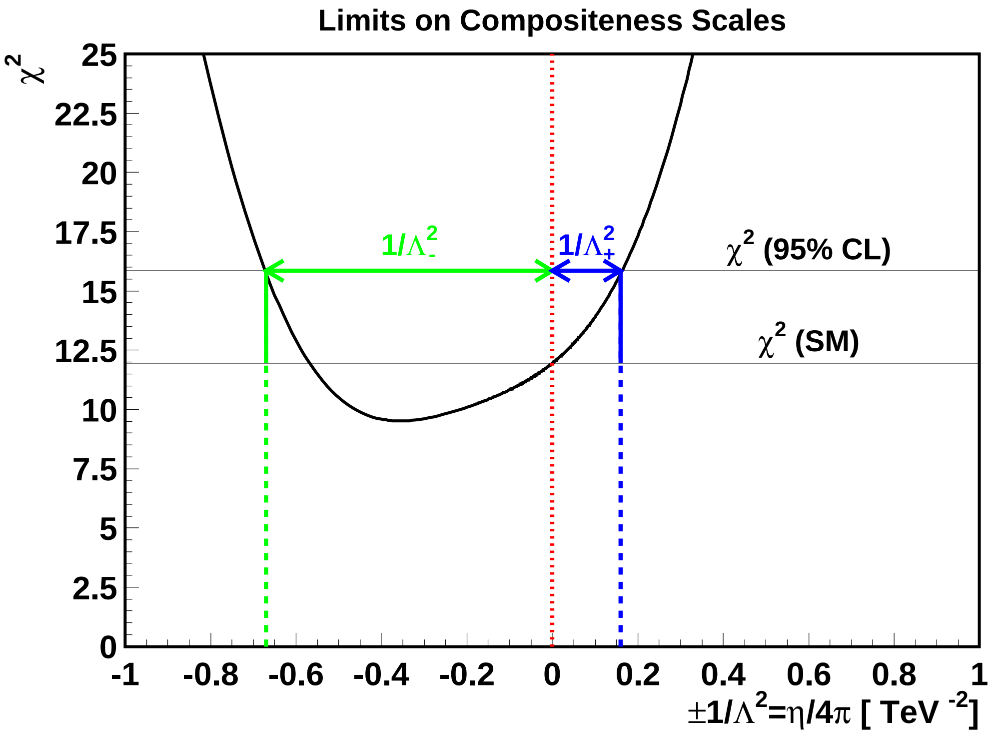
$$P(x>X_{lim}) = 1 - CL$$

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





Limits on Compositeness Scales
χ^2
25
22.5
20
17.5
15
12.5
10
7.5
5
2.5
0
$1/\Lambda_-^2$
$1/\Lambda_+^2$
χ^2 (95% CL)
χ^2 (SM)
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
$\pm 1/\Lambda^2 = \eta/4\pi$ [TeV $^{-2}$]

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)

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.

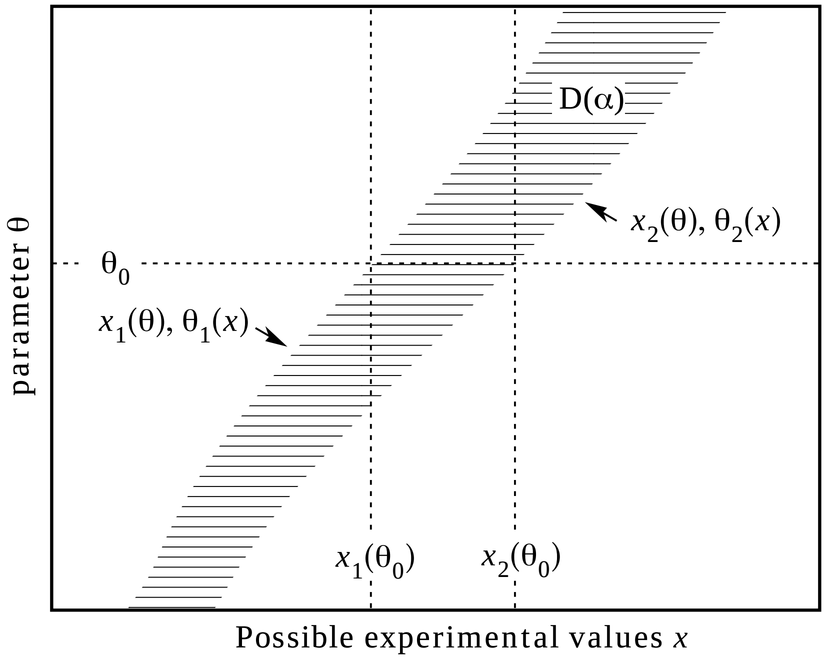
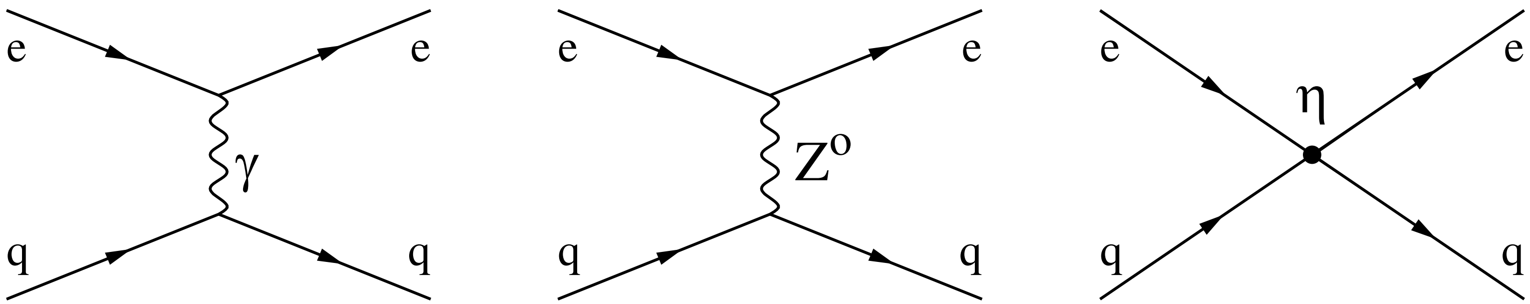


Figure 32.3: Construction of the confidence belt (see text).

Models

Contact Interactions

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



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

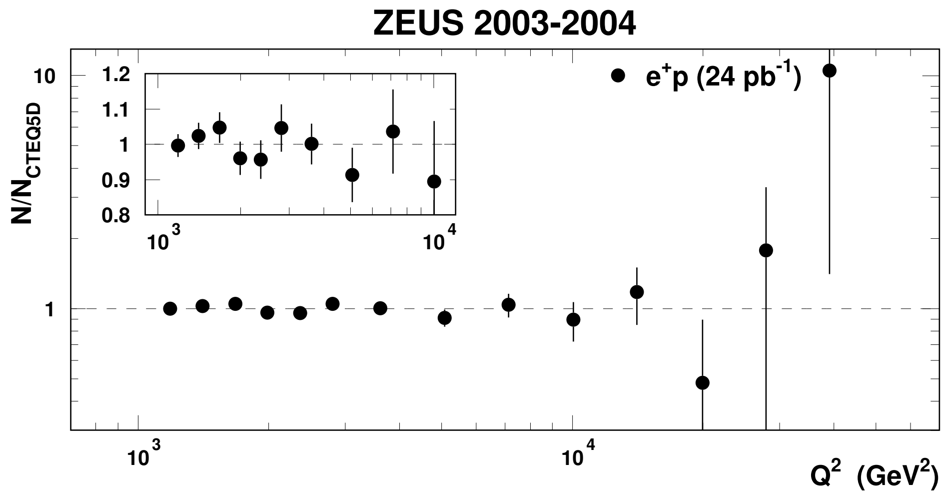
γ $\qquad Z^\circ$ $\qquad$?

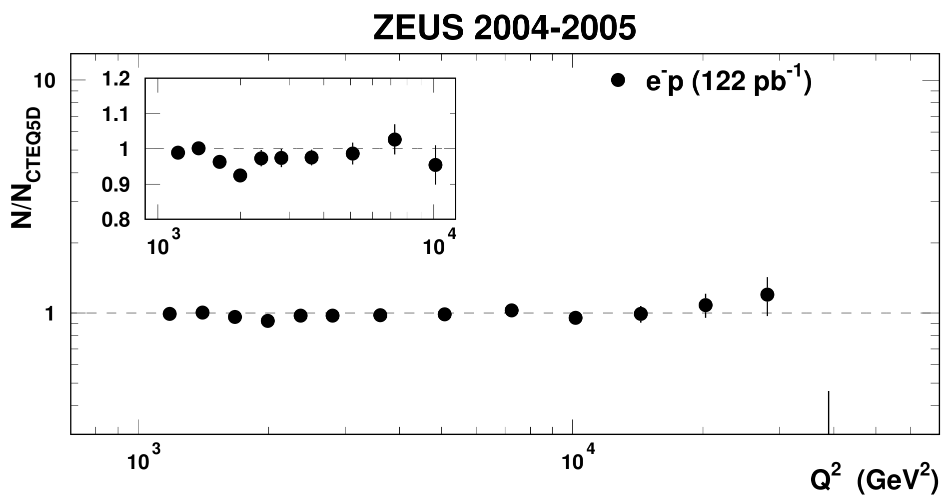
$\eta^{eq}_{\alpha\beta}$ - 4 possible couplings for every flavor q

Different models assume different helicity structure of new interactions

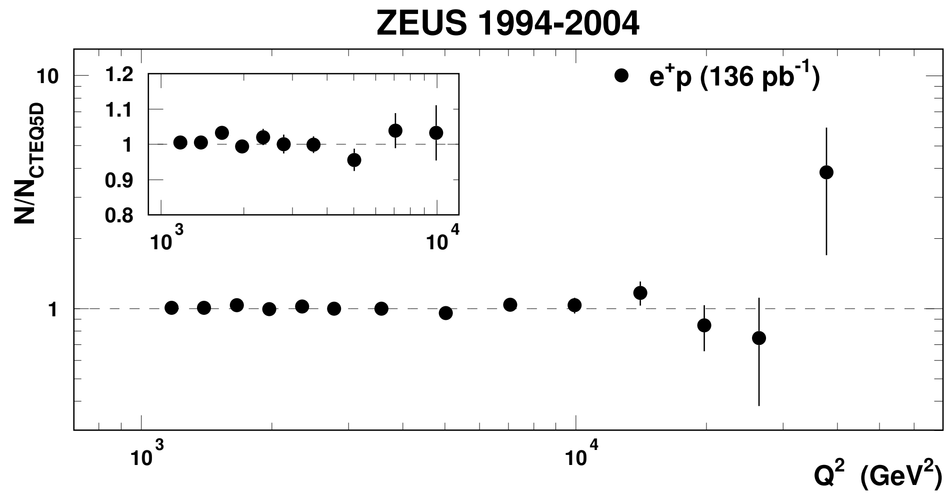
Data and analysis

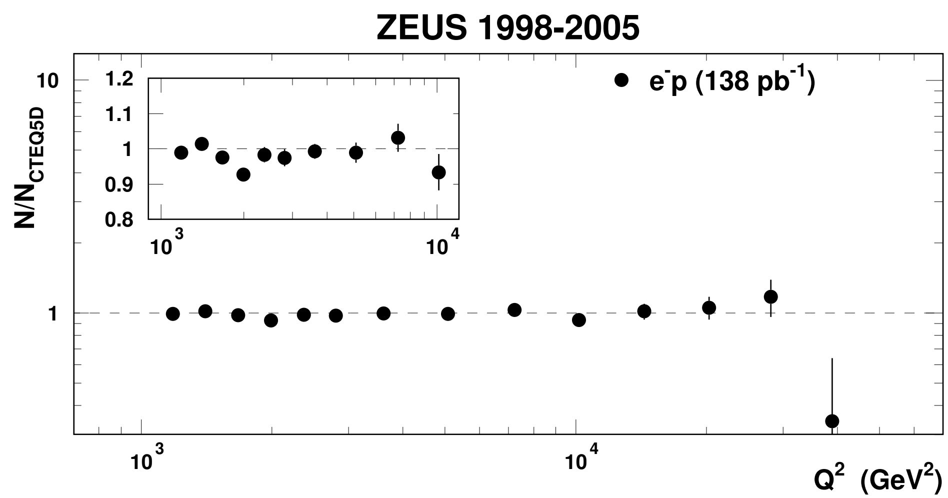
HERA-II data





All HERA data





Analysis

Probability function

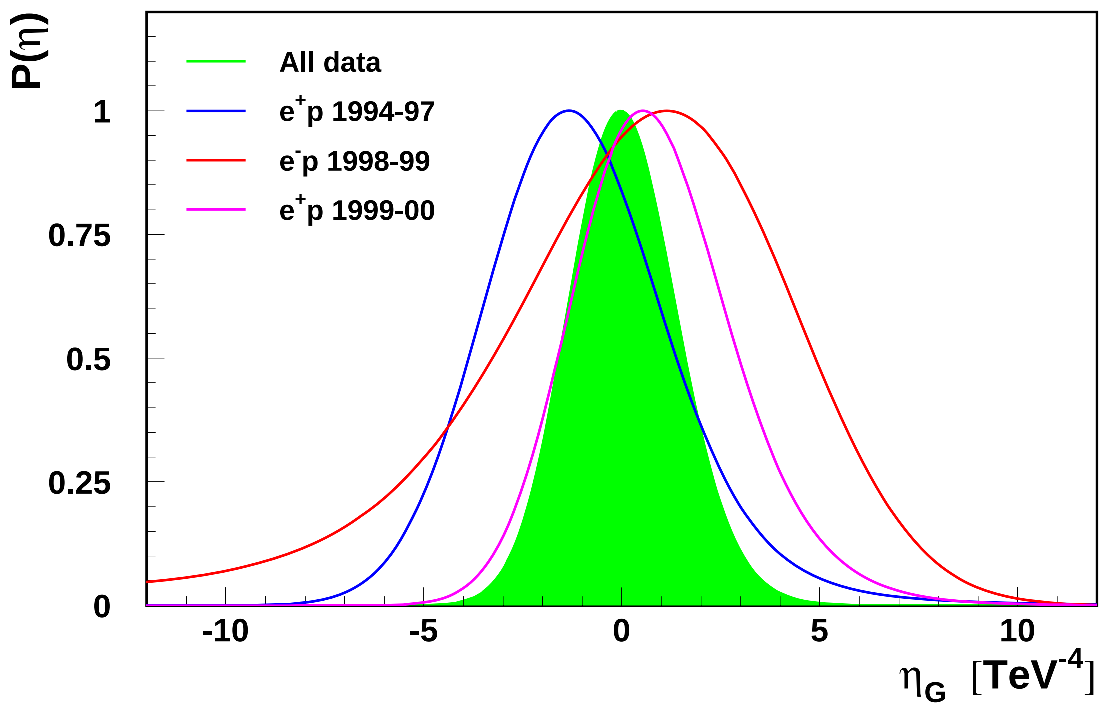
Observed numbers of events in Q^2 bins n_i are compared with the CI model expectations $\mu_i(\eta_G)$ using the probability function:

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

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

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

Resulting probability function for the nominal data:



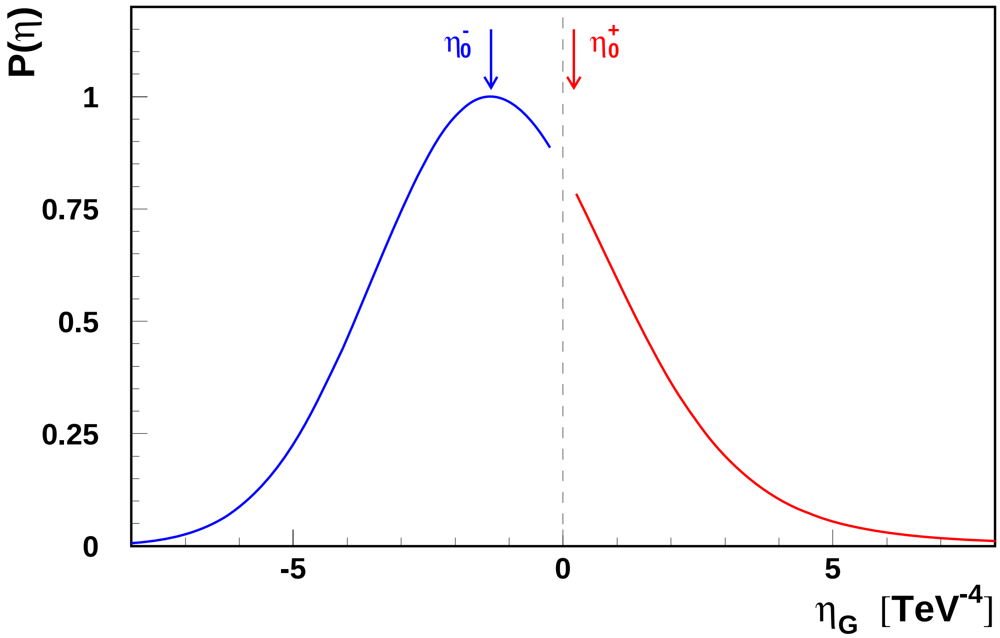
normalized to $\max_\eta P(\eta_G) = 1$

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:

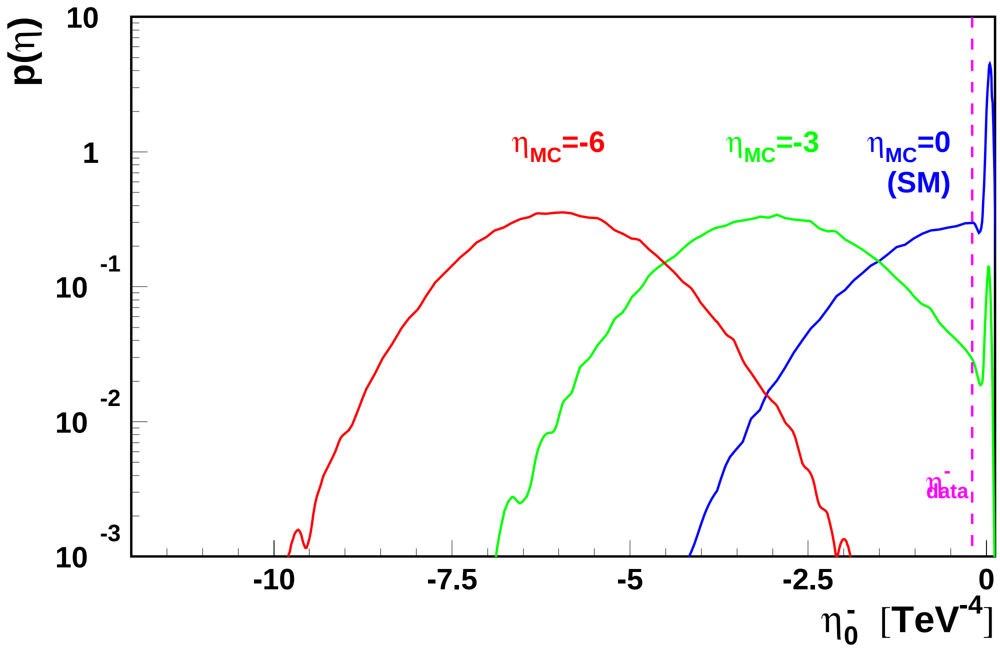
$$\eta_\circ^- = -0.02 \text{ TeV}^{-4}$$
$$\eta_\circ^+ = 0$$

very good agreement with the Standard Model

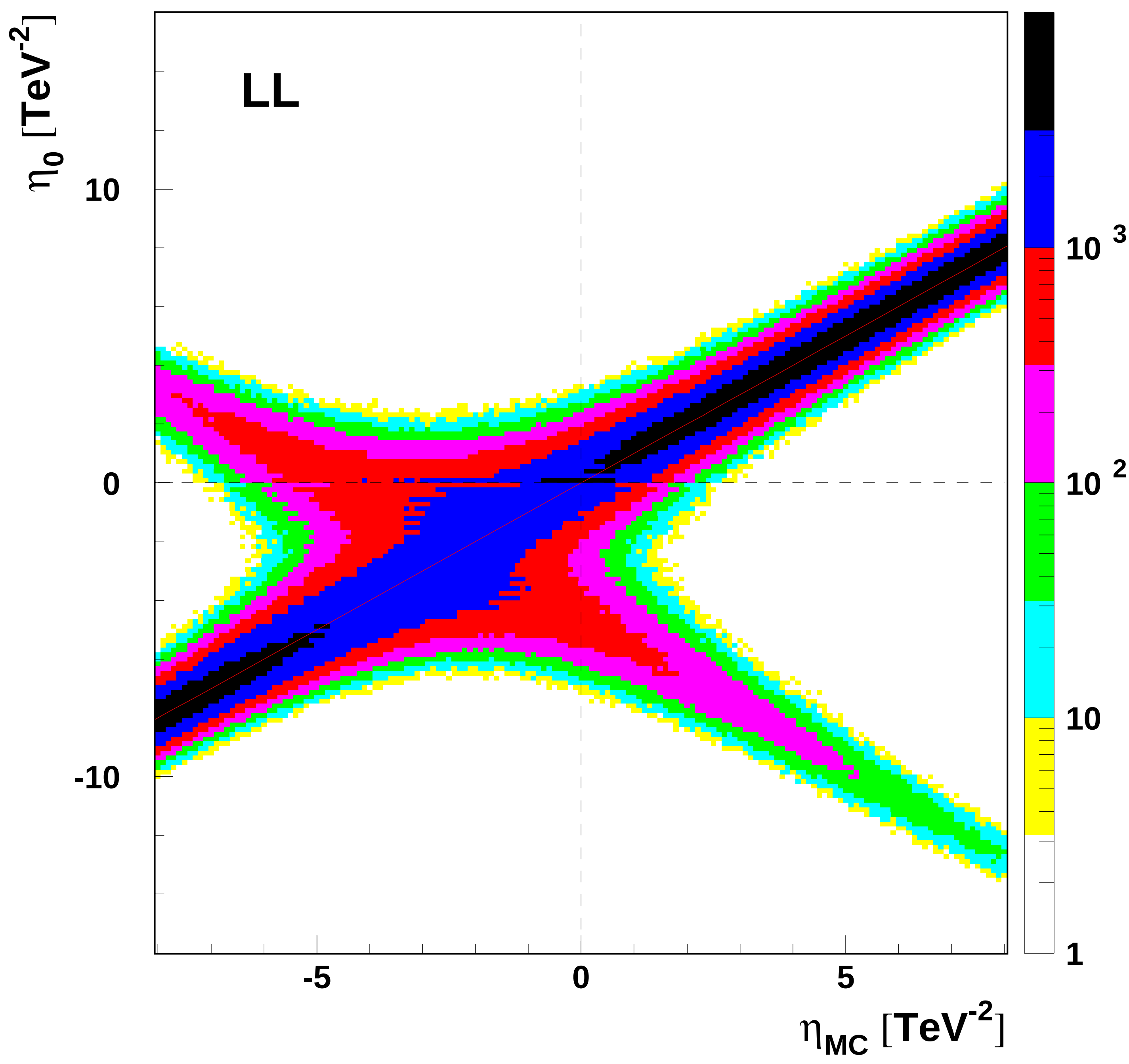
Analysis

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.



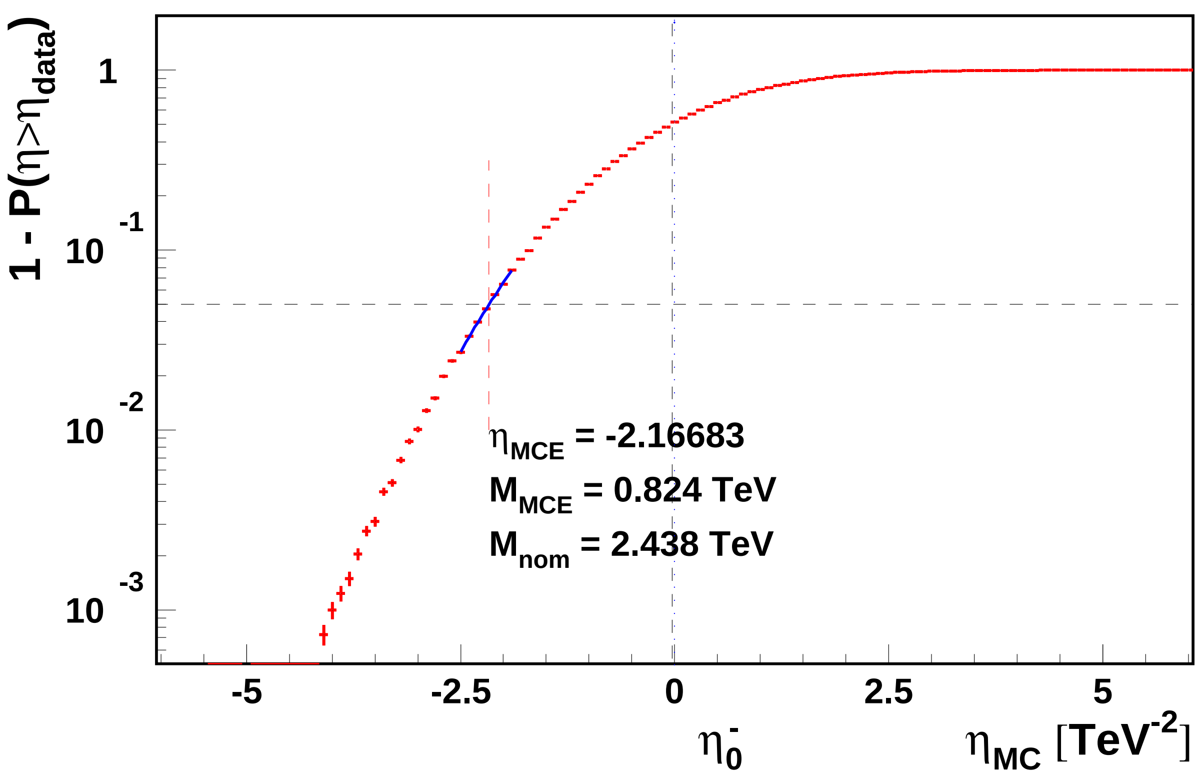
LL
η_0 [TeV^{-2}]
10
0
-10
-5
0
5
η_{MC} [TeV^{-2}]
10^3
10^2
10
1

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



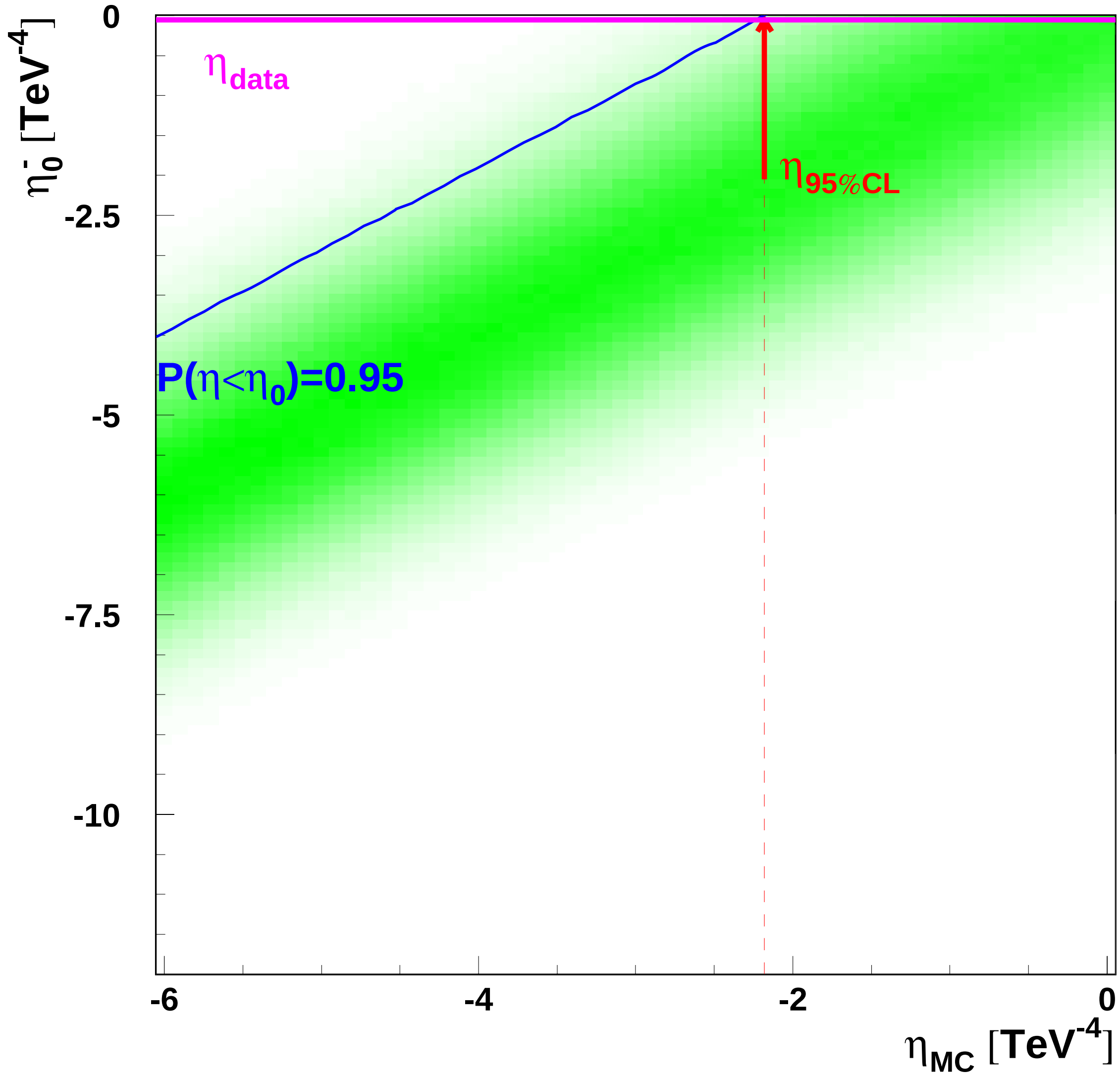
$\eta_G < -2.167$ TeV^{-4} on 95% CL

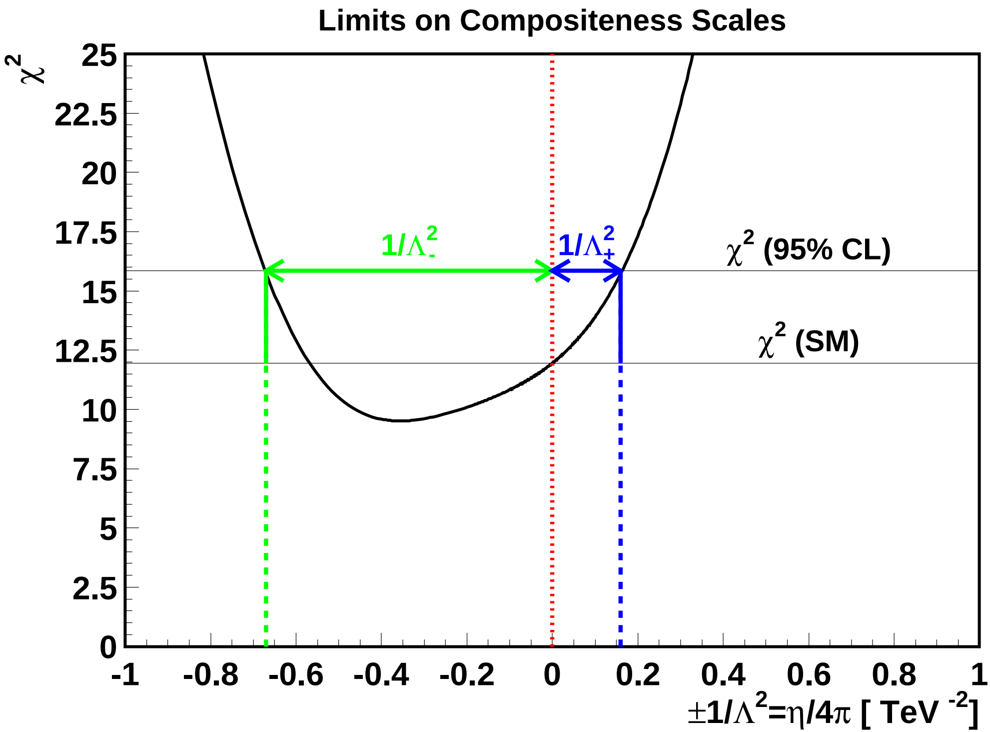
Analysis

Limit setting

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

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





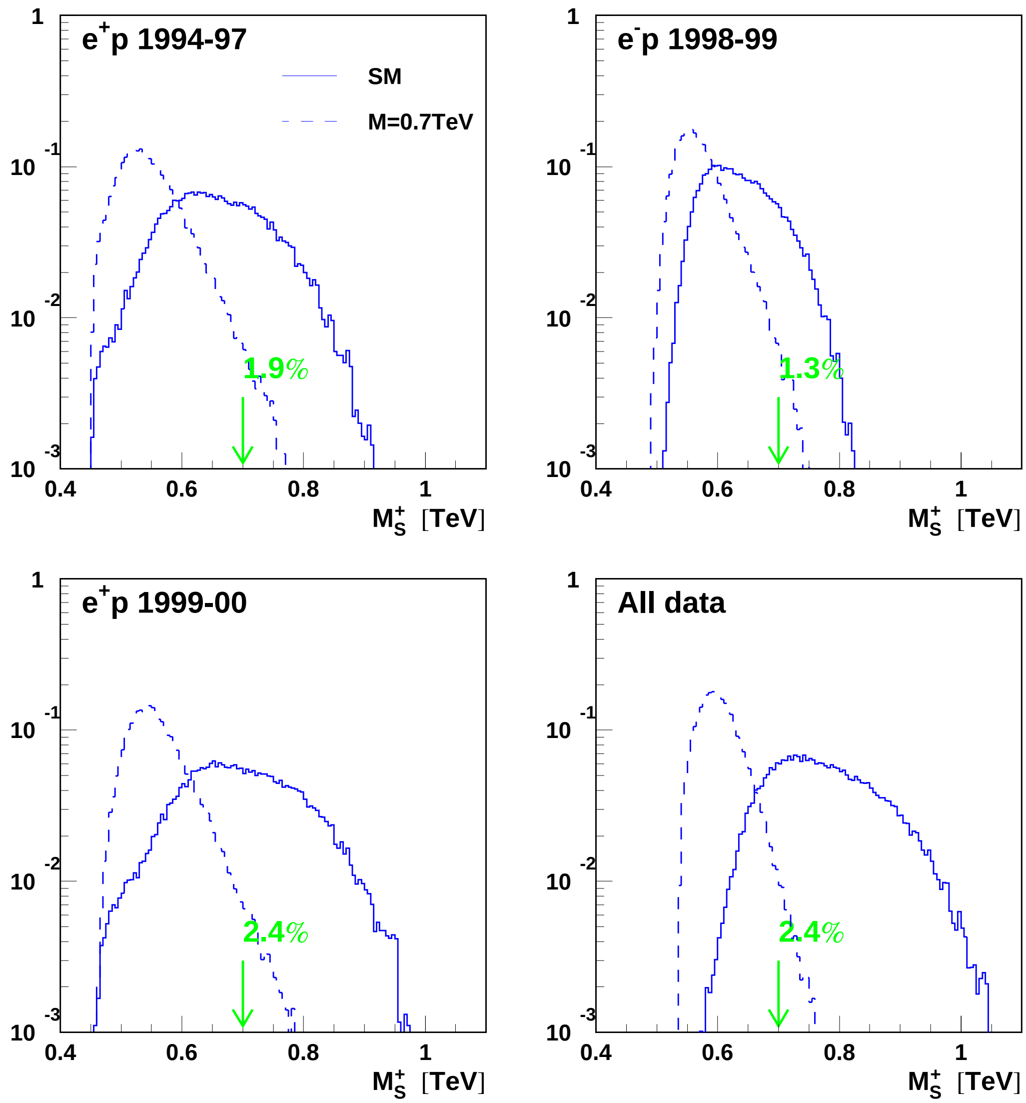
Limits on Compositeness Scales
χ2
25
22.5
20
17.5
15
12.5
10
7.5
5
2.5
0
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
±1/Λ2=η/4π [TeV -2]
1/Λ2-
1/Λ2+
χ2 (95% CL)
χ2 (SM)

Comparison with expectations

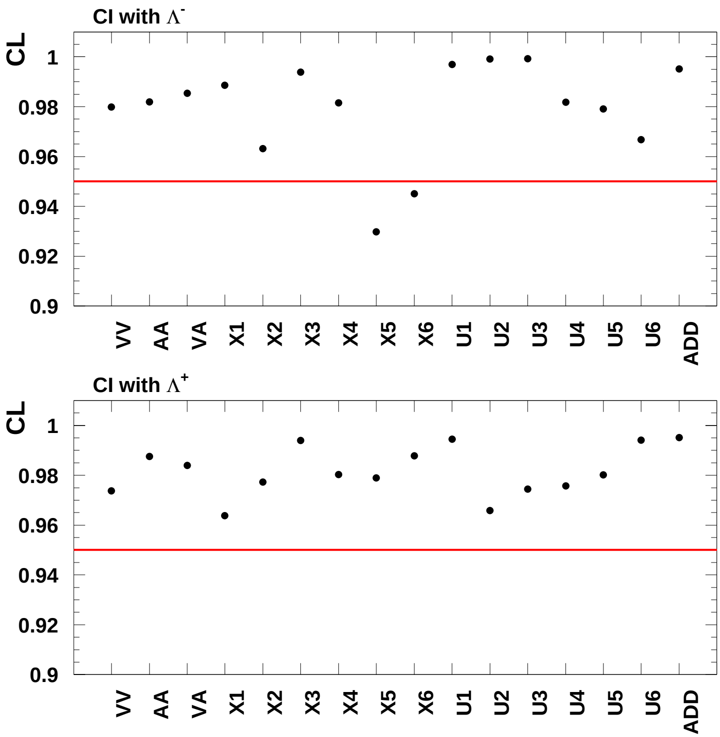
Large Extra Dimensions

Limits expected for M_S^{+} = 0.7 TeV

H1 method



$\Rightarrow$ H1 method results rather in 97.5% CL limits !?



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.