Υπολογιστική Φυσική Στοιχειωδών Σωματιδίων

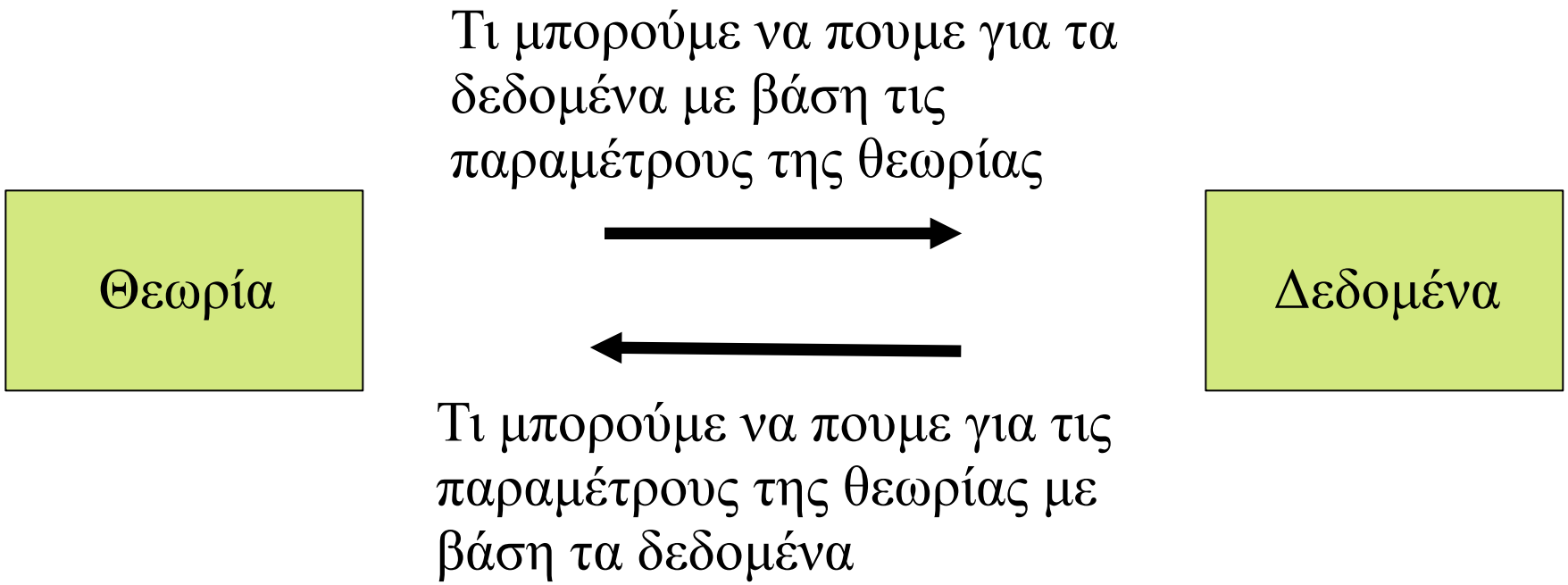
Εκτιμητές-Estimators

Υπολογισμός των παραμέτρων & προσαρμογή (fitting) των δεδομένων

- Ενδιαφέρει ο υπολογισμός των παραμέτρων της συνάρτησης πυκνότητας πιθανότητας (pdf), π.χ της f(x,θ) από ένα set μετρήσεων
- Οι συχνότερα χρησιμοποιούμενες μέθοδοι ειναι:
  - της μεγιστοποίησης της πιθανότητας (Maximum Likelihood)
  - των ελαχίστων τετραγώνων (Least squares)
- Πλεονεκτήματα και μειονεκτήματα των δύο μεθόδων
- Θα αναφερθούμε στα πολύ βασικά κάθε μεθόδου

Σύγκριση δεδομένων με την θεωρία Εκτιμητές των παραμέτρων κατανομών

- Περί της διαδικασίας 'εκτίμησης' :



Τι είναι ένας εκτιμητής

$$\hat{\mu}(\{x\}) = \frac{1}{N}\sum_i x_i$$

$$\hat{\mu}(\{x\}) = \frac{x_{\max} + x_{\min}}{2}$$

$$\hat{V}(\{x\}) = \frac{1}{N}\sum_i (x_i - \hat{\mu})^2$$

$$\hat{V}(\{x\}) = \frac{1}{N-1}\sum_i (x_i - \hat{\mu})^2$$

- Η διαδικασία που μας παρέχει την τιμή για μία παράμετρο ή μία ιδιότητα μιάς κατανομής, σαν συνάρτηση των δεδομένων μας: $\hat{\theta}$

Πότε ένας εκτιμητής είναι αξιόπιστος?

- Όταν έχει συνέπεια (consistent) $\underset{N\to\infty}{Limit}(\hat{a}) = a$
- Όταν είναι αμερόληπτος (unbiased) b= E[$\hat{\alpha}$]-α

$$\langle \hat{a} \rangle = \iiint \ldots \hat{a}(x_1, x_2, \ldots) P(x_1; a) P(x_2; a) P(x_3; a) \ldots dx_1 dx_2 \ldots = a$$

- Όταν η απόκλιση είναι ελάχιστη

$$V(\hat{a}) = \left\langle \left(\hat{a} - \langle \hat{a} \rangle\right)^2 \right\rangle \qquad \text{(minimum)}$$

- Θα πρέπει να είναι γρήγορος (efficient)

Η συνάρτηση πιθανότητας (Likelihood function)

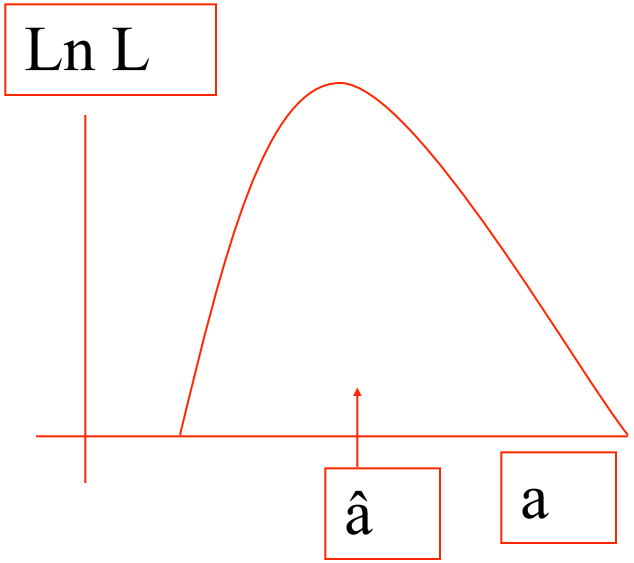
- Σύνολο δεδομένων $\{x_1, x_2, x_3, \ldots x_N\}$ (κάθε x_ι μπορεί να είναι πολυδιάστατο)
- Η πιθανότητα για κάθε x_ι εξαρτάται από παράμετρο α που επίσης μπορεί να είναι πολυδιάστατη
- Η συνολική πυκνότητα πιθανότητας είναι:

$$P(x_1;a)\, P(x_2;a)\, P(x_3;a) \ldots P(x_N;a) = L(x_1, x_2, x_3, \ldots x_N ;a)$$

Likelihood

Υπολογισμός της Maximum Likelihood (ML)

- Από το σύνολο των δεδομένων $\{x_1, x_2, x_3, \ldots x_N\}$ υπολογίζουμε την τιμής της παραμέτρου α που μεγιστοποιεί την Likelihood
- Στην πράξη συνήθως μεγιστοποιούμε τον λογάριθμο της L

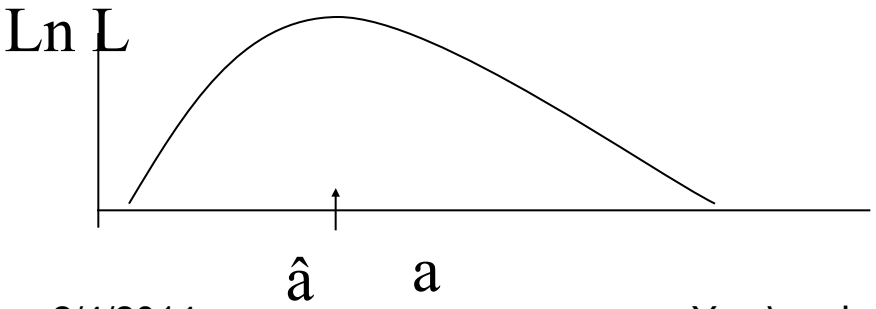


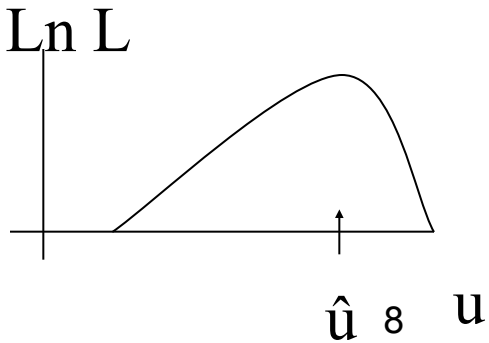
$$\left.\frac{dL}{dA}\right|_{a=\hat{a}} = 0$$

Η ML έχει μερικές χρήσιμες ιδιότητες

Ιδιότητες της ML

- Έχει συνέπεια
- Για μεγάλο σύνολο δεδομένων N είναι 'αμερόληπτη' (unbiased)
- Είναι αποδοτική (efficient) για μεγάλο N (η απόκλιση $\mathbb{V}$ είναι ελάχιστη)
- Για μικρό N πρέπει να είμαστε προσεκτικοί (biased)
- Αν χρησιμοποιήσουμε την συνάρτηση u(a), τότε και û=u(â) (invariant, but does not follow that û unbiased if â unbiased)





Περισσότερα για την (ML)

- Δίνει την πιό πιθανή τιμή για τα δεδομένα που έχουμε -όχι κατ'ανάγκη και την πραγματική τιμή της παραμέτρου που προσαρμόζουμε

- Συνήθως απαιτούνται αριθμητικές μέθοδοι για την ελαχιστοποίηση/μεγιστοποίηση

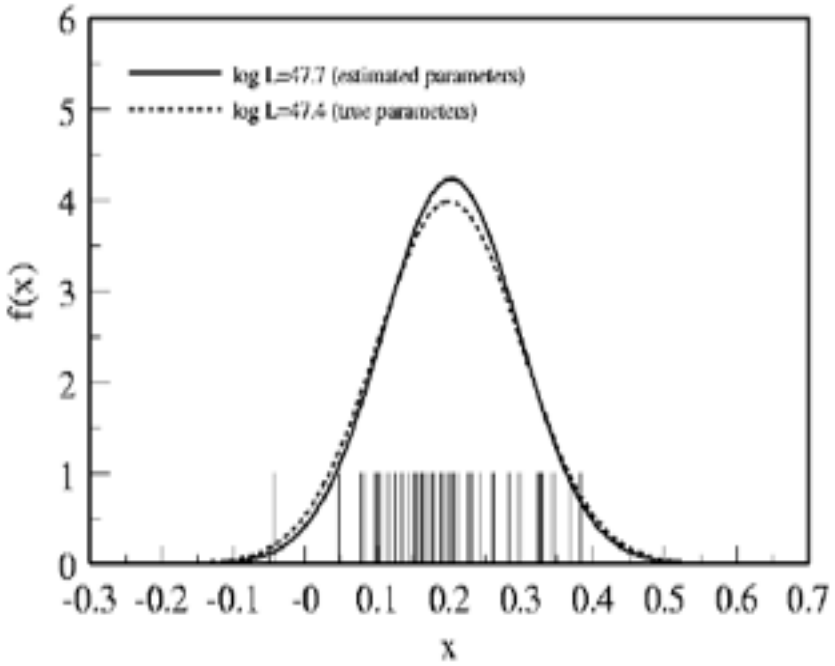
- Για περισσότερες από μία μεταβλητές η ελαχιστοποίηση/ μεγιστοποίηση δεν είναι εύκολη
- Χρησιμοποιούμε MINUIT αλλά προσοχή στο αρνητικό πρόσημο του λογαρίθμου!

- Αν η συνάρτηση $P(x;\alpha)$ της πιθανότητας που χρησιμοποιούμε είναι λάθος ο υπολογισμός της α ΔΕΝ έχει νόημα!

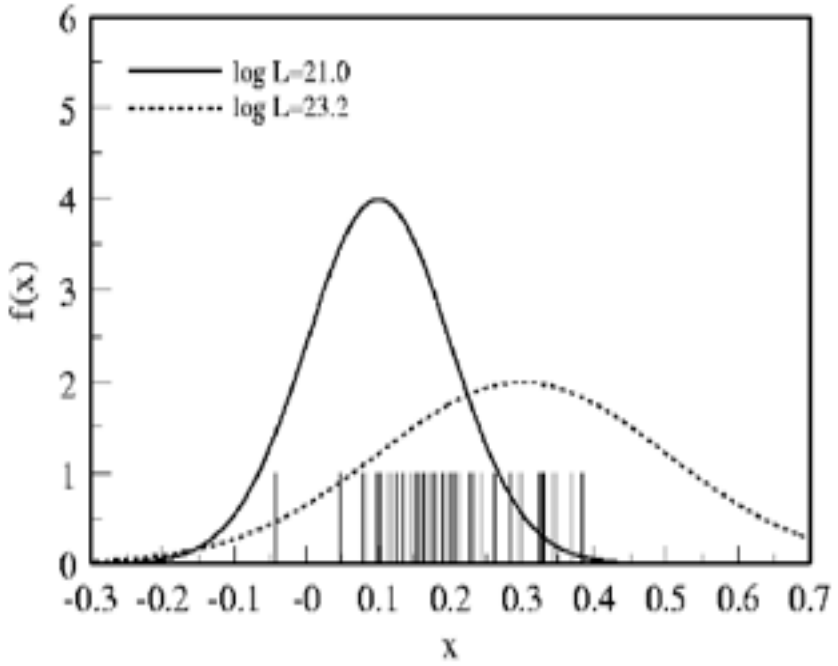
Παράδειγμα ML

- Έστω 50 σημεία μιας μεταβλητής x με κατανομή gauss: G(μ=0.2, σ=0.1)

Οι καλύτερες τιμές για μ και σ και οι πραγματικές



Τιμές των μ και σ μακριά από το μεγιστο της logL



Maximum Likelihood

- Το μέγιστο βρίσκεται είτε γραφικά από τις τιμές του logL για διάφορα α είτε από διαφόριση της logL που άλλοτε μπορεί να γίνει αναλυτικά και άλλοτε με αριθμητικές μεθόδους
- Παράδειγμα : ο υπολογισμός του χρόνου ζωής τ μιάς κατάστασης από σύνολο N μετρήσεων του χρόνου $\{t_i\}$. Η lnL συνάρτηση είναι:

$$\ln L = \sum \ln(\frac{1}{\tau} e^{-t_i/\tau}) = \sum -\frac{t_i}{\tau} - \ln\tau$$

- Διαφορίζουμε την lnL καί βρίσκουμε την τιμή τ που μηδενίζει την παράγωγο

$$\frac{d\ln L}{d\tau} = \sum(\frac{t_i}{\tau^2} - \frac{1}{\tau}), \sum(\frac{t_i}{\tau^2} - \frac{1}{\tau}) = 0 \Rightarrow$$

$$\hat{\tau} = \frac{1}{N}\sum t_i$$

Υπολογισμός σφαλμάτων στον εκτιμητή ML

- Αν το πλήθος των μετρήσεων N πολύ μεγάλο τότε η διαφορά ($\hat{a}-a_0$) έχει κατανομή gauss με μ=0 και απόκλιση σ :

$$\sigma_a^2 = V(\hat{a}) = -\frac{1}{\left(\frac{d^2 \ln L}{da^2}\Big|_{a_0}\right)} = -\frac{1}{\left\langle \frac{d^2 \ln L}{da^2} \right\rangle}$$

=minimum variance bound (MVB)=το όριο ακριβείας του εκτιμητή δεν μπορει να ειναι μικρότερο του V

- Η κατανομή πιθανότητας για το $\hat{a}$ είναι gauss αποδεικνύεται από το CLT

Υπολογισμός σφαλμάτων στον εκτιμητή ML

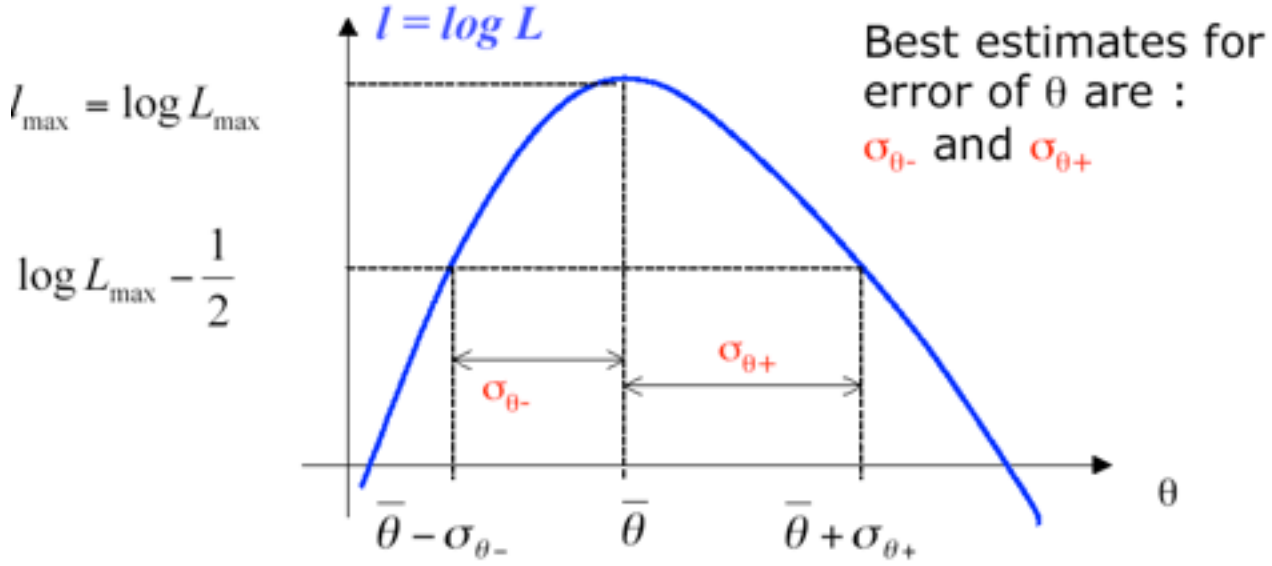
- Αποδεικνυεται οτι για μεγάλο N η συνάρτηση L (likelihood) είναι gauss και άρα η συναρτηση logL είναι παραβολή.
- Επομένως το σφάλμα στον εκτιμητή το βρίσκουμε γραφικά : από την γραφική παράσταση της log likelihood που πήραμε από τα δεδομένα
- 1σ από την κορυφή $\log L = \log L_{max} - 0.5$

  2σ από την κορυφή $\log L = \log L_{max} - 2.0$

  3σ από την κορυφή $\log L = \log L_{max} - 4.5$ …κλπ

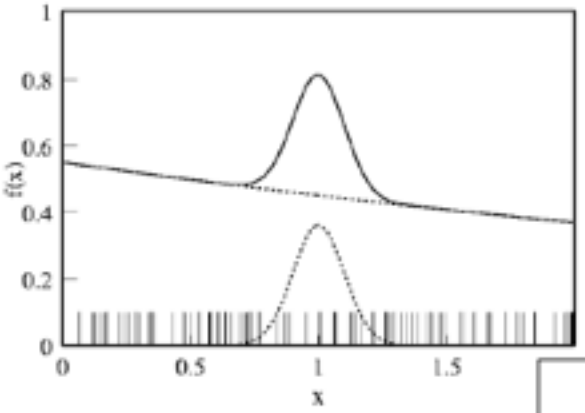
Υπολογισμός σφαλμάτων στον εκτιμητή ML (ασυμμετρικά σφάλματα)

- Εαν το πλήθος των γεγονότων N είναι μικρό, η L ΔΕΝ είναι *gaussian* και η logL ΔΕΝ είναι παραβολή
- Αποδεικνύεται ότι το 1σ του εκτιμητή βρίσκεται πάλι στο $\log L = \log L_{max} - 0.5$



Ειδικές περιπτώσεις

- Στα δεδομένα υπάρχει συνεισφορά από διάφορες πηγές. (Συνεισφορά από διάφορες pdfs) π.χ.
  - Στην αναζήτηση νέων σωματιδίων τα δεδομένα περιέχουν – μετά από επιλογή των γεγονότων– σήμα και υπόβαθρο
  - Στον υπολογισμό της μάζας ενός συντονισμού υπάρχει και συμβολή από υπόβαθρο λόγω πολλαπλών συνδυασμών



number of different contributions

$$f(x;\theta) = \sum_{i=1}^{m} \theta_i f_i(x)$$

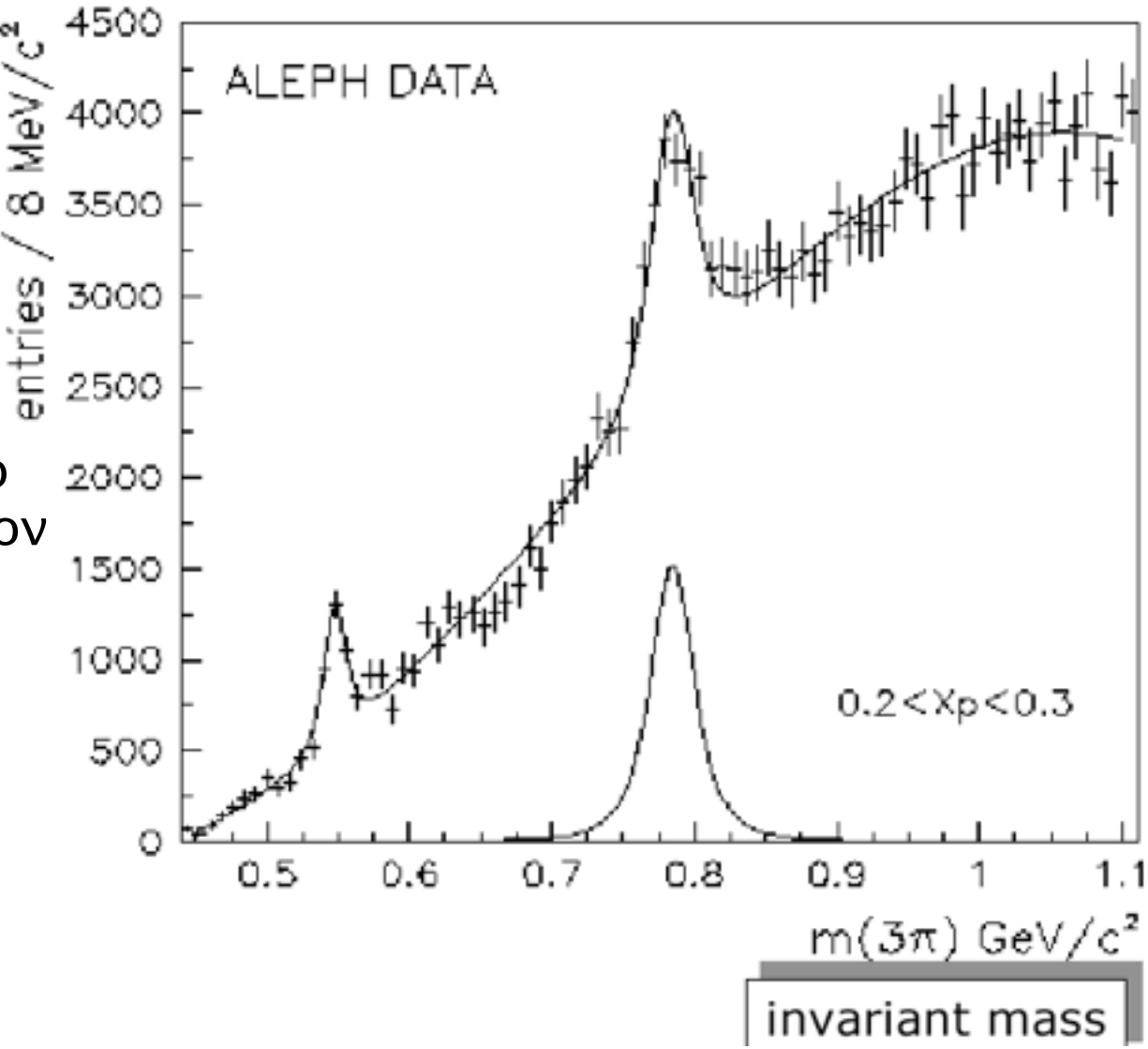
can depend on further parameters

relative fractions :
a priori known (from analytical calc. or Monte Carlo), or to be fitted!

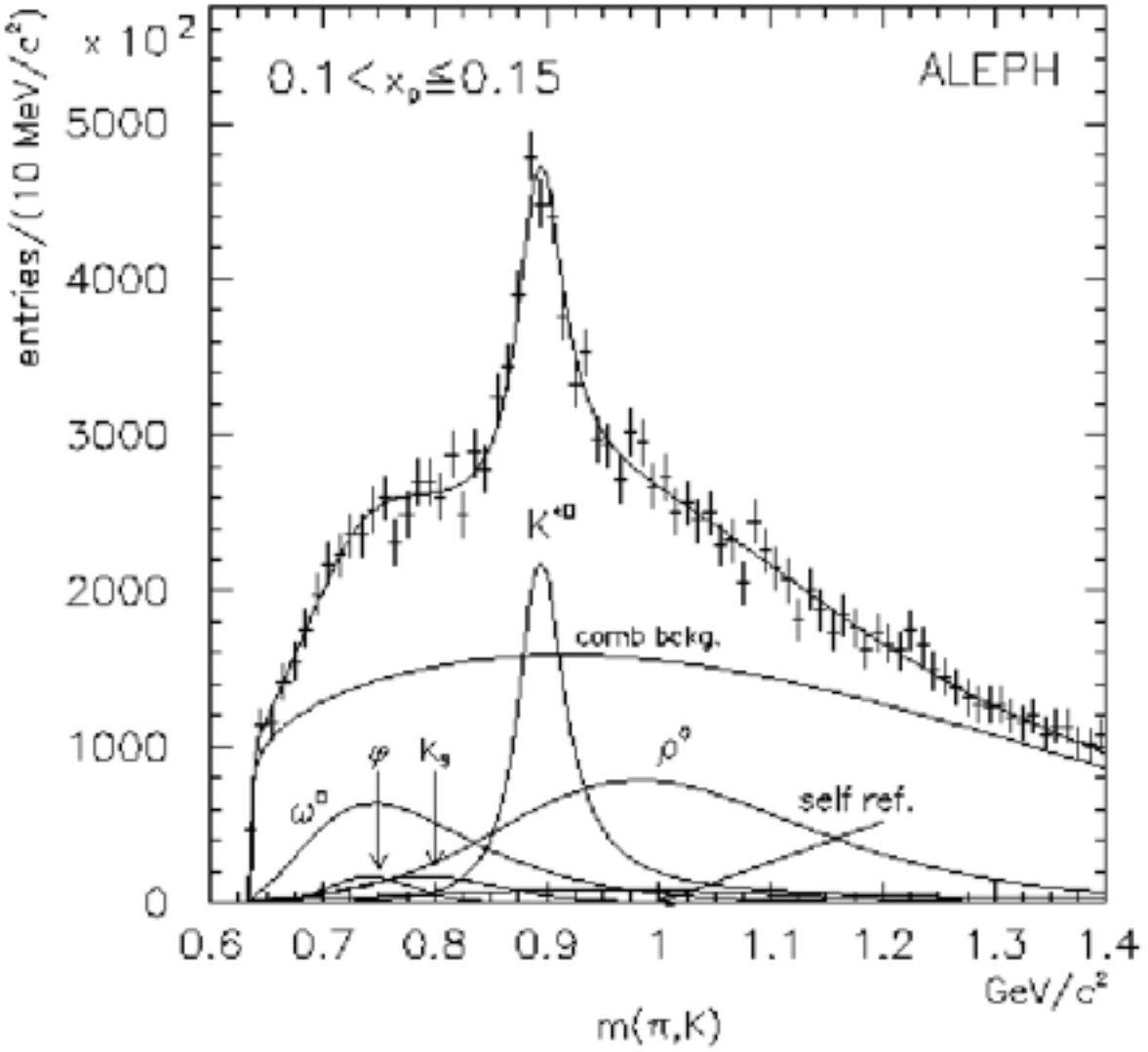
Ειδικές περιπτώσεις

- Παράδειγμα συντονισμού με υπόβαθρο από πολλαπλούς συνδυασμούς

Στα αδρονικά γεγονότα σε e^+e^- στο LEP έχουμε κατά μέσον όρο 20 φορτισμένα σωμάτια. Ο συντονισμός ω(782) που μπορεί να δημιουργηθεί διασπάται συχνά σε $\pi^+\pi^-\pi^0$.



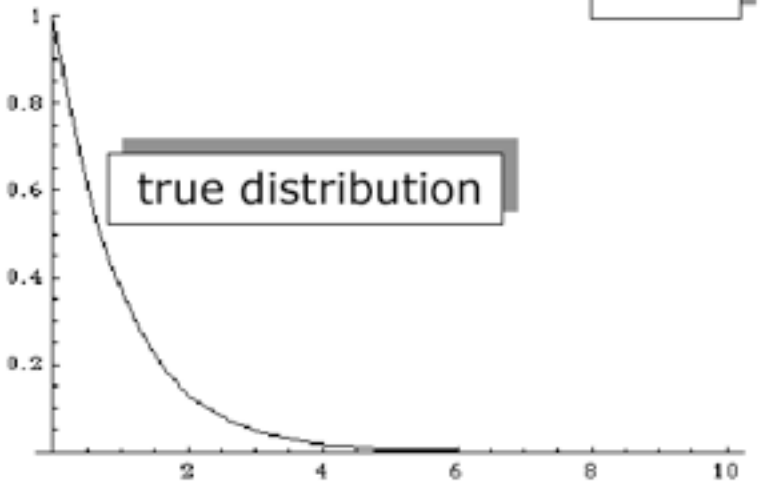
Ειδικές περιπτώσεις

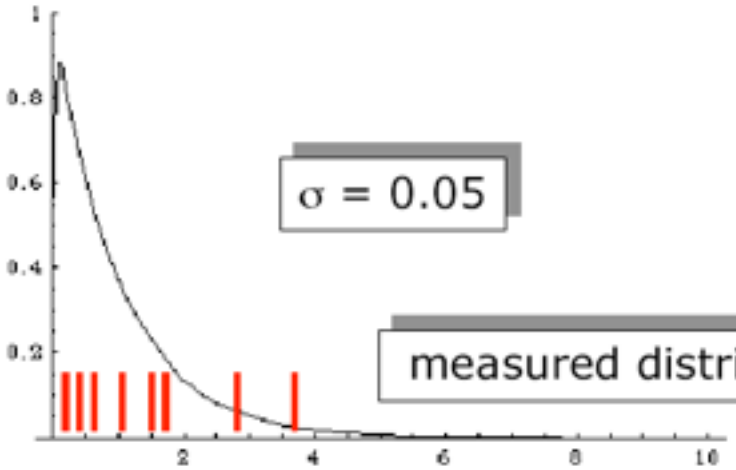


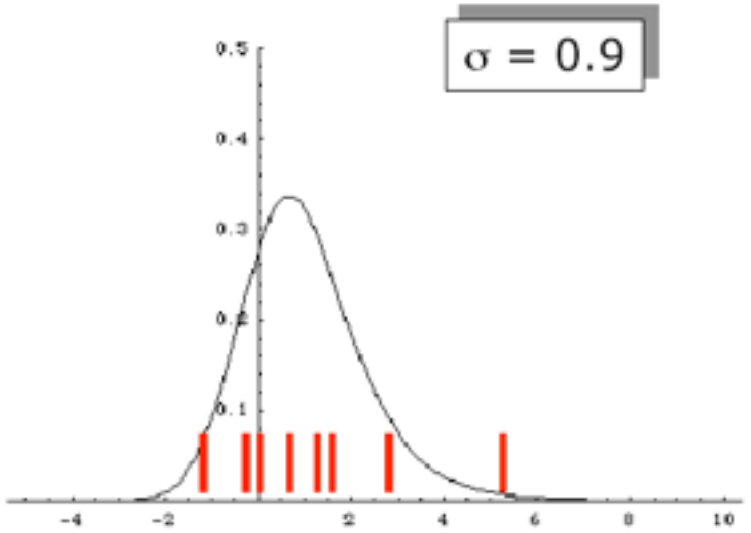
Ειδικές περιπτώσεις

- πως παίρνουμε υπ'όψη τον ανιχνευτή στα αποτελέσματα? (π.χ. διακριτική ικανότητα)
- Παράδειγμα : η μέτρηση του μέσου μήκους διάσπασης σωματιδίου γίνεται με πεπερασμένη ακρίβεια
- Η 'αληθινή' pdf είναι εκθετική με μέση τιμή τ
- Η διακριτική ικανότητα του ανιχνευτή είναι κατανομή gauss με μέση τιμή t_i, (την μέτρηση) και σφάλμα σ
- Η κατανομή που μετρούμε είναι η συνέλιξη των δύο κατανομών

$\tau = 1$





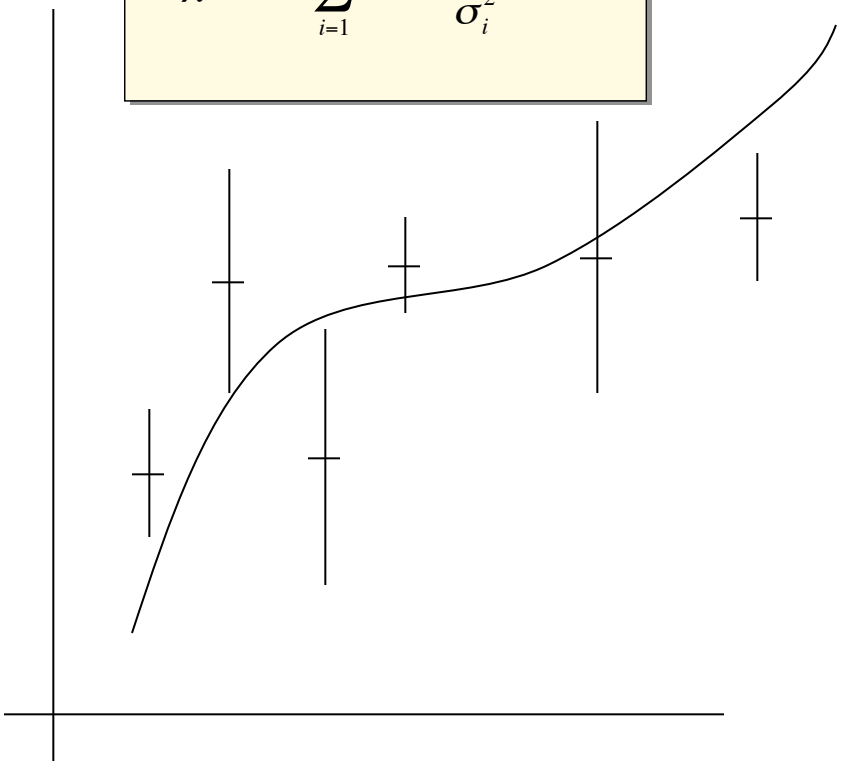


so: can measure negative values!

Take this into account when fitting for τ

Μεθοδος ελαχίστων τεραγώνων (least squares)

$$\chi^2(a) = \sum_{i=1}^{N} \frac{[y_i - f(x_i;a)]^2}{\sigma_i^2}$$



- Μέτρηση του y σε διάφορα x με σφάλμα σ και πρόβλεψη f(x;a)
- Η πιθανότητα:

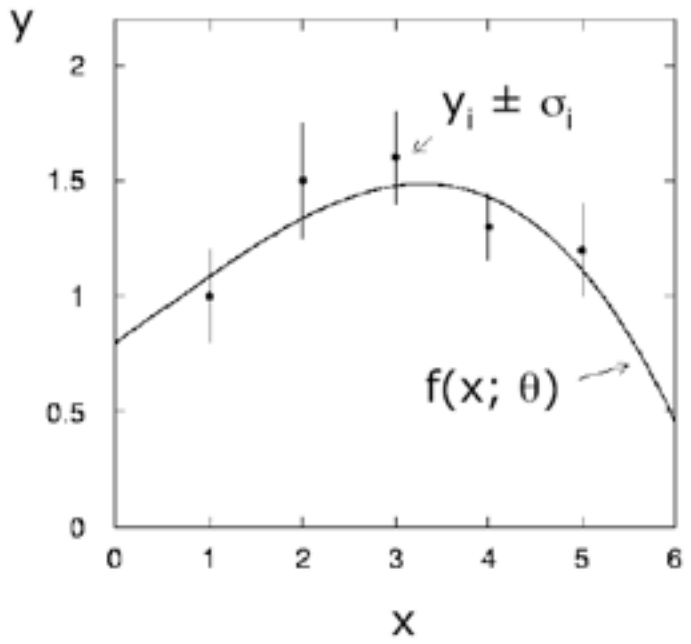
$$\propto e^{-(y-f(x;a))^2/2\sigma^2}$$

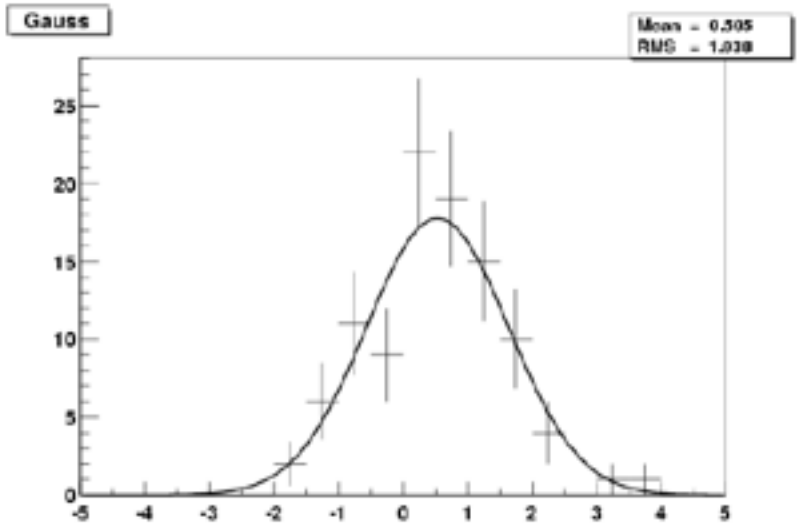
- (y_i gaussian random variables)
- Η lnL(a) είναι:

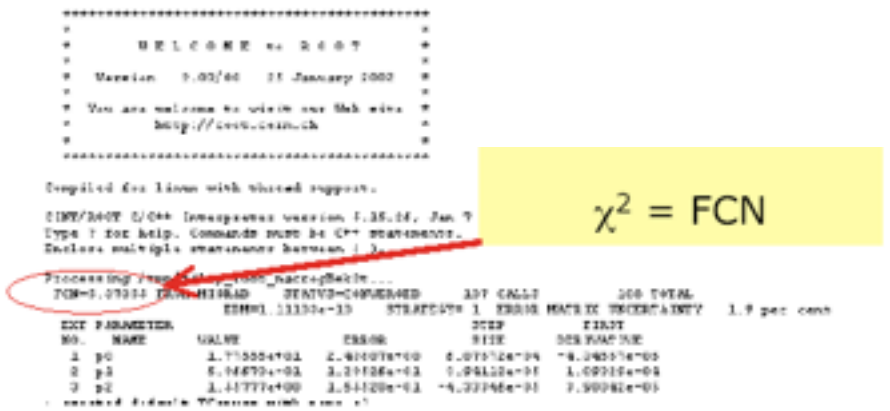
$$\log L(a) = -\frac{1}{2}\sum_{i=1}^{N} \frac{[y_i - f(x_i;a)]^2}{\sigma_i^2}$$

- Maximize lnL=>minimize χ^2

Μεθοδος ελαχίστων τεραγώνων– παραδείγματα







- Definitions, Errors on parameters

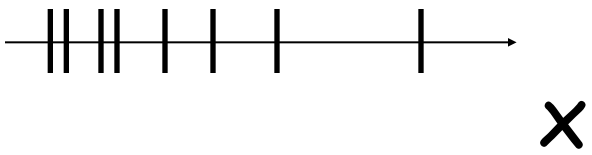
Μεθοδος ελαχίστων τεραγώνων
Η ποιότητα της προσαρμογής

- Σύγκριση της προσαρμογής (fit) με τα δεδομένα
- Πρέπει το χ^2 της προσαρμογής: $\chi^2 \approx 1$ per data point οι βαθμοί ελευθερίας του συστήματος ndof :

  $N_{degrees\ Of\ Freedom} = N_{data\ points} - N_{parameters}$
- Αν τα δεδομένα σε bins:

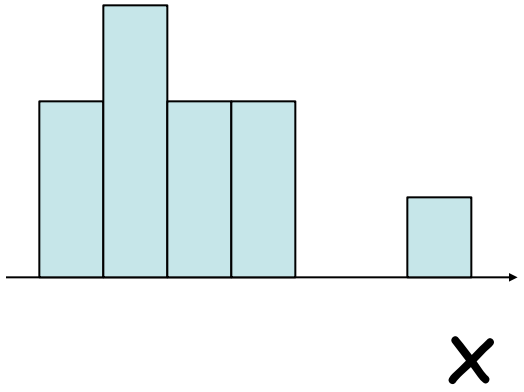
  $N_{degrees\ Of\ Freedom} = N_{data\ bins} - N_{parameters}$
- Αν $\chi^2 >> ndof$: bad fit, bad theory, underestimated errors, bad luck!
- Αν $\chi^2 << ndof$: errors too big, good luck!

Fitting Histograms



- Συνήθως βάζουμε τα $\{x_i\}$ σε bins
- Οπότε τα data είναι $\{n_j\}$, n_j Poisson distributed

  mean $f(x_j)=P(x_j)\Delta x$



- 4 τεχνικές:
  - Full ML
  - Binned ML
  - Proper χ^2
  - Simple χ^2

Τι μεγιστοποιούμε/ελαχιστοποιούμε

- Full ML

$$\ln L = \sum_i \ln P(x_i; a\ )$$

- Binned ML

$$\ln L = \sum_j \ln Poisson(n_j; f_j) \approx \sum_j n_j \ln f_j - f_j$$

- Proper χ^2 approx. by Poisson $\sigma^2=\mu$) $\sum_j \frac{(n_j - f_j)^2}{f_j}$ (contents of each bin are

- Simple χ^2 number of entries observed $\sum_j \frac{(n_j - f_j)^2}{n_j}$ (approx. variance with

Τι χρησιμοποιούμε?

- Full ML: χρησιμοποιεί όλη την πληροφορία αλλα είναι πιο περίπλοκη. Την προτιμούμε για μικρό αριθμό γεγονότων
- Binned ML: λιγότερο περίπλοκη.Χάνουμε πληροφορία αν το bin size πολυ μεγάλο
- Proper χ^2: λιγότερο περίπλοκη και δίνει την ποιότητα του fit απ'ευθείας. Αν n_j large Poisson-->Gauss
- Simple χ^2: n_j πρέπει να είναι μεγάλο

Τι χρησιμοποιούμε?

- Likelihood ratio για να υπολογίσουμε πόσο : signal-like ή background-like είναι το πειραματικό μας αποτέλεσμα (π.χ. Higgs discovery) κατασκευάζουμε έναν estimator (estimator function) X
- Θεωρούμε ότι όταν ο estimator είναι "positive" είναι "more background-like" οταν ειναι "negative" είναι "more signal-like"

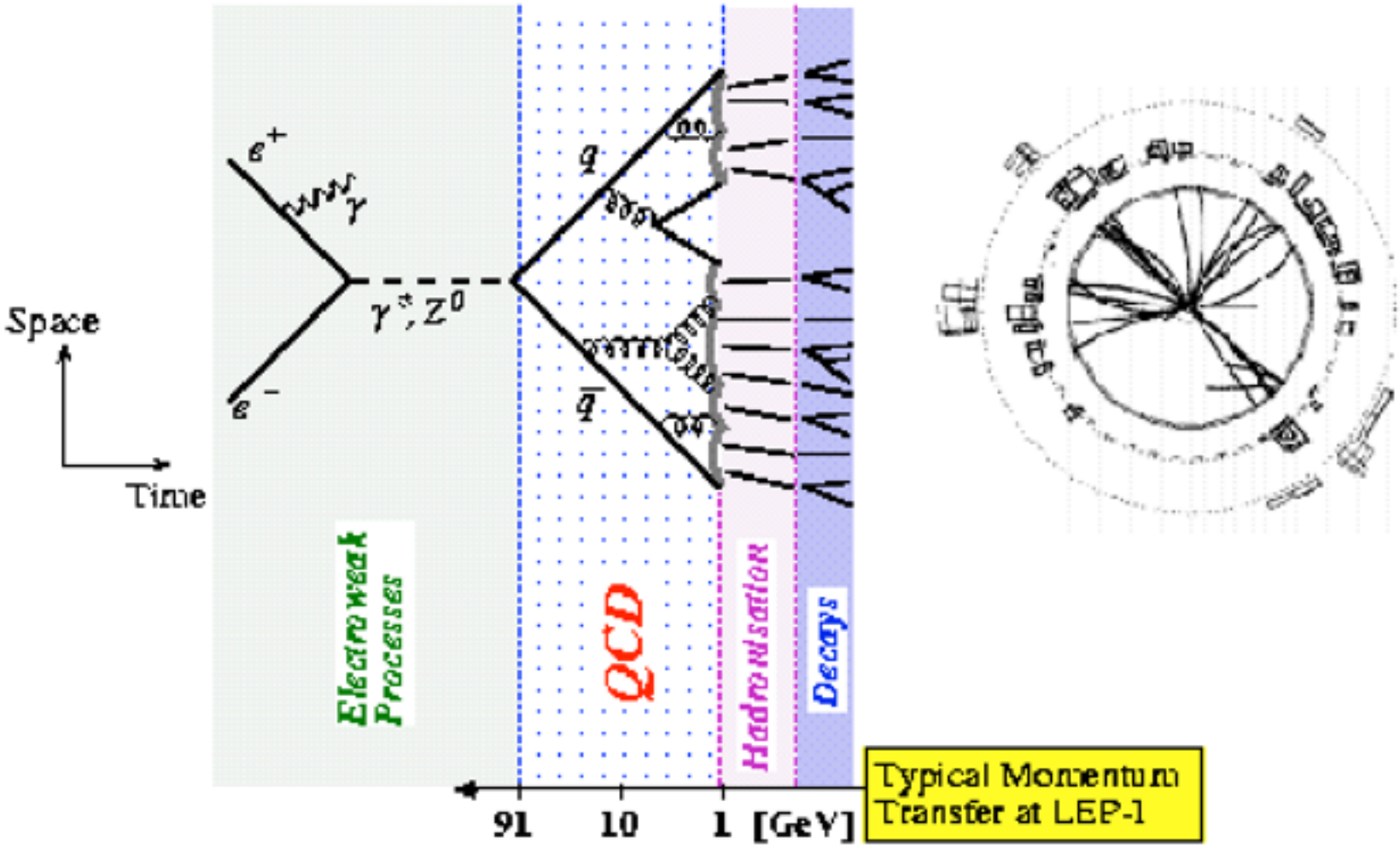
$$X(x) = -2\ln Q(x) = -2\ln \frac{\mathcal{L}(x|s+b)}{\mathcal{L}(x|b)}$$

$\mathcal{L}(x|b)$ →Η Likelihood του x με την υπόθεση ότι μόνο το background μπορεί να μας δώσει x

- Το x είναι χαρακτηριστική ιδιότητα των δεδομένων (συνήθως το πλήθος των γεγονότων)

$$\mathcal{L}(x|b) = \prod_{i=1}^{\text{number of bins}} \frac{e^{-b_i} b_i^{x_i}}{x_i!}$$

Είναι γινόμενο Poisson πιθανοτήτων για όλα τα bins της κατανομής x


Space
Time
e^+
e^-
γ
γ^*, Z^0
q
$\bar{q}$
Electroweak Processes
QCD
Hadronisation
Decays
91
10
1 [GeV]
Typical Momentum Transfer at LEP-1