Cauchy- (=Breit–Wigner-) distribution

\[ f(x) = \frac{1}{\pi} \frac{1}{1 + x^2} \]

\[ f(x; \Gamma, x_0) = \frac{1}{\pi} \frac{\Gamma/2}{\Gamma^2/4 + (x - x_0)^2} \]

Expectation value \( E(x) = x_0 \)

(aber schlecht definiert)

Width at half max (FWHM) = \( \Gamma \)

standard deviation \( \sigma(x) = \infty \)

Resonance phenomena, Cauchy is Fourier transform (in frequency- = energy space) of the exponential distribution (in time t).

Uncertainty relation: Resonance width = \( h / \text{lifetime} \)
Finding signals...

Search for small resonance signal in large background.
Using selection cuts (also applying multivariate methods):
Maximize signal efficiency
Minimize background

If \( \text{Signal} > 3\sigma \) Signifikanz \( \rightarrow \) evidence
If \( n \text{Signal} > 5\sigma \) Signifikanz \( \rightarrow \) discovery

If \( \text{Signal} < 3\sigma \) Signifikanz \( \rightarrow \) no evidence \( \rightarrow \) Upper Limit
Finding signals (2)

“Good” upper limits when no signal candidates
“Good” evidence, when as many as possible signal candidates

Psychological bias?
Limit area difficult and dangerous:
Too low limits, too large first evidences.

→ Optimisation by cuts must be reasonable! Should never be “optimised” on data. But on simulation before one has looked to real data.
→ Way out: “Blind analyses” wherever possible.
Example:

Signal?

Cuts on additional variables 2 and 3:
=> Higher significance

Fit: 3 $\sigma$ significance.
Evidence?
Example for a wrong evidence

Truth: completely random Monte-Carlo distribution
thus just statistical fluctuations.
No correlations with other variables 2 and 3.

Probability of $3\sigma$ should be quite rare ($<0.3\%$).

Yes, but without a priori- models this is a search in many bins (and many histograms).

Statistical fluctuations of other variables can be exploited through cuts to increase the "significance" of the "signal".
Examples for wrong evidences

\( \zeta(8.3) \) Crystal Ball 1983
über hundert Theorie-Papiere
u.a. Higgs-Kandidat

ALEPH 4-Jet Signal
bei LEP 1.5
Kandidat für hA
Already discovered:
Splitting of $a_2$ meson
Top quark around 70 GeV
Substructure of W and Z
Substructure of quarks
Supersymmetry
...
Cold fusion

Careful, especially when you first read about it in the Washington Post.
Also a confirmation of a second experiment is no 100% guarantee.
Do Pentaquarks exist (2004)?

LEPS

Diana

Noon plus Deuterium

CLAS

Statistical Methods BND 2011
Michael Feindt
Do Pentaquarks exist?

\[ \gamma p \rightarrow nK^+K_S^0 \]

\( \Theta^+ (1540) \)

\( M(nK^+) \)

\( M=1527 \pm 2.3 \) (stat) MeV
\( \Gamma=0.2 \pm 2 \) (stat) MeV

\( M=1528 \pm 2.6 \) (stat) MeV
\( \Gamma=8 \pm 2 \) (stat) MeV

\( \chi^2 / \text{dof}=39 / 44 \)
peak=1521.5 \pm 1.5 \) MeV
width=6.1 \pm 1.6 \) MeV
events=221 \pm 48

\( M(\pi^+\pi^-p) [\text{GeV}] \)
Do Pentaquarks exist?

\[ SVD - 2 \]
\[ pA \rightarrow pK_S^0 + X \]

\[ COSY - TOF \]
\[ pp \rightarrow \Sigma^+ pK_S^0 \]

PDG 2004: Status ***
But... Which mass?

\[ M_{\Theta} \]

(LEPS) SPring-8
SAPHIR
CLAS-1
CLAS-2
DIANA

\[ nK^+ \]

\[ pK_S^0 \]

\[ nK^+ \rightarrow pK^0_s \]

How many Pentaquarks?

JINR

\[ np \rightarrow npK^+K^- \quad P_n = 5.20 \text{ GeV/c} \]

\[ M_{nk}^c (\text{GeV}^2) \]

\[ -0.70 \leq \cos \Theta_n^c \leq 0.70 \]
\[ 2.05 \leq M_{nk^+K^-} \leq 2.15 \]
\[ 2.24 \leq M_{nk^+K^-} \leq 2.28 \]

\[ M_{nk^c} (\text{GeV}^2) \]

\[ -0.70 \leq \cos \Theta_n^c \leq 0.70 \]
\[ 2.10 \leq M_{nk^+K^-} \leq 2.24 \]
\[ 2.28 \leq M_{nk^+K^-} \leq 2.50 \]

\[ M_{nk^c} (\text{GeV}^2) \]

\[ 2.24 \leq M_{nk^+K^-} \leq 2.29 \]
More Pentaquarks?

H1

\[ \text{D}^*\text{p} + \text{D}^*\bar{\text{p}} \]

- Signal + bg. fit
- Bg. only fit

charm...

<table>
<thead>
<tr>
<th>NA49</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ ssdd\bar{u} ]</td>
</tr>
<tr>
<td>[ \Xi^-\pi^- ]</td>
</tr>
<tr>
<td>[ ssd\bar{ud} ]</td>
</tr>
<tr>
<td>[ \Xi^-\pi^+ ]</td>
</tr>
<tr>
<td>[ \bar{s}s\bar{d}\bar{ud} ]</td>
</tr>
<tr>
<td>[ \Xi^+\pi^- ]</td>
</tr>
<tr>
<td>[ \bar{s}s\bar{d}\bar{du} ]</td>
</tr>
<tr>
<td>[ \Xi^+\pi^+ ]</td>
</tr>
</tbody>
</table>

Statistical Methods BND 2011
Michael Feindt
Many negative searches…

Despite intensive search nothing found in large experiments:
ALEPH, DELPHI, L3, OPAL, CDF, BELLE, BABAR, BES,
H1 in ZEUS-channel, ZEUS in H1-channel
(but partly huge signals for established states)

Attention Bayes-Theorem: Prior probability has changed!
In exotic channels you should be very sure that it
is not a statistical fluctuation before claiming a signal.
For pentaquarks there are many new channels, masses
unknown, danger of combinatorics

My opinion 2004: $P_{\text{subjective}}(\text{Pentaquark exists}) < 50\%$
$P_{\text{subjective}}(\text{all Pentaquarks exist}) < 1\%$
(I was almost killed for that opinion at that time, PDG **** rating)
.Today: $P(\text{Pentaquark exists}) < 0.01\%$
Historical developments I (PDG)
Historical developments II (PDG)

Statistical Methods BND 2011
Michael Feindt
A recent example

Dissertation Felix Wick (KIT, 2011)
CDF Collaboration: Mass of orbitally excited $\Lambda_c$

Much more statistics than previous experiments. More precise.

But: Many sigma away from PDG value. Inconsistent!
A recent example

Dissertation Felix Wick (KIT, 2011), CDF Collaboration:
Mass of orbitally excited $\Lambda_c$

Simple Breit Wigner fit does not describe data well

Complicated lineshape due to narrow subresonance thresholds on Dalitz plot necessary. Changes results!
Einführung in die Datenanalyse: Parameterschätzung

Prof. Dr. Michael Feindt
Vorstandsmitglied CETA
Institut für Experimentelle Kernphysik
Universität Karlsruhe

Herbstschule für Hochenergiephysik
Kloster Maria Laach 2004
Parameter estimation (Fitting)

Maximum Likelihood

Least Squares

Statistical Methods BND 2011
Michael Feindt
Parameter estimation with least squares

Several measurements $y_i$ with known variances $\sigma_i^2$
Linear model with parameter vector $a$
$E(y) = A \cdot a$

Find best estimator for $a$ and its uncertainty.

Least squares principle:
Minimize sum $S$ of quadratic deviations between model and measurements.

Solution: Derivatives $\frac{dS}{da} = 0$
Linear least squares

\[ A \cdot a = y \]

Linear equation system
One unique solution

\[ n = p \]

Overconstrained equation system
Try to find best compromise
\[ n > p \]

\[ \approx \]

\[ n - p \] degrees of freedom

Statistical Methods BND 2011
Michael Feindt
Parameter estimation with least squares

Identische Fehler: \( S = \sum_{i=1}^{n} \Delta y_i^2 \)

Unterschiedliche Fehler: \( S = \sum_{i=1}^{n} \left( \frac{\Delta y_i}{\sigma_i} \right)^2 \)

Korrelierte Messungen mit Kovarianzmatrix \( V \): \( S = \Delta y^T V^{-1} \Delta y \)

Beispiel: Mittelwert von \( n \) Messungen \( y_i \):

\[
S = \sum_{i=1}^{n} (y - y_i)^2 = \text{Minimum} \\
\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i
\]
Solution of linear optimisation problem

Modell für Lineare Kleinste Quadrate: \( y = A a \)

\[
\begin{align*}
    r &= y - A a & \text{Residuenvektor} \\
    W &= V[y]^{-1} & \text{Gewichtsmatrix = Inverse der Kovarianzmatrix}
\end{align*}
\]

Prinzip der Kleinsten Quadrate: Minimiere den Ausdruck

\[
S(a) = r^T W r = (y - A a)^T W (y - A a)
\]

bezüglich \( a \), d.h. \( dS/da = 0 \).

Lösung ist lineare Funktion der Messwerte \( y \):

\[
\hat{a} = \left( A^T W A \right)^{-1} A^T W y = B y
\]

Kovarianzmatrix von \( a \) durch Fehlerfortpflanzung:

\[
V[\hat{a}] = B V[y] B^T = \left( A^T W A \right)^{-1} = \text{Inverse der 2. Ableitung von } S
\]

Statistical Methods BND 2011
Michael Feindt
Dependence on measurement error (residual) distribution

Straight line fit to 20 data points (ndf=20-2=18)
Three different distribution functions,
all mean 0, standard deviation=0.5
Parameter estimation with least squares

25000 Monte-Carlo tests:
All parameter-distributions are Gaussian, with width compatible to expectation from error propagation (for both parameters)
Parameter estimation with least squares

Mittlere $\chi^2$-Werte sind alle gleich NDF=20-2=18
Nur bei Gauß'schen Messfehlern folgt $S$ einer $\chi^2$-Verteilung, d.h. Prob($\chi^2$) = const., sonst anderen Verteilungen.

Methode der kleinsten Quadrate funktioniert unabhängig von der Fehlerverteilung, wenn → Daten unverzerrt sind (d.h. 1. Moment bekannt) und → Kovarianzmatrix (d.h. 2. Moment) (korrekt) bekannt ist.
Robust least squares

Outlyers in data (z.B. typing error, wrong hits on a track...) can influence fit very strongly (due to quadratic dependence) and can lead to completely nonsense solutions.

Simple recipe for robustification:

1. Normal least squares fit, delivers residuals
2. Modify data through limitation of residuals to $c\sigma$. A good choice is $c=1.5$.
3. Redo fit with pseudo-measurements instead of origina ones.

There are other, robust loss-functions (z.B. Huber-funktion), but not any more analytically solvable.
Häufige Fehler bei $\chi^2$-Minimierung

z.B. Kalibration eines Kalorimeters im Teststrahl der Energie $E$, $N$ Messungen $y_i$, Definition von effektivem $\chi^2$

$$\chi^2_{falsch} = \frac{1}{N} \sum_{i=1}^{N} (a \cdot y_i - E)^2$$

führt NICHT zum gewünschten Ergebnis $a = E/\bar{y}$, sondern zu

$$a = \frac{E \cdot \bar{y}}{(\sum_k y_i^2)/N} = \frac{E \cdot \bar{y}}{\bar{y}^2 + \sigma^2} \neq E/\bar{y}$$

Bias-Korrektur nötig!
Häufige Fehler bei \( \chi^2 \)-Minimierung (Forts.)

Die inverse Konstante kann verzerrungsfrei bestimmt werden mittels:

\[
\chi^2 = \frac{1}{N} \sum_{k=1}^{N} (y_i - a_{inv}E)^2
\]

Allgemein gilt:
In einem \( \chi^2 \)-Ausdruck nicht die gemessenen Werte \( y_i \), sondern nur die theoretische Erwartung verändern!
Vorsicht bei kleinen Zahlen!

MINUIT (in PAW und root) benutzt bei Histogrammfits per Default Gauss-Statistik und $\sigma = \sqrt{n}$. In einem leeren Bin ist $\sigma = 0$. Damit keine Unendlichkeiten auftreten, werden diese Bins werden einfach weggelassen!
Attention with small numbers!

Das ist nicht richtig! Besser: Poisson-Statistik und Likelihood-Fit, wenn Anzahl der Ereignisse in einem Bin kleiner als 10.
Non linear least squares

Often non-linear function $f(x,a)$ appears in model.

Then linearize this. For that you need good STARTING VALUES in order to perform a TAYLOR EXPANSION.

After solution of linearized problem ITERATE to perform a Taylor expansion around the new solution.

If iteration doesn't improve, shorten Change of solution vector.

Converge criteria are needed.
Least squares with additional constraints

Example: Kinematic fitting in $\pi^0$ decay

Measured: $E, \cos(\theta), \phi$ mit Kovarianzmatrix von zwei Photonen.

A priori knowledge: Photons stem from $\pi^0$ decay.

Constraint: Invariant mass $m(\gamma \gamma) = m(\pi^0)$

Change measured parameters such that constraints are exactly fulfilled and $\chi^2$ is minimal. This can improve e.g. $E(\pi^0)$ resolution considerably.
Least squares with constraints (2)

Write conditions in form \( f_i(x, a) = 0 \).

Solution by introduction of Lagrange multipliers and minimization of

\[
L(a, \lambda, \Delta y) = S(a, \Delta y) + 2 \cdot \sum_i \lambda_i f_i(x, a)
\]

w.r.t. \( a \) und \( \lambda \). Am Minimum ist \( dL/d\lambda = f = 0 \), also constraint is fulfilled

Beispielapplikation: APLCON von V. Blobel
Maximum Likelihood Parameter Estimation

Aufgabenstellung:
Several independent measurements $x_i, i = 1, \ldots, n$ of a quantity $x$ were performed.
We have a model for the probability density (PDF) $p(x; \vec{a})$ with parameter vector $\vec{a}$.
Find best estimator for $\vec{a}$ and its uncertainty.

Als PDF ist $p(x; \vec{a})$ positiv und normiert (in $x$, aber nicht in $\vec{a}$):

$$p(x; a) \geq 0 \quad \text{mit} \quad \int_{\Omega} p(x; a) dx = 1$$
Likelihood function:
Values of probability density for n independent measurements (product) to observe just the values $x_i$ is called likelihood function:

$$\mathcal{L}(\alpha) = p(x_1|\alpha) \cdot p(x_2|\alpha) \cdots p(x_n|\alpha) = \prod_{i=1}^{n} p(x_i|\alpha).$$

It only depends on parameters $a$ (measurements are fixed!)

Maximum likelihood principle:
The best estimator of $a$ is the one maximizing the likelihood function $\mathcal{L}(\alpha)$.
Maximum Likelihood in Practice

Technical and theoretical reasons:
Minimize (negative) logarithm of likelihood function

\[ F(a) = - \ln \mathcal{L}(a) = - \sum_{i=1}^{n} \ln p(x_i|a) \]

Likelihood equation defines estimator \( \hat{\alpha} \): \( \frac{dF(\hat{\alpha})}{d\alpha_j} = 0 \).

Combination of several (uncorrelated) experiments is easy:

\[ \mathcal{L}(a) = \mathcal{L}_2(a) \cdot \mathcal{L}_2(a) \quad \text{multipliziere Likelihood-Funktionen} \]

\[ F(a) = F_1(a) + F_2(a) \quad \text{addiere log. Likelihood-Funktionen} \]
Maximum Likelihood–Error Estimation

F(a) approximately quadratic around minimum

First derivative approx linear, =0 at minimum

Second derivative almost constant
Standard deviation = 1 / curvature

\[ F(a) - F(\hat{a})_{\text{min}} = \frac{1}{2} \quad \iff \quad |a - \hat{a}| = 1\sigma \]

\[ F(a) - F(\hat{a})_{\text{min}} = \frac{n^2}{2} \quad \iff \quad |a - \hat{a}| = n\sigma \]
Maximum Likelihood-- Applications

binned maximum likelihood

model-PDF in analytic form

Often: model-PDF only known by Monte-Carlo

⇒ need >10-fold MC- statistics
⇒ in 1D: smooth MC-prediction
⇒ or take into account finite MC-statistics explicitly
  with the method of Barlow
When the success probability for a certain Bernoulli (yes or no) event is \( p \), then the probability to achieve exactly \( k \) successes in \( n \) experiments is given by the binomial distribution:

\[
P_n(k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \ldots, n
\]

The factor \( \binom{n}{k} \) is the number of combinations to select \( k \) elements from \( n \) different elements.

**Expectation value**

\[ E[k] = n p \]

**Variance**

\[ V[k] = n p (1 - p) \]
Maximum Likelihood Fit for \( p \) of a binomial distribution

A coin is thrown \( N \) times. What can be stated about the probability \( P \) to throw „head“?

0 trials: We don't know anything. \( p \) flat in [0,1].

1. trial: head. Then probability of \( p(\text{Kopf})=0 \) is zero.

\( N \) trials: \( p(\text{head}) \) distribution gets more precise, and Gaussian.
Maximum Likelihood mit Bayes-Prior
Maximum Likelihood binomial fit
Probability distributions (2)

Uncertainty gets smaller with $\frac{1}{\sqrt{N}}$
Maximum Likelihood binomial fit with different a priori assumptions

Different a priori-Annahmen:

1. $p$ flat in $[0,1]$.
2. Coin is probably oky. Gaussian around 0.5.
3. Coin is probably manipulated, but I don't know in which direction.

Most professional non-informative Bayesian prior (Jeffreys prior) for binomial likelihood: $\text{Beta}(0.5,0.5)$
Maximum Likelihood binomial fit with different a priori assumptions

Influence of prior gets negligible with more precise likelihood. Choose vague priors (do not exclude anything!) for faster convergence (The narrow Gaussian here was worst!)

Statistical Methods BND 2011
Michael Feindt