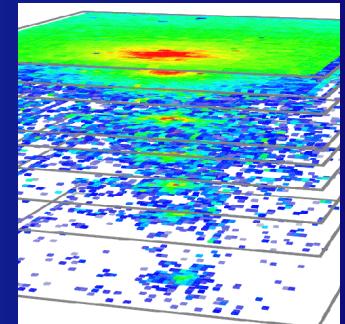
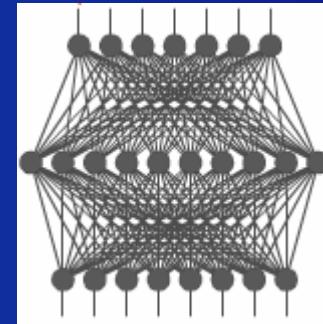
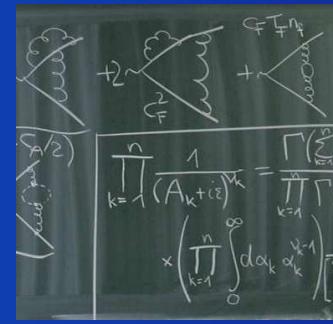
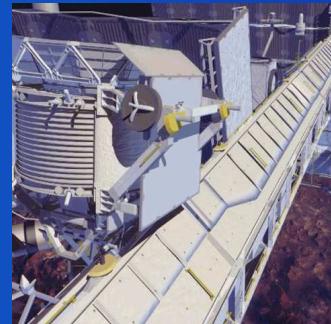




# Einführung in die Datenanalyse

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

Herbstschule für Hochenergiephysik  
Kloster Maria Laach 2004



# Inhalt

## Determinismus und Zufall

Teilchenphysik-Experimente

Wahrscheinlichkeitslehre und Statistik

Ein paar Fallen

## Simulationen

Monte-Carlo-Methoden, Pseudo-, Quasizufallszahlen

## Parameter-Schätzung

Maximum Likelihood, Least Squares, Kalman-Filter,  
robuste Schätzungen

## Künstliche Intelligenz und Anwendungen

Neuronale Netzwerke, NeuroBayes, Phi-T  
Support-Vektor-Maschinen

# Danksagung

V. Blobel

persönlich, Lecture Notes Sommergraduiertenschule Uni Heidelberg

G. Cowan

Lecture Notes div. Vorlesungen

K. Rinnert

Folien zum Kalman-Filter

M. Moch

root Animationen

A. Scheurer

Scannen

J. Bossert

Quasi-Zufallszahlen

# Literatur

V. Blobel, E. Lohrmann

Statistische und numerische Methoden der Datenanalyse,  
Teubner, Stuttgart 1998

G. Cowan

Statistical Data Analysis, Clarendon, Oxford, 1998

D.S. Sivia

Data Analysis – A Bayesian Tutorial, Clarendon, Oxford 1996

+viele mehr, z.B. Brandt-Dahmen

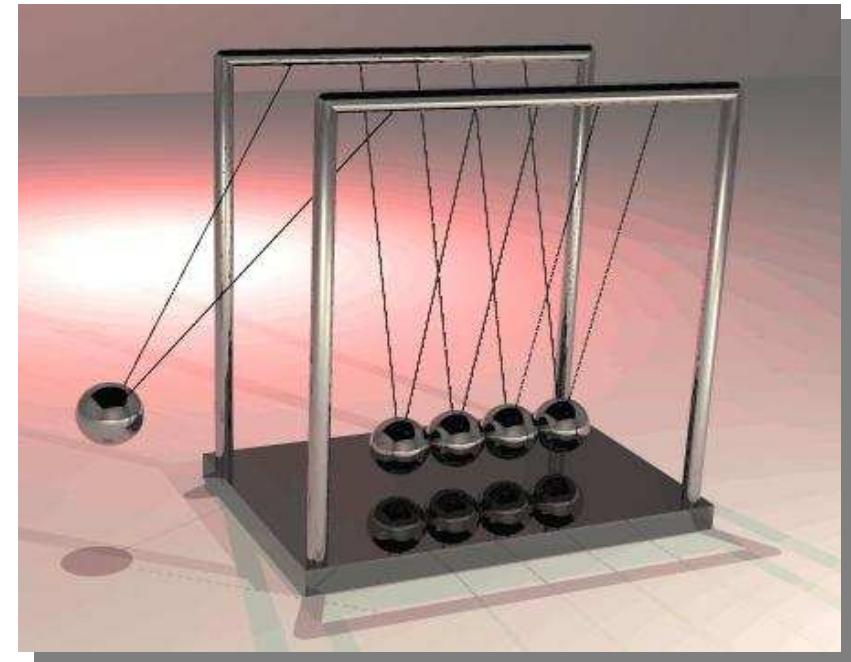
# Vorhersehbar

Bei einfachen klassischen physikalischen Prozessen ist das Ergebnis exakt vorhersagbar

(eine Ursache erzeugt eine eindeutige Wirkung,  
Determinismus)

Beispiele hierfür sind:

Pendel, Planetenbahnen,  
Billard, Elektromagnetismus...



# Zufall

Rein zufällige Ereignisse sind prinzipiell nicht vorhersagbar  
(auch bei genauer Kenntnis der Ausgangssituation!)

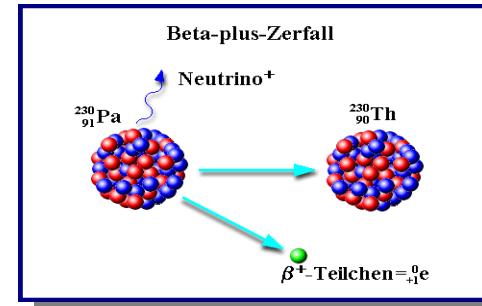
Beispiele hierfür sind:

Lottozahlen

(Zu viele Einflussgrößen,  
deterministisches Chaos)



radioaktiver Zerfall  
(Quantenmechanik)



Elektronisches Rauschen

# Wahrscheinlichkeit

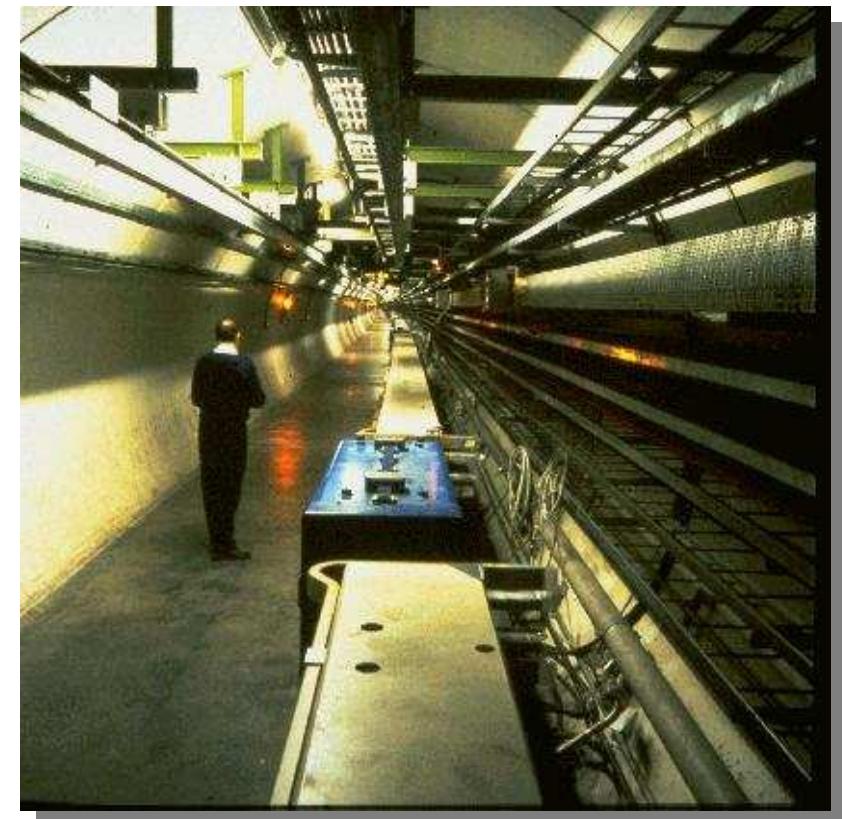
Viele Systeme: Mischung aus vorhersagbarer Komponente und Zufallskomponente.

à Wahrscheinlichkeitsaussage, Statistik.

Extraktion der vorhersagbaren Komponente  
Bestimmung von Modell-Parametern aus Messdaten

# Elementarteilchenphysik-Experimente

Viele 1.000.000.000 identische Versuche über Jahre:  
Kollisionen zwischen z.B. Elektronen und Positronen

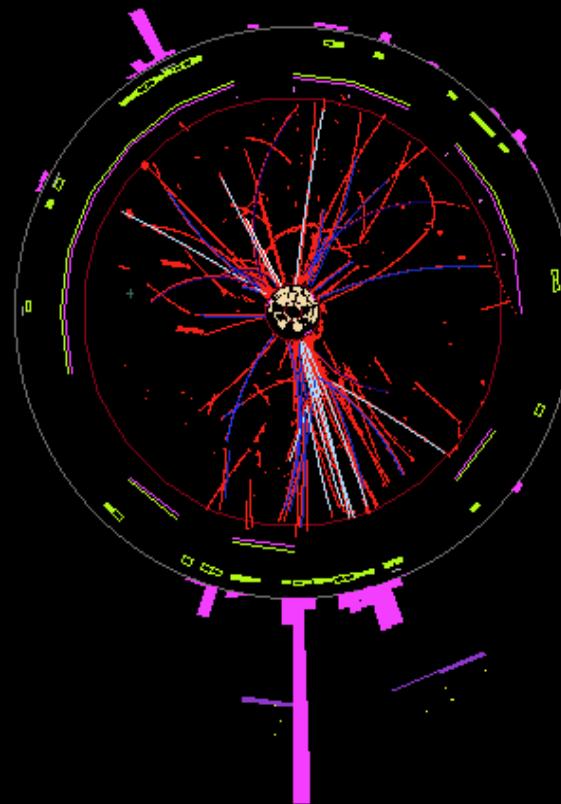


Quantenmechanik:  
Jedes Mal  
passiert etwas  
anderes!

OPAL Experiment am LEP

Run 13816 Event 9618

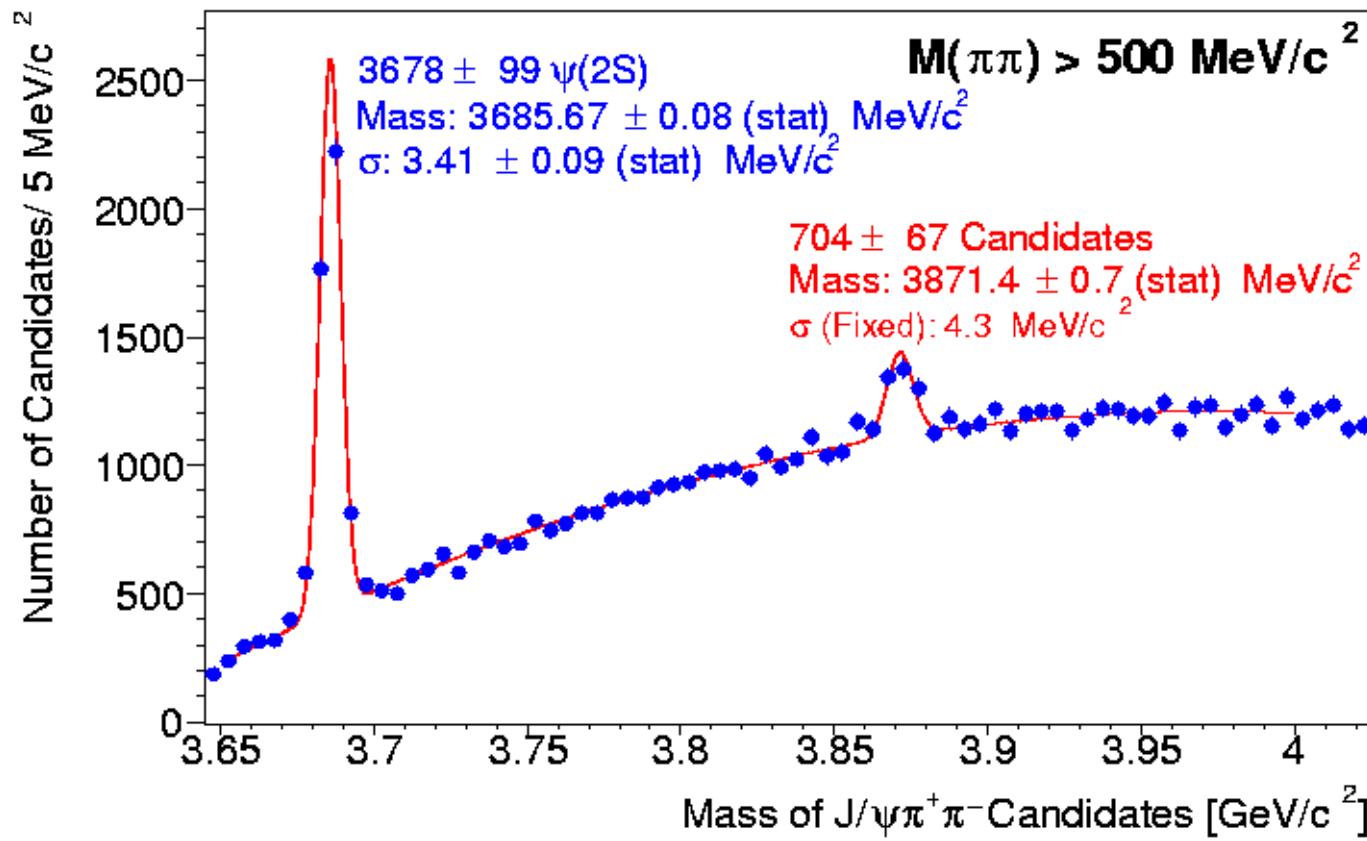
16 Jun 2000 05:34:19



Centre-of-Mass Energy 205 GeV

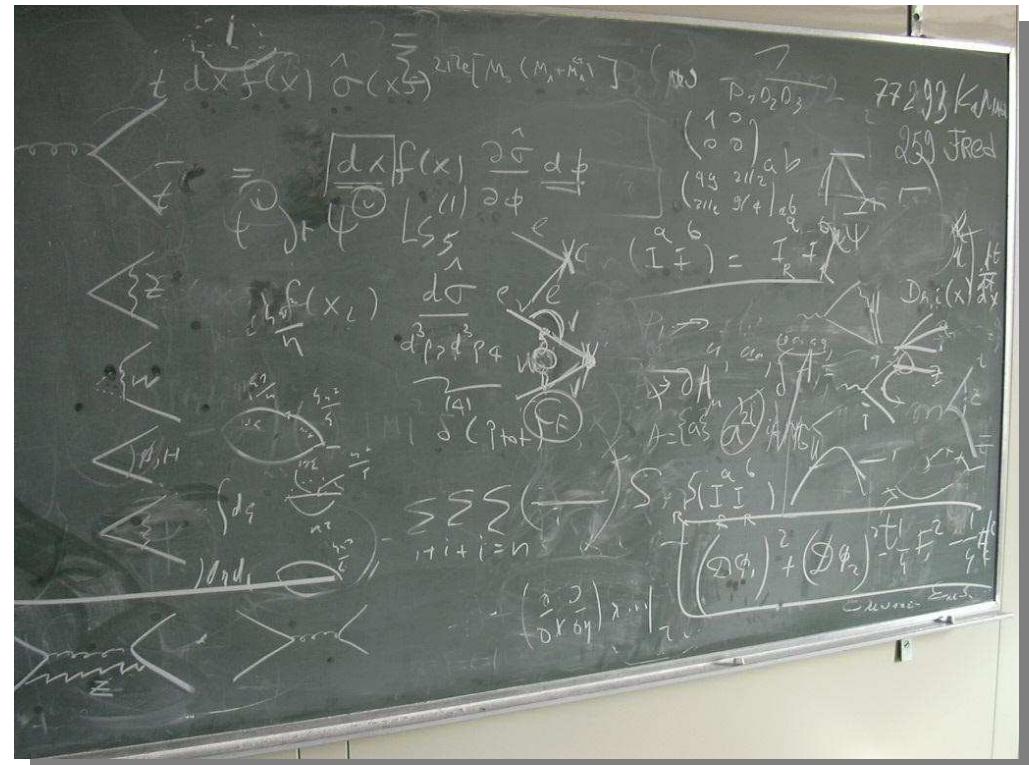


# Experiment: Messe Häufigkeitsverteilungen



# Theorie: Relativistische Quantenfeldtheorie

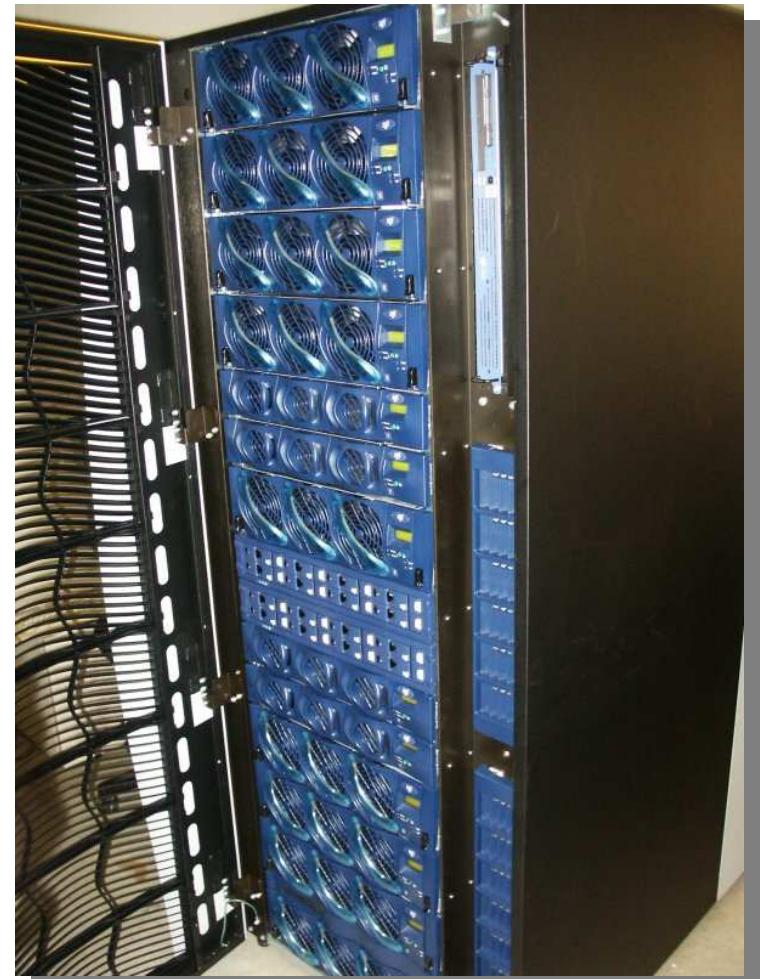
**Theoretiker berechnen  
Übergangsmatrixelemente  
in QED, QCD, SUSY etc.  
z.B. mit Hilfe von Feynman-  
Diagrammen.  
Diese enthalten Parameter  
(z.B. Massen,  
Kopplungskonstanten,  
Mischungswinkel)**



Durch Quadrieren der Matrixelemente und (teilweise) Integration über den Phasenraum (z.B. mit Monte-Carlo-Methode) können Vorhersagen über die Wahrscheinlichkeitsverteilungen im Experiment in Abhängigkeit der freien Parameter gemacht werden.

$$\frac{d\sigma}{d\text{LIPS}} = \|\mathcal{M}\|^2$$

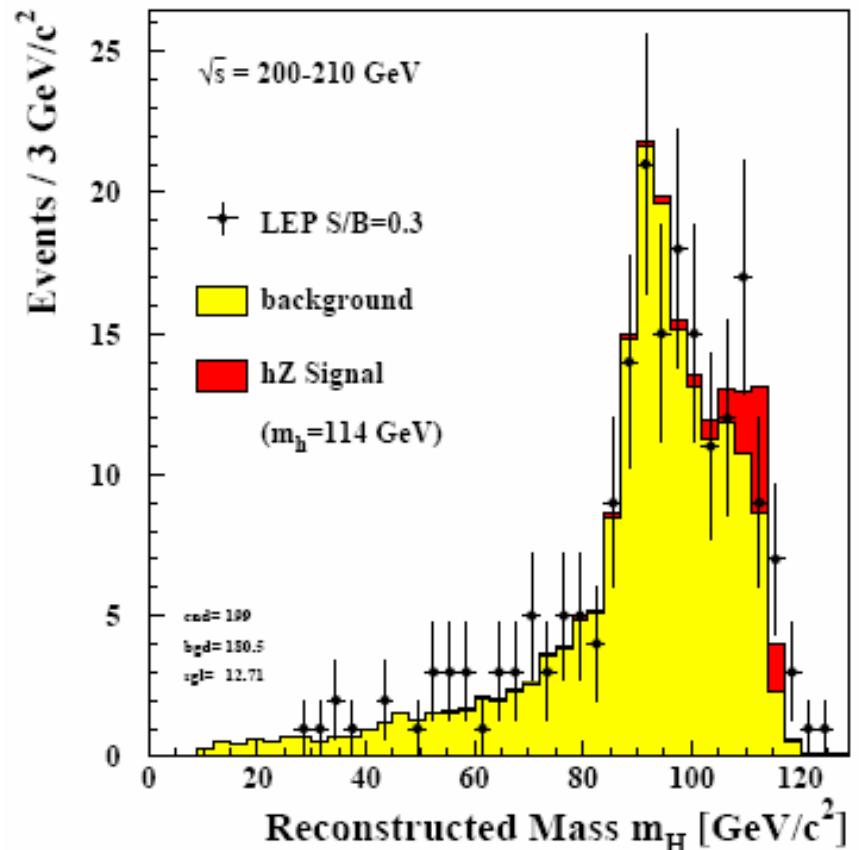
$$\sigma = \int \|\mathcal{M}\|^2 d\text{LIPS}$$



# Experimentalphysiker:

**Simuliere Theorie mit Hilfe von Monte Carlo, vergleiche Daten und Simulation,  
Bestimme so die physikalischen Parameter (z.B. Raten, Massen, Kopplungskonstanten, Mischungswinkel)**

**Wichtigstes  
Mathematisches Tool: Dreisatz  
Aber: manchmal viel mehr nötig!**



# Definition Wahrscheinlichkeit

Frequentist-Wahrscheinlichkeit = „objektive“ Definition  
für beliebig wiederholbare Ereignisse oder bei Vohandensein von  
Symmetrien anwendbar

Bayes-Wahrscheinlichkeit = „subjektive“ Definition  
auch für einmalige Ereignisse anwendbar

Streit der Schulen zwischen Frequentisten und Bayesianern

# Definition Wahrscheinlichkeit 2

## Frequentist-Definition von Wahrscheinlichkeit

### Kombinatorische Definition:

Wenn ein Ereignis in  $n$  verschiedenen Arten auftreten kann, die alle gleiche Wahrscheinlichkeit haben, und wenn  $k$  Ereignisse davon die Eigenschaft A aufweisen, ist die Wahrscheinlichkeit für A:  $P(A) = k/n$

### Empirische Definition:

Eine Beobachtung ist unter identischen Bedingungen unabhängig voneinander  $n$  mal wiederholt.

Wenn Eigenschaft A dabei  $k$  mal beobachtet wird, ist das Verhältnis  $k/n$  die empirische Wahrscheinlichkeit. Die Wahrscheinlichkeit  $P(A)$  wird definiert als der Grenzwert für unendlich viele Beobachtungen  $n$ .

### Beide Definitionen können kritisiert werden:

Kombinatorisch: Schlange, die sich in den Schwanz beißt.

Empirisch: Grenzwert kann in der Praxis nie erreicht werden.

Viele Probleme: Experimente nicht wiederholbar

# Definition Wahrscheinlichkeit 3

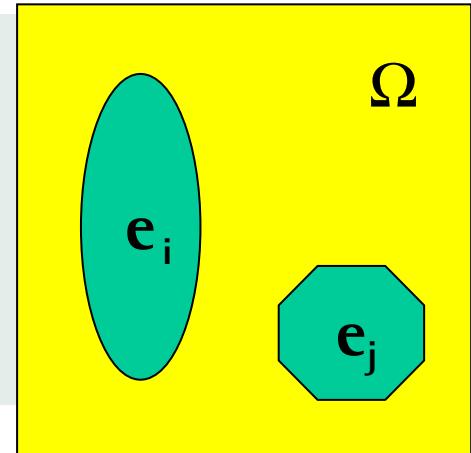
## Formale Definition von Wahrscheinlichkeit: Kolmogorov-Axiome (1931)

Elementary events  $e_i$  are considered, which *exclude* each other:

$e_i$  = elementary event

$\Omega$  = set of all elementary events

$P(e_i)$  = probability of event  $e_i$



- |   |          |
|---|----------|
| 1. $P(e_i) \geq 0$ for all $i$                        | positiv  |
| 2. $P(e_i \vee e_j) = P(e_i) + P(e_j)$ ( $\vee$ = or) | additiv  |
| 3. $\sum_{\Omega} P(e_i) = 1$                         | normiert |

# Kombinationen von Wahrscheinlichkeiten

$$P(\mathcal{A} \text{ or } \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B}) - P(\mathcal{A} \text{ and } \mathcal{B})$$

$$P(\mathcal{A} \text{ and } \mathcal{B}) = 0$$

$$P(\mathcal{A} \text{ or } \mathcal{B}) = P(\mathcal{A}) + P(\mathcal{B})$$

$$P(\mathcal{A} \text{ or } \overline{\mathcal{A}}) = P(\mathcal{A}) + P(\overline{\mathcal{A}}) = 1$$

$$P(\mathcal{A} \text{ and } \mathcal{B}) = P(\mathcal{A}) \cdot P(\mathcal{B}|\mathcal{A}) = P(\mathcal{B}) \cdot P(\mathcal{A}|\mathcal{B})$$

$$P(\mathcal{A} \text{ and } \mathcal{B}) = P(\mathcal{A}) \cdot P(\mathcal{B})$$

$\mathcal{A}$  or  $\mathcal{B}$

if  $\mathcal{A}$  and  $\mathcal{B}$  exclude each other

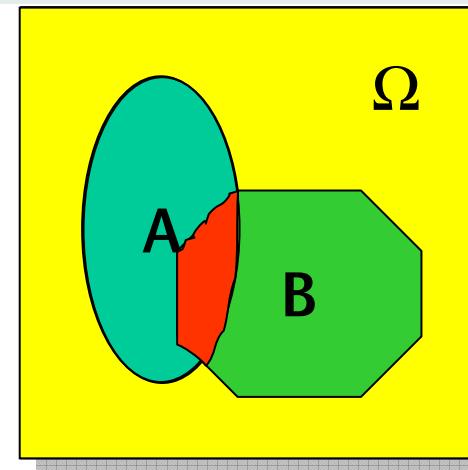
if  $\mathcal{A}$  and  $\mathcal{B}$  exclude each other

( $\overline{\mathcal{A}}$  = not  $\mathcal{A}$ )

$\mathcal{A}$  and  $\mathcal{B}$

if  $\mathcal{A}$  and  $\mathcal{B}$  independent

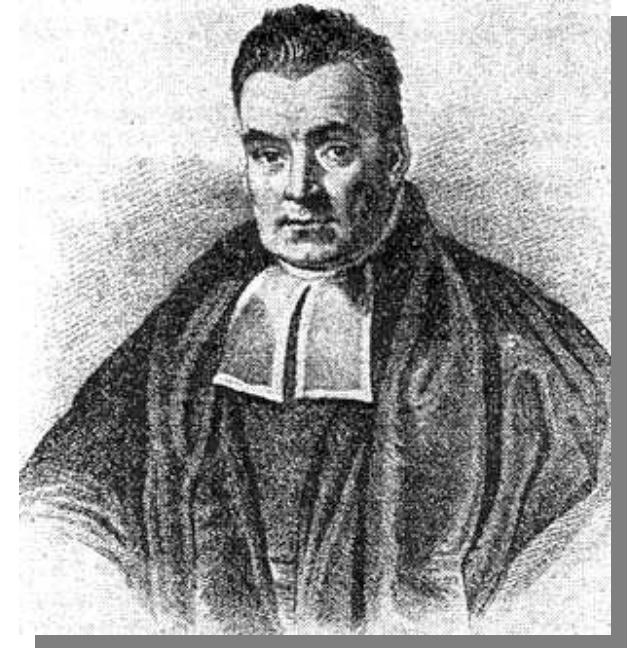
Bedingte Wahrscheinlichkeit, dass  
 $A$  wahr ist, wenn  $B$  wahr ist.



# Reverend Thomas Bayes (1702 – 1761)

Wahrscheinlichkeit ist der Grad des Glaubens, dass ein Experiment ein bestimmtes Ergebnis haben wird.

-Subjektive Wahrscheinlichkeit-  
(erfüllt Kolmogorov-Axiome !)



*Essay Towards Solving a Problem in the Doctrine of Chances* (1763), posthum veröffentlicht in *Philosophical Transactions of the Royal Society of London*.

# Beispiele für Bayes-Wahrscheinlichkeit

Frequenz-Aussagen oft nicht möglich. Dann ist Bayes- Interpretation die einzige mögliche:

Wahrscheinlichkeit ist der Grad des Glaubens, dass eine Aussage zutrifft:

Das Teilchen in diesem Ereignis ist ein Positron.

Die Natur ist supersymmetrisch.

Es wird morgen regnen.

Deutschland wird 2006 Fussball-Weltmeister.

Es hat am 8. März 1792 in Kairo geregnet.

Oft kritisiert, weil „subjektiv“ und „unwissenschaftlich“. Beruht jedoch auf einfacher Wahrscheinlichkeitsrechnung und ist, richtig angewendet, nicht im Widerspruch zu Frequentist-Ansatz.

# Bayes' Theorem (1)

Bedingte (conditional) Wahrscheinlichkeiten:

$$P(B|A) = \frac{P(B \cap A)}{P(A)}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

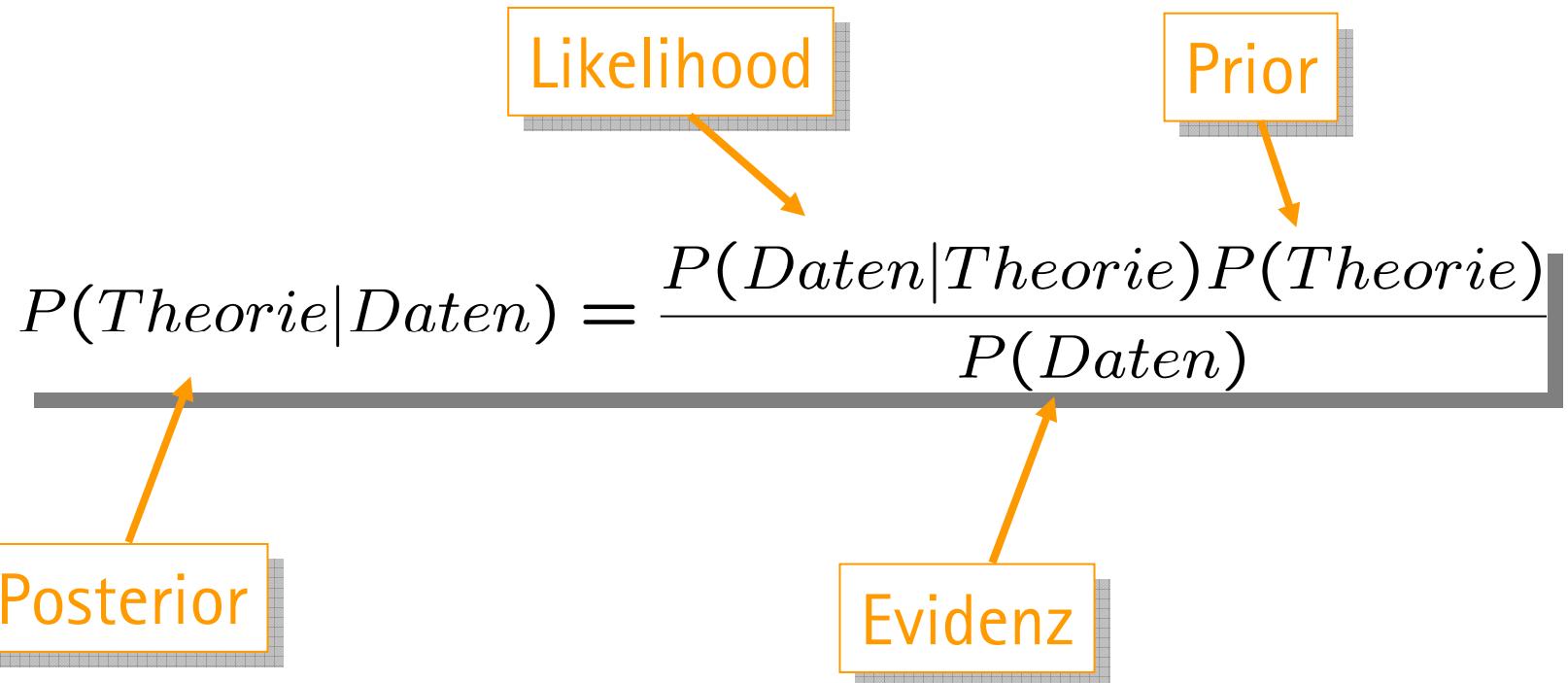
Wegen  $P(A \cap B) = P(B \cap A)$  gilt:

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

Bayes' Theorem

# Bayes' Theorem (2)

Besonders wichtig durch die Interpretation A=Theorie B=Daten

$$P(\text{Theorie}|\text{Daten}) = \frac{P(\text{Daten}|\text{Theorie})P(\text{Theorie})}{P(\text{Daten})}$$


Likelihood

Prior

Posterior

Evidenz

# Bayes' Theorem (3): Beispiel AIDS-Test

Wahrscheinlichkeit in  
allgemeiner Bevölkerung:

$$P(\text{AIDS}) = 0.001$$

$$P(\text{no AIDS}) = 0.999$$

a priori-Wissen

Ziemlich zuverlässiger AIDS-Test  
(Resultat + oder -):

$$P(+|\text{AIDS}) = 0.98$$

$$P(-|\text{AIDS}) = 0.02$$

$$P(+|\text{no AIDS}) = 0.03$$

$$P(-|\text{no AIDS}) = 0.97$$

Messung,  
Likelihoods

Wie besorgt sollte man sein, wenn man ein positives Testresultat hat?  
d.h. wie groß ist (die a posteriori-Wahrscheinlichkeit)  $P(\text{AIDS}|+)$ ?

## Bayes' Theorem (4): Beispiel AIDS-Test (2)

$$\begin{aligned} P(\text{AIDS}|+) &= \frac{P(+|\text{AIDS}) P(\text{AIDS})}{P(+|\text{AIDS}) P(\text{AIDS}) + P(+|\text{no AIDS}) P(\text{no AIDS})} \\ &= \frac{0.98 \times 0.001}{0.98 \times 0.001 + 0.03 \times 0.999} \\ &= 0.032 \end{aligned}$$

Die Posterior-Wahrscheinlichkeit  $P(\text{AIDS}|+)$  beträgt nur 3,2%!  
Warum? Wegen der kleinen Prior-Wahrscheinlichkeit von 0.01% und  
der nicht vernachlässigbaren Mißidentifikationswahrscheinlichkeit!  
Vorsicht: Prior ist nicht richtig, wenn man zu einer Risikogruppe gehört!

# Bayes'sche vs. klassische Statistik

Klassische Statistik ist nur Sonderfall der Bayes-Statistik:

$$P(\text{Theorie}|\text{Daten}) = \frac{P(\text{Daten}|\text{Theorie})P(\text{Theorie})}{P(\text{Daten})}$$

Maximieren der Likelihood statt  
der a posteriori-Wahrscheinlichkeit  
heisst:

Implizite Annahme, dass die  
Prior-Wahrscheinlichkeit flach  
verteilt ist, d.h. jeder Wert ist  
gleich wahrscheinlich.

Hört sich vernünftig an, ist aber falsch!  
Heisst nicht, dass man nichts weiss!

# Nicht-informativer Prior

Wenn man über eine dimensionsbehaftete Größe  $x$  nichts (nicht einmal die Größenordnung) weiß, dann ist

$$f(x) = \text{const.} \quad \text{FALSCH!}$$

Vielmehr ist

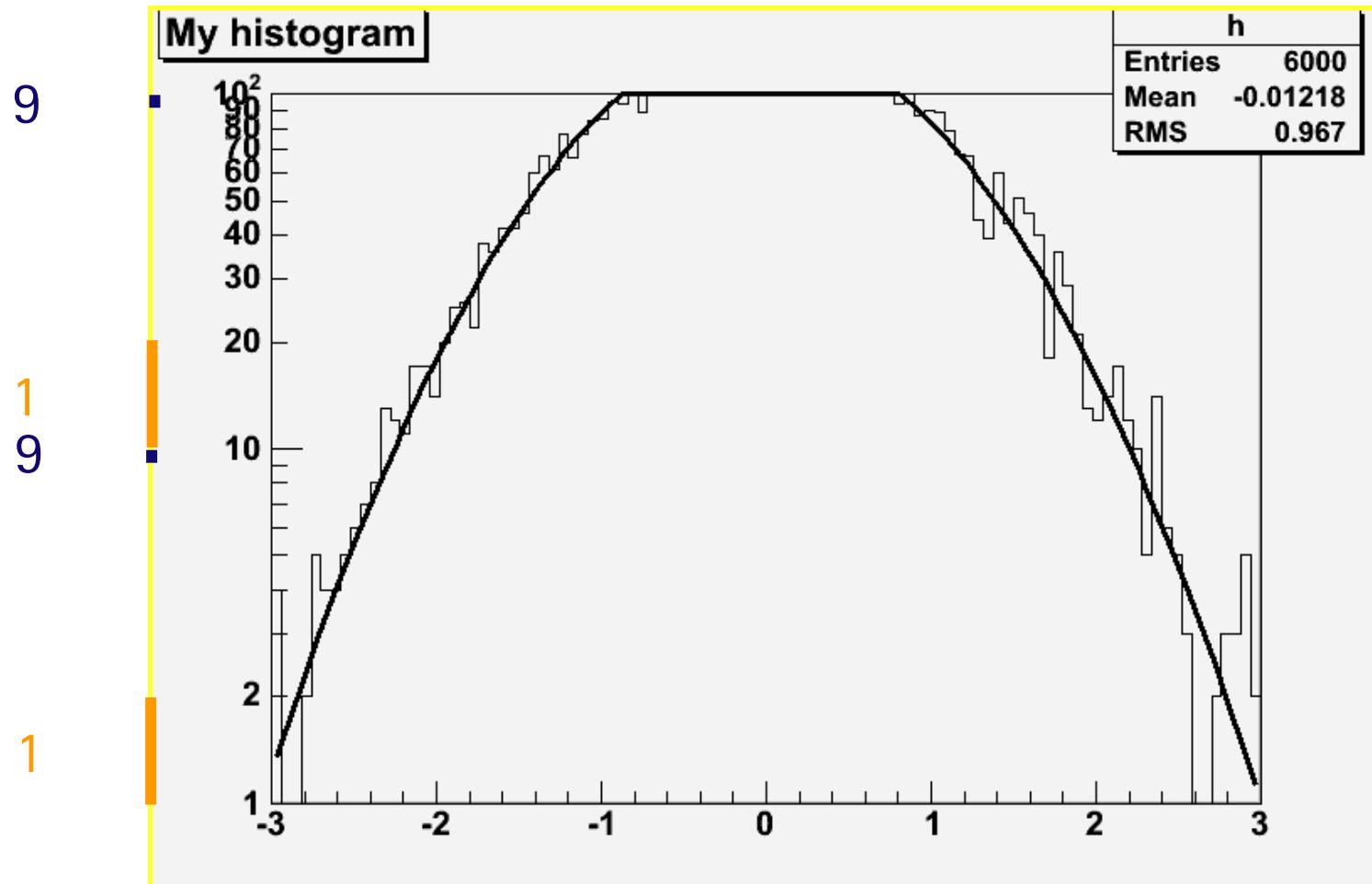
$$f(\ln(x)) = \text{const.} \quad \text{RICHTIG!}$$

Das ist gleichbedeutend mit

$$f(x) \propto 1/x.$$

Bradfordsche Zahlengesetz: Die Ziffer 1 ist viel öfter die signifikanteste Ziffer einer Zahl als die 9. Wie auf logarithmischer Skala. Übrigens: Finanzamt weiß Bescheid: Steuererklärungen richtig fälschen!

# Nicht-informativer Prior

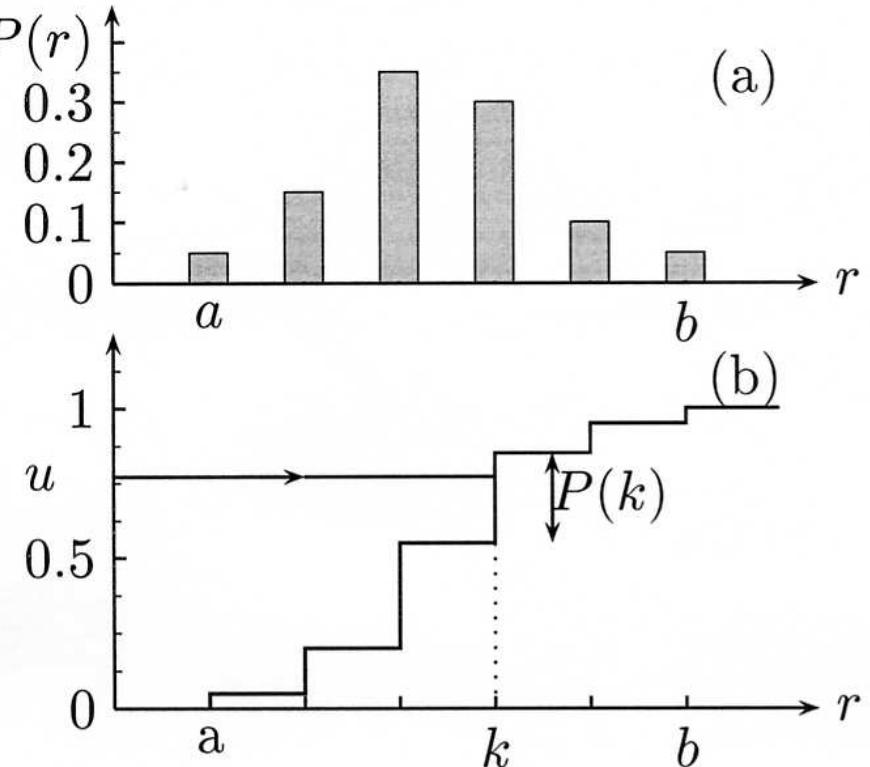


# Diskrete Zufallszahlen

Eine diskrete Variable  $k$  kann eine endliche oder unendliche Anzahl von Werten  $k_i$  with  $i = a, a+1, \dots, b$  annehmen. Die Wahrscheinlichkeit für den Wert  $k_i$  ist  $P(k_i)$  hat die Normierung

$$\sum_{i=a}^b P(k_i) = 1.$$

Kumulierte Verteilungsfunktion  
(=Wahrscheinlichkeit, Wert  $k_i$  oder einen kleineren zu beobachten)



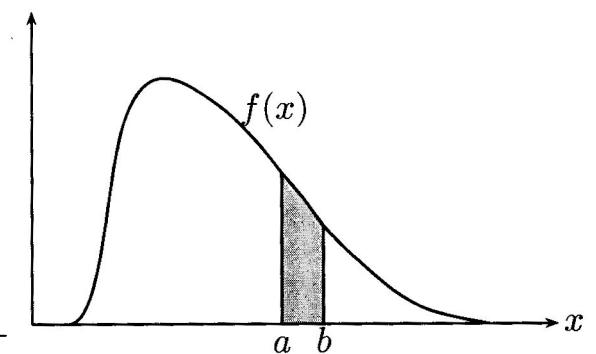
# Wahrscheinlichkeitsdichte

Eine kontinuierliche Zufallsvariable  $x$  kann reellwertige Werte annehmen. Die Wahrscheinlichkeit, dass  $x$  in das Intervall  $a \leq x < b$  fällt, ist

$$P(a \leq x < b) = \int_a^b f(x)dx$$

Dabei ist  $f(x)$  die *Wahrscheinlichkeitsdichte* (Probability density function, pdf) der Zufallsvariablen  $x$ ; Die Dichte ist nicht-negativ und auf 1 normiert:

$$f(x) \geq 0 \quad \int_{-\infty}^{+\infty} f(x)dx = 1$$



# Verteilungsfunktion

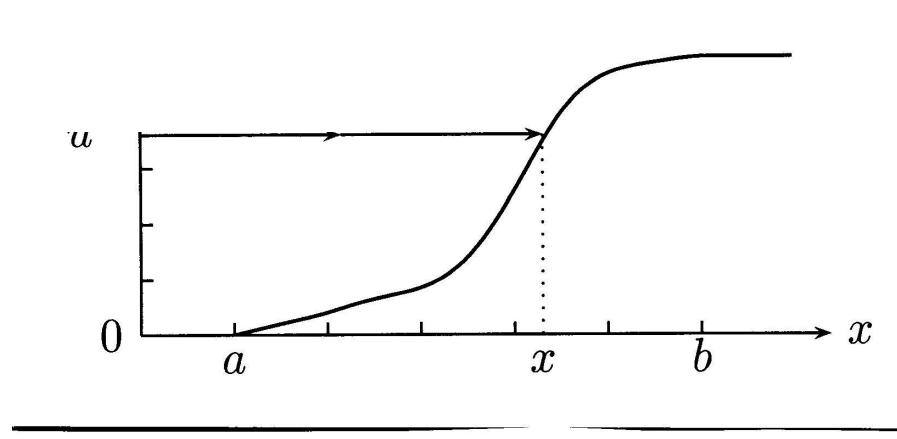
Die *Verteilungsfunktion* (distribution function)  $F(x)$  ist als kumulierte Wahrscheinlichkeit, einen kleineren Wert als  $x$  zu finden, definiert:

$$F(x) = \int_{-\infty}^x f(x')dx'$$

mit  $F(-\infty) = 0$  und  $F(+\infty) = 1$ .

Die Wahrscheinlichkeitsdichte ist also die Ableitung der Verteilungsfunktion:

$$f(x) = dF(x)/dx.$$



# Erwartungswert und Varianz

Der *Erwartungswert* einer Funktion  $h(x)$ , wobei  $x$  eine Zufallsvariable mit einer Wahrscheinlichkeitsdichte  $f(x)$  ist, ist definiert durch

$$E[h] = \int_{-\infty}^{+\infty} h(x) f(x) dx \quad E[h] = \sum_{i=a}^b h(r_i) P(r_i).$$

Wichtige Spezialfälle sind der Erwartungswert für  $x$  bzw.  $k$  (der Mittelwert) und die *Varianz* von  $x$  bzw.  $k$ :

$$E[x] = \int_{-\infty}^{+\infty} x f(x) dx = \langle x \rangle \quad E[k] = \sum_{i=a}^b k_i P(k_i) = \langle k \rangle$$

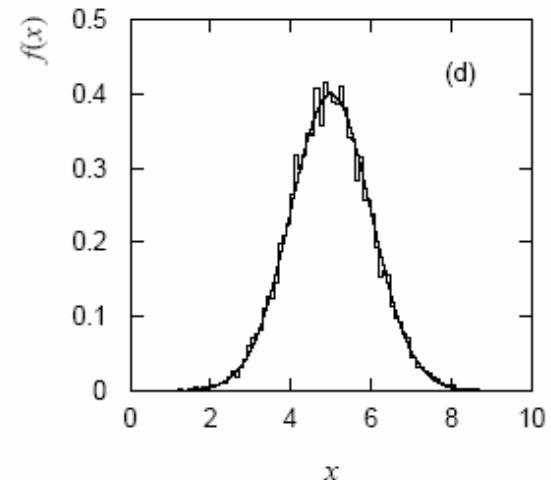
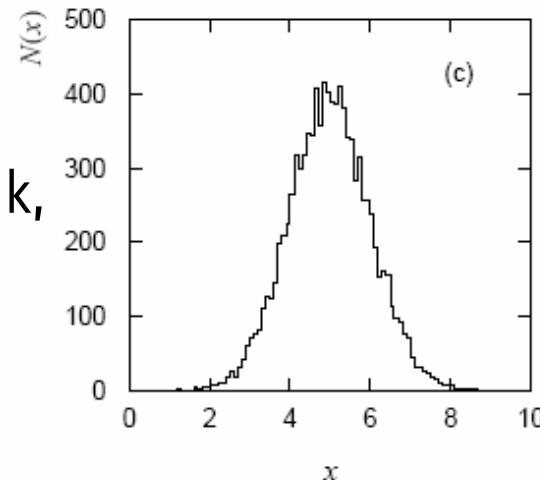
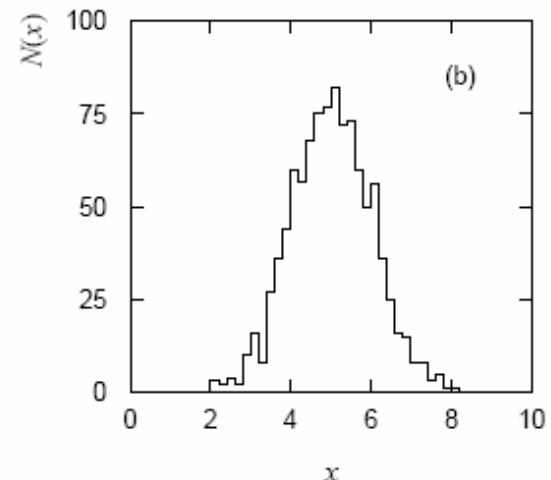
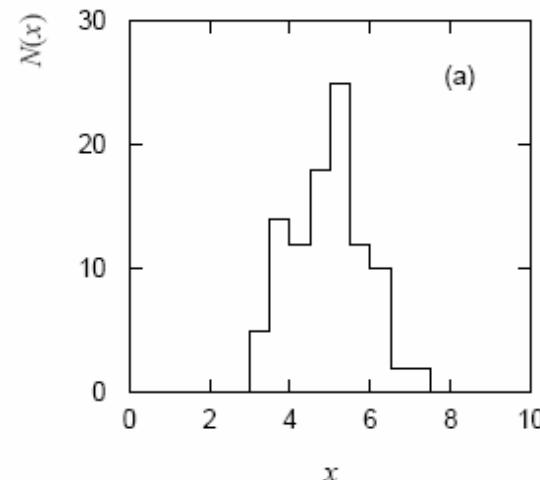
$$V[x] = \int_{-\infty}^{+\infty} (x - \langle x \rangle)^2 f(x) dx \quad V[k] = \sum_{i=a}^b (k_i - \langle k \rangle)^2 P(k_i)$$

Standardabweichung

$$\sigma = \sqrt{V}$$

# Histogramme

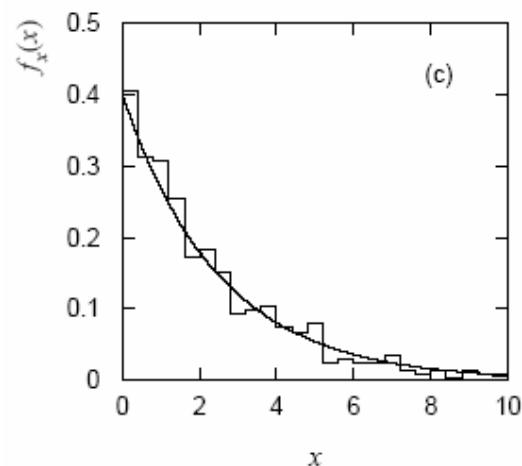
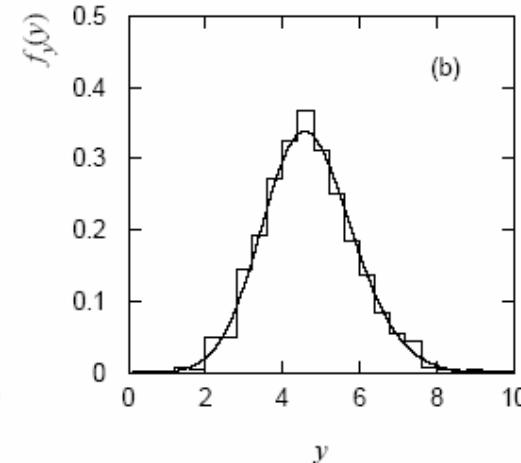
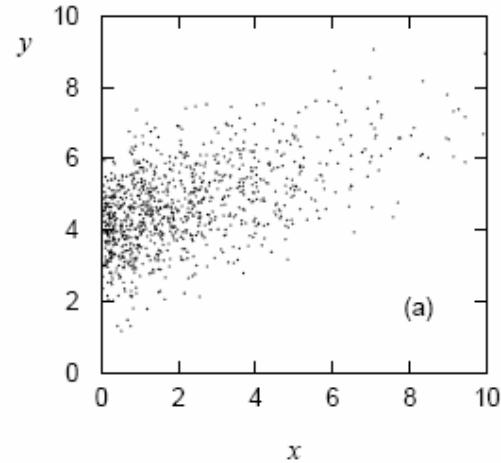
Häufigkeitsverteilung:  
Anzahl Ereignisse  
in endlichen Intervallen  
(Bins)



PDF  $f(x)$  = Histogram  
mit unendlicher Statistik,  
Binbreite Null,  
normiert auf Fläche 1

# Mehrdimensionale Zufallszahlen

2-dim. PDF:  
 $f(x,y)$   
(Scatter-Plot)

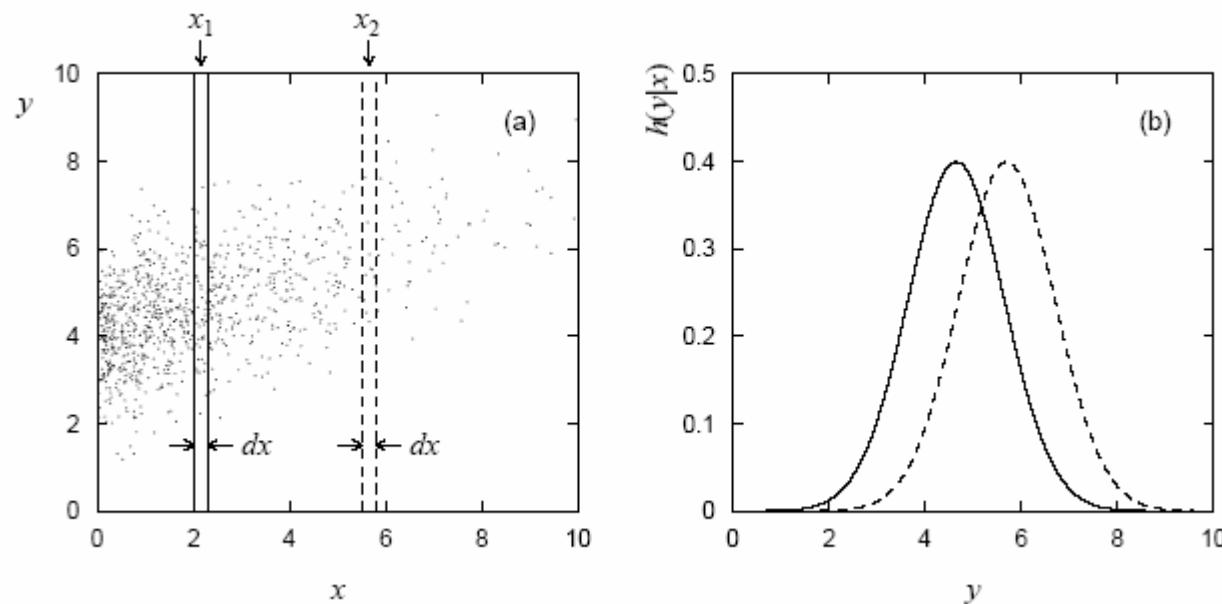


Randverteilungen  
=Projektionen auf Achsen

$$f_x(x) = \int f(x, y) dy$$

$$f_y(y) = \int f(x, y) dx$$

# Mehrdimensionale Zufallszahlen (2)



Bedingte Wahrscheinlichkeitsdichten

$$h(y|x) = \frac{f(x,y)}{f_x(x)}$$

# Applets (CUWUStat)



Verknüpfung mit datalist.lnk

Daten, Histogramme



Verknüpfung mit boxmodel.lnk

Münze, Würfel



Verknüpfung mit Games.lnk

Ziegenproblem, Poker

# Binomialverteilung (1)

Wenn die Wahrscheinlichkeit für ein bestimmtes Ereignis („Erfolg“)  $p$  ist, ist die Wahrscheinlichkeit, in  $n$  Experimenten exakt  $k$  Erfolge zu beobachten, durch die Binomialverteilung gegeben:

$$P_n(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, \dots n$$

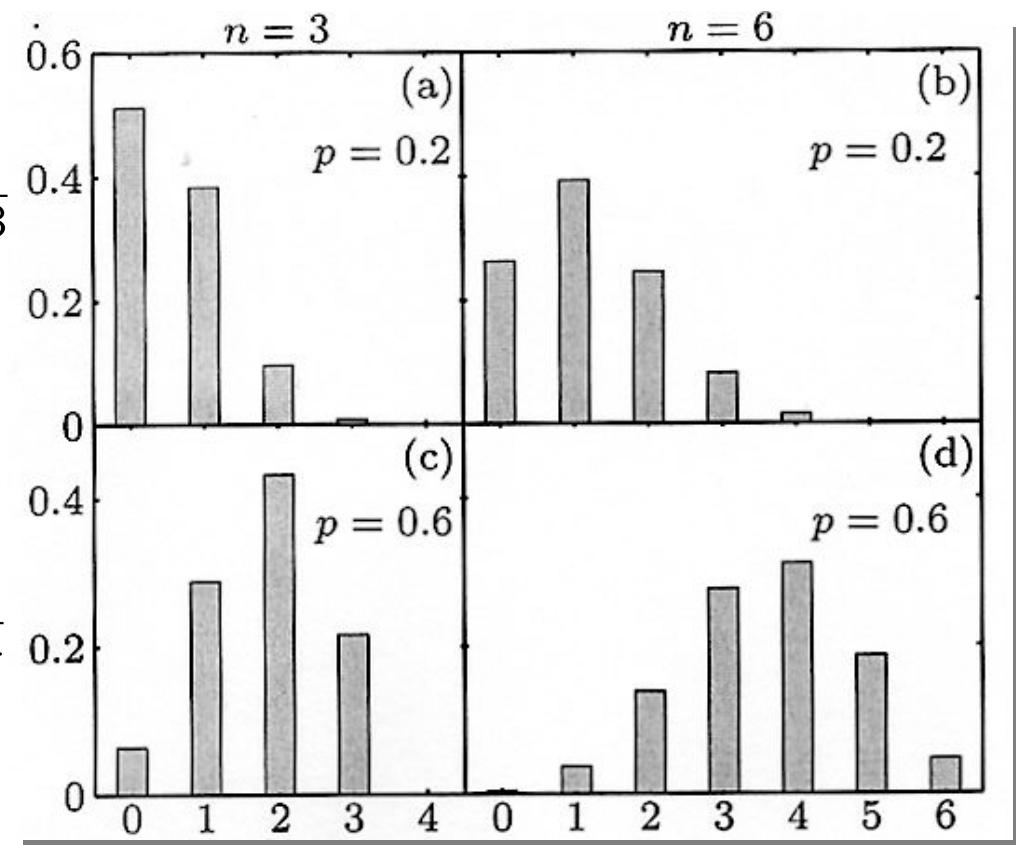
Der Faktor  $\binom{n}{k}$  ist die Anzahl der Kombinationen,  $k$  Elemente aus  $n$  unterschiedlichen Elementen zu selektieren.

$$\text{Erwartungswert} = E[k] = np$$

$$\text{Varianz} = V[k] = np(1-p)$$

# Binomialverteilung (2)

$$\begin{aligned}\mu &= 0.6 \\ \sigma &= \sqrt{3 \cdot 0.2 \cdot 0.8} \\ \sigma &\approx 0.7\end{aligned}$$



$$\begin{aligned}\mu &= 1.2 \\ \sigma &= \sqrt{6 \cdot 0.2 \cdot 0.8} \\ \sigma &\approx 0.98\end{aligned}$$

$$\begin{aligned}\mu &= 3.6 \\ \sigma &= \sqrt{6 \cdot 0.6 \cdot 0.4} \\ \sigma &= 1.2\end{aligned}$$

# Poissonverteilung

Betrachte Binomialverteilung im Grenzfall  $n \rightarrow \infty$ ,  $p \rightarrow 0$ , aber  $n \cdot p \rightarrow \mu$  (fester Erwartungswert):

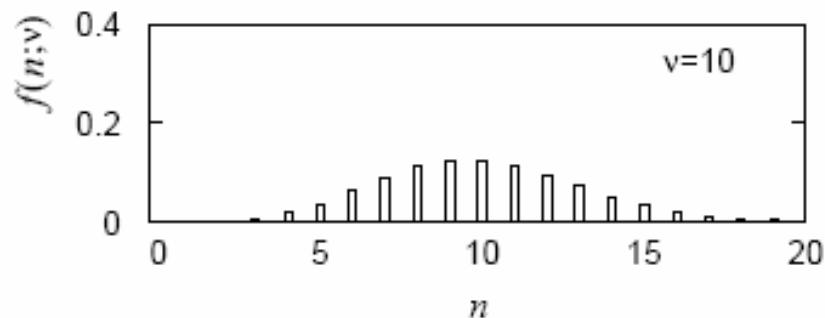
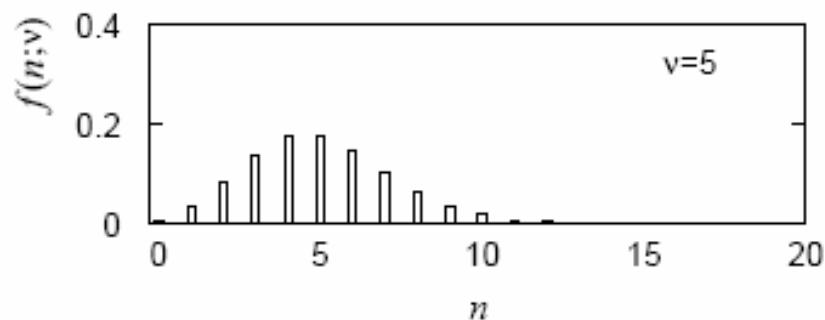
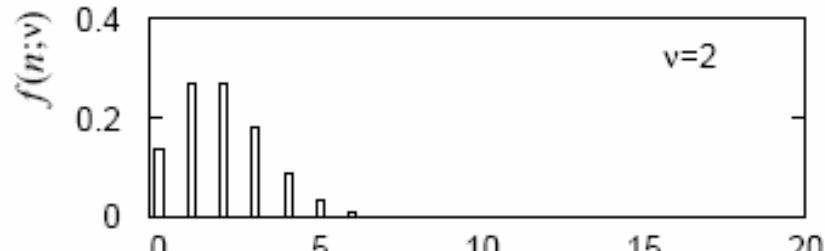
$$P_\mu(k) = \frac{\mu^k}{k!} e^{-\mu} \quad k = 0, 1, \dots$$

$$\text{Erwartungswert } E[k] = \mu \qquad \qquad \text{Varianz } V[k] = \mu$$

Für in der Zeit  $t$  konstante Ereignisraten folgt die Anzahl der in einem festen Zeitintervall beobachteten Ereignisse einer Poisson-Verteilung, und die Zeitdifferenz zwischen zwei Ereignissen einer Exponentialverteilung:

$$f(t; \tau) dt = \frac{1}{\tau} e^{-t/\tau} dt$$

# Poissonverteilung



Erwartungswert  $\mu$   
Standardabweichung  $\sqrt{\mu}$

Grenzfall für große  $\mu$ :  
Normalverteilung

Anwendung:  
Bininhalt in einem  
Histogramm

# Normalverteilung

Die Normalverteilung (oder Gauß-Verteilung) ist die wichtigste kontinuierliche Verteilung.

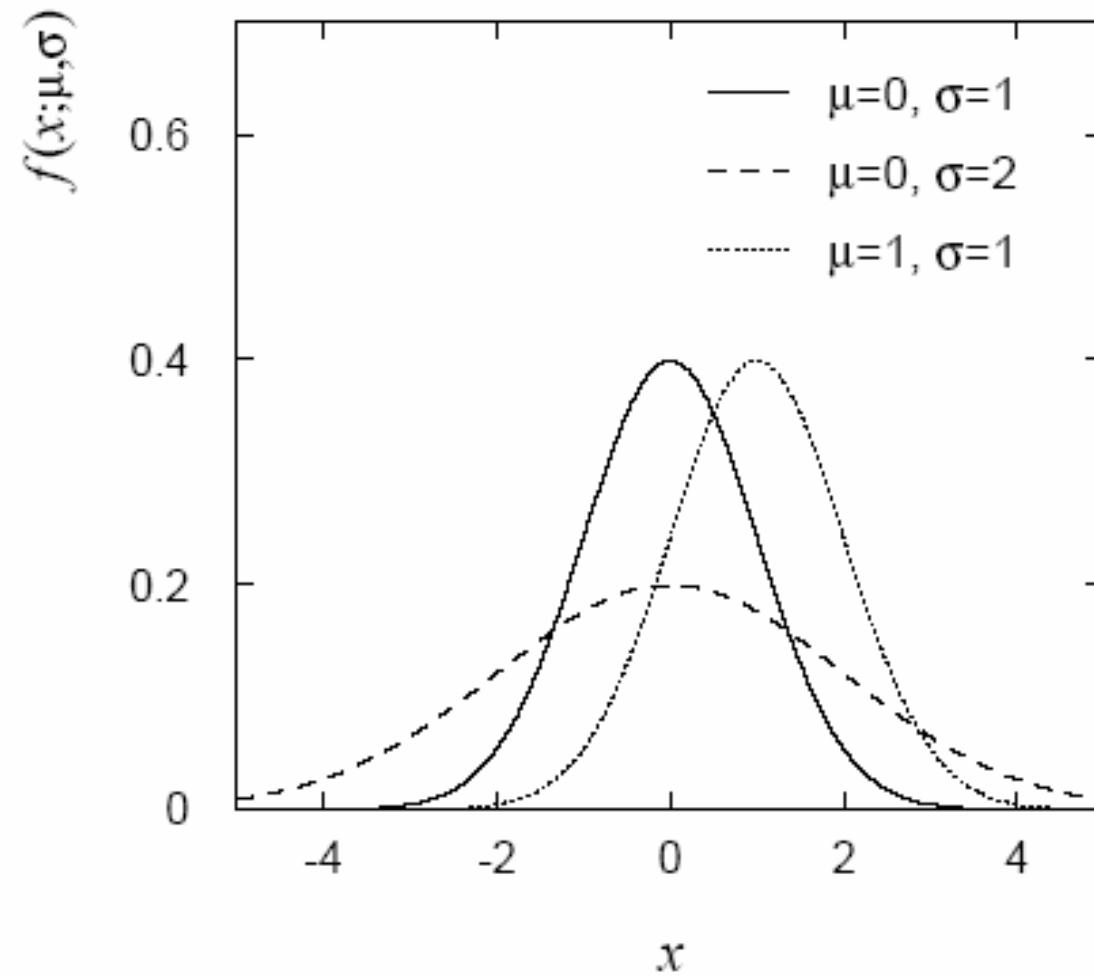
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad x \in (-\infty, \infty)$$

Erwartungswert  $E[x] = \mu$       Varianz  $V[x] = \sigma^2$

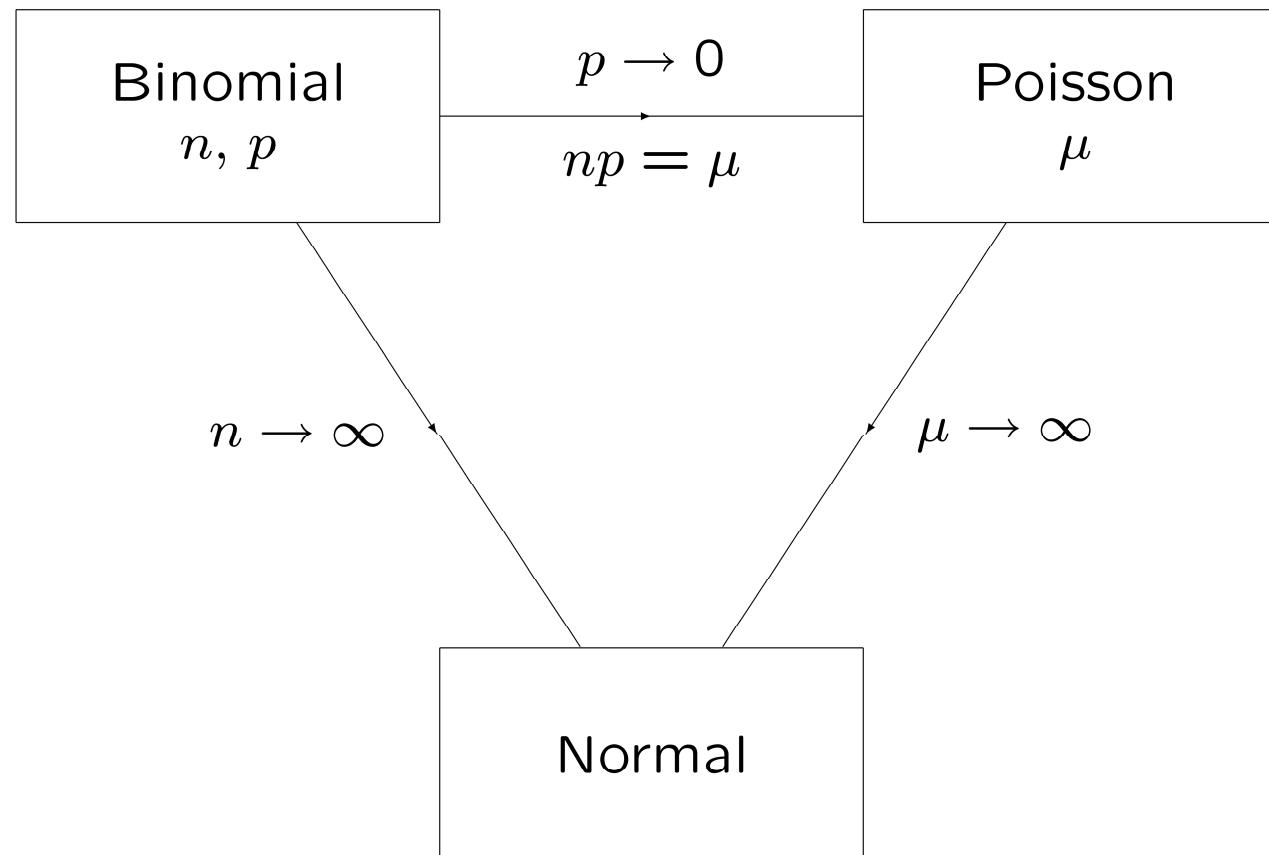
$$\begin{aligned} P(|x - \mu| < 1 \cdot \sigma) &= 68.26\% \\ P(|x - \mu| < 2 \cdot \sigma) &= 95.45\% \\ P(|x - \mu| < 3 \cdot \sigma) &= 99.73\% \end{aligned}$$

Für große  $\mu$  nähert sich die Poisson-Verteilung einer Gauss-Verteilung mit  $\sigma = \sqrt{\mu}$  an.

# Normalverteilung



# Standardverteilungen



# Zentraler Grenzwertsatz

Im Grenzfall von großen  $n$  ist die Summe  $\sum_i x_i$  von unabhängigen Zufallszahlen eine Zufallszahl mit Erwartungswert

$$\mu = \sum_i \mu_i$$

und Varianz

$$\sigma^2 = \sum_i \sigma_i^2$$

die der *Normalverteilung* folgt, *unabhängig von der Verteilungsfunktion der individuellen Variablen  $x_i$* .

Demonstration

Verknüpfung mit Gaussian-Distribution

# Chi-Quadrat-Verteilung

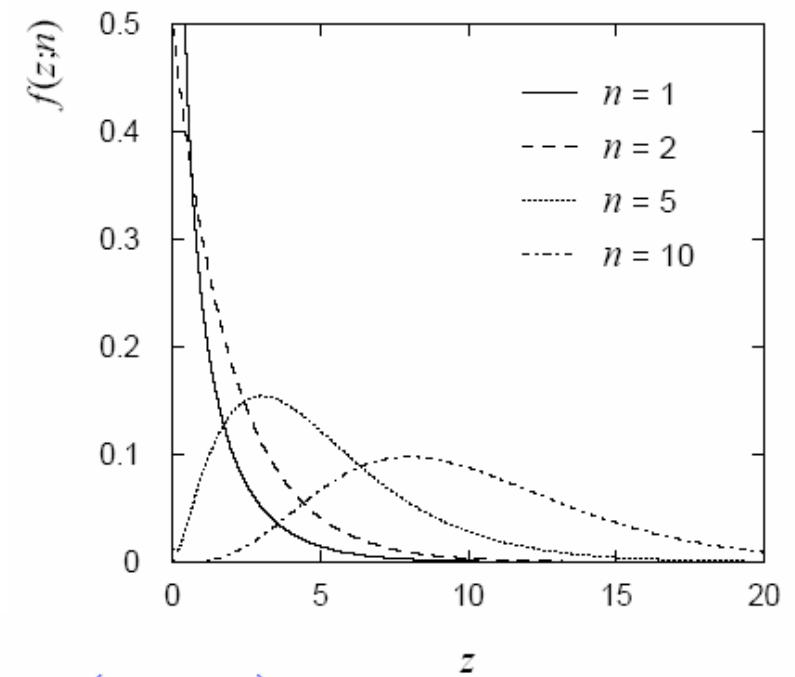
Beschreibt die Summe der quadratischen Abweichungen vom Erwartungswert einer n-dimensionalen Normalverteilung.

$$\text{Erwartungswert } E(z) = E(\chi^2) = n$$

$$\text{Varianz } V(z) = 2n$$

$$\text{Standardabweichung } \sigma(z) = \sqrt{2n}$$

$$f(z; n) = \frac{1}{2^{n/2}\Gamma(n/2)} z^{n/2-1} e^{-z/2} \quad (z \geq 0)$$



(=Quadrat der Radien von n-dimensionalen Vektoren.  
In hochdimensionalen Räumen steckt das Volumen einer Kugel fast vollständig nahe der „Oberfläche“.)

# Cauchy- (=Breit-Wigner-) Verteilung

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

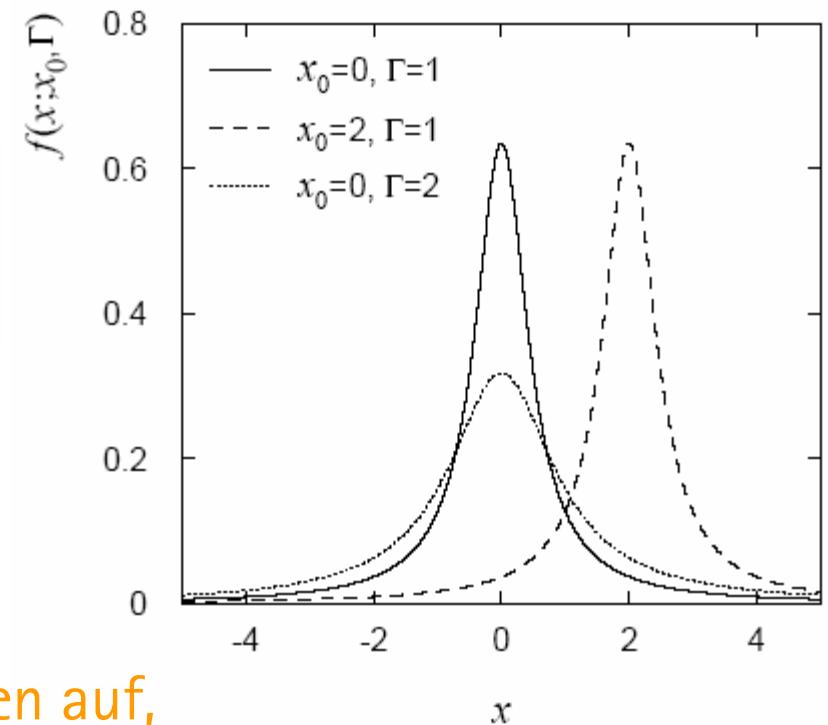
$$f(x; \Gamma, x_0) = \frac{1}{\pi} \frac{\Gamma/2}{\Gamma^2/4 + (x - x_0)^2}$$

Erwartungswert  $E(x) = x_0$   
 (aber schlecht definiert)

Halbwertsbreite (FWHM) =  $\Gamma$   
 Standardabweichung  $\sigma(x) = \infty$

Tritt bei allen Resonanzphänomenen auf,  
 ist Fouriertransformierte (im Frequenz- = Energieraum)  
 der Exponentialverteilung (in Zeit t).

Unschärferelation: Resonanzbreite =  $h/\text{Lebensdauer}$



# Signale finden...

Suche kleines Resonanzsignal in großem Untergrund.  
Maximiere Signal-Effizienz  
Minimiere Untergrund

Wenn Signal  $> 3\sigma$  Signifikanz  $\rightarrow$  Evidenz  
Wenn Signal  $> 5\sigma$  Signifikanz  $\rightarrow$  Entdeckung

Wenn Signal  $< 3\sigma$  Signifikanz  $\rightarrow$  keine Evidenz  $\rightarrow$  Upper Limit

## Signale finden (2)

„Gute“ Upper Limits, wenn gar keine Signal-Kandidaten

„Gute“ Evidenz, wenn möglichst viele Signal-Kandidaten

→ Psychologisches Bias ?

Grenzbereich schwierig und gefährlich:

zu kleine obere Schranken, zu große erste Evidenzen

Optimieren durch Cuts muss sinnvoll sein!

Soll nicht an Daten „optimiert“ werden, sondern an Simulationen, bevor die Daten gesichtet werden.

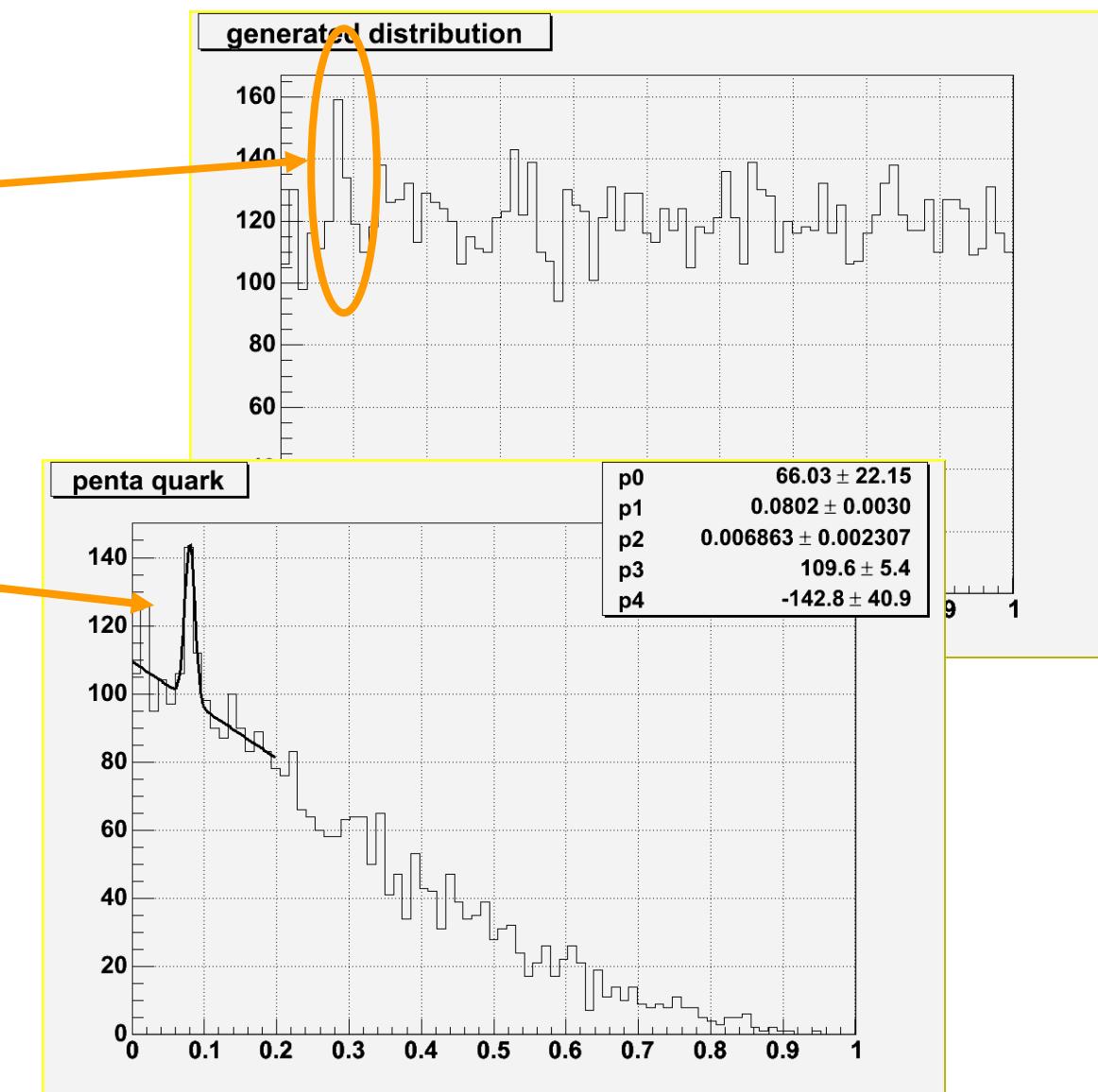
(Sicher: Blind-Analysen)

# Beispiel:

Signal?

Cuts auf weitere  
Variablen 2 und 3:  
=> Signifikanz erhöht!

Fit:  $3\sigma$  Signifikanz.  
Evidenz?



# Beispiel für falsche Evidenz

Wahrheit: völlig zufällige Monte-Carlo-Verteilung  
also nur statistische Fluktuation.

Keine Korrelation mit anderen Variablen 2 und 3.

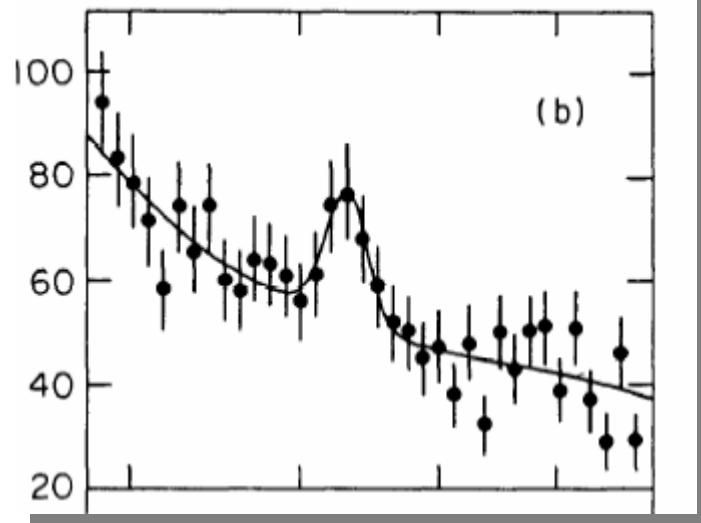
Wahrscheinlichkeit von  $3 \sigma$  sollte sehr selten sein ( $<0.3\%$ ).

Ja, aber ohne a priori- Modell Suche in vielen Bins (und Histogrammen),

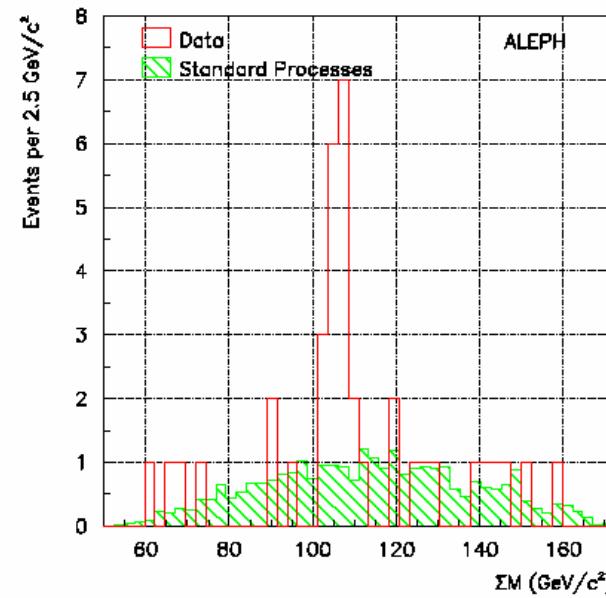
Statistische Fluktuationen der anderen Variablen können durch Cuts benutzt werden, um die Signifikanz des „Signals“ zu erhöhen.

Bayes Theorem -- Subjektive Wahrscheinlichkeit??

# Beispiele für falsche Evidenzen



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



ALEPH 4-Jet Signal  
bei LEP 1.5  
Kandidat für hA

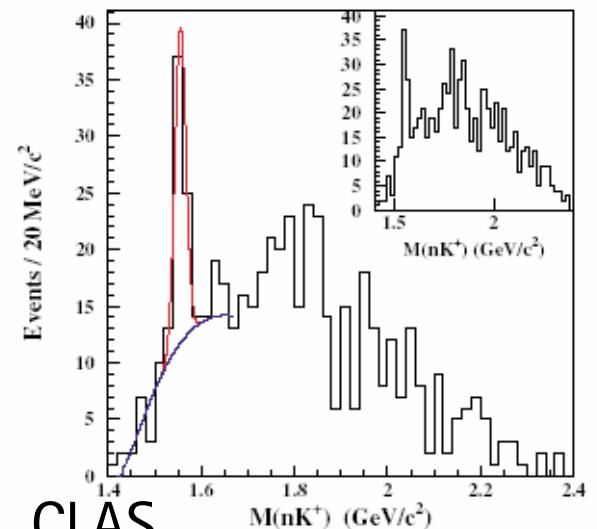
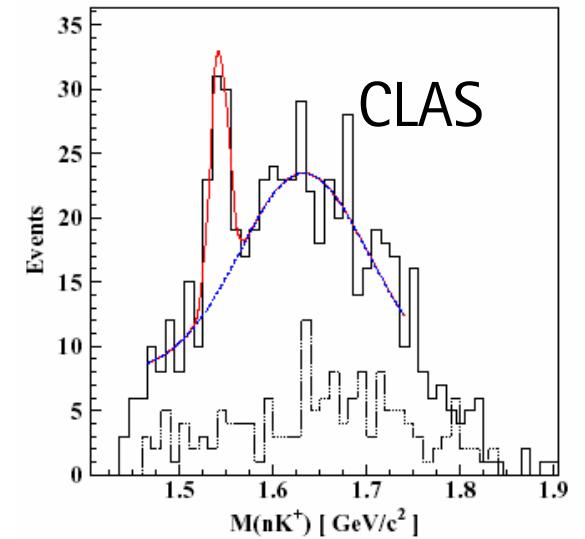
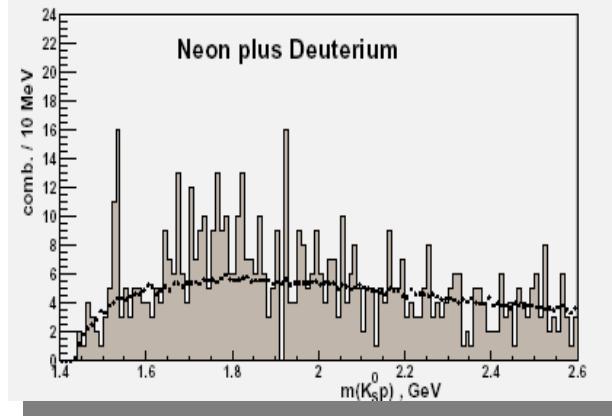
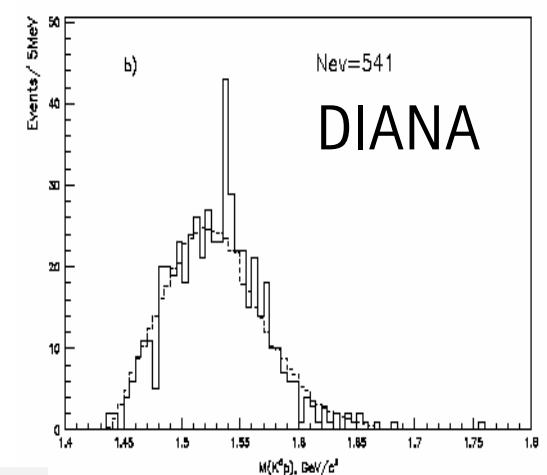
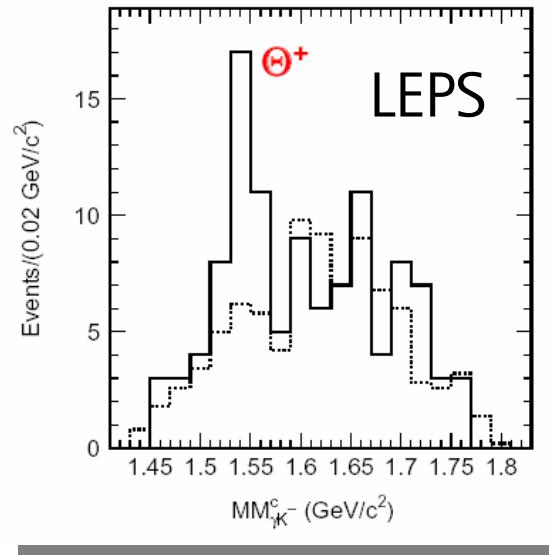
# Beispiele für falsche Evidenzen

Es wurden bereits entdeckt:

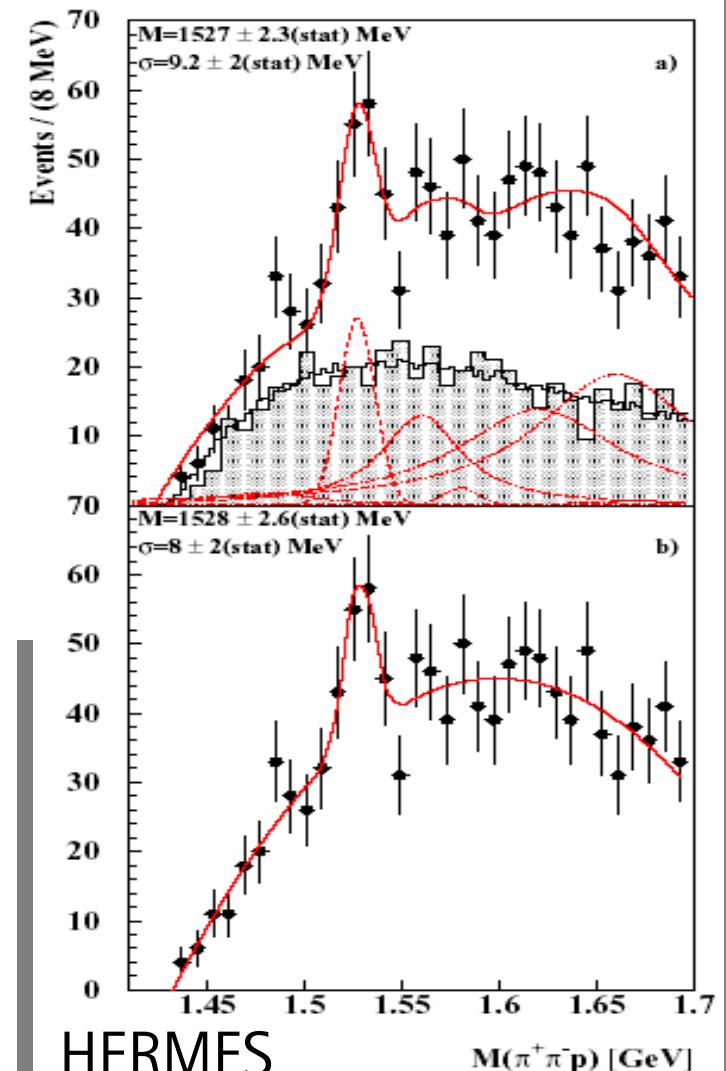
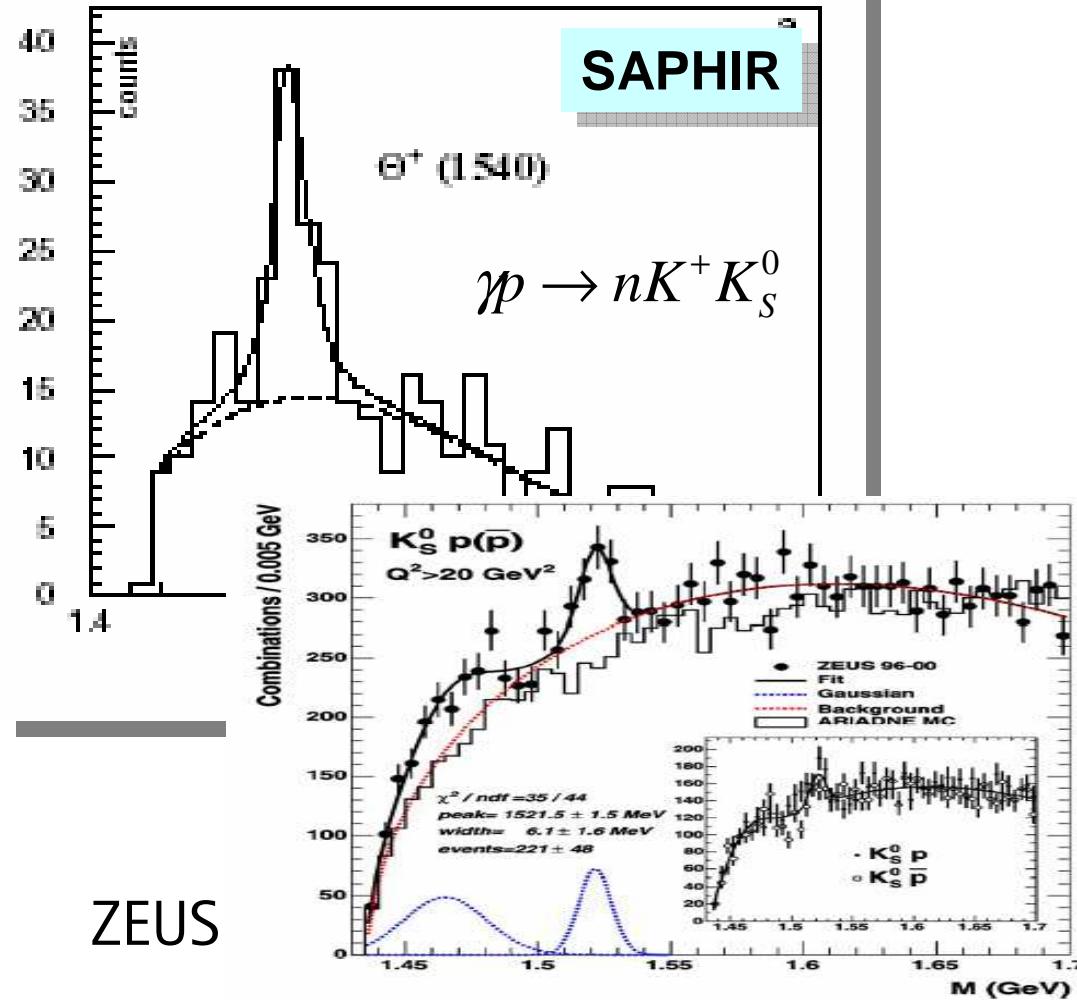
- Aufspaltung des  $a_2$ -Mesons
- top Quark bei 1/2 der wahren Masse
- Substruktur von W und Z
- Substruktur von Quarks
- Supersymmetrie
- ...
- Kalte Kernfusion

Vorsicht! Insbesondere, wenn man die Entdeckung in der Washington Post veröffentlicht. Auch eine Bestätigung eines zweiten Experiments ist keine 100%-Garantie!

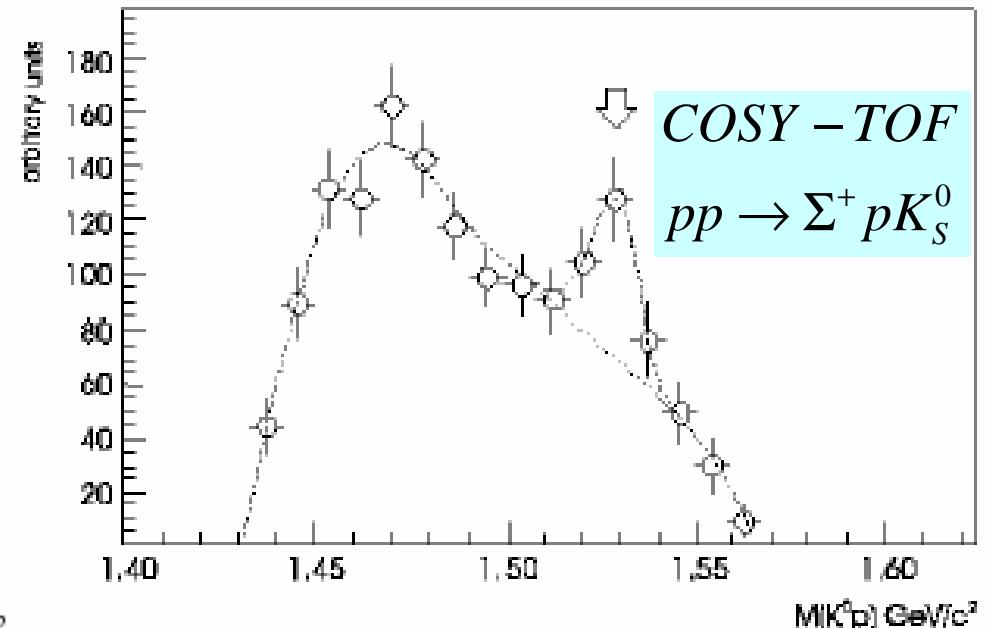
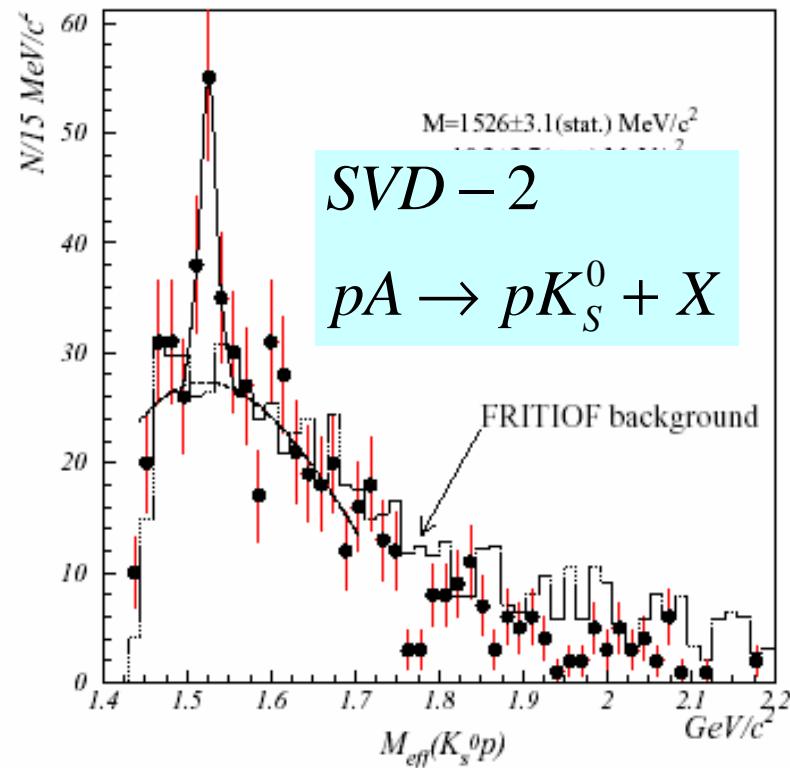
# Existieren Pentaquarks?



# Existieren Pentaquarks?

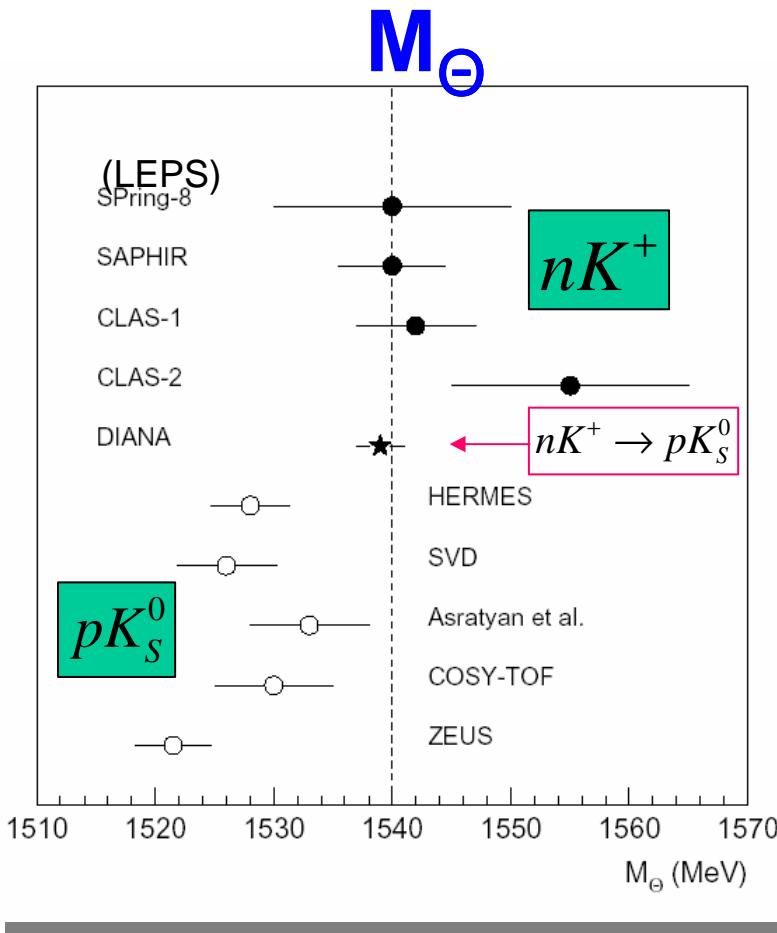


# Existieren Pentaquarks?



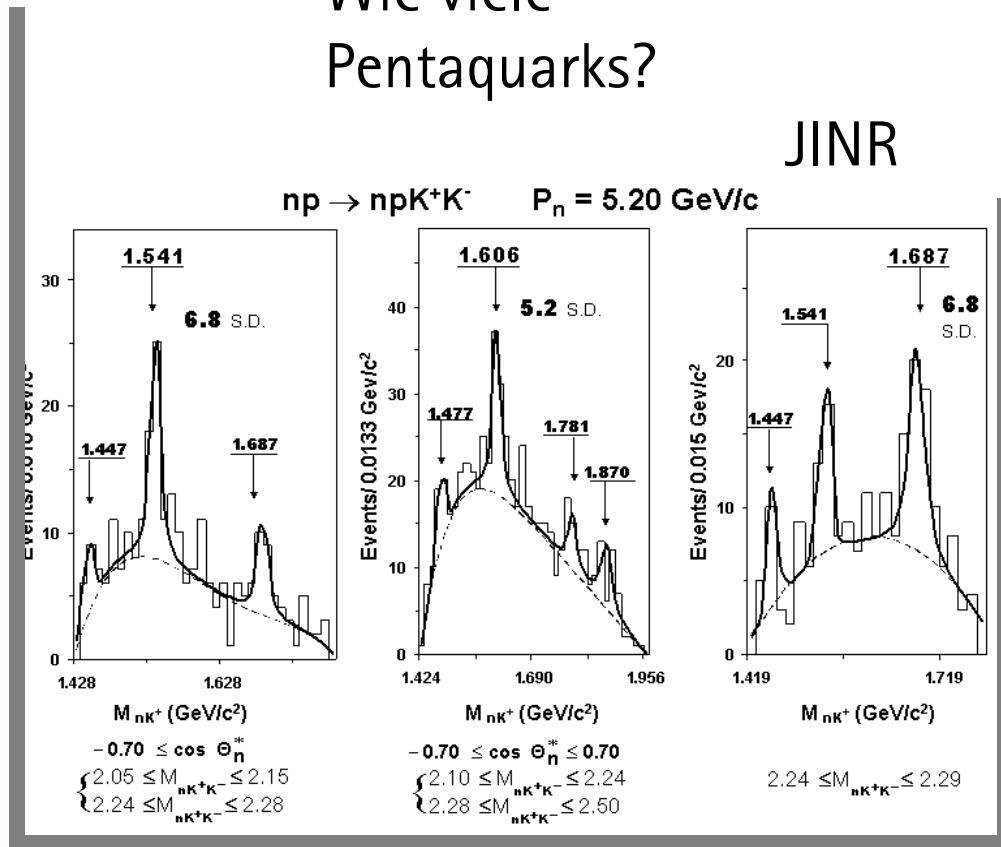
PDG 2004: Status \*\*\*

# Aber... Welche Masse?



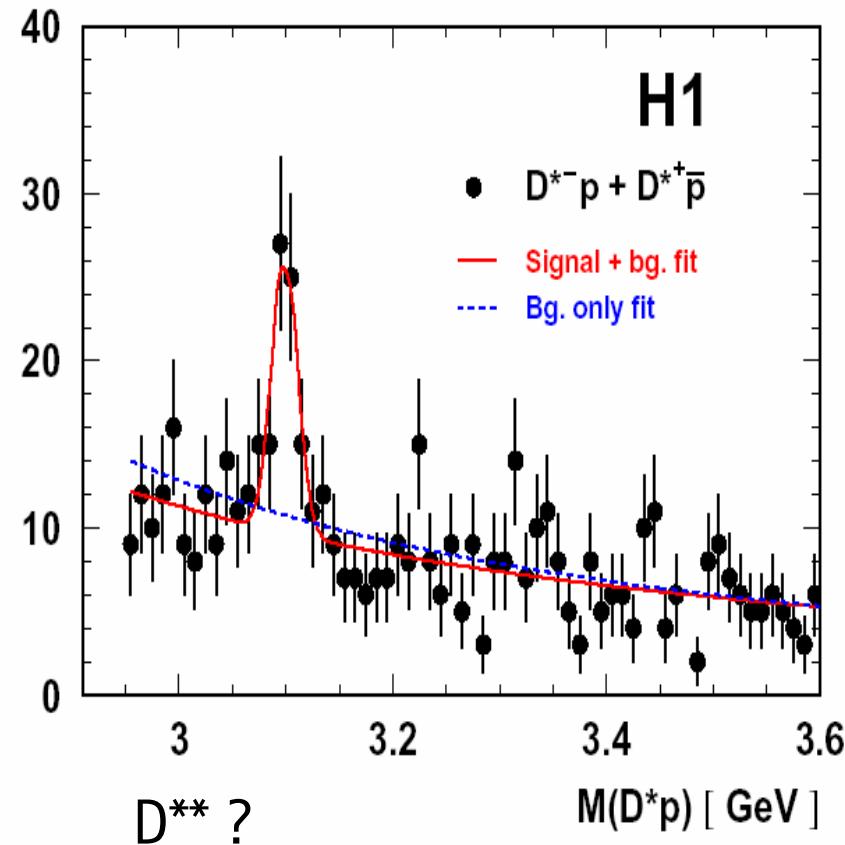
Wie viele  
Pentaquarks?

JINR



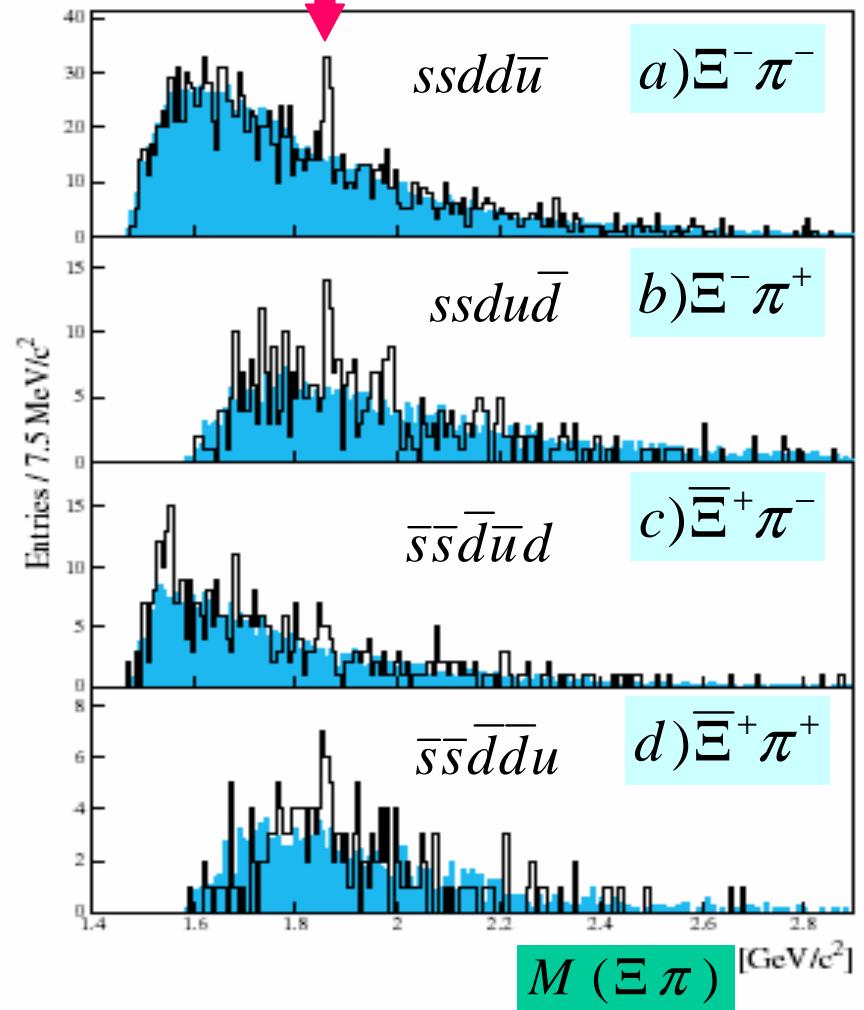
# Weitere Pentaquarks?

charm...



doppelt-strange

NA49



# Viele negative Suchen...

Trotz intensiver Suche nichts gefunden in großen Experimenten:

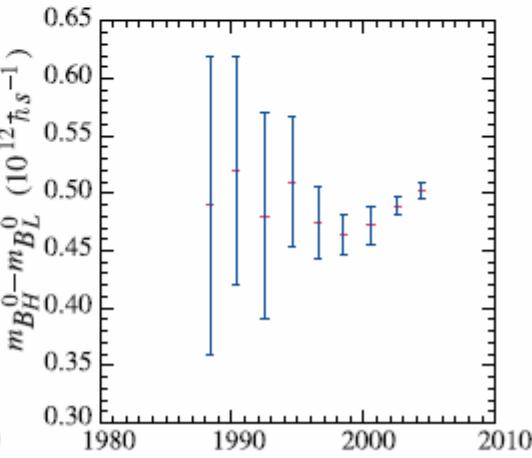
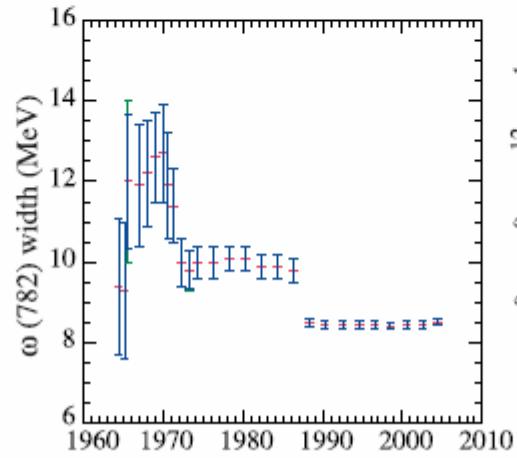
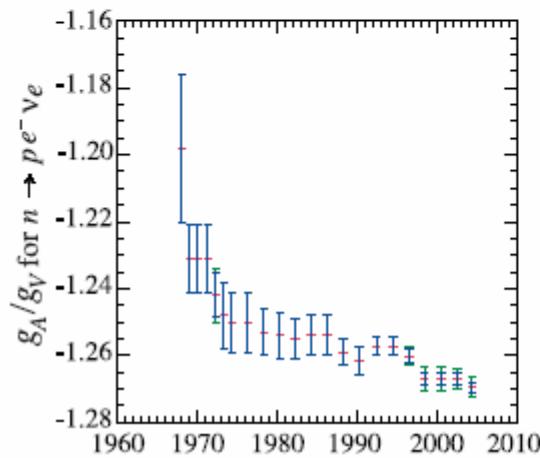
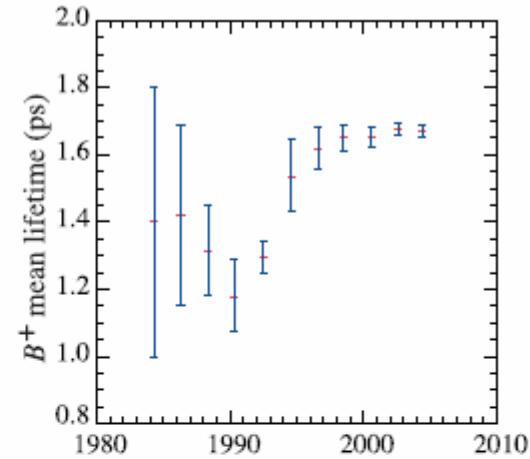
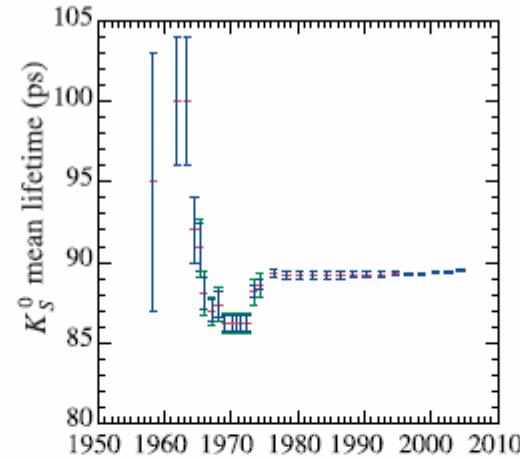
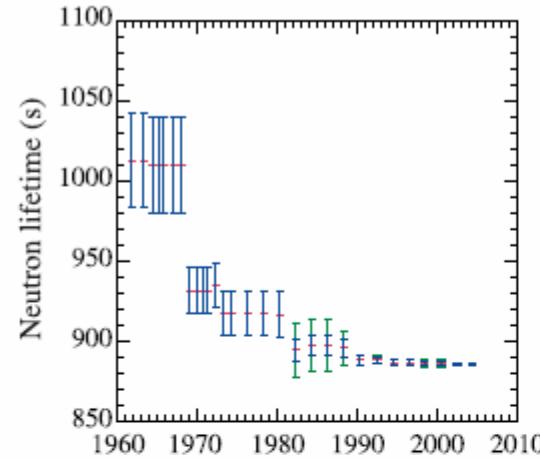
ALEPH, DELPHI, L3, OPAL, CDF, BELLE, BABAR, BES,  
H1 im ZEUS-Kanal, ZEUS im H1-Kanal  
(z.T. aber große Signale für etablierte Zustände)

Achtung Bayes-Theorem: Prior-Wahrscheinlichkeit hat sich geändert!

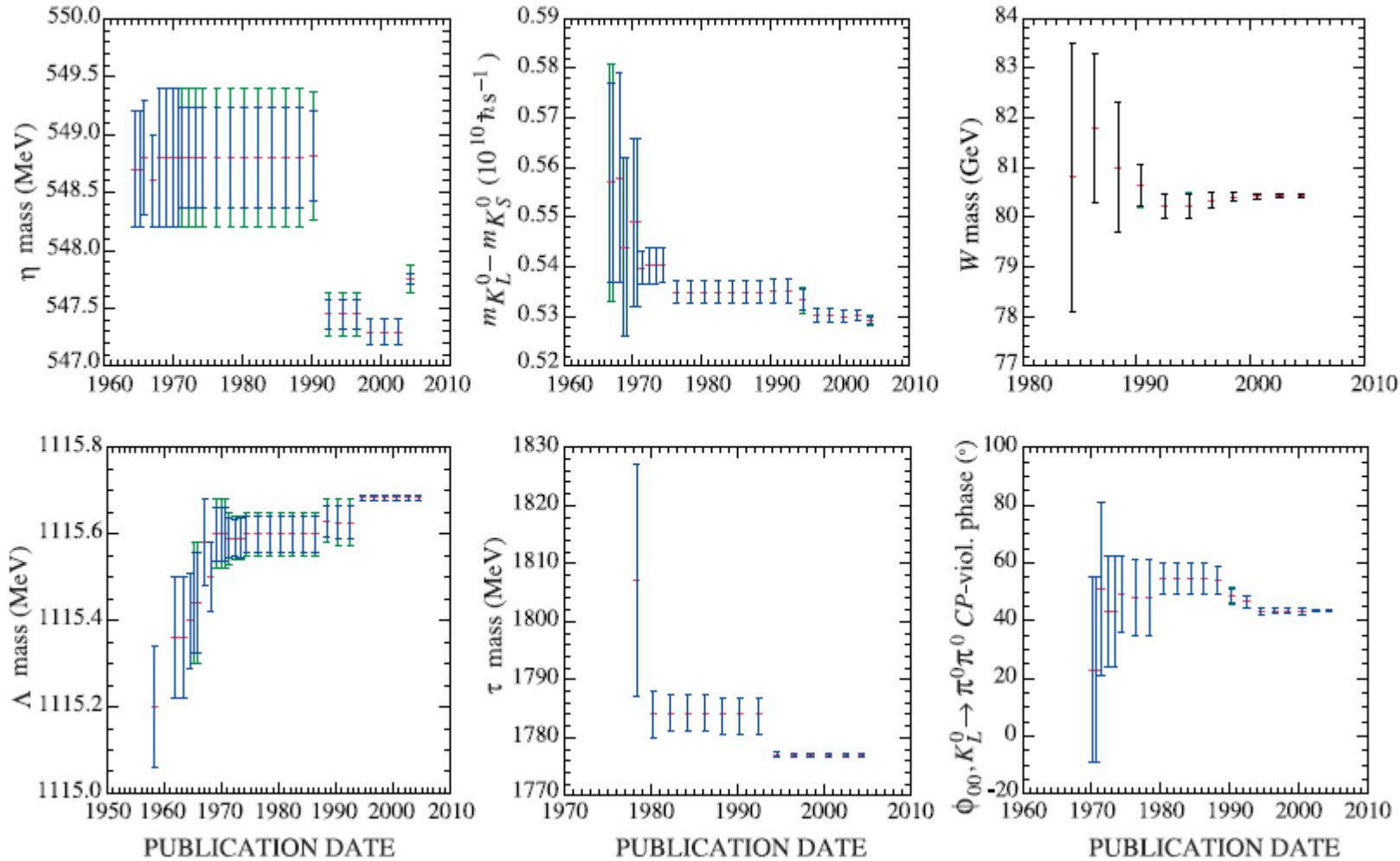
Früher hätte man sich in exotischen Kanälen sehr sicher sein müssen, dass es sich nicht um eine statistische Fluktuation handelt. Für Pentaquarks gibt es sehr viele neue Kanäle, Massen nicht bekannt: Kombinatorik!

Meine Meinung: sehr kritisch!  $P_{\text{subjektiv}}(\text{Pentaquark existiert}) < 50\%$   
 $P_{\text{subjektiv}}(\text{alle behaupteten Pentaquarks existieren}) < 1\%$

# Historische Entwicklungen I (PDG)



# Historische Entwicklungen II (PDG)

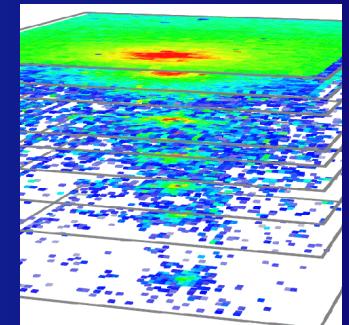
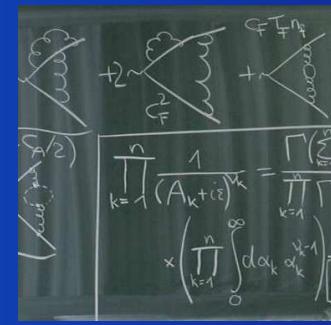




# Einführung in die Datenanalyse: Parameterschätzung

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

Herbstschule für Hochenergiephysik  
Kloster Maria Laach 2004



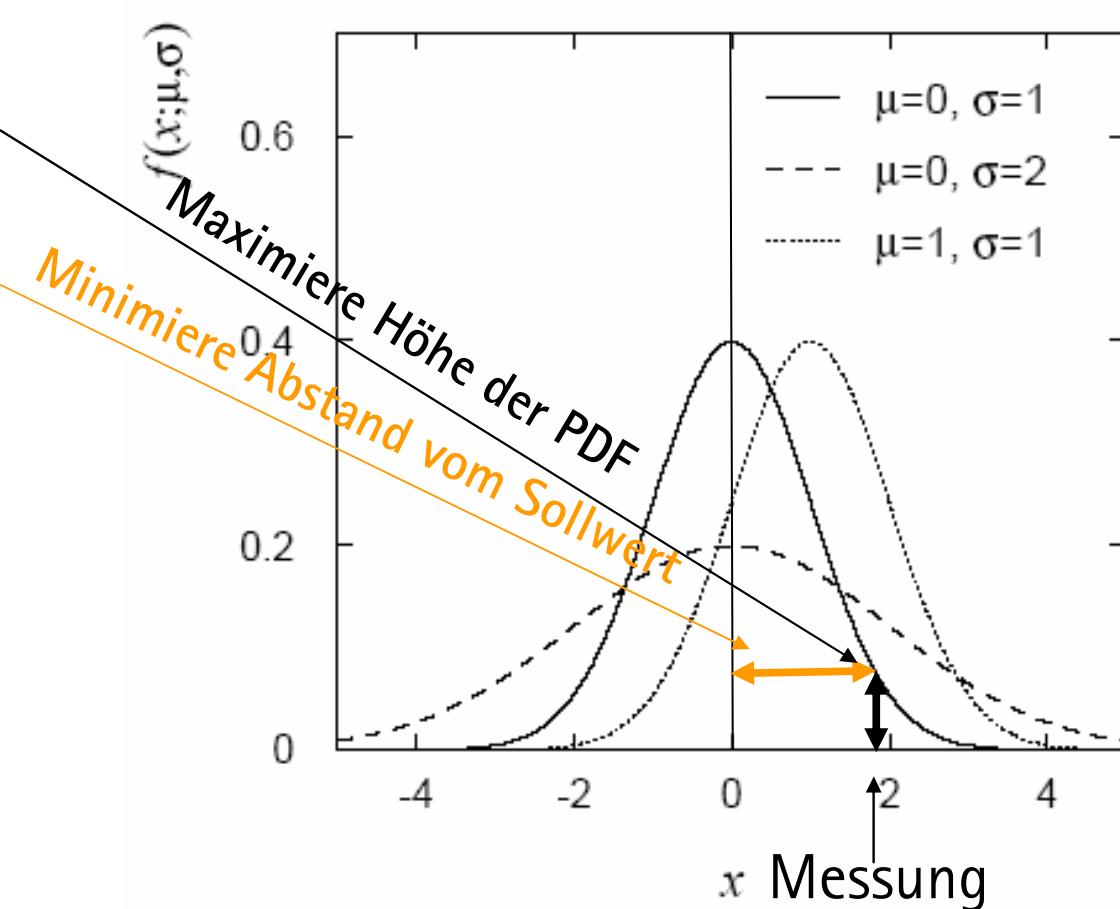
# Inhalt

## Parameterschätzung -- Fitting

Maximum Likelihood

Kleinste Quadrate

Kalman-Filter



# Parameterschätzung mit kleinsten Quadraten

Aufgabenstellung:

Mehrere Messungen  $y_i, i = 1, \dots, n$  wurden durchgeführt.

Ein lineares Modell mit Parametern  $\vec{a}$  macht Vorhersagen für die Messungen. Die Messungen sind unverzerrt, d.h.  $E[\vec{y}] = A\vec{a}_{wahr}$  und man kennt die Varianzen jeder einzelnen Messung  $\sigma_i^2$ .

Zu finden ist der beste Schätzwert für die unbekannten Parameter  $\vec{a}$  und seine Unsicherheit.

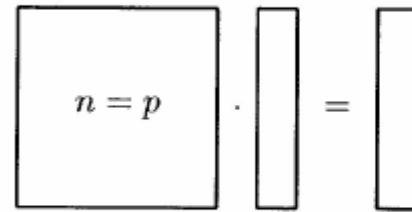
Prinzip der kleinsten Quadrate:

Minimiere die Summe  $S$  der quadratischen Abweichungen  $\Delta y_i$  zwischen Modell und Messungen.

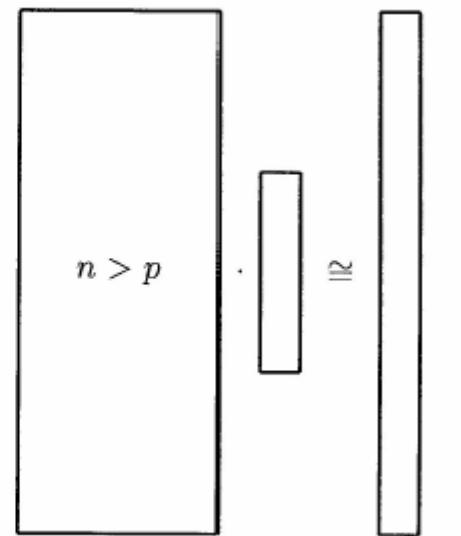
Lösung: Ableitungen  $\frac{dS}{da} = 0$

# Lineare kleinste Quadrate

$$\mathbf{A} \cdot \mathbf{a} = \mathbf{y}$$


$$\begin{matrix} n = p \\ \square \end{matrix} \cdot \begin{matrix} | \\ | \\ | \end{matrix} = \begin{matrix} | \\ | \\ | \end{matrix}$$

$$\mathbf{A} \cdot \mathbf{a} \cong \mathbf{y}$$


$$\begin{matrix} n > p \\ \square \end{matrix} \cdot \begin{matrix} | \\ | \\ | \end{matrix} \cong \begin{matrix} | \\ | \\ | \end{matrix}$$

Lineares Gleichungssystem  
eindeutig lösbar

Überbestimmtes Gleichungssystem  
Ausgleichsrechnung,  
 $n-p$  Freiheitsgrade

# Parameterschätzung mit kleinsten Quadraten

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  $y_i$ :

$$S = \sum_{i=1}^n (y - y_i)^2 = \text{ Minimum} \quad \hat{y} = \sum_{i=1}^n y_i / n$$

# Lösung des linearen Optimierungsproblems

Modell für Lineare Kleinste Quadrate:  $y = Aa$

$r = y - Aa$  Residuenvektor

$W = V[y]^{-1}$  Gewichtsmatrix = Inverse der Kovarianzmatrix

Prinzip der Kleinsten Quadrate: Minimiere den Ausdruck

$$S(a) = r^T W r = (y - Aa)^T W (y - Aa)$$

bezüglich  $a$ , d.h.  $dS/da = 0$ .

Lösung ist lineare Funktion der Messwerte  $y$ :

$$\hat{a} = (A^T W A)^{-1} A^T W y = B y$$

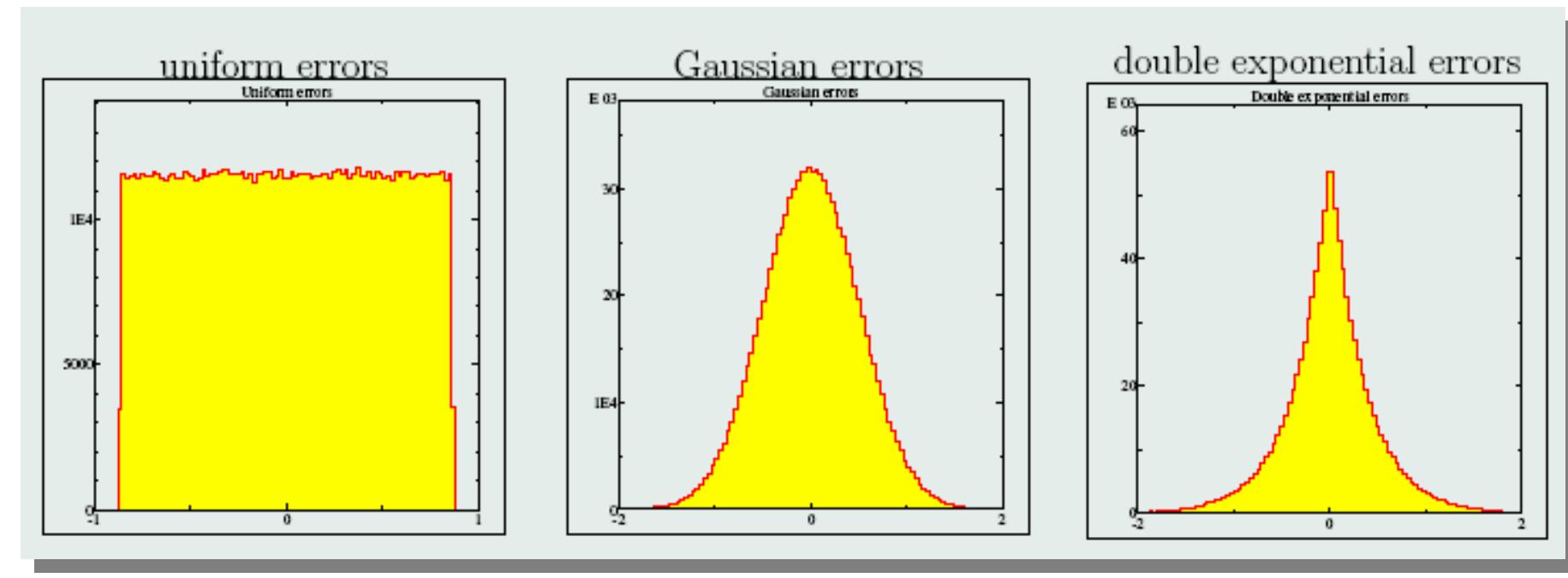
Kovarianzmatrix von  $a$  durch Fehlerfortpflanzung:

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

# Abhängigkeit von der Messfehler-Verteilung

Geradenfit an 20 Datenpunkte ( $\text{ndf}=20-2=18$ )

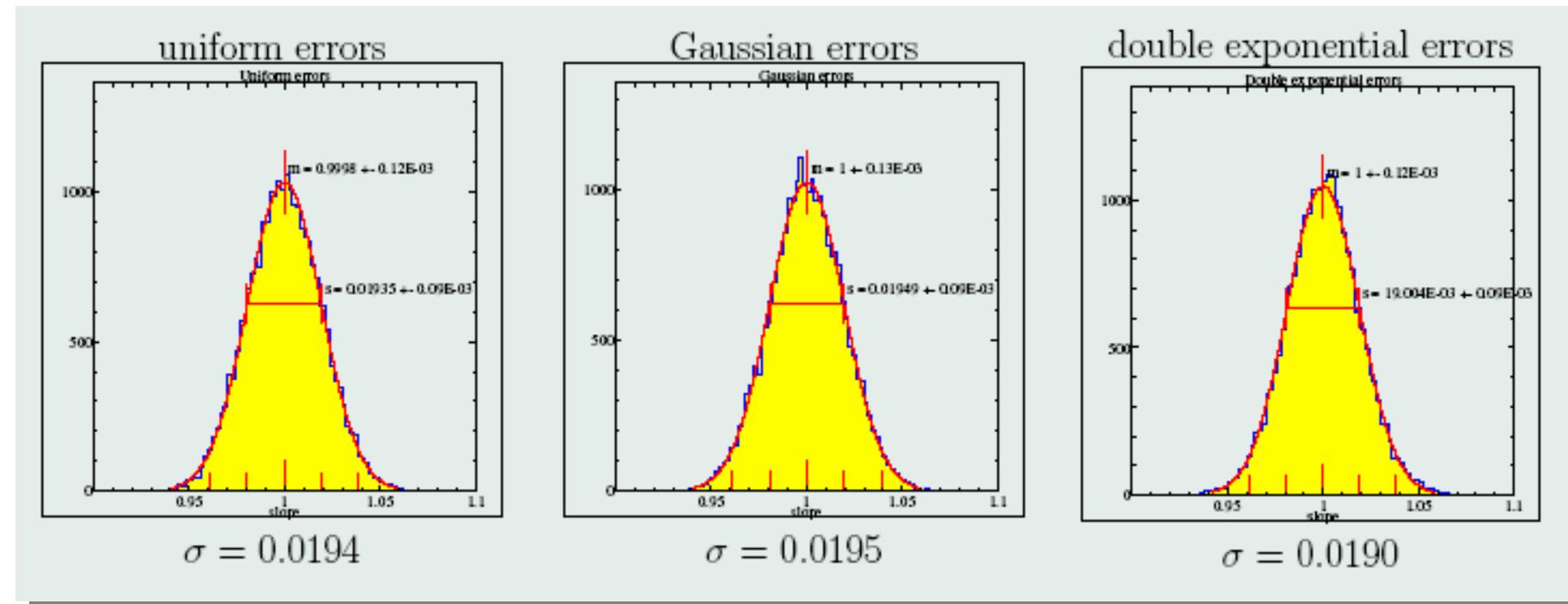
Drei verschiedene Verteilungsfunktionen der Einzelmessungen,  
alle Mittelwert 0, Standardabweichung=0.5



# Parameterschätzung mit kleinsten Quadraten

25000 Monte-Carlo-Tests:

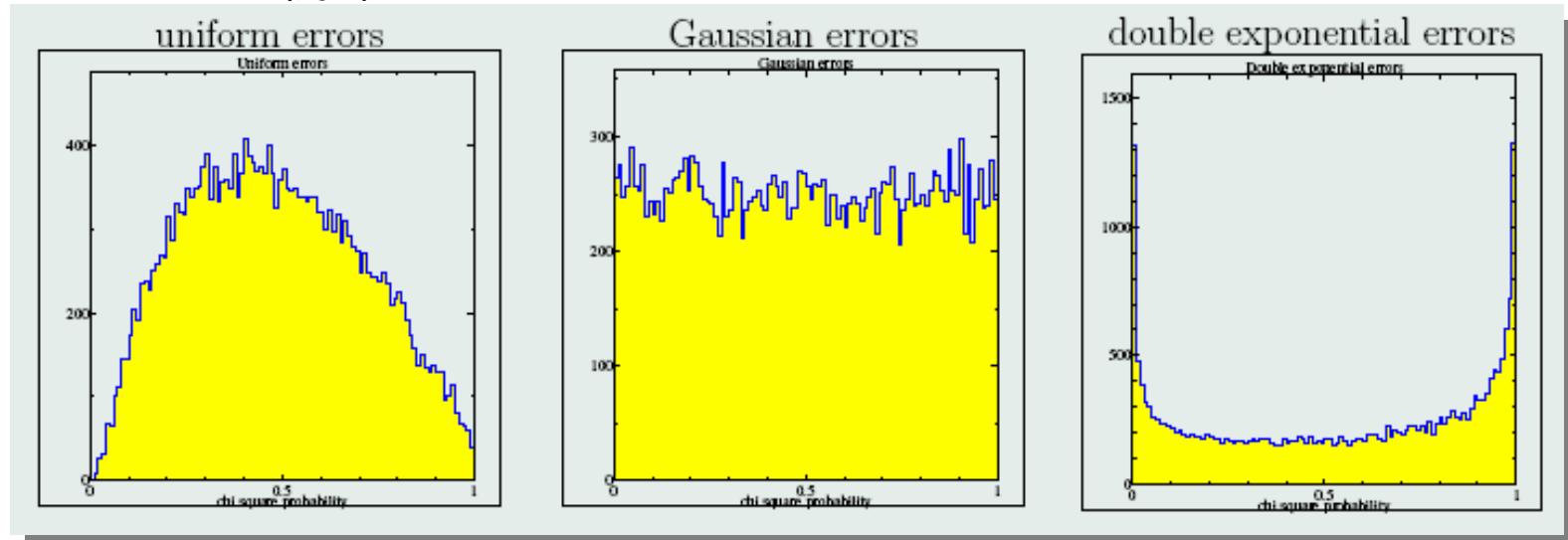
Alle Parameter-Verteilungen sind Gaussisch, die Breite kompatibel  
Zur Erwartung aus Fehlerfortpflanzung (für beide Parameter)



# Parameterschätzung mit kleinsten Quadraten

Mittlere  $\chi^2$ -Werte sind alle gleich NDF=20-2=18

Nur bei Gauß'schen Messfehlern folgt S einer  $\chi^2$ -Verteilung,  
d.h.  $\text{Prob}(\chi^2) = \text{const.}$ , 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.

# Robuste kleinste Quadrate

Ausreisser in den Daten (z.B. Eingabefehler, falsche Messpunkte auf Spur) können Fit wegen quadratischer Abhängigkeit sehr stark beeinflussen und zu völlig sinnlosen Lösungen führen.

Rezept zur Robustifizierung:

1. Normaler Kleinste Quadrate-Fit, liefert Residuen.
2. Modifiziere Daten durch Limitieren der Residuen auf  $c\sigma$ . Eine gute Wahl ist  $c=1.5$ .
3. Wiederhole Fit mit Pseudo-Messungen statt Originalmessungen.

Es existieren auch andere Loss-Funktionen (z.B. Huber-Funktion), aber Nicht mehr analytisch lösbar.

# 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:

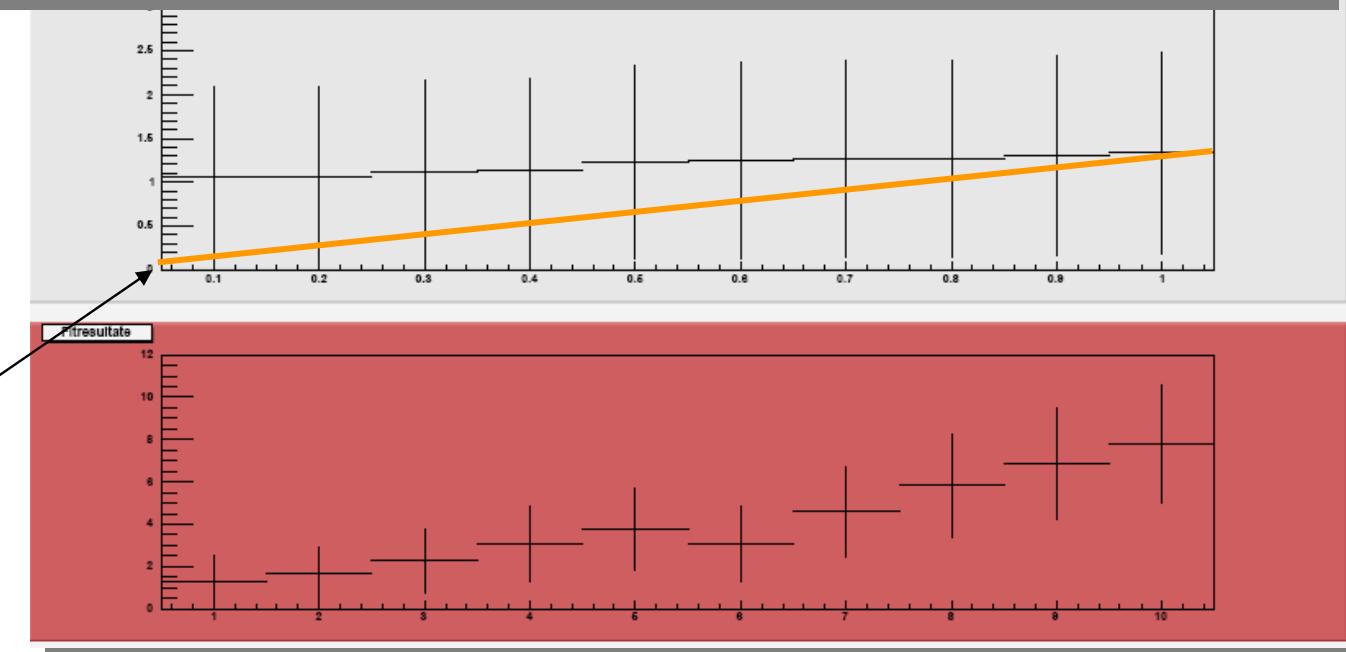
In einem  $\chi^2$ -Ausdruck nicht die gemessenen Werte  $y_i$ , sondern nur die theoretische Erwartung verändern!

# 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!

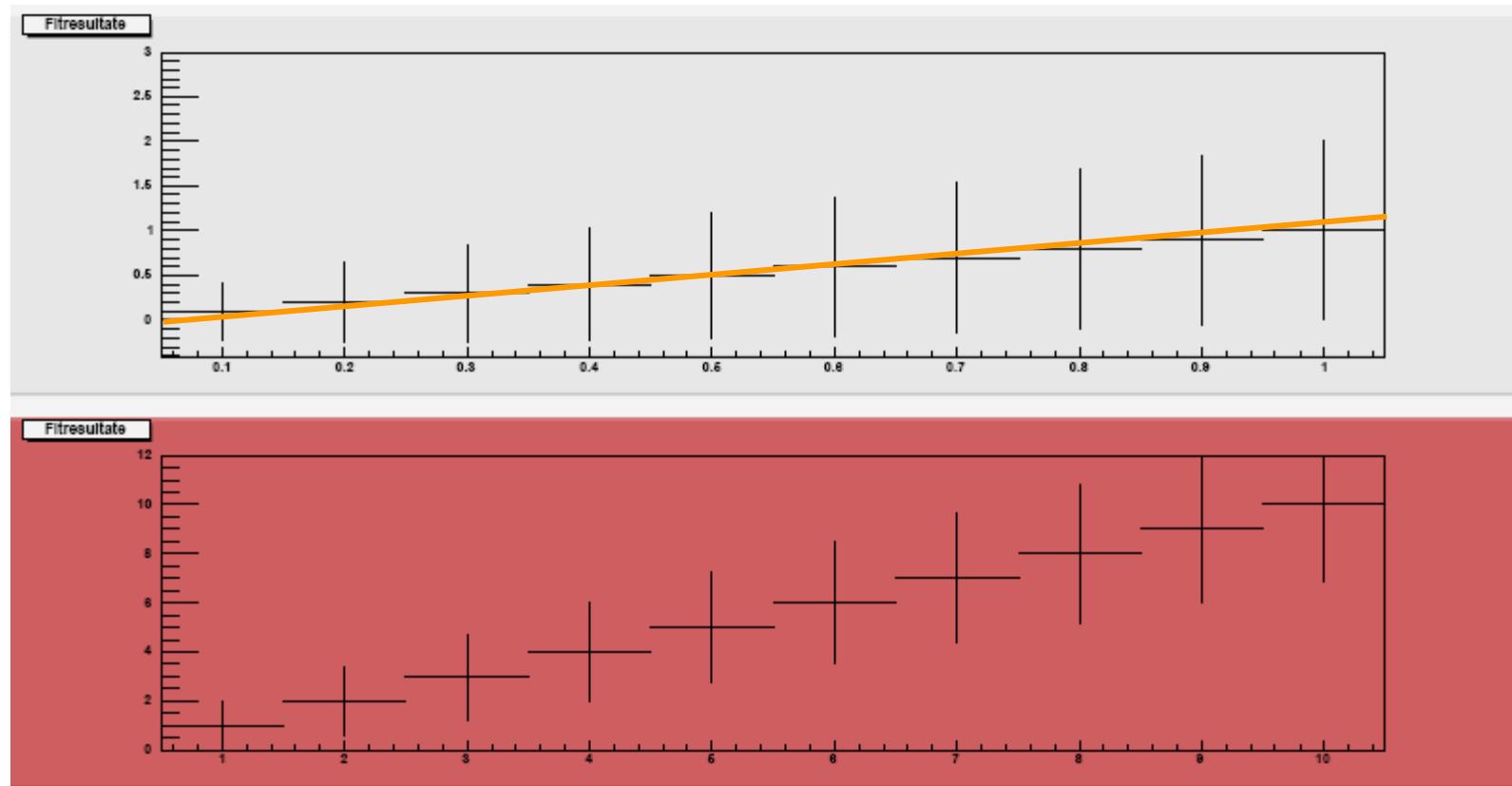
Demo  
Verknüpfung mit flat\_fit.lnk

Soll-Resultat



# Vorsicht bei kleinen Zahlen!

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



# Nichtlineare kleinste Quadrate

In der Praxis treten oft nichtlineare Funktionen  $f(x, a)$  auf. Diese sollten dann LINEARISIERT werden. Dazu sind vernünftige STARTWERTE notwendig, um die eine Taylor-Entwicklung durchgeführt wird.

Nach Lösung des linearisierten Problems wird iteriert und eine Taylor-Entwicklung um die neue Lösung durchgeführt.

Falls die Iteration nicht zu einer Verbesserung führt, wird die Veränderung des Lösungsvektors iterativ halbiert, bis es zu einer Verbesserung kommt.

Es sind geeignete Konvergenzkriterien festzulegen.

# Kleinste Quadrate mit Nebenbedingungen

Beispiel: Kinematisches Fitten im  $\pi^0$ -Zerfall.

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

A-priori-Wissen: Photonen stammen aus dem Zerfall eines  $\pi^0$ -Mesons.

Constraint: Invariante Masse  $m(\gamma\gamma) = m(\pi^0)$ .

Verändere gemessene Parameter so, dass die Nebenbedingung bei minimalem  $\chi^2$  erfüllt ist. Weitere Beispiele: Impuls- und Energieerhaltung!

# Kleinste Quadrate mit Nebenbedingungen (2)

Schreibe Nebenbedingungen in der Form  $f_i(x, \mathbf{a}) = 0$ .

Lösung durch Einführen von Lagrange-Multiplikatoren  $\lambda_i$  und Minimieren von

$$L(\mathbf{a}, \lambda, \Delta\mathbf{y}) = S(\mathbf{a}, \Delta\mathbf{y}) + 2 \cdot \sum_i \lambda_i f_i(x, \mathbf{a})$$

bezüglich  $\mathbf{a}$  und  $\lambda$ . Am Minimum ist  $dL/d\lambda = f = 0$ , also die Nebenbedingung erfüllt.

Beispielapplikation: APLCON von V. Blobel

# Maximum Likelihood Parameterschätzung

Aufgabenstellung:

Mehrere unabhängige Messungen  $x_i$ ,  $i = 1, \dots, n$  einer Größe  $x$  wurden durchgeführt.

Man weiß, dass die Daten durch eine Wahrscheinlichkeitsdichte (PDF)  $p(x; \vec{a})$  beschrieben wird, die Parameter  $\vec{a}$  enthält. Zu finden ist der beste Schätzwert für die unbekannten Parameter  $\vec{a}$  und seine Unsicherheit.

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

$$p(x; a) \geq 0 \quad \text{mit} \quad \int_{\Omega} p(x; a) dx = 1$$

# Maximum Likelihood- Prinzip

Likelihood-Funktion:

Werte der Wahrscheinlichkeitdichte für die n unabhängigen Messungen (Produkt), gerade die Werte  $x_i$  zu messen, ist die Likelihood-Funktion

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

die nur noch von den Parametern  $\vec{a}$  abhängt.

Maximum-Likelihood-Prinzip:

Der beste Schätzwert  $\hat{\vec{a}}$  ist der, der die Likelihood-Funktion  $\mathcal{L}(\vec{a})$  maximiert.

# Maximum Likelihood in der Praxis

Technische und theoretische Gründe:

Minimiere negativen Logarithmus der Likelihood-Funktion:

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

Likelihood-Gleichung definiert den Schätzwert  $\hat{a}$ :  $\frac{dF(\vec{a})}{da_j} = 0$ .

Kombination verschiedener Messungen (Experimente) einfacher:

$$\mathcal{L}(a) = \mathcal{L}_1(a) \cdot \mathcal{L}_2(a) \quad \text{multiplizierte Likelihood-Funktionen}$$

$$F(a) = F_1(a) + F_2(a) \quad \text{addiere log. Likelihood-Funktionen}$$

# Maximum Likelihood- Fehlerabschätzung

$F(a)$  näherungsweise quadratisch um das Minimum

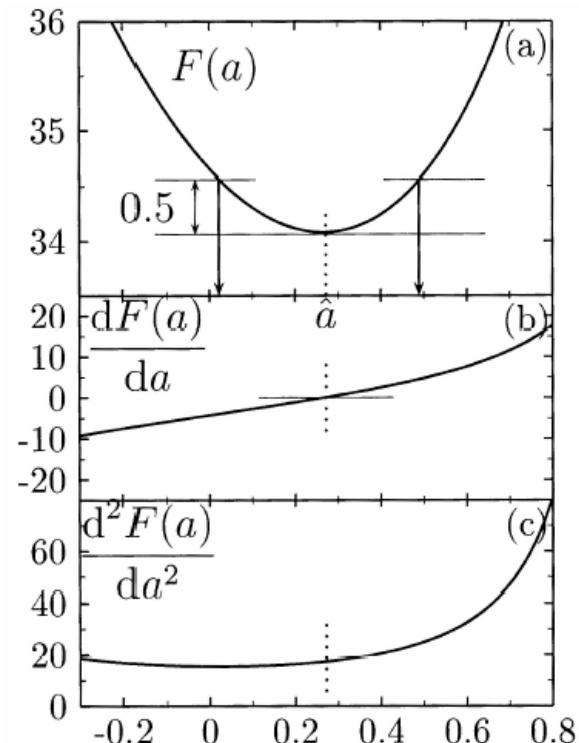
Erste Ableitung näherungsweise linear,  $=0$  am Minimum

Zweite Ableitung näherungsweise konstant:

Standardabweichung=1/Krümmung

$$F(a) - F(\hat{a})_{min} = \frac{1}{2} \quad \Leftrightarrow \quad |a - \hat{a}| = 1\sigma$$

$$F(a) - F(\hat{a})_{min} = \frac{n^2}{2} \quad \Leftrightarrow \quad |a - \hat{a}| = n\sigma$$



# Maximum Likelihood-- Anwendungen

Gebinnte Maximum Likelihood

Modell-PDF in analytischer Form

Oft: Modell-PDF nur durch Monte-Carlo bekannt

- ⤒ benötige >10-fache MC- Statistik
- ⤒ In 1D: Glätten der MC-Vorhersage
- ⤒ oder berücksichtige endliche MC-Statistik explizit mit der Methode von Barlow

# Binomialverteilung (1)

Wenn die Wahrscheinlichkeit für ein bestimmtes Ereignis („Erfolg“)  $p$  ist, ist die Wahrscheinlichkeit, in  $n$  Experimenten exakt  $k$  Erfolge zu beobachten, durch die Binomialverteilung gegeben:

$$P_n(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, \dots n$$

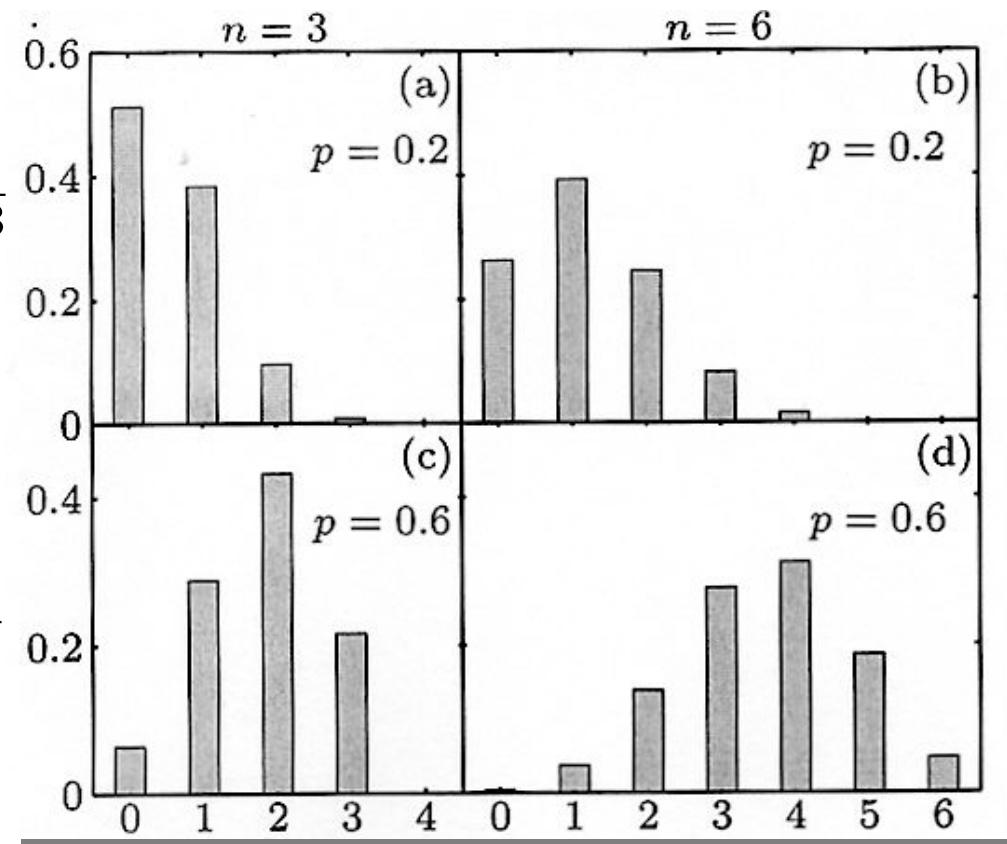
Der Faktor  $\binom{n}{k}$  ist die Anzahl der Kombinationen,  $k$  Elemente aus  $n$  unterschiedlichen Elementen zu selektieren.

$$\text{Erwartungswert} = E[k] = np$$

$$\text{Varianz} = V[k] = np(1-p)$$

# Binomialverteilung (2)

$$\begin{aligned}\mu &= 0.6 \\ \sigma &= \sqrt{3 \cdot 0.2 \cdot 0.8} \\ \sigma &\approx 0.7\end{aligned}$$



$$\begin{aligned}\mu &= 1.2 \\ \sigma &= \sqrt{6 \cdot 0.2 \cdot 0.8} \\ \sigma &\approx 0.98\end{aligned}$$

$$\begin{aligned}\mu &= 3.6 \\ \sigma &= \sqrt{6 \cdot 0.6 \cdot 0.4} \\ \sigma &= 1.2\end{aligned}$$

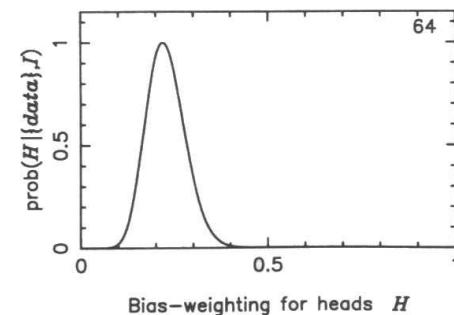
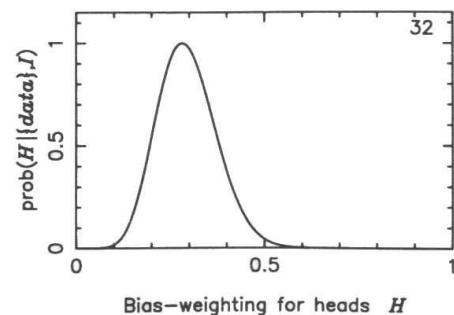
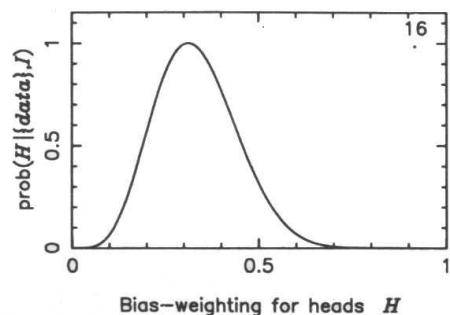
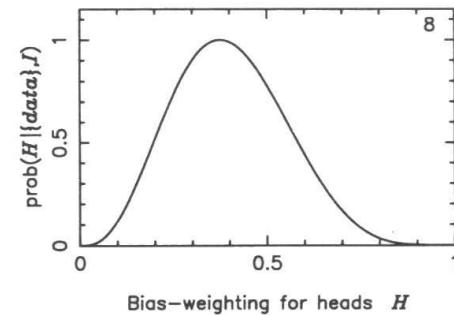
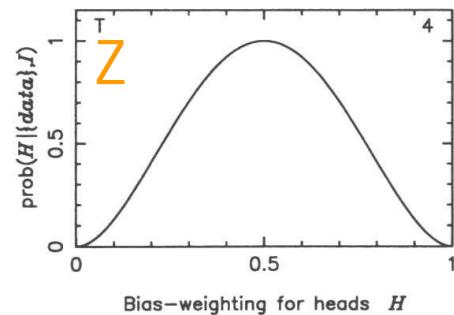
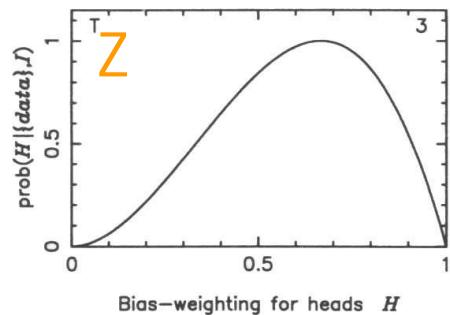
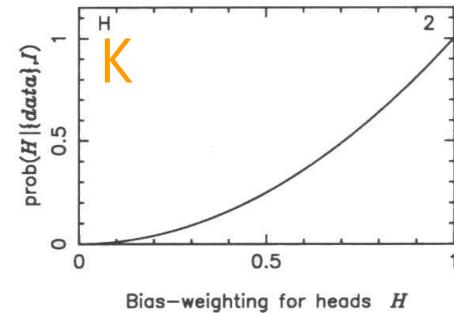
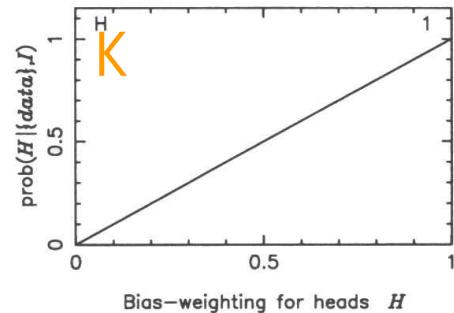
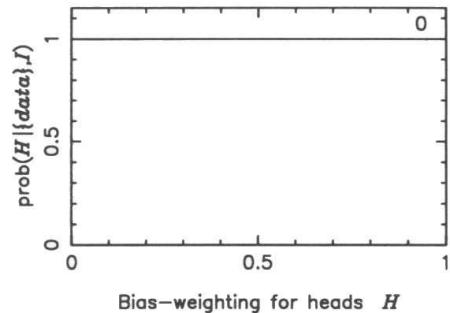
# Maximum Likelihood Fit für $p$ einer Binomialverteilung

Eine Münze wird  $N$  mal geworfen. Was kann man über die Wahrscheinlichkeit  $p$ , dass „Kopf“ geworfen wird, sagen?

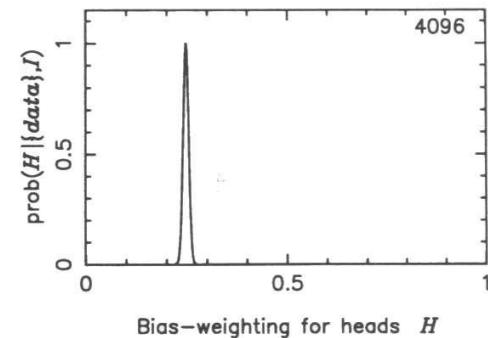
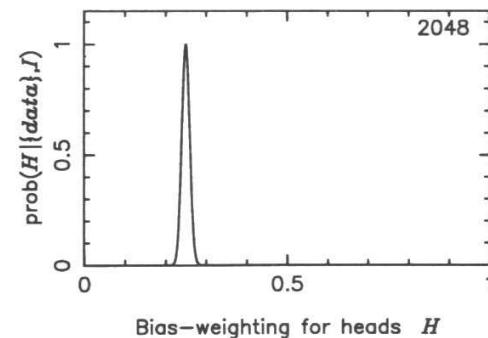
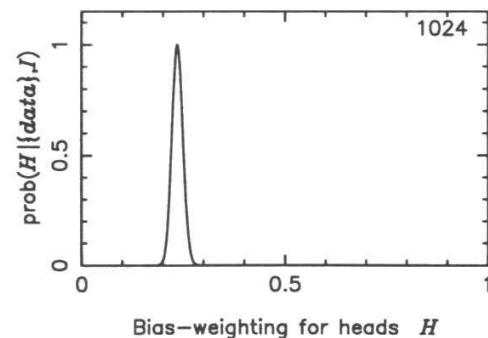
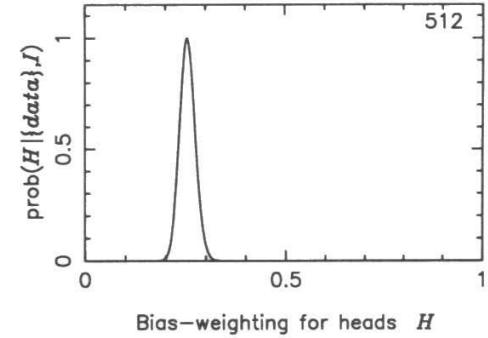
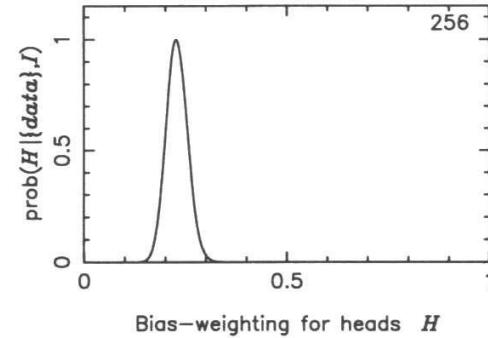
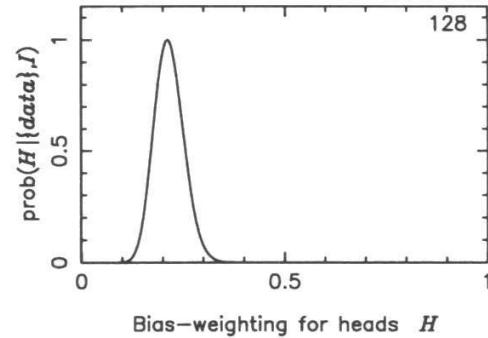
0 Versuche: Wir wissen nichts.  $p$  flach in  $[0,1]$ .

1. Versuch: Kopf. Die Wahrscheinlichkeit für  $p(\text{Kopf})=0$

$N$  Versuche:  $p(\text{Kopf})=$  wird immer genauer, Gaussförmig verteilt.



# Maximum Likelihood-Binomialfit Wahrscheinlichkeitsverteilungen

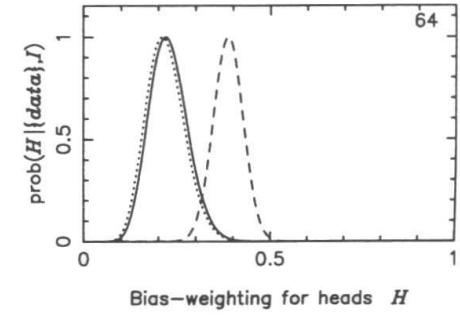
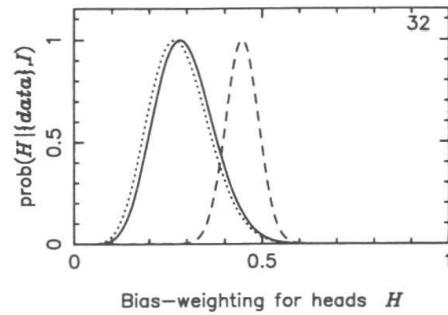
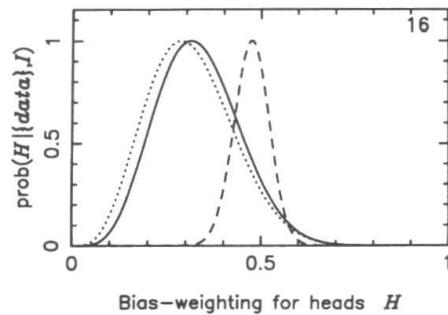
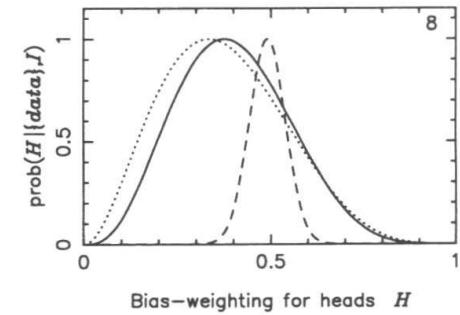
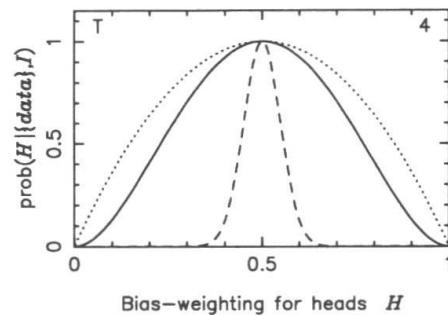
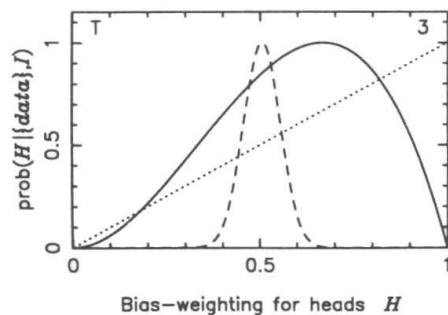
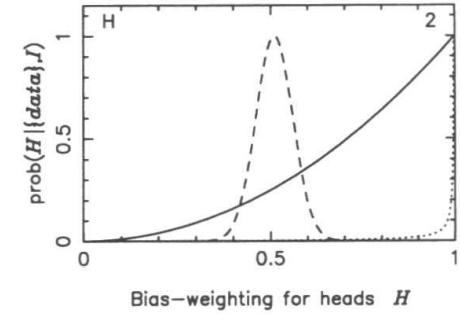
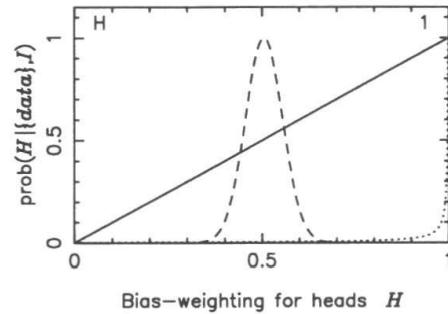
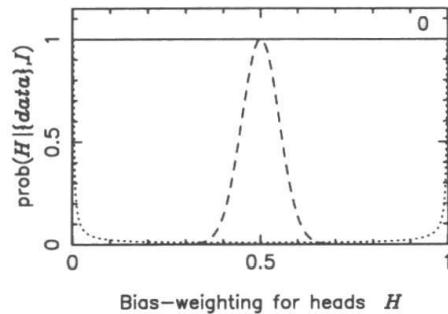


Fehler wird mit  $1/\sqrt{N}$  kleiner

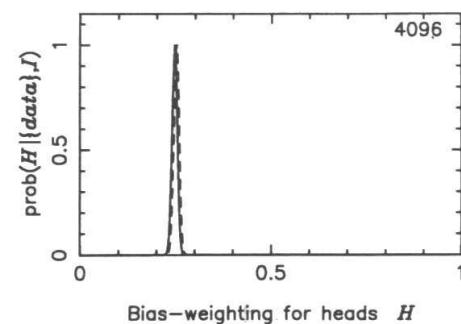
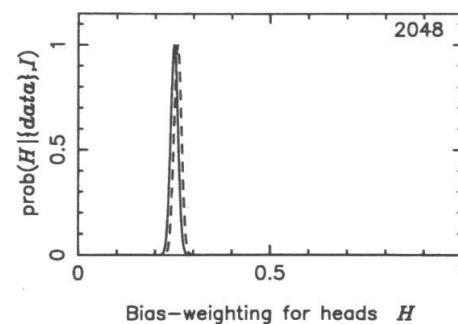
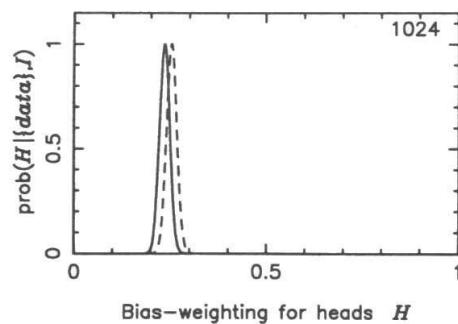
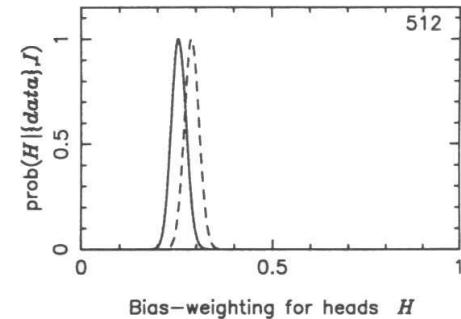
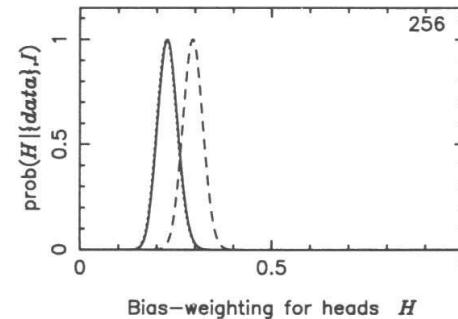
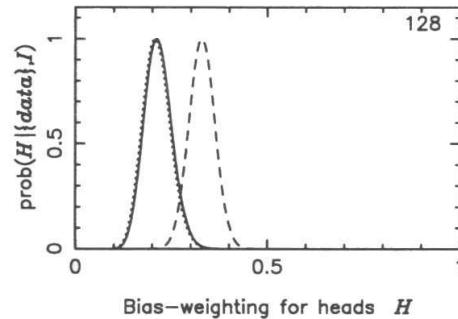
# Maximum Likelihood-Binomialfit mit verschiedenen a priori-Annahmen

Verschiedene a priori-Annahmen:

1. Wir wissen nichts.  $p$  flach in  $[0,1]$ .
2. Die Münze ist wahrscheinlich okay. Gauss um 0.5.
3. Die Münze ist wahrscheinlich gefälscht,  
ich weiss aber nicht, in welche Richtung.



# Maximum Likelihood-Binomialfit mit verschiedenen a priori-Annahmen



Fehler wird mit  $1/\sqrt{N}$  kleiner  
Einfluss des Priors wird vernachlässigbar

# Maximum Likelihood vs. Kleinstes Quadrate

Wenn Messwerte um Erwartungswert Gauß-verteilt sind, sind Maximum Likelihood und kleinste Quadrate identisch:

$$f(x_i; a) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp \left[ -\frac{(y_i - f(x_i; a))^2}{2\sigma_i^2} \right]$$

Neg. log-Likelihood-Funktion:

$$F(a) = \frac{1}{2} \sum_{i=1}^n \frac{(y_i - f(x_i; a))^2}{\sigma^2} + \text{const.}$$

(bis auf einen Faktor 1/2).

Wenn das Modell stimmt, folgt  $2F(a)$  am Minimum einer  $\chi^2$ -Verteilung mit  $n-p$  Freiheitsgraden ( $n$ =Anzahl Messungen,  $p$ =Anzahl der unbekannten Parameter)

# Maximum Likelihood vs. Kleinstes Quadrate

Vergleich:

	Maximum Likelihood	Kleinstes Quadrate
Voraussetzung	PDF exakt bekannt	Mittelwert und Varianz bekannt
Basis	Höhe der PDF	Abweichung vom Mittelwert der PDF
Effizienz	maximal	maximal unter linearen Schätzern
Komplexität	aufwändig, meist nichtlinear	linear, oft exakt lösbar
Robustheit	nein (PDF muss exakt sein)	nein (Ausreißer)
Spezialfall	identisch bei Gaußschen Fehlern	

## Kalman-Filter



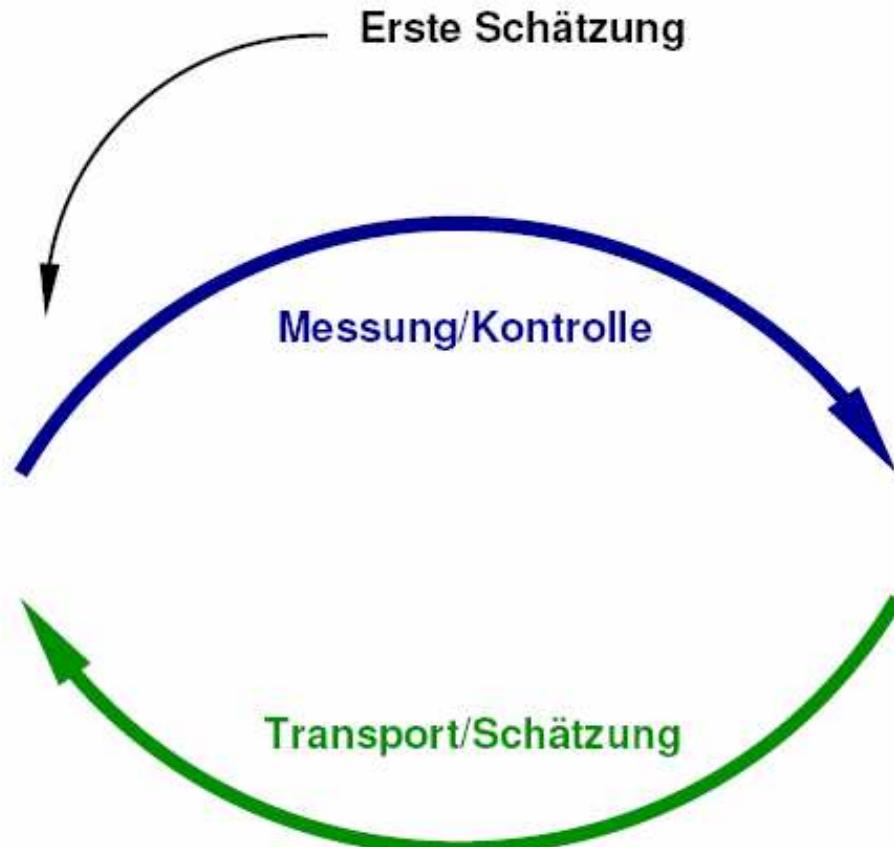
Rudolf Emil Kalman

(ein spezielles lineares  
Kleinste-Quadrat-Verfahren)

Kalman, R.E. 1960: "A New Approach to Linear Filtering and Prediction Problems", Transaction of the ASME – Journal of Basic Engineering, pp. 35-45 (March 1960)

## Was ist ein Kalman-Filter?

- optimal
- rekursiv
- nutz alle Informationen

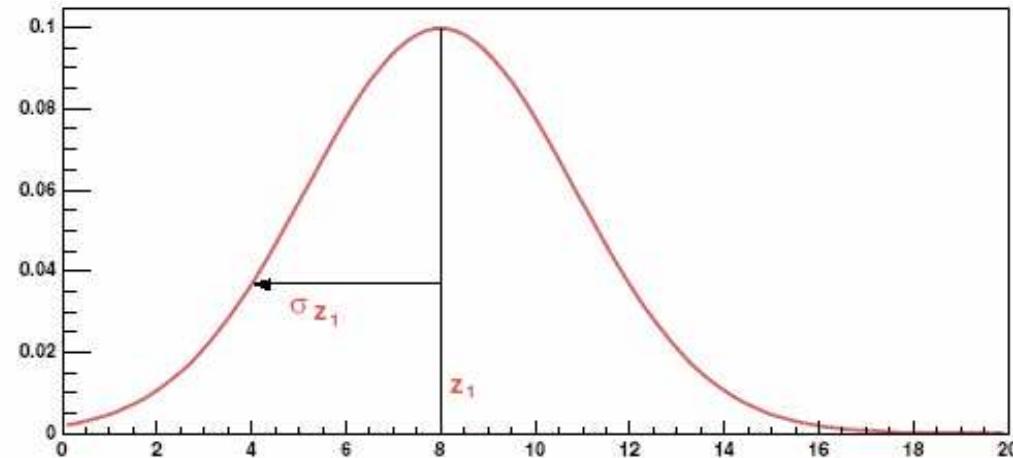


## Voraussetzungen

1. Weißes Rauschen (insbesondere keine Zeitreihenkorrelationen)
2. Gaussverteilte Fehler ( $\langle x \rangle = x_{0.5} = x_{max}$ )
3. Lineares Transportmodell (Unverzerrtheit)

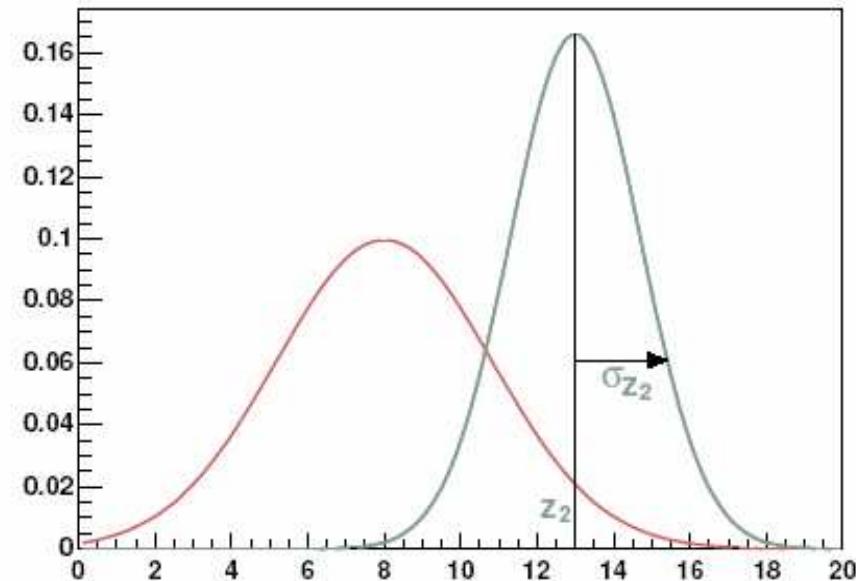
## Einfache Problemstellung

- Problem: Navigation (z.B. Feststellung des Breitengrads)
- Einfach, da nur eine Messgröße
- Es gibt einen Messfehler (Sextanten sind schwer zu bedienen!)
- Beste Schätzung:  $\hat{x}(t_1) = z_1$
- Varianz der Schätzung:  $\sigma_x^2(t_1) = \sigma_{z_1}^2$



## Die Zweite Messung

- Ein besserer Navigator mißt bei  $t_2 \approx t_1$
- Neue Schätzung:  $z_2$
- Neue Varianz:  $\sigma_{z_2}^2$
- Mit  $\sigma_{z_2} < \sigma_{z_1}$

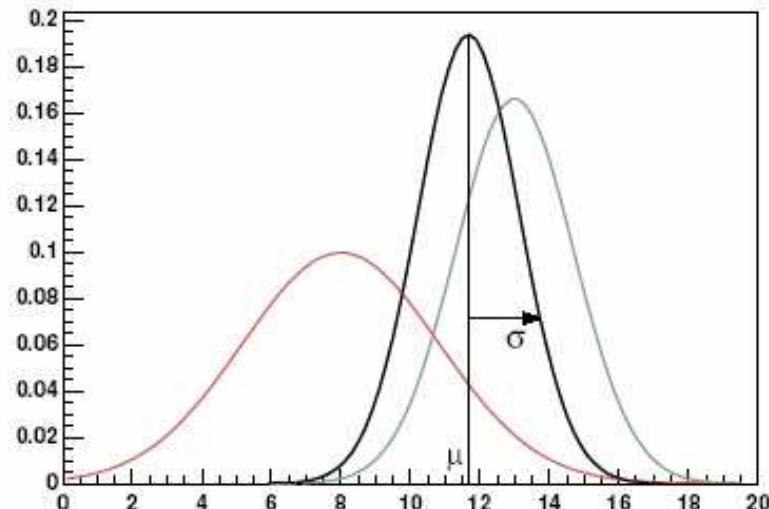


## Informationsgewinn

Beide Messungen werden kombiniert:

$$\mu = \frac{\sigma_{z_2}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2} z_1 + \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2} z_2$$

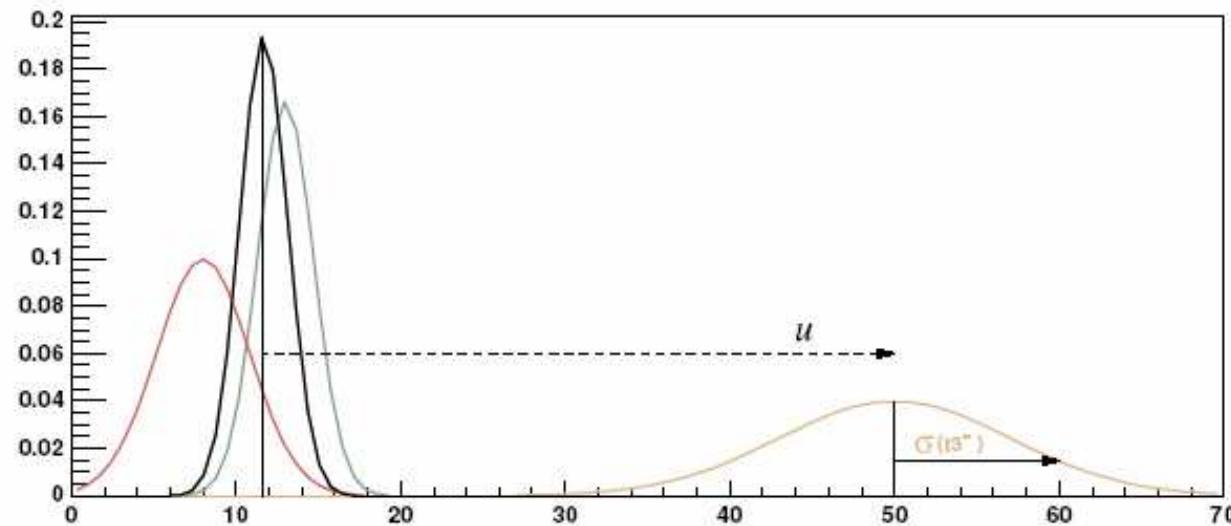
$$\frac{1}{\sigma^2} = \frac{1}{\sigma_{z_1}^2} + \frac{1}{\sigma_{z_2}^2}$$



- Gewichtetes Mittel: Beitrag gemäß Vertrauenswürdigkeit.
- Selbst schlechte Messungen enthalten Information!
- $\sigma \rightarrow \infty \Rightarrow$  nicht gemessen.

## Transport

- Zeitabhängigkeit:  $x \equiv x(t) \rightarrow$  Transportmodell
- Z.B. konstante Geschwindigkeit:  $dx/dt = u + w$ 
  - $u$  ist die (wahre) Geschwindigkeit
  - $w$  ist ein Fehlerterm (normalverteilt);  $N(0, \sigma_w)$
- Keine weitere Messung  $\Rightarrow \sigma$  wird größer  $\Rightarrow$  Unsicherheit wächst:



## “a priori” Größen

Aus dem Transport Modell erhalten wir für den Zeitpunkt  $t_3^- \approx t_3$ :

- *a priori* Schätzung:

$$\hat{x}(t_3^-) = \hat{x}(t_2) + u \cdot (t_3 - t_2)$$

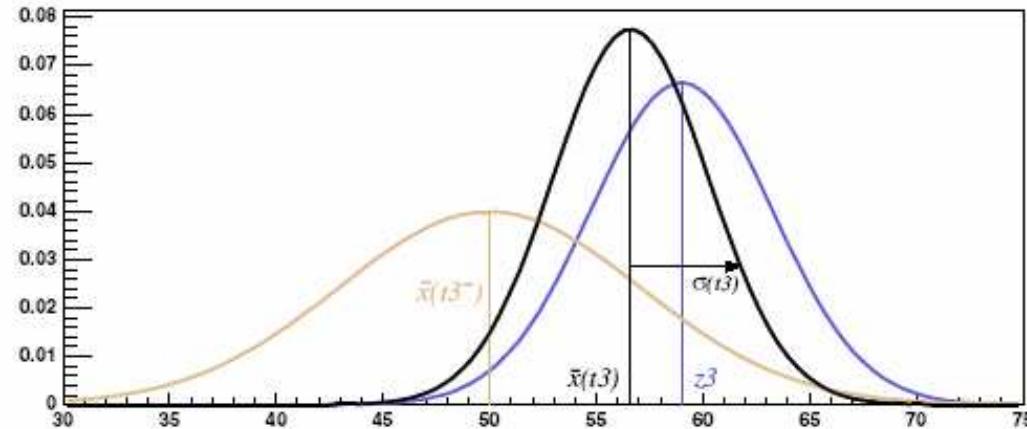
- *a priori* Varianz:

$$\sigma_x^2(t_3^-) = \sigma_x^2(t_2) + \sigma_w^2 \cdot (t_3 - t_2)^2$$

“*A priori*”, da vor der nächsten Messung zum Zeitpunkt  $t_3$ .

## Messung und “a posteriori” Größen

Nun neue Messung zum Zeitpunkt  $t_3$  mit Messwert  $z_3$  und Varianz  $\sigma_{z_3}$ :



- *a posteriori* Schätzung:

$$\hat{x}(t_3) = \hat{x}(t_3^-) + K(t_3) \cdot (z_3 - \hat{x}(t_3^-))$$

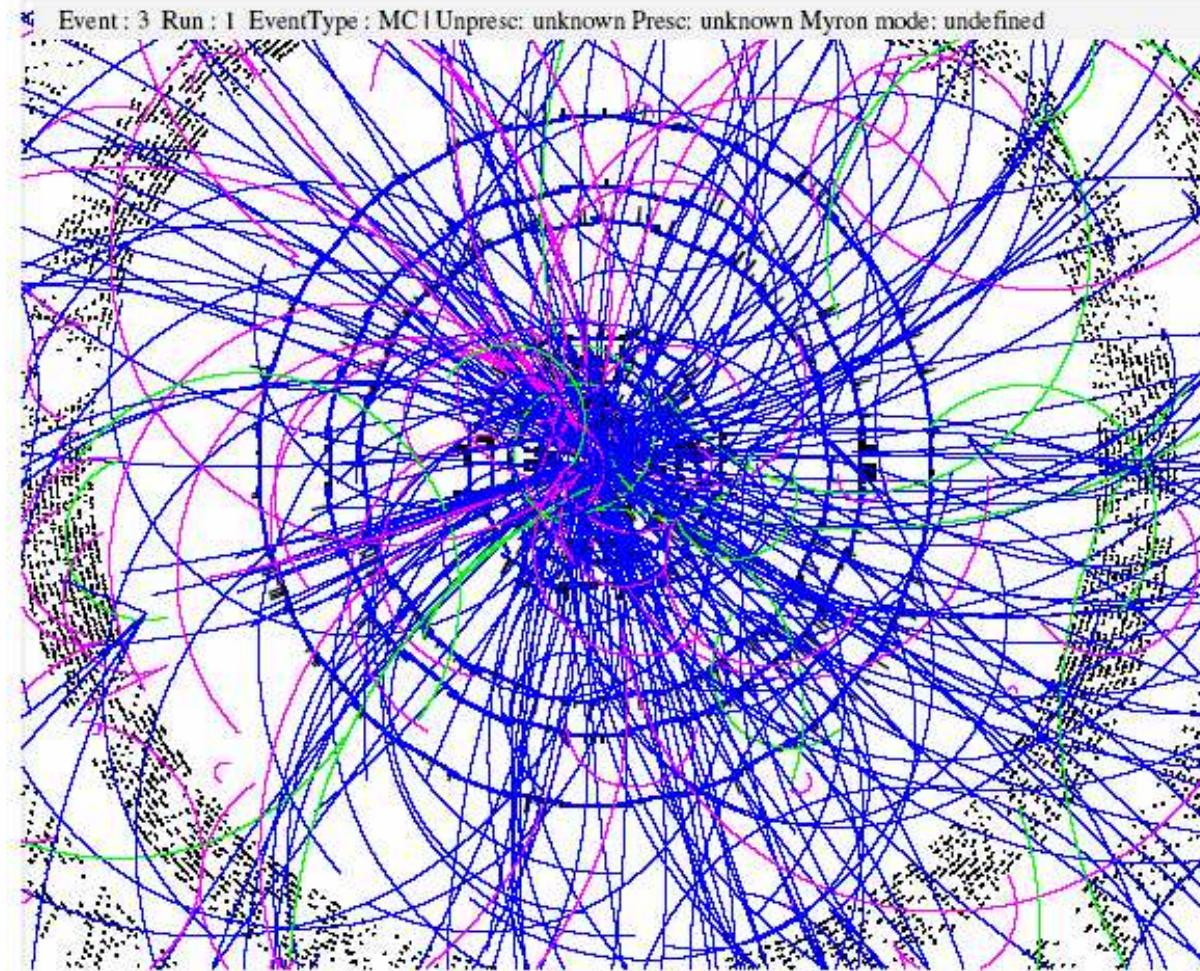
- *a posteriori* Varianz:

$$\sigma_x^2(t_3) = \sigma_x^2(t_3^-) - K(t_3) \cdot \sigma_x^2(t_3^-)$$

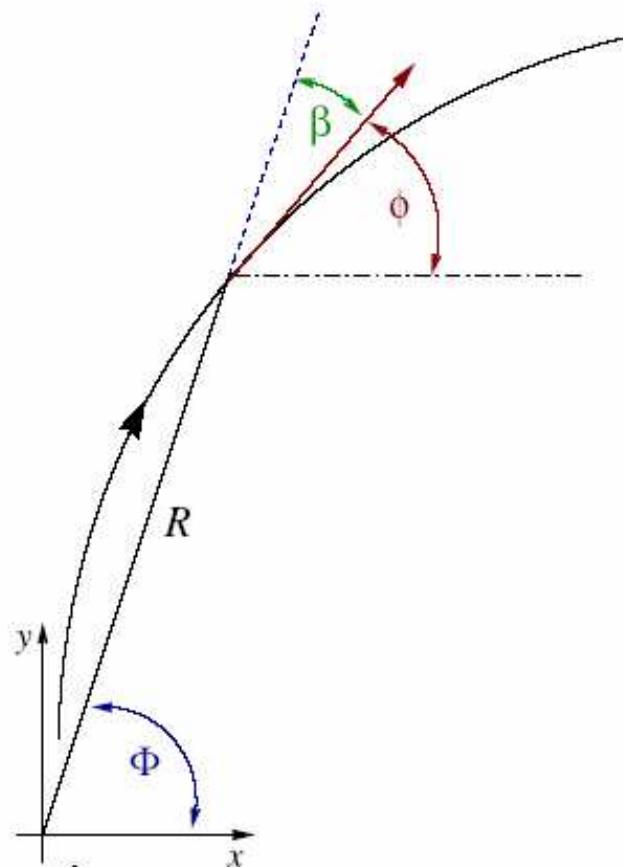
- Mit dem *Kalman-Gain*:

$$K(t_3) = \frac{\sigma_x^2(t_3^-)}{\sigma_x^2(t_3^-) + \sigma_{z_3}^2}$$

## Reales Problem: Rekonstruktion von Teilchenspuren



## Spurparameter



- Zu Bestimmen:
  - $\phi$  (Richtung)
  - $\beta = \Phi - \phi$
  - $\kappa = 1/R$  (Krümmung)
- Messung(en):
  - $\Phi, R$

## Linearisierung

- Lineares Problem, Bestimmung von  $\mathbf{a}$ :

$$y(x) = f(x, \mathbf{a}) \equiv a_1 f_1(x) + a_2 f_2(x) + \dots + a_p f_p(x)$$

- Nichtlineares Problem, Bestimmung von  $\mathbf{a}$ :

$$y(x) = f(x, \mathbf{a}); \text{ beliebig}$$

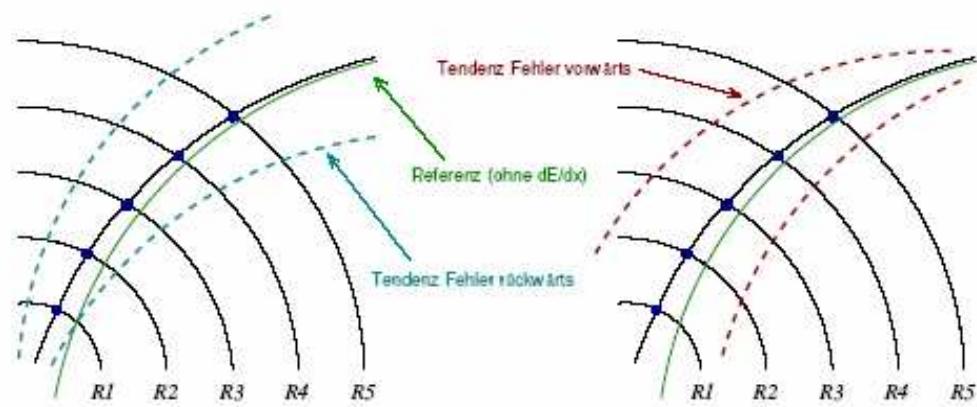
- Linearisiertes Problem, Bestimmung von  $\Delta \mathbf{a} = \mathbf{a} - \mathbf{a}^*$ :

$$y(x) = f(x, \mathbf{a}) \approx f(x, \mathbf{a}^*) + \sum_{j=1}^p \frac{\partial f}{\partial a_j} \Bigg|_{a_j=a_j^*} \cdot (a_j - a_j^*)$$

## Vorteile des Kalman-Filters in der Spurrekonstruktion

- Behandlung von Energieverlust und Vielfachstreuung
- Schnelles Testen verschiedener Hypothesen
- Möglichkeit zur Überprüfung der Detektoranordnung und Fehlerabschätzung mittels “Unbiased Residuals”

## Überprüfung von Detektoranordnung und Fehlerabschätzung



"Unbiased Residual" =  $(\text{Messung} - \text{Gewichtetes Mittel Vorwärts/Rückwärts})/\text{Fehler}$   
Erwartung:  $N(0, 1)$

- $\langle x \rangle \neq 0 \rightarrow$  Detektorelement verschoben
- $\sigma \neq 1 \rightarrow$  Fehlerabschätzung falsch

## WWW Resourcen

Folien zu diesem Vortrag: **(von Kurt Rinnert, EKP, Uni Karlsruhe)**

<http://www-ekp.physik.uni-karlsruhe.de/~rinnert/kalman/index.html>

Gutes Portal zum Kalman-Filter:

<http://www.cs.unc.edu/~welch/kalman>

# Monte Carlo

Simulation und Integration mit Hilfe von Zufallszahlen.  
Numerische Integration:

$$\int_a^b f(x)dx \approx (b - a) \cdot \frac{1}{N} \sum_i f(x_i)$$

Auswahl der Punkte:

Regelmäßiges Gitter: Unsicherheit  $\propto 1/n^{2/d}$

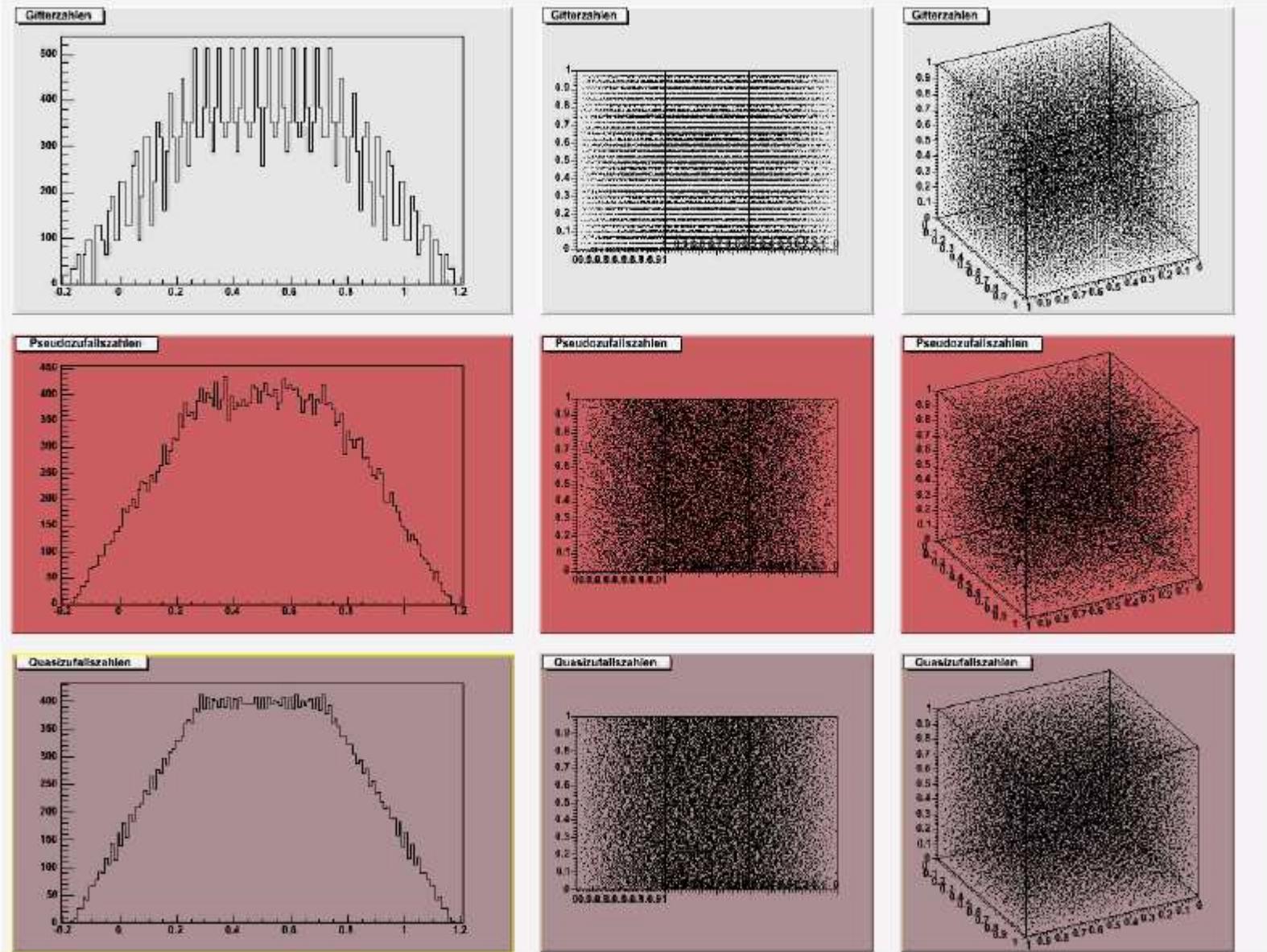
Pseudo-Zufallszahlen: Unsicherheit  $\propto 1/\sqrt{n}$  unabhängig von d!

Ab 4 Dimensionen ist Monte Carlo immer besser!

In 2- ca. 14 Dimensionen zur Integration noch besser:

Quasi-Zufallszahlen (siehe J. Bossert, IEKP-KA/2001-11)

# Quasi-Zufallszahlen Pseudo-Zufallszahlen regelmäßiges Gitter

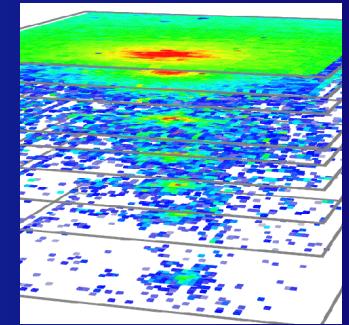
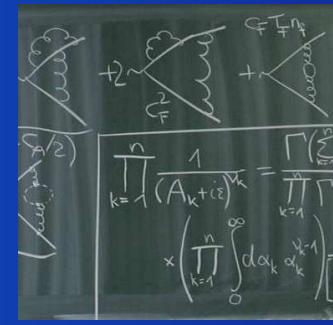




# Einführung in die Datenanalyse Klassifikation und Neuronale Netze

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

Herbstschule für Hochenergiephysik  
Kloster Maria Laach 2004



# Inhalt

Hypothesentest/Klassifikation

Neuronale Netzwerke

Klassifikation

Wahrscheinlichkeitsdichten

NeuroBayes

Anwendungen

Support-Vektor-Maschinen

# Hypothesentests

Schnitte in einer Test-Statistik:

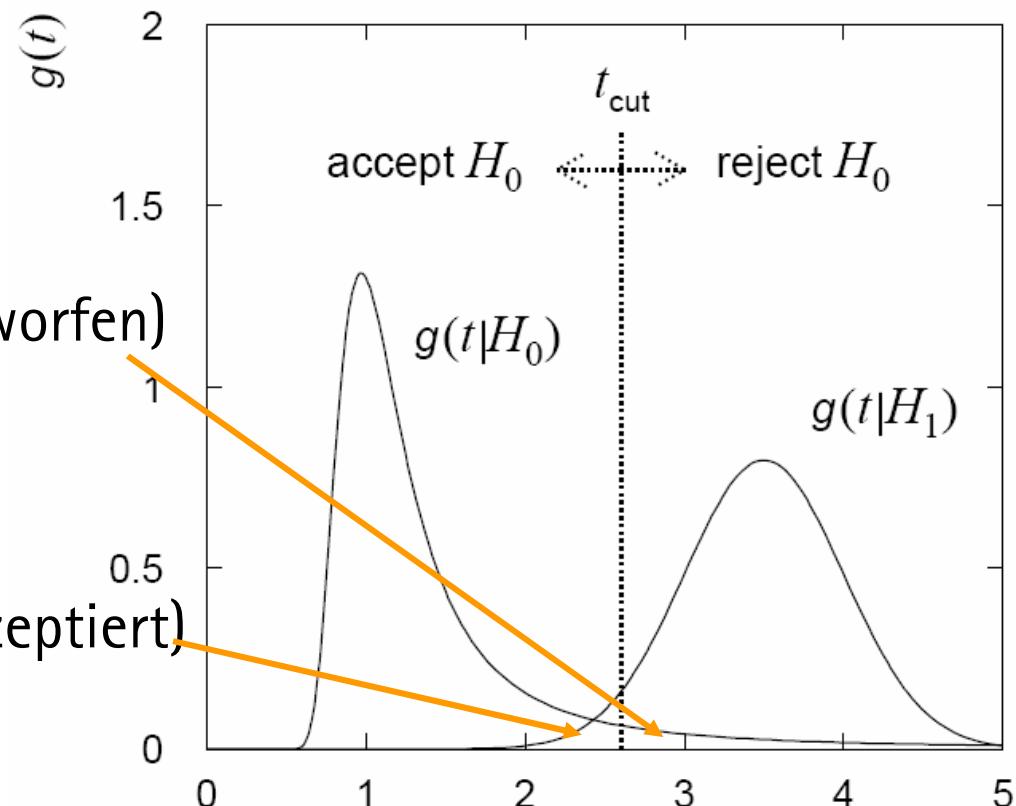
Akzeptiere Hypothese  $H_0$ ,  
wenn  $t < t$

Fehler 1. Art:

$P_1$ (wahre Hypothese wird verworfen)

Fehler 2. Art:

$P_2$ (falsche Hypothese wird akzeptiert)



# Hypothesentests

Wichtige Größen für alle Klassifikationsverfahren:

Effizienz:

$$\varepsilon = \frac{P(\text{selektiert und wahr})}{P(\text{wahr})} = 1 - P1$$

Reinheit (Purity):

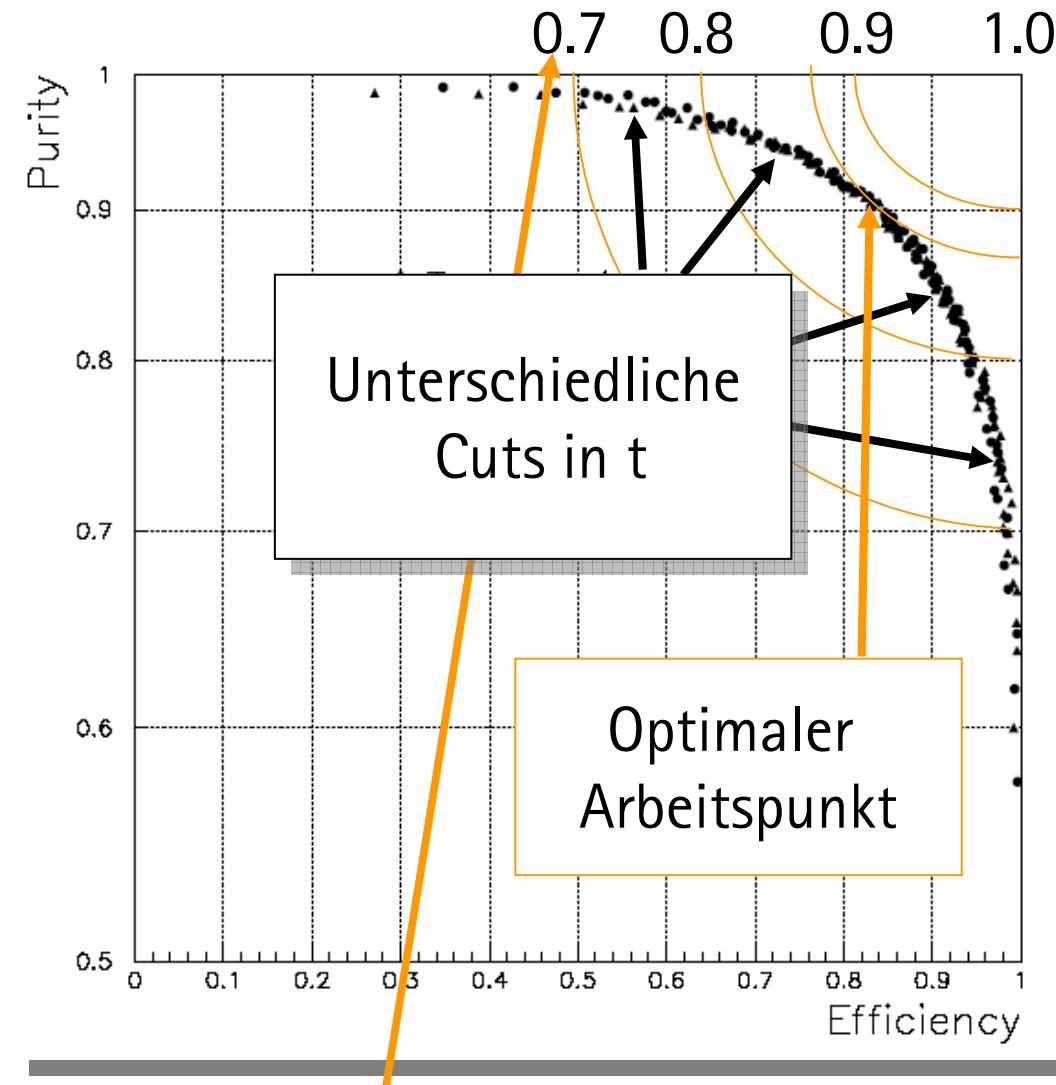
$$\mathcal{P} = \frac{P(\text{selektiert und wahr})}{P(\text{selektiert})} = 1 - P2$$

Wahl des Arbeitspunktes (d.h. des Cut-Werts) je nach Anwendung.

# Hypothesentests

Ein statistisches Verfahren ist um so besser, desto näher es an den Punkt (1,1) im Purity-Efficiency-Plot kommt

Optimale Wahl des Arbeitspunktes je nach Aufgabe:  
Wie skaliert der Gesamtfehler der Analyse mit  $\varepsilon$  und  $P$ ?



Flavour-Tagging in Mixing-Analysen:  $Signifikanz \propto \sqrt{\varepsilon} \cdot (2P - 1)$

# Konstruktion einer Test-Statistik

Lineare Trennung von korrelierten Input-Variablen durch Hyperebene im n-dim. Raum

Fischer-Diskriminante:

maximiere Abstand der Mittelwerte der beiden Klassen in Einheiten der Summe ihrer Varianzen

Neyman-Pearson-Lemma: n unkorrelierte Variablen werden optimal durch das Likelihood ratio getrennt:

$$t = \sum_i^n \frac{L_{H_0}(x_i)}{L_{H_1}(x_i)}$$

# NeuroBayes®-Training



- > 18140 Länderspiele in Datenbank  
(seit ~1900)
- > Aus zeitl. Entwicklung der Einzelergebnisse:  
**24 abgeleitete Größen** (z.B. technische Indikatoren wie moving averages, exponential moving averages , ...)
- > Training für jedes Ergebnis: 0:0, 0:1, 0:2, 0:3, 1:0, 1:1, 1:2, 1:3, 2:0, 2:1, 2:2, 2:3, 3:0, 3:1, 3:2, 3:3
- > Überraschung: 0:1 bzw. 1:0 unabhängig von Mannschaft am Wahrscheinlichsten
- > Tipps für Sonderwertung: aus aktueller Ausgabe des **Kicker-Magazins**

# Gibt es Zusammenhang zwischen Länderspielen von vor 50 Jahren und Spielen?

NEIN !

aber: Vorhersagbarkeit von Spielen durch jüngere Vergangenheit.

Zusammenhang zw. jüngerer Historie und nächstem Spiel war vor 50 Jahren ähnlich wie heute.



# Ergebnis des WM2002 Tipp-Spiels



Fußball-Experte

<phi-t® Mitarbeiter



Rang	Name	Punkte
1.	Jörgen Dalmau	104
2.	Marcel Stanitzki	94
3.	Andreas Heiss <phi-t>	92
4.	Alexander Skiba	90
5.	Valeria Bartsch	88
6.	Dominic Hirschbuehl	86
7.	Phi-T	86
8.	Nicole Lichtenberger	82
9.	Thilo Wüst	80
10.	Yves Kemp	78
11.	Zoltan Albrecht	76
12.	Christian Haag <phi-t>	72
13.	Alexander Feindt	72
14.	Gary Barker	70
15.	Ludger Ramler <phi-t>	70
16.	Christian Sander	68
17.	Hartmut Stadie	68
18.	Markus Moch	66
19.	Alexander Dierlamm	66
20.	Ulrich Kerzel	64
21.	Thorsten Scheidle	64
22.	Wolfgang Wagner	64
23.	Jochen Bossert <phi-t>	56
24.	Thomas Allmendinger	52
25.	Martin Erdmann	52
26.	Michael Wermuth	48



# Ergebnis des WM2002 Tipp-Spiels (ohne Sonder- wertung)



Fußball-Experte

<phi-t><sup>®</sup> Mitarbeiter



Rang	Name	Punkte
1.	Jörgen Dalmau	92
2.	Marcel Stanitzki	82
3.	Phi-T	80
4.	Alexander Skiba	78
5.	Zoltan Albrecht	76
6.	Valeria Bartsch	76
7.	Thilo Wüst	76
8.	Andreas Heiss <phi-t>	76
9.	Dominic Hirschbühl	74
10.	Yves Kemp	72
11.	Nicole Lichtenberger	70
12.	Gary Barker	70
13.	Christian Haag <phi-t>	66
14.	Alexander Feindt	66
15.	Wolfgang Wagner	64
16.	Ulrich Kerzel	64
17.	Thorsten Scheidle	64
18.	Ludger Ramler <phi-t>	64
19.	Hartmut Stadie	62
20.	Christian Sander	62
21.	Markus Moch	60
22.	Alexander Dierlamm	58
23.	Jochen Bossert <phi-t>	56
24.	Thomas Allmendinger	52
25.	Martin Erdmann	46
26.	Michael Wermuth	36



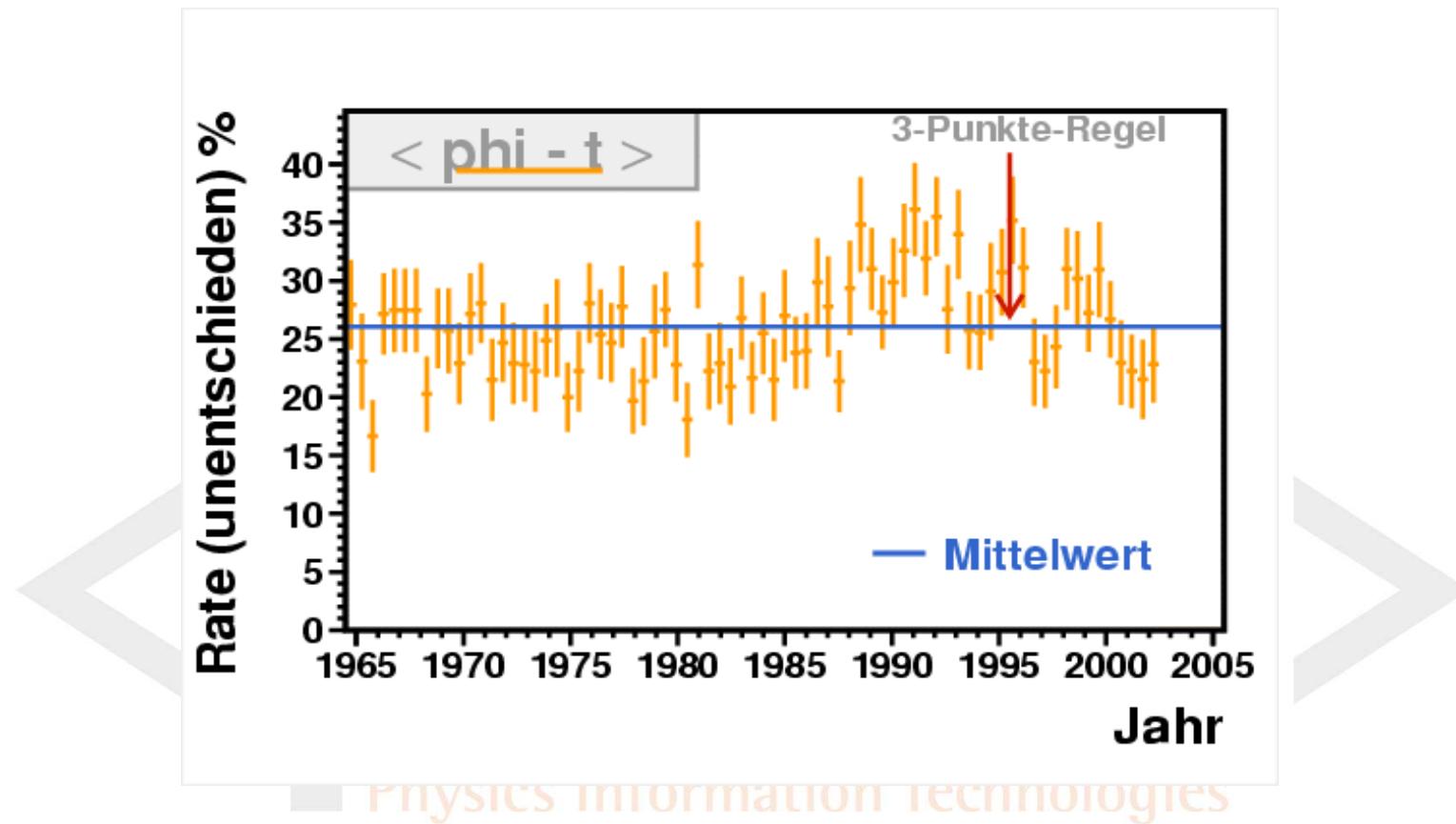


## Datenbank-Auszug

- > Saison 63/64 bis 1. Hälfte 03/04
- > Insgesamt **12335** Begegnungen
- > Mehr als **100** abgeleitete Größen
- > **3** Trainings: Heim-Sieg, Heim-Niederlage, Unentschieden

Saison	Spieltag	Heim-Gast	Ergebnis (Halbzeit)
2001/2002	27.10.2001	TSVM-BALE	1: 4 ( 1: 1 )
2001/2002	27.10.2001	VFBS-STPA	2: 0 ( 1: 0 )
2001/2002	27.10.2001	ENCO-BODO	0: 2 ( 0: 0 )
2001/2002	27.10.2001	SVWB-VFLW	1: 0 ( 0: 0 )
2001/2002	27.10.2001	FCKO-FCBM	0: 2 ( 0: 1 )
2001/2002	27.10.2001	FCKA-SC04	0: 0 ( 0: 0 )
2001/2002	27.10.2001	BOMO-SCFR	2: 2 ( 1: 1 )
2001/2002	28.10.2001	FCNU-HBSC	1: 3 ( 0: 1 )
2001/2002	28.10.2001	HASV-HARO	0: 1 ( 0: 1 )
2001/2002	03.11.2001	FCBM-HASV	3: 0 ( 0: 0 )
2001/2002	03.11.2001	VFLW-FCNU	5: 0 ( 2: 0 )
2001/2002	03.11.2001	BODO-VFBS	1: 0 ( 0: 0 )
2001/2002	03.11.2001	STPA-TSVM	0: 3 ( 0: 1 )
2001/2002	03.11.2001	BALE-FCKA	2: 1 ( 1: 1 )
2001/2002	03.11.2001	SC04-FCKO	3: 1 ( 1: 1 )
2001/2002	03.11.2001	HARO-SVWB	0: 1 ( 0: 0 )
2001/2002	04.11.2001	HBSC-BOMO	3: 0 ( 2: 0 )
2001/2002	04.11.2001	SCFR-ENCO	3: 1 ( 1: 1 )
2001/2002	17.11.2001	TSVM-BODO	1: 3 ( 0: 0 )
2001/2002	17.11.2001	VFBS-SCFR	3: 0 ( 1: 0 )
2001/2002	17.11.2001	HBSC-VFLW	2: 0 ( 1: 0 )
2001/2002	17.11.2001	SVWB-FCBM	1: 0 ( 1: 0 )
2001/2002	17.11.2001	HASV-SC04	0: 0 ( 0: 0 )
2001/2002	17.11.2001	FCKO-BALE	1: 2 ( 0: 0 )
2001/2002	17.11.2001	FCKA-STPA	5: 1 ( 2: 1 )
2001/2002	18.11.2001	FCNU-HARO	2: 0 ( 0: 0 )

# Erkenntnis: Einführung der 3-Punkte-Regel hat keine Auswirkung auf Häufigkeit von Unentschieden



# Vorhersage für 23. Spieltag



# Weitere Anwendung im Bereich Fußball: Vorhersage von Verletzungsrisiken von Spielern...

AC Milan wendet  
neuronale Netze bereits  
erfolgreich an

4 Jahre Datenerhebung:

- > biomedizinische Daten
- > Fitness-Daten
- > Psychologische Gutachten

Computer predicts sports injuries

Italian soccer powerhouse looks to artificial intelligence for edge on the field

Arsenal's Robert Pires receives treatment on his injured right knee during FA Cup soccer action in March. The injury forced Pires out of the game.



Nick Potts / AP file

By Linda Carroll  
MSNBC CONTRIBUTOR

# Das <phi-t> Mausspiel:

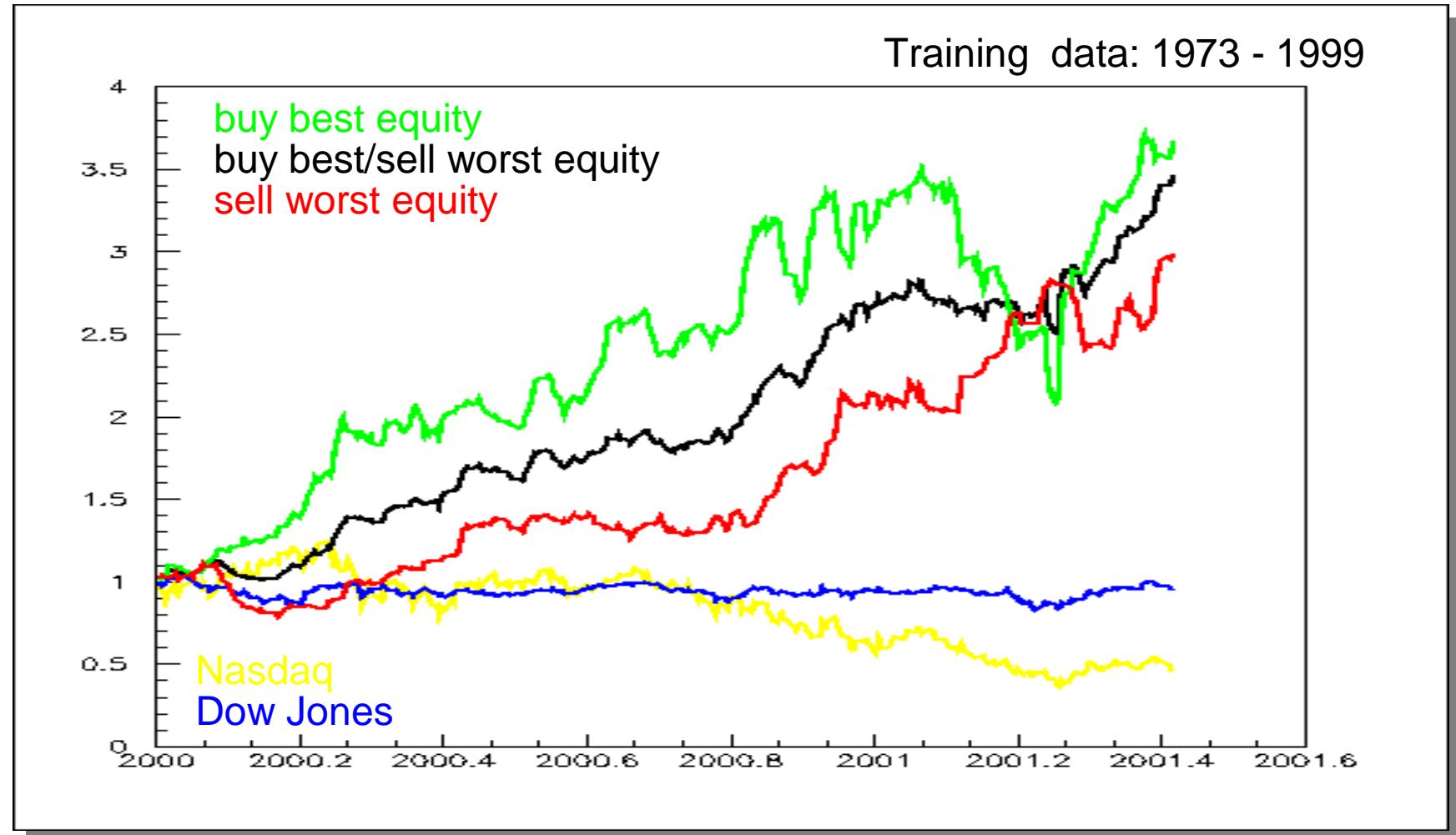
oder: wir können sogar Ihren freien Willen  
vorhersagen...



[www.phi-t.de/mousegame](http://www.phi-t.de/mousegame)

Physics Information Technologies

# Aktienkursvorhersage



# Ameritrade Börsenspiel

Netscape: Investors Cup

File Edit View Go Communicator Help

- Transactions  
Player Performance  
- Top 20 Weekly  
...Gainers  
- Top 20 Daily  
...Gainers  
- Relative  
...Performance  
Education  
- Education Center  
Personal Finance 

Contact Us  
Register  
Official Rules

Your Portfolio at a Glance  
Net Account Value  
\$61,304.84  
Buying Power  
\$11,736.09  
State Rank Rank  
1 356

  
25 FREE TRADES OPEN YOUR ACCOUNT NOW! GO!

**Ameritrade Investors Cup**

Rank	Name	Percentile Rank	Account Value	Percentage Gain
1	JEN	100.00%	\$124,905.36	149.81%
2	fatcat1	99.99%	\$113,448.24	126.90%
3	BigAI	99.99%	\$101,452.42	102.90%
4	bags	99.98%	\$101,019.72	102.04%
5	bowlingout	99.97%	\$100,638.61	101.28%
351	3base4eva	97.63%	\$61,445.48	22.89%
352	Josevi	97.63%	\$61,392.65	22.79%
353	AlainFlores	97.62%	\$61,375.48	22.75%
354	Illwah	97.61%	\$61,346.37	22.69%
355	KaoLee	97.61%	\$61,320.34	22.64%
356	phi-t	97.60%	\$61,304.84	22.61%
357	Buffdaddy17	97.59%	\$61,297.63	22.60%
358	pishi	97.58%	\$61,296.77	22.59%
359	idwise	97.58%	\$61,287.46	22.57%
360	Unclebubba63	97.57%	\$61,285.48	22.57%
361	ADFUTURES	97.56%	\$61,270.95	22.54%
14778	Steve007	0.03%	\$11,024.97	-77.95%
14779	kamil z. madi	0.03%	\$10,348.72	-79.30%
14780	Buster Jigs	0.02%	\$8,736.80	-82.53%
14781	ZIO	0.01%	\$5,894.09	-88.21%
14782	bongowally	0.01%	\$5,566.66	-88.87%

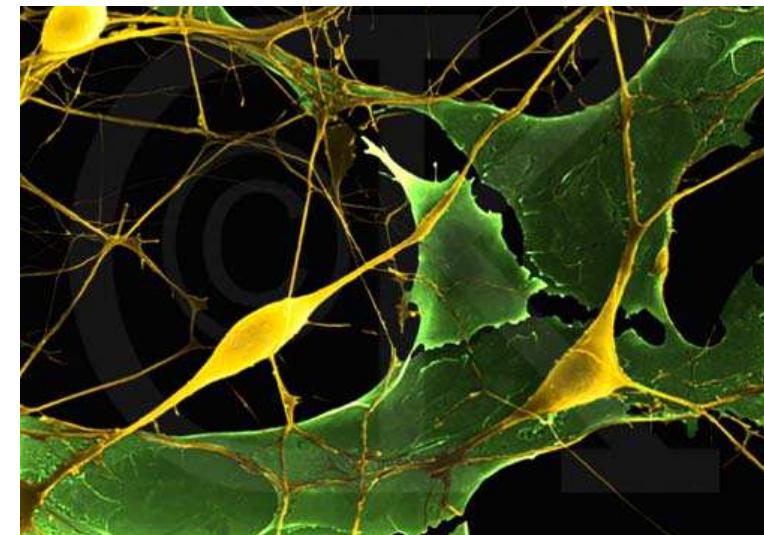
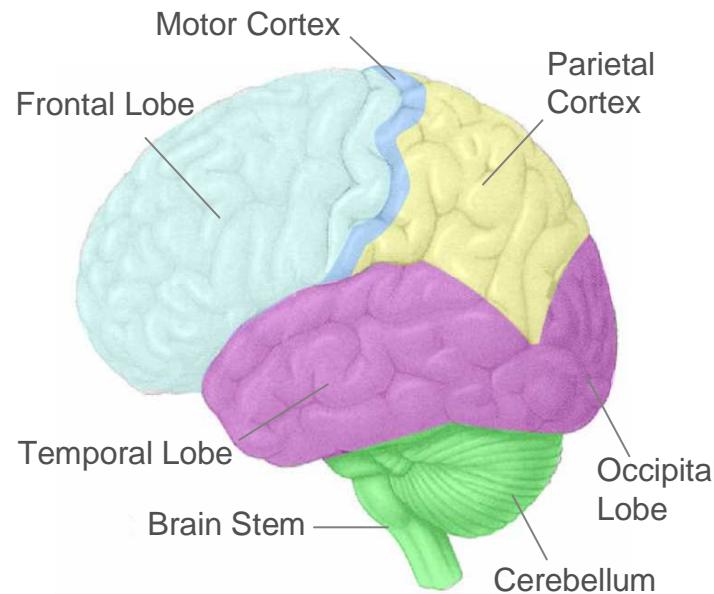
SnapQuote   For Options, use "OP"(+BBMAT). For an Index, use "SI"(\$INDU).  
Quotes provided by S&P Comstock.



Jan-Mar 2002:  
Performance +22.61%  
Platz 356 von 14782  
(beste 2.5%)

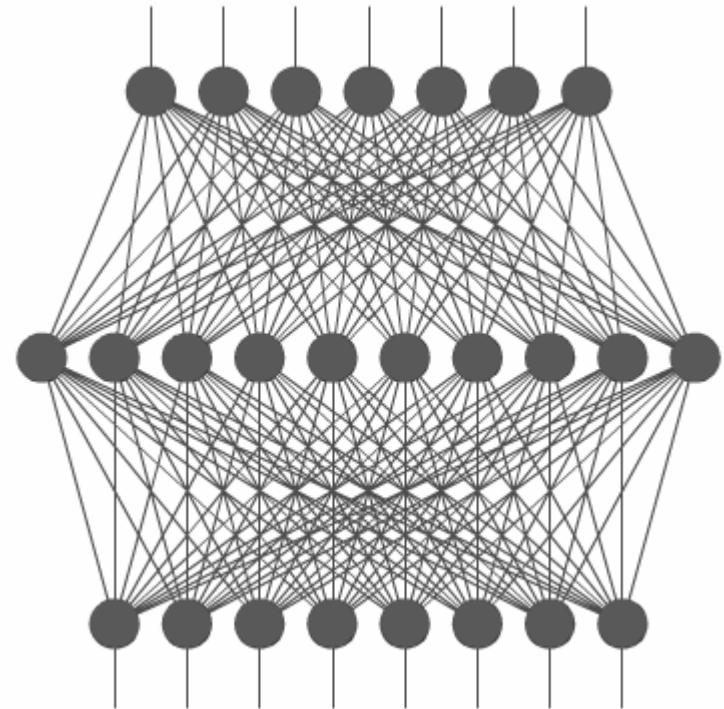
# <phi-t> NeuroBayes®

**Neuronale Netzwerke:**  
Selbstlernende Verfahren, der Natur nachempfunden



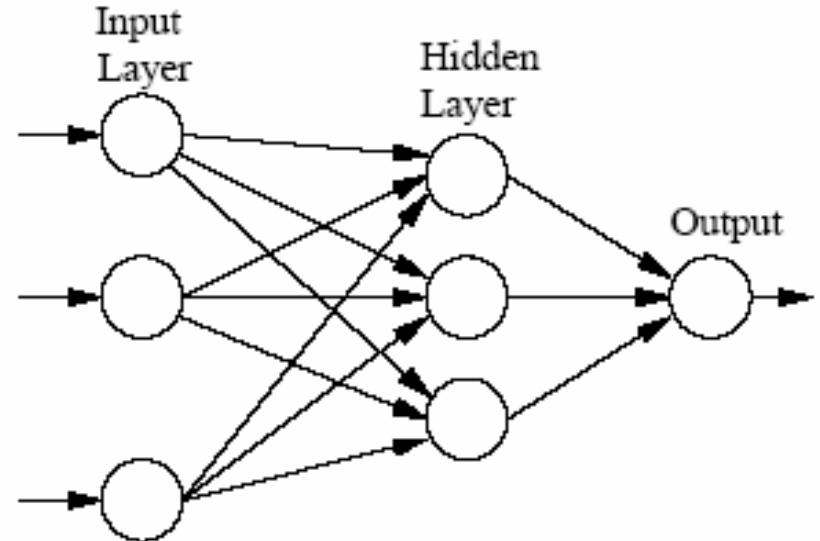
# Neuronale Netzwerke

Die Information  
(das Wissen, die Expertise)  
steckt in den Verbindungen  
zwischen den Nervenzellen



# Neuronale Netzwerke

## Basisfunktion



Output des Knotens  $j$  in Lage  $n$  wird aus gewichteter Summe der Outputs aller Knoten der Lage  $n - 1$  berechnet:

$$x_j^{(n)} = f\left(\sum_i w_{i,j}^{(n)} x_i^{(n-1)} + w_{0,j}^{(n)}\right)$$

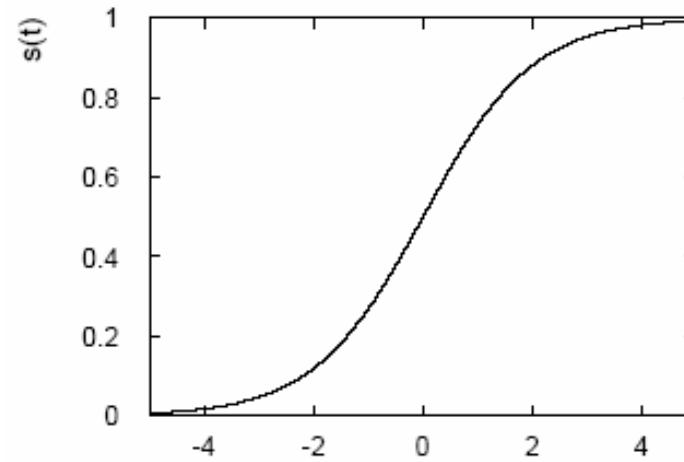
Jeder Verbindung wird ein Gewicht  $w_{i,j}^{(n)}$ , jedem Knoten ein Bias  $w_{0,j}^{(n)}$  zugeordnet.

# Neuronale Netzwerke – Transferfunktion

Als Transferfunktion  $f(x)$  wird eine nichtlineare monotone Funktion mit eingeschränktem Wertebereich benutzt, z.B. die Sigmoidfunktion:

$$f(x) = \frac{1}{1 + \exp(-x)}$$

Sie bildet das Intervall  $(-\infty, \infty)$  auf  $(0, 1)$  ab.



# Neuronale Netzwerke – Training

Als Training bezeichnet man den Minimierungsprozess einer Loss-Funktion, bei dem die Gewichte so angepasst werden, dass die Abweichung vom Soll-Output für ein Set von Input-Vektoren minimiert wird.

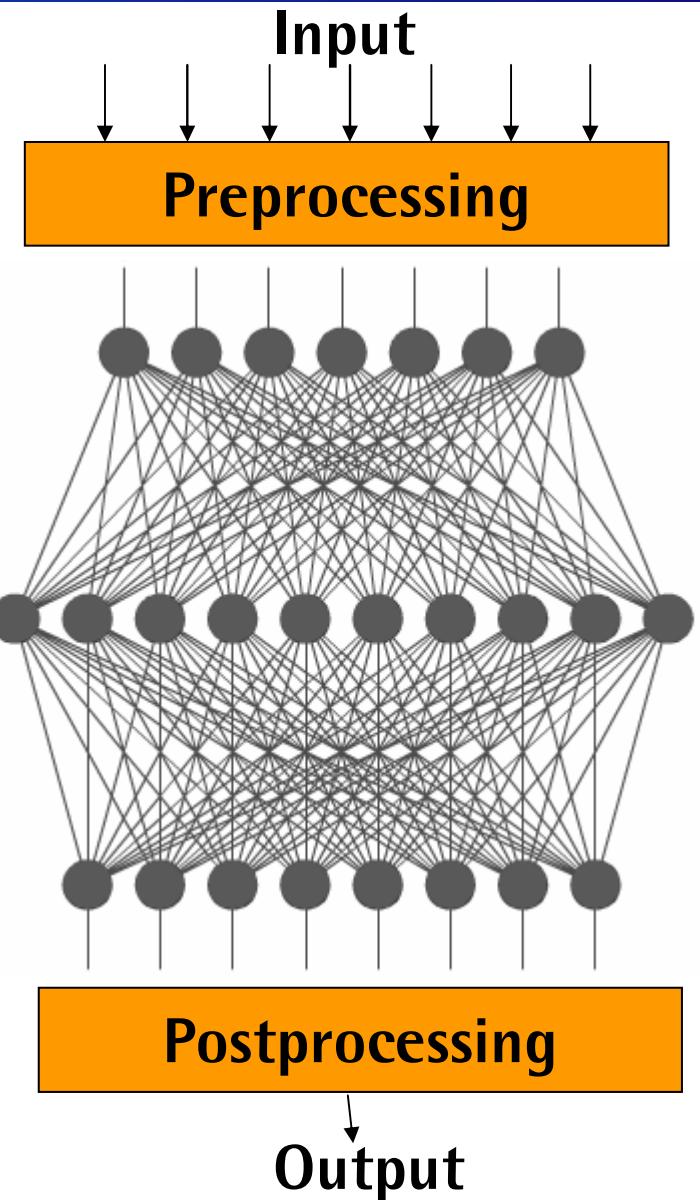
Mögliche Loss-Funktionen:  
Summe der quadratischen Abweichungen  
oder Entropie(Maximum Likelihood)

# NeuroBayes® Prinzip

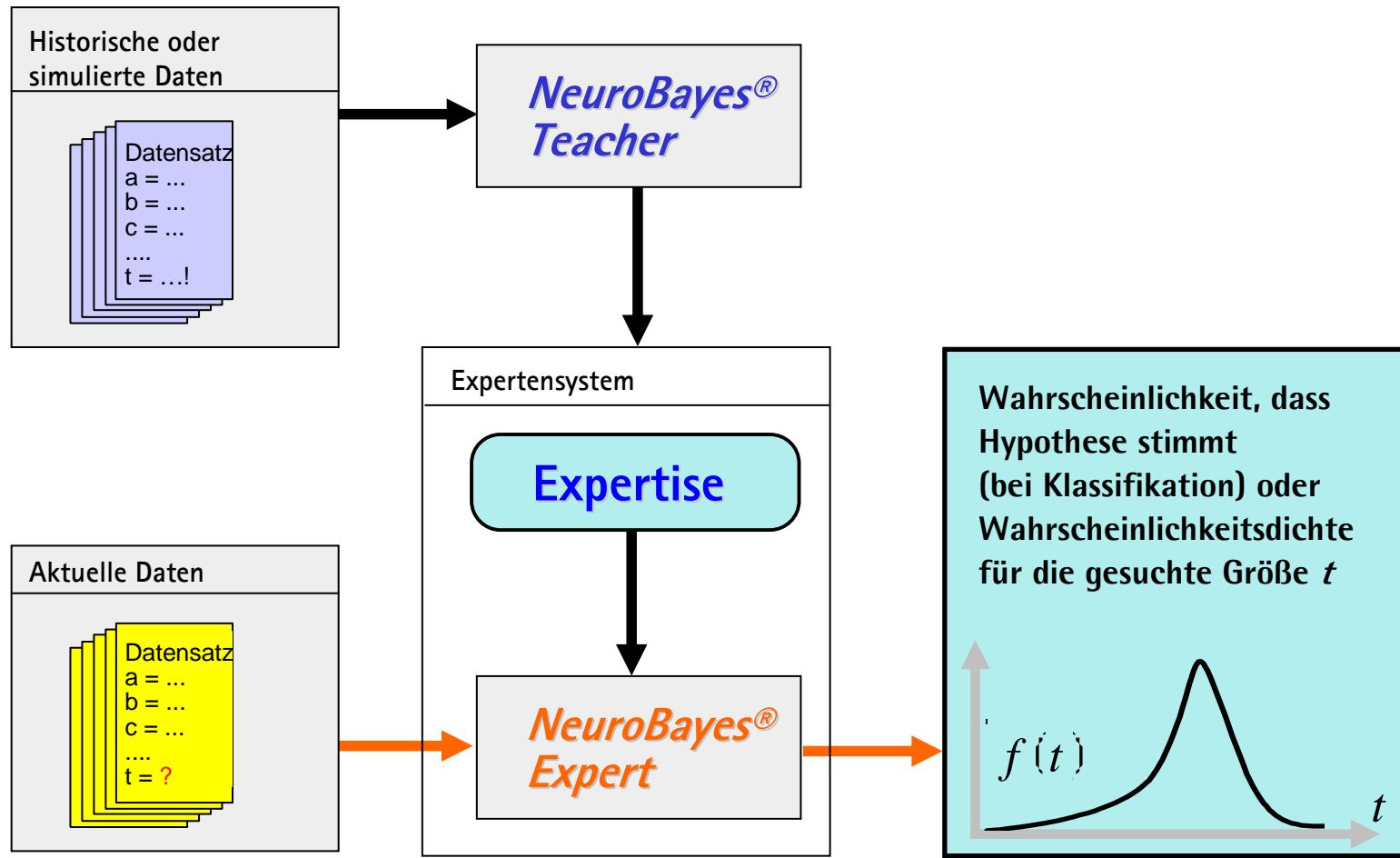
**NeuroBayes® Teacher:**  
Lernen von komplizierten  
Zusammenhängen aus  
bestehenden Datenbanken

**NeuroBayes® Expert:**  
Prognosen für unbekannte Daten

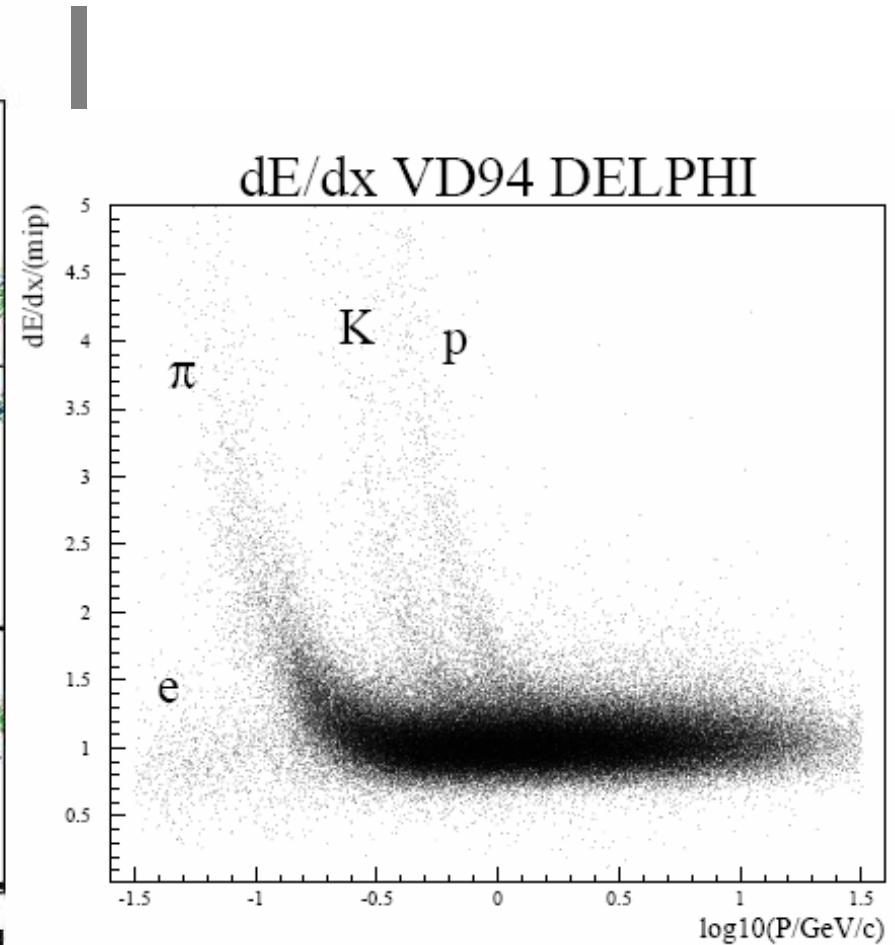
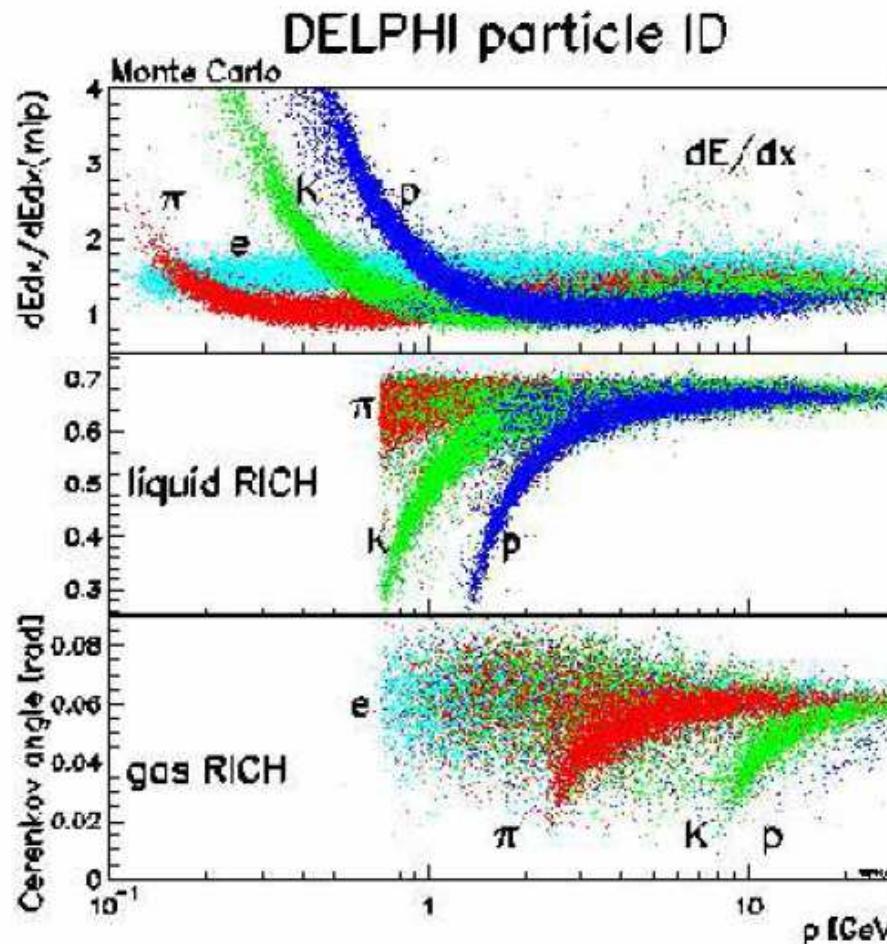
Signifikanzkontrolle



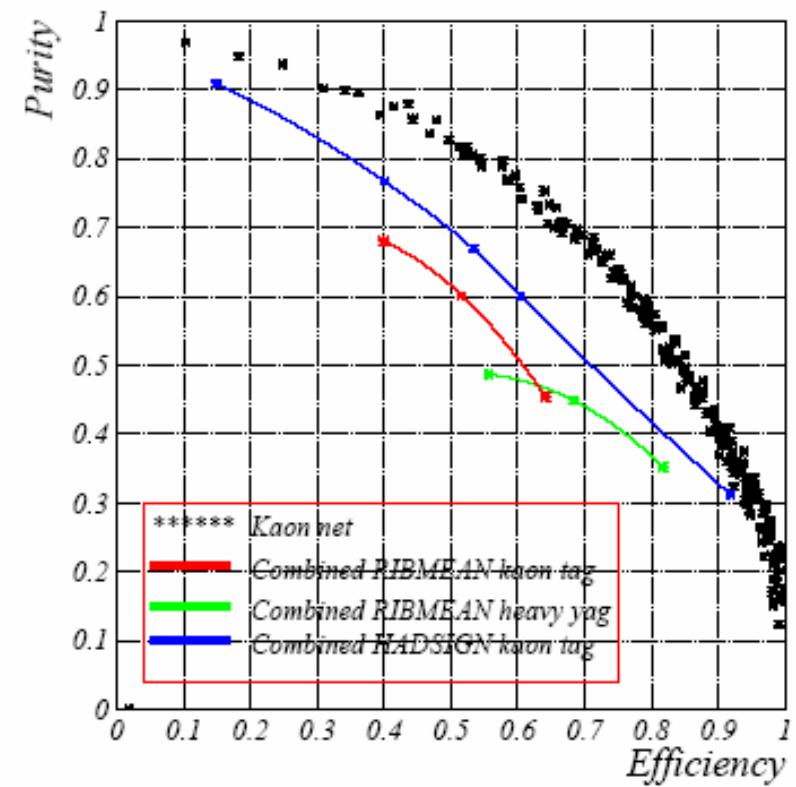
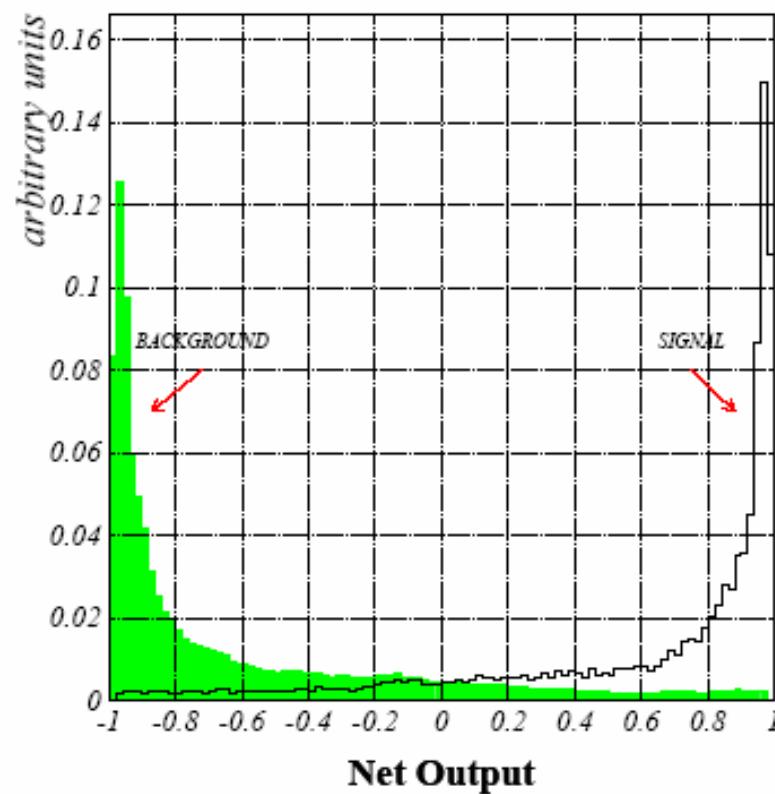
# Funktionsweise: Training und Anwendung



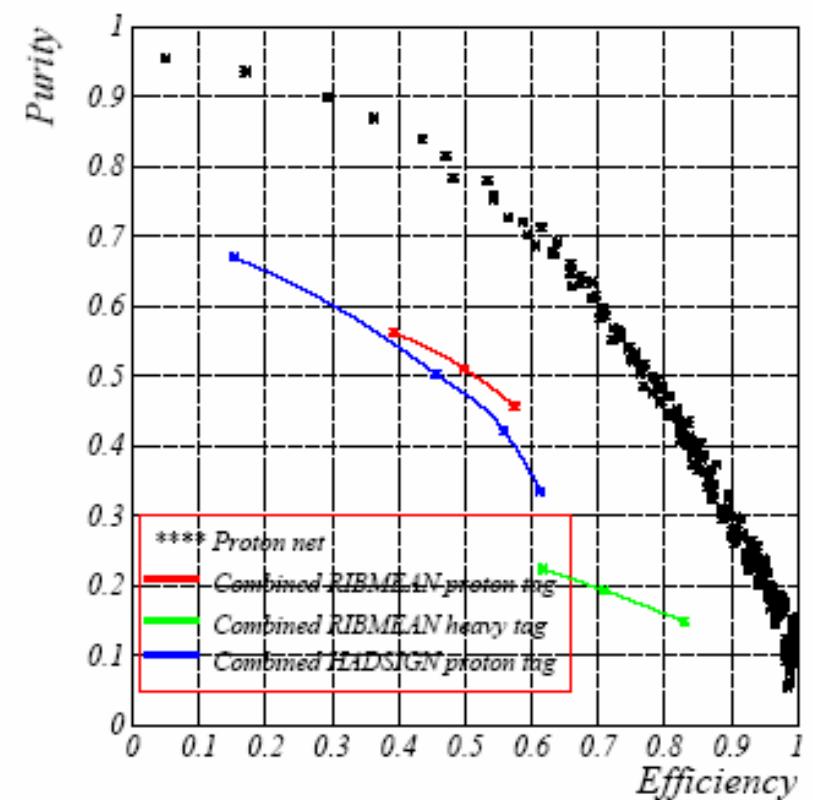
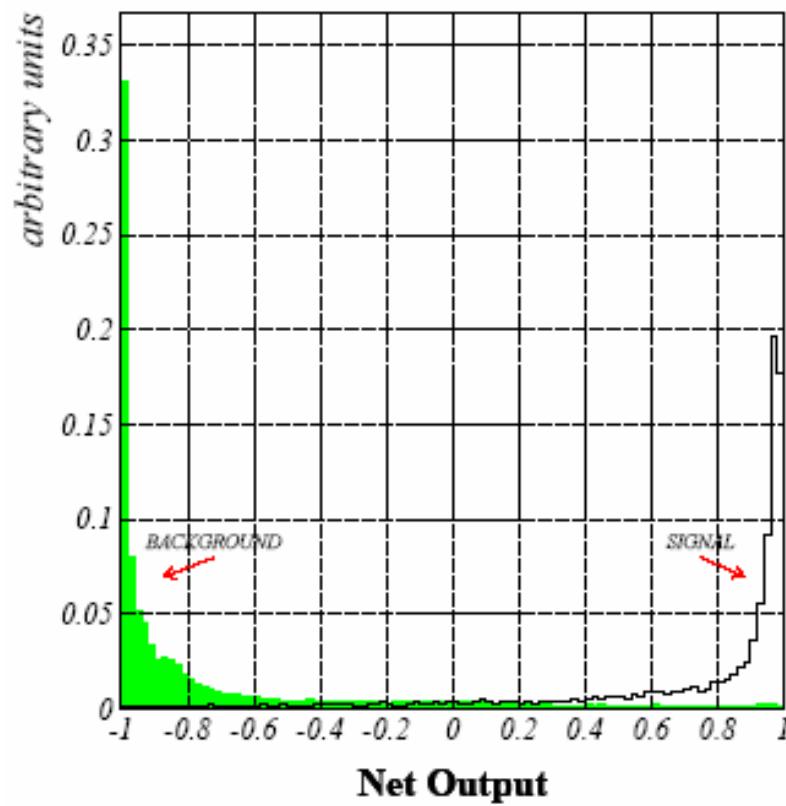
# Beispiele: DELPHI Particle ID (MACRIB)



# MACRIB Kaon ID

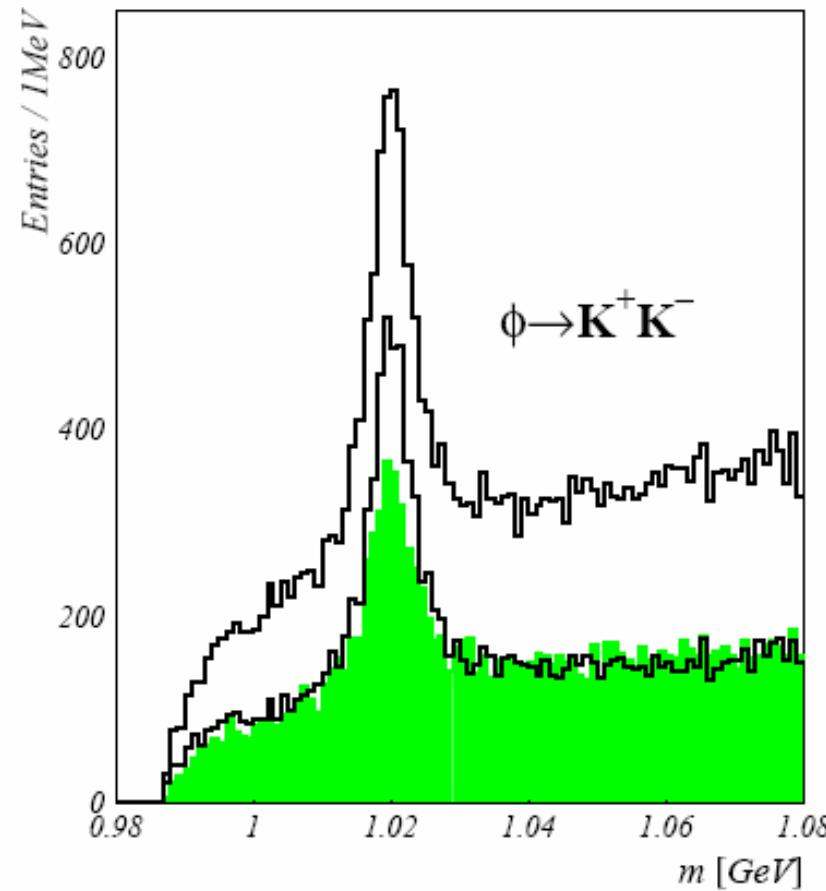


# MACRIB Proton ID

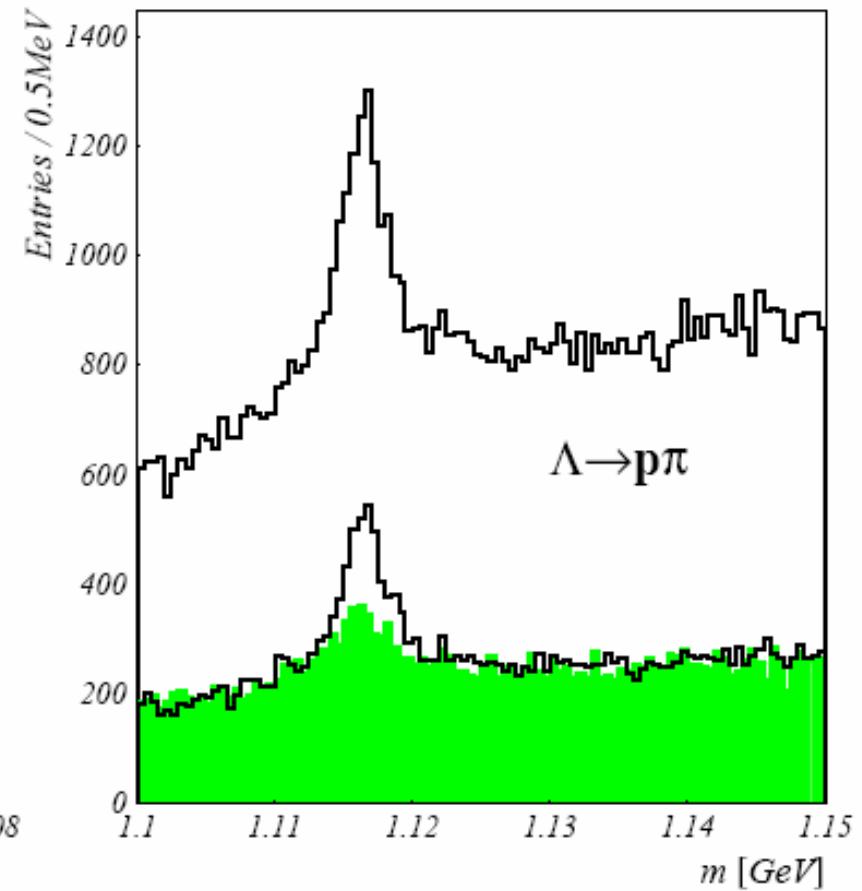


# MACRIB Konsequenzen für Analysen

300% Phi-Mesonen



300% Lambda-Baryonen

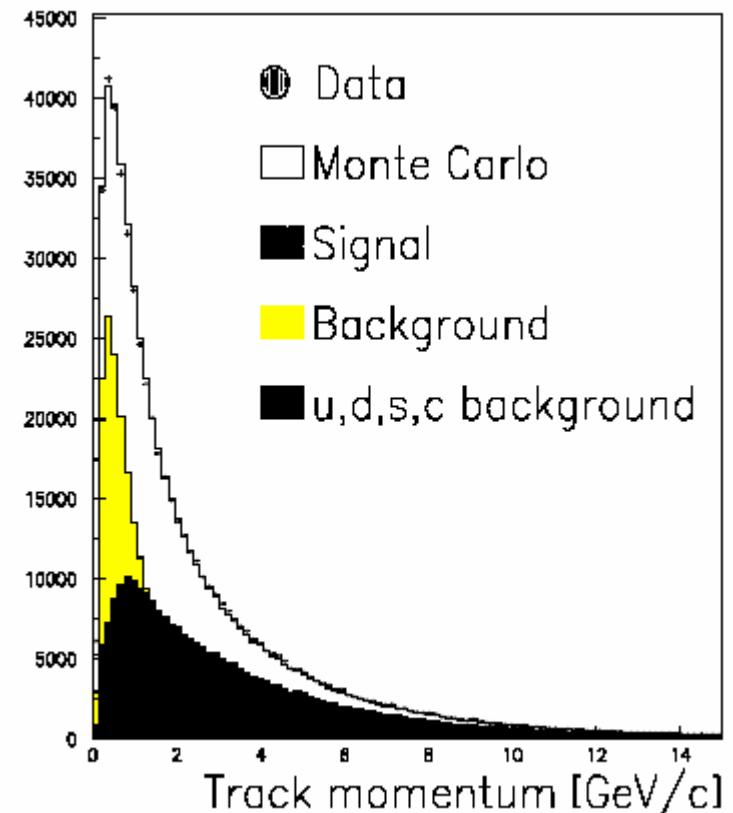
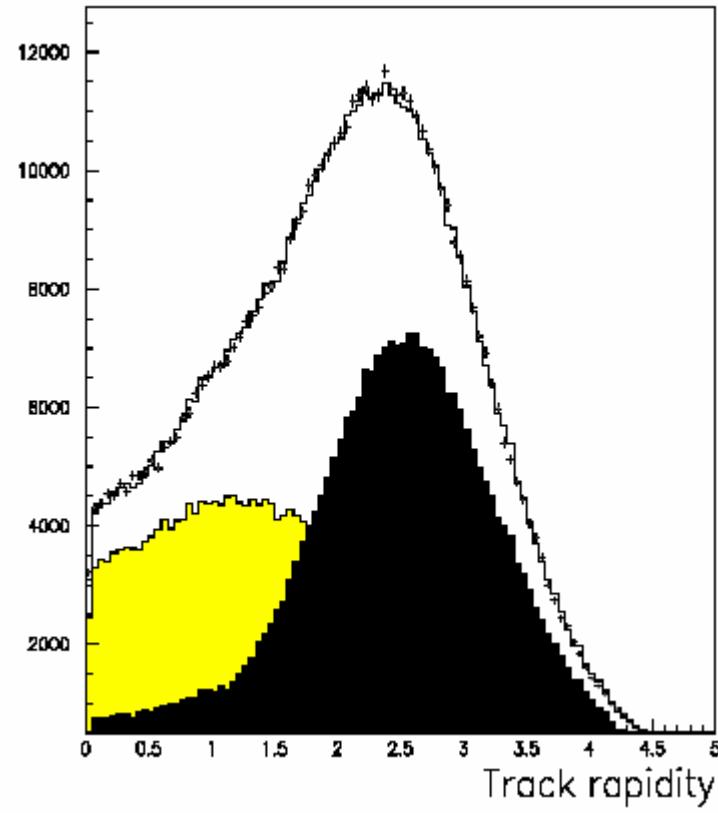




# BSAURUS inklusive B-Rekonstruktion (DELPHI)

Expertensystem mit > 20 Neuronalen Netzwerken

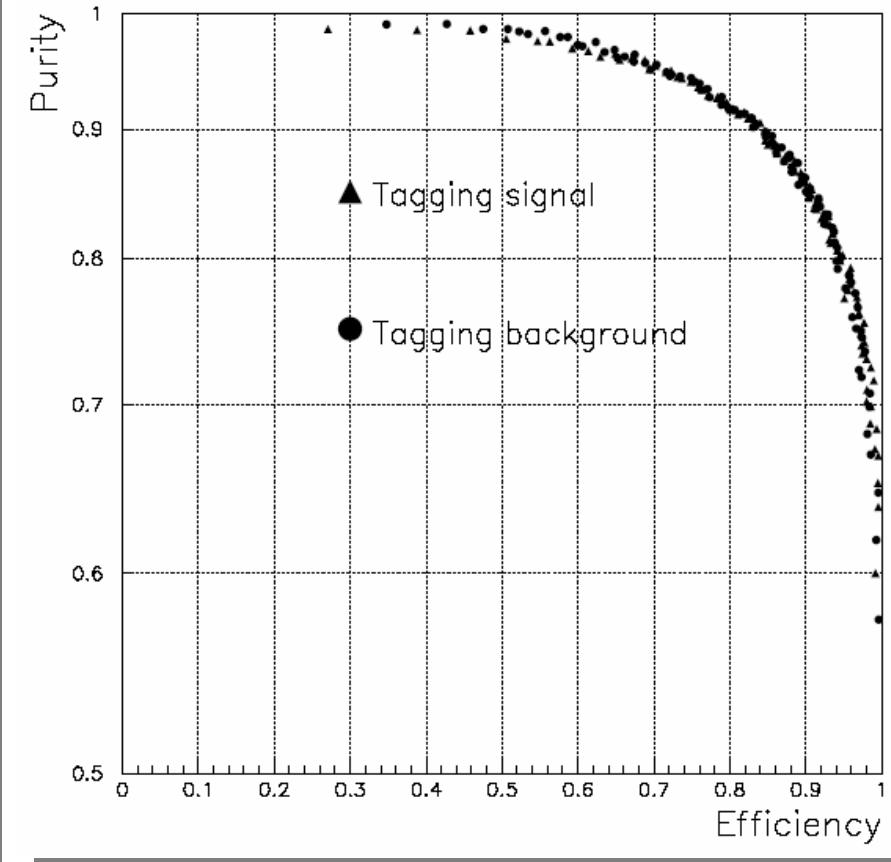
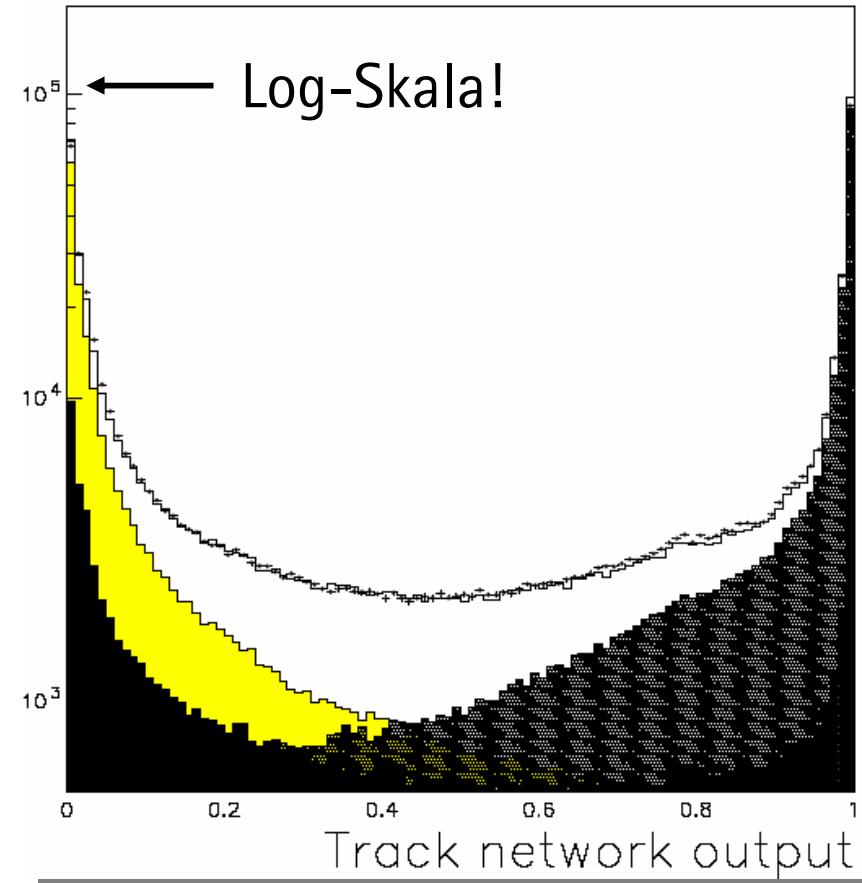
Trainiert mit Monte Carlo . MC muss gut sein!





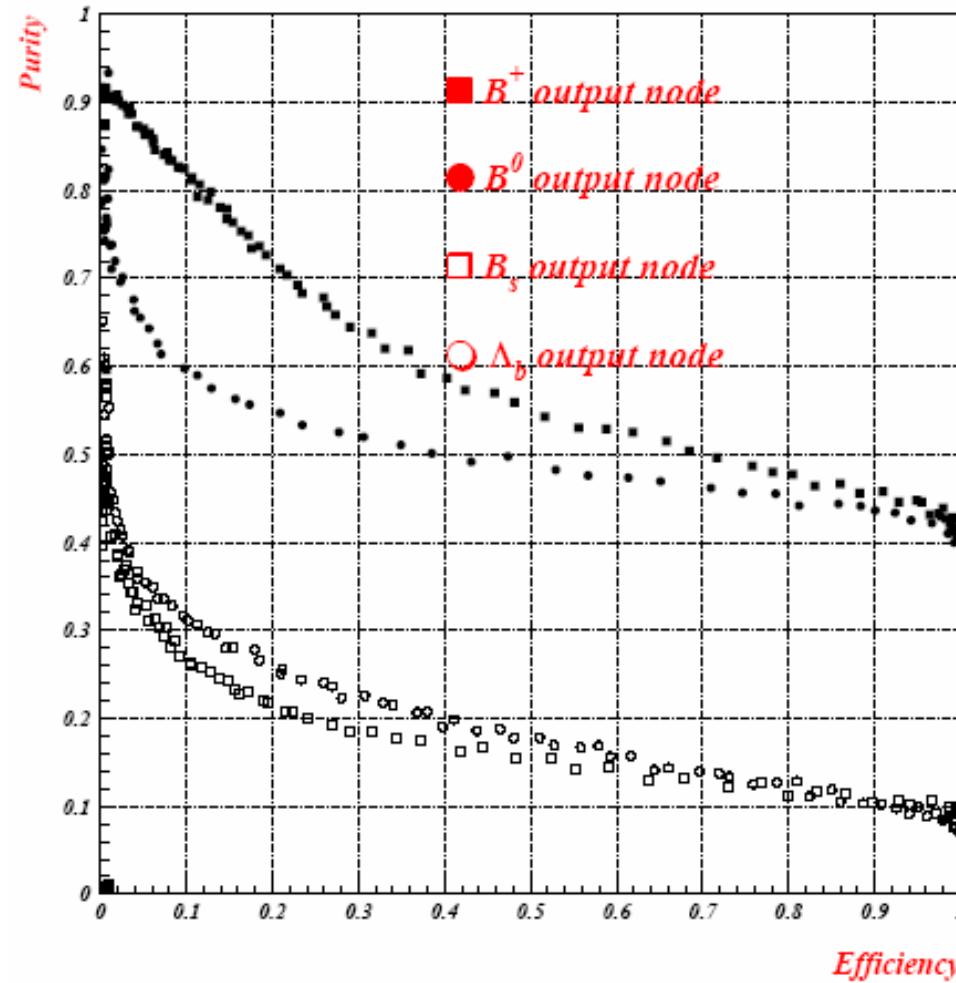
# BSAURUS inklusive B-Rekonstruktion (DELPHI)

Track-Net: Kommt eine Spur aus B-Zerfall oder nicht?





# BSAURUS inklusive B-Rekonstruktion (DELPHI)

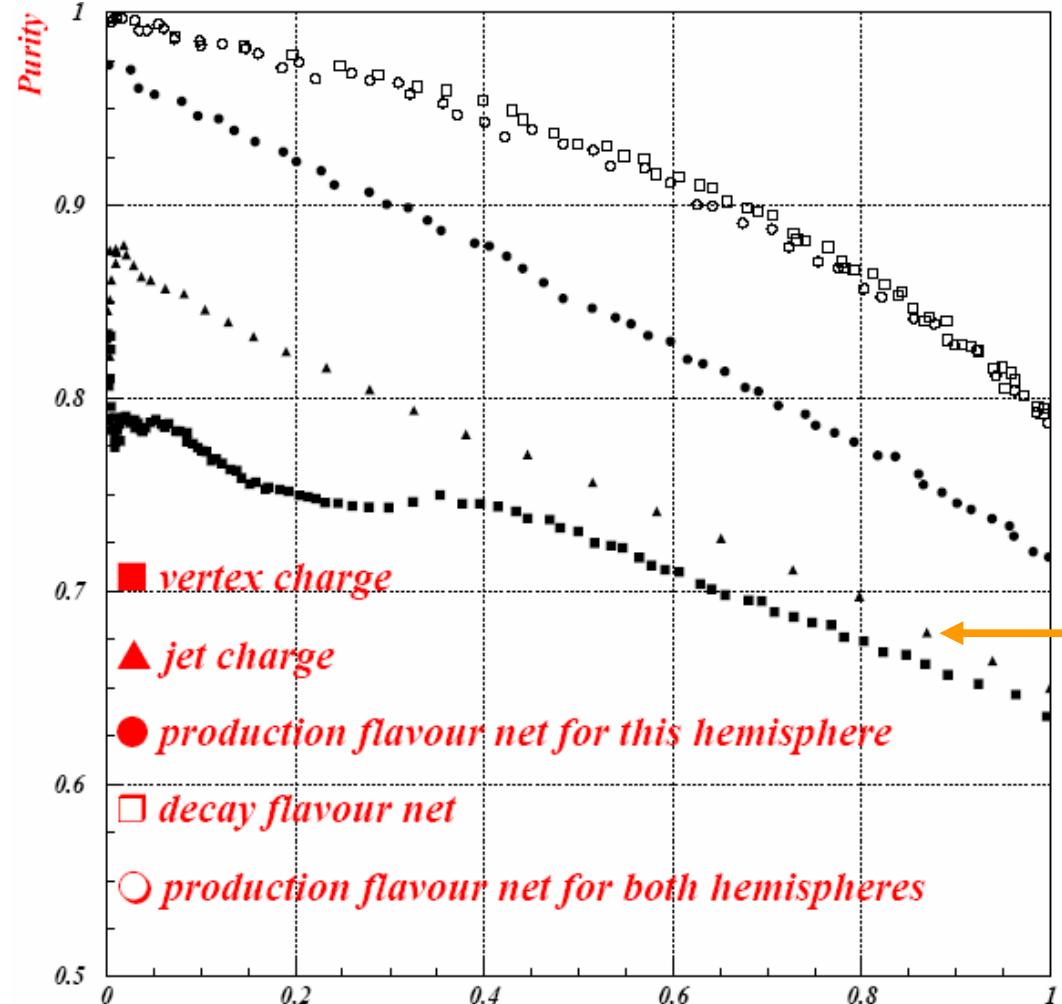


Ist das B-Hadron in dieser Hemisphäre ein  $B^+$ , ein  $B^0$ , ein  $B_s$  oder ein  $\Lambda_b$ ?

a priori-Wahrscheinlichkeiten



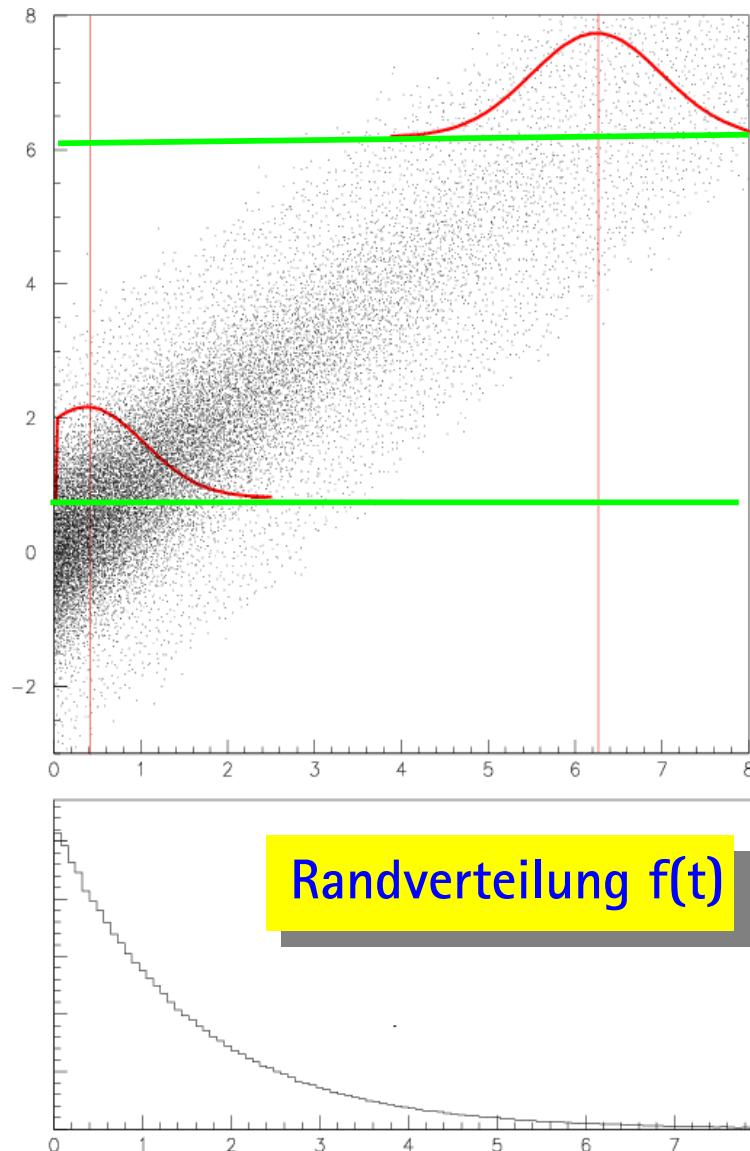
# BSAURUS inklusive B-Rekonstruktion (DELPHI)



B- Flavour-Tag  
zum Erzeugungszeitpunkt

Standard-Jet-Ladung

Flavour-Tagging in  
B-Mixing-Analysen:  
 $Signifikanz \propto \sqrt{\varepsilon} \cdot (2P - 1)$



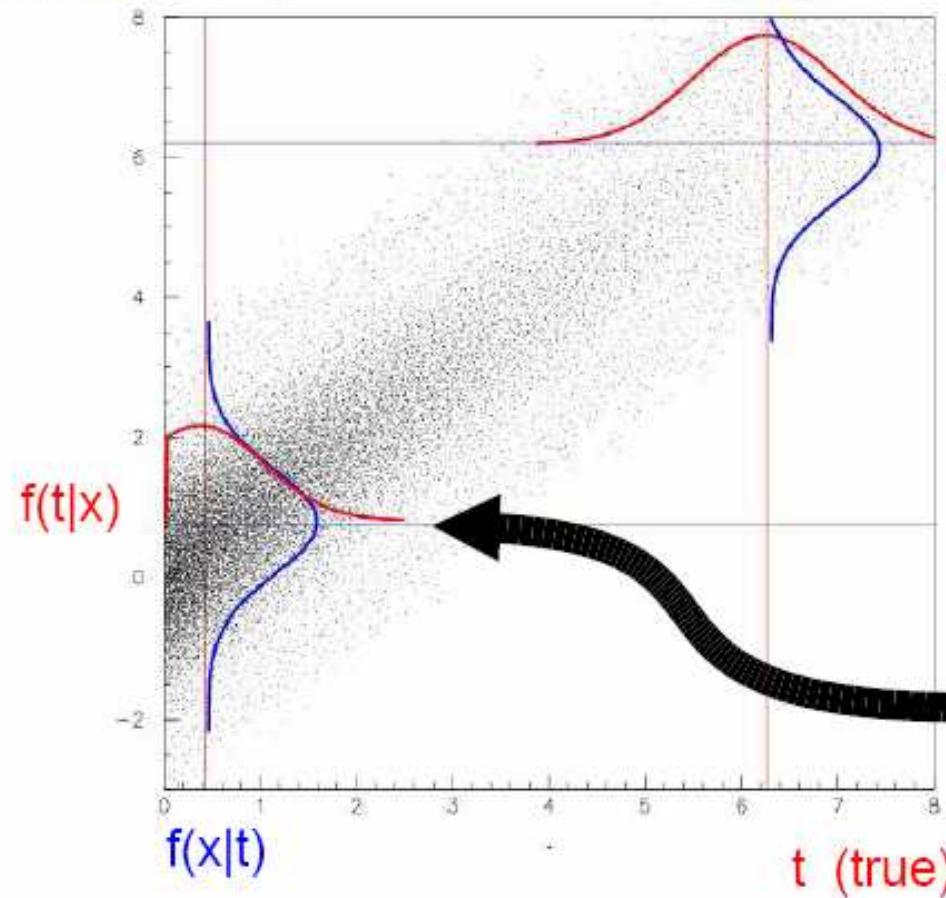
## Bedingte Wahrscheinlichkeitsdichten $f(t|x)$

Bedingte Wahrscheinlichkeitsdichte  
für einen speziellen Fall  $x$   
(Bayes'scher Posterior)

Bedingte Wahrscheinlichkeitsdichten  
 $f(t|x)$  sind Funktionen von  $x$ , hängen aber  
auch von der Randverteilung  $f(t)$  ab.

Gesamtverteilung  
(Bayes'scher Prior)

$x$  (measured)



$f(x|t)$

$f(t|x)$

$t$  (true)

Klassischer Ansatz:  
 $f(x|t)=f(t|x)$   
 näherungsweise korrekt bei guter Auflösung ohne nahe physikalische Grenzen

Bayes'scher Ansatz:  
 berücksichtigt a priori-Wissen  $f(t)$ :  

- Lebensdauer niemals negativ
- Wahre Lebensdauerverteilung exponentialverteilt

# Das Ziel

Ziel:

Bayes'sche Schätzung  $f(t | \vec{x})$  für eine einzelne  
multidimensionale Messung  $\vec{x}$ .

- Komponenten von  $\vec{x}$  können korreliert sein.
- Komponenten von  $\vec{x}$  sollten zu t oder seiner Unsicherheit  
korreliert sein.
- Das alles soll automatisch aus Datenbanken mit Monte-Carlo-  
Simulationen oder historischen Daten gelernt werden.

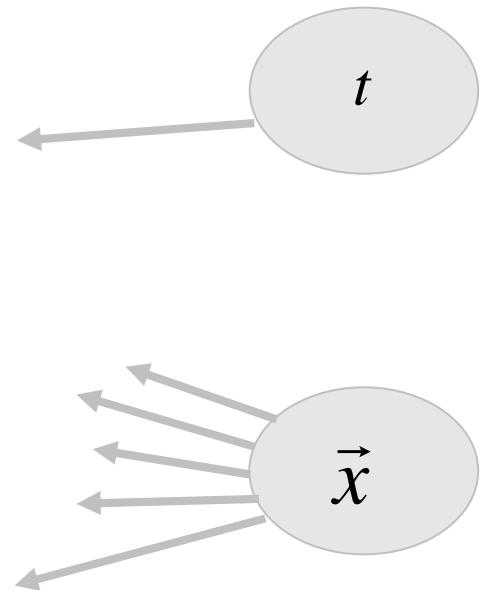
Beachte:

Konditionierte Wahrscheinlichkeitsdichte enthält viel mehr Informationen  
als nur der **Mittelwert**, der in einer Regressionanalyse bestimmt wird.  
Sie gibt auch Auskunft über die **Unsicherheit** und die **Form** der Verteilung,  
insbesondere **nicht-Gauss'sche Ausläufer**.

# Bedingte Wahrscheinlichkeitsdichten in der Teilchenphysik

Was ist die Wahrscheinlichkeitsdichteverteilung von der wahren B-Meson-Energie  $E$  in diesem Ereignis, das mit dem DELPHI-Detektor bei LEP II bei dieser Strahlenergie, mit dieser effektiven Schwerpunktenergie  $s'$ , diesen  $n$  Spuren mit jenen Impulsen und Rapiditäten in der Hemisphäre, die diesen Vertex bilden, mit dieser Zerfallslänge und Wahrscheinlichkeit, dieser Anzahl nicht gut rekonstruierter Spuren etc. pp. aufgenommen wurde?

$$f(t | \vec{x})$$



# NeuroBayes solution ansatz

Discretize  $f(t)$  into  $N$  intervals of same area by equalisation (nonlinear transformation  $t \rightarrow s$ )



Train a neural network with  $N$  output nodes to the  $N$  binary decisions:  
The true  $t$  is larger than / lower than threshold  $i$



Fit smooth function (cubic spline) through  $N$  net outputs:  
= cumulated conditional probability in transformed variable  $s$



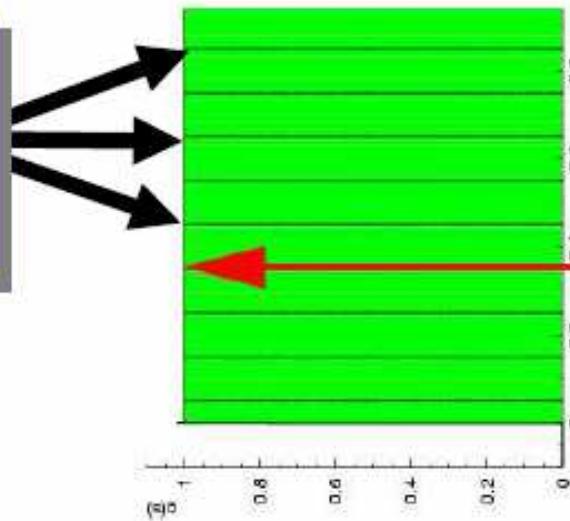
Analytic differentiation returns probability density function in transformed variable  $s$



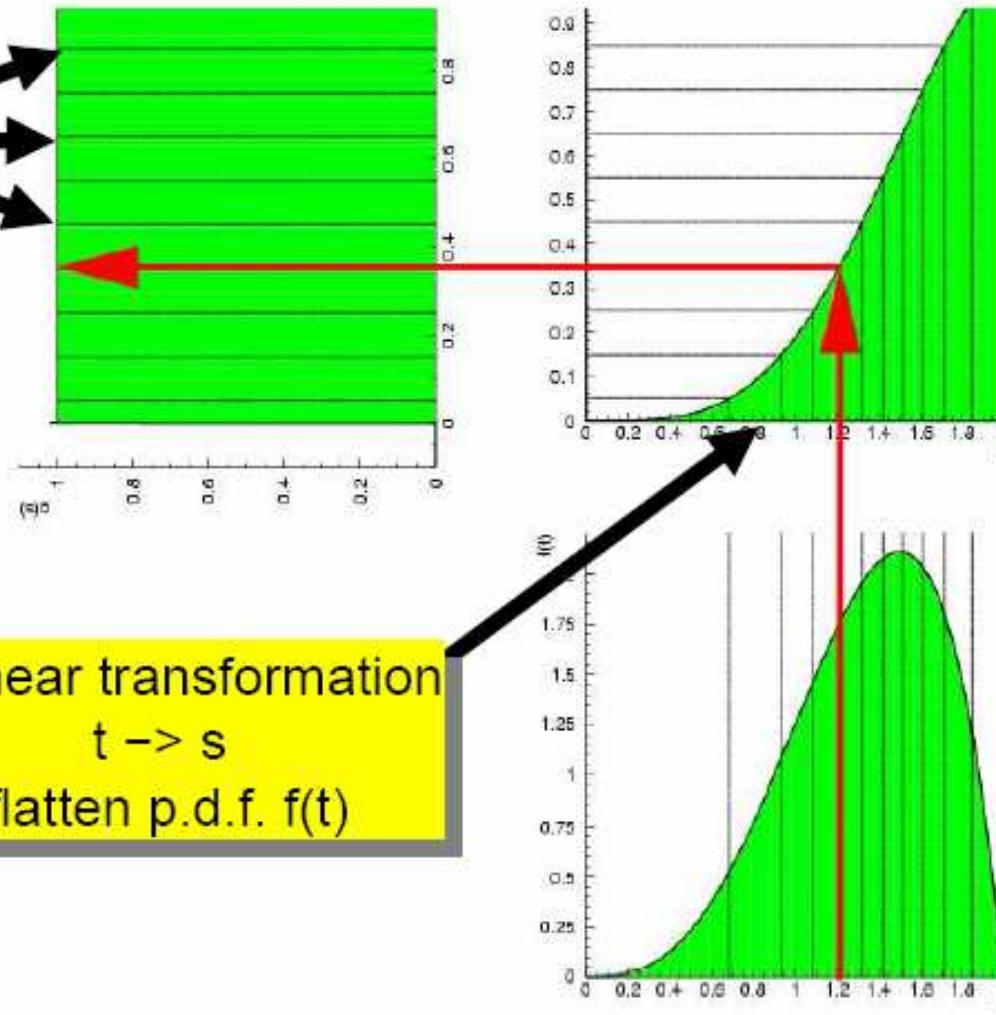
Back transformation to variable  $t$  returns  $f(t|x)$

# Equalisation and discretisation

discretization  
of  $f(t)$   
into  $N$  intervals  
of same area



nonlinear transformation  
 $t \rightarrow s$   
to flatten p.d.f.  $f(t)$



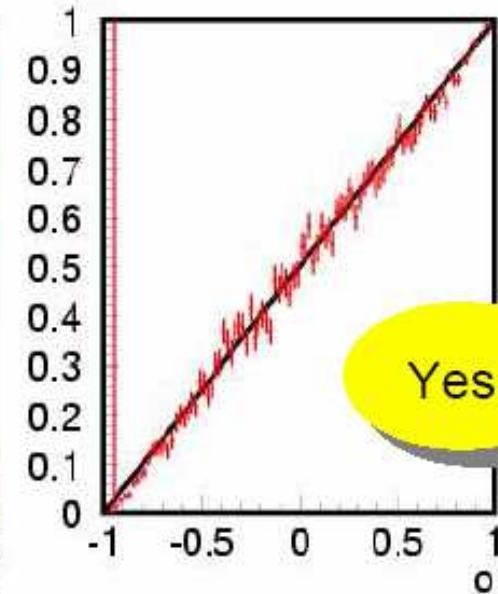
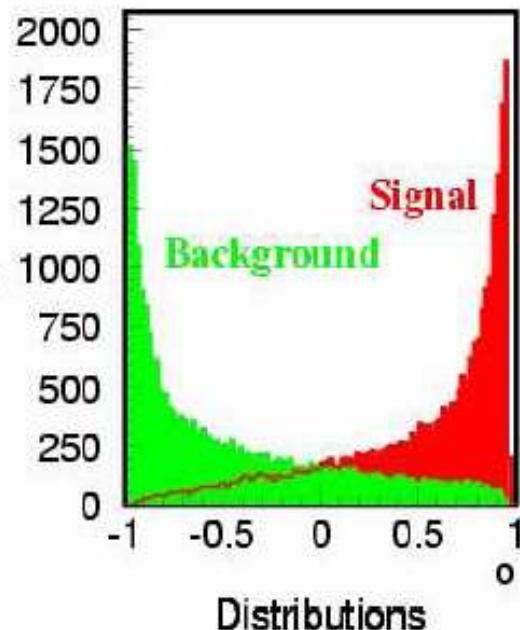
# Theoretical basis

(Under some controllable conditions...)

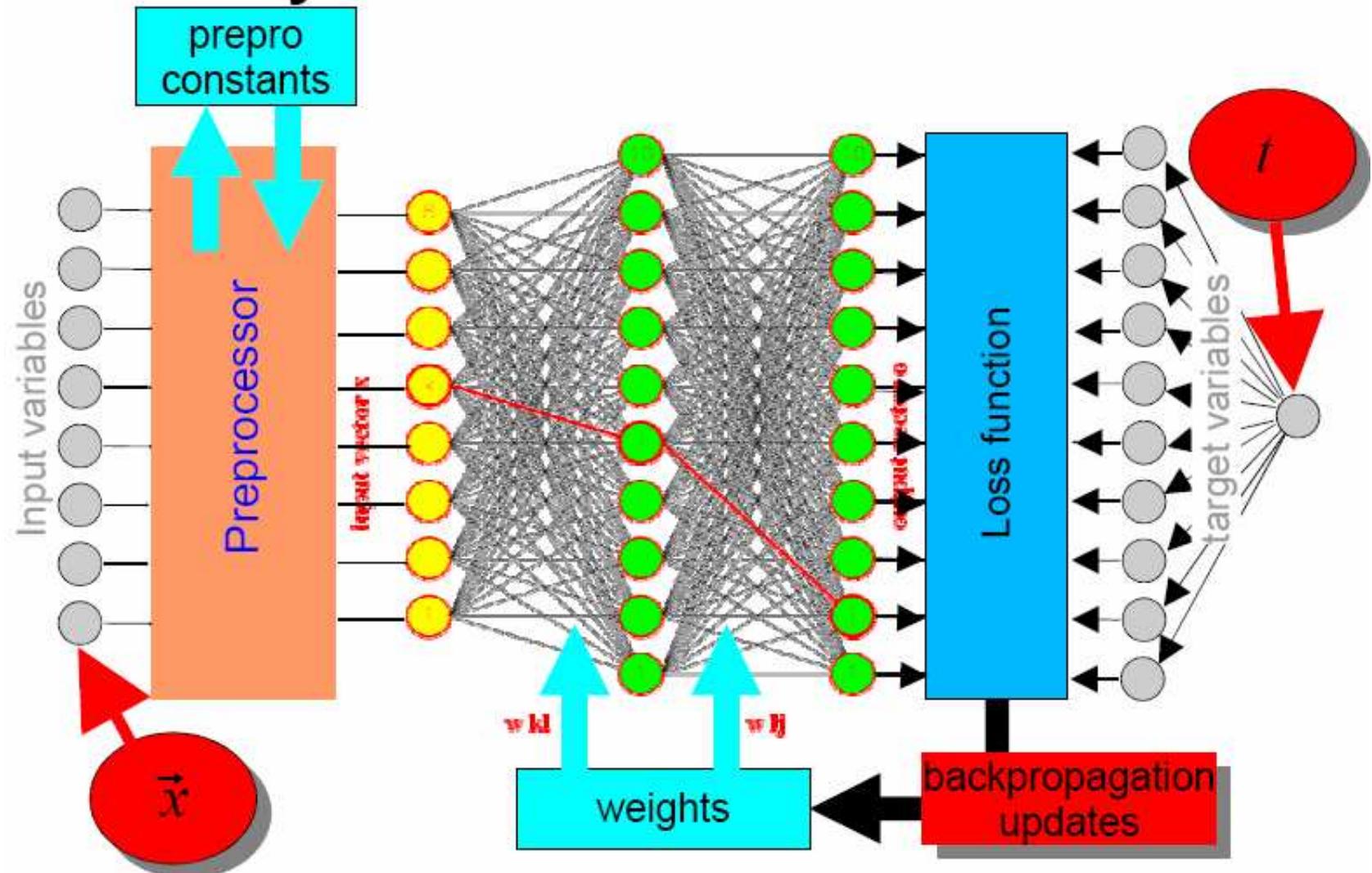
neural network outputs can be interpreted  
as Bayesian a posteriori probability  
that the classification is correct

Purity of a given  
output is  
linear function  
of the  
output value

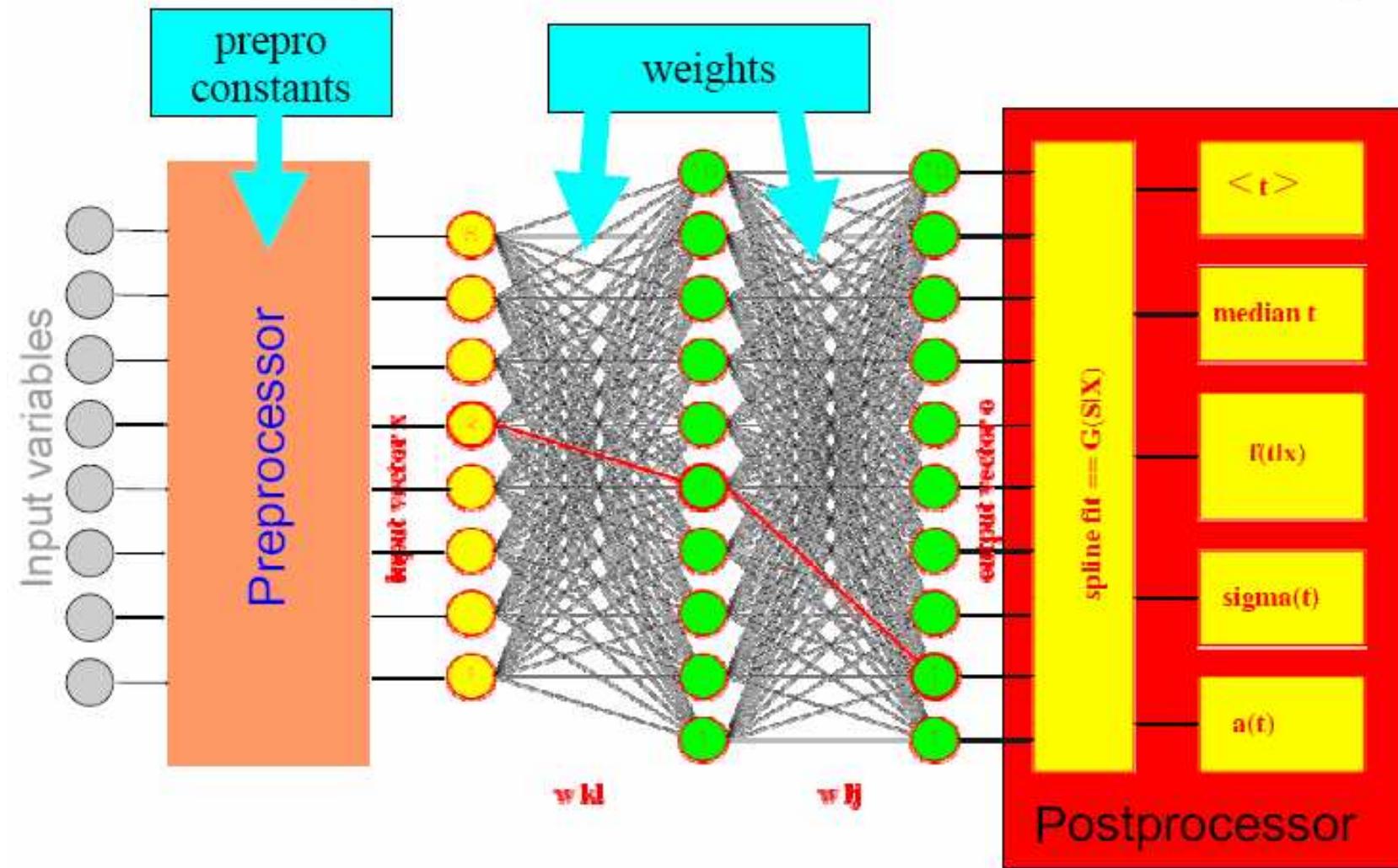
Neural Net Output (node 10 of 20)



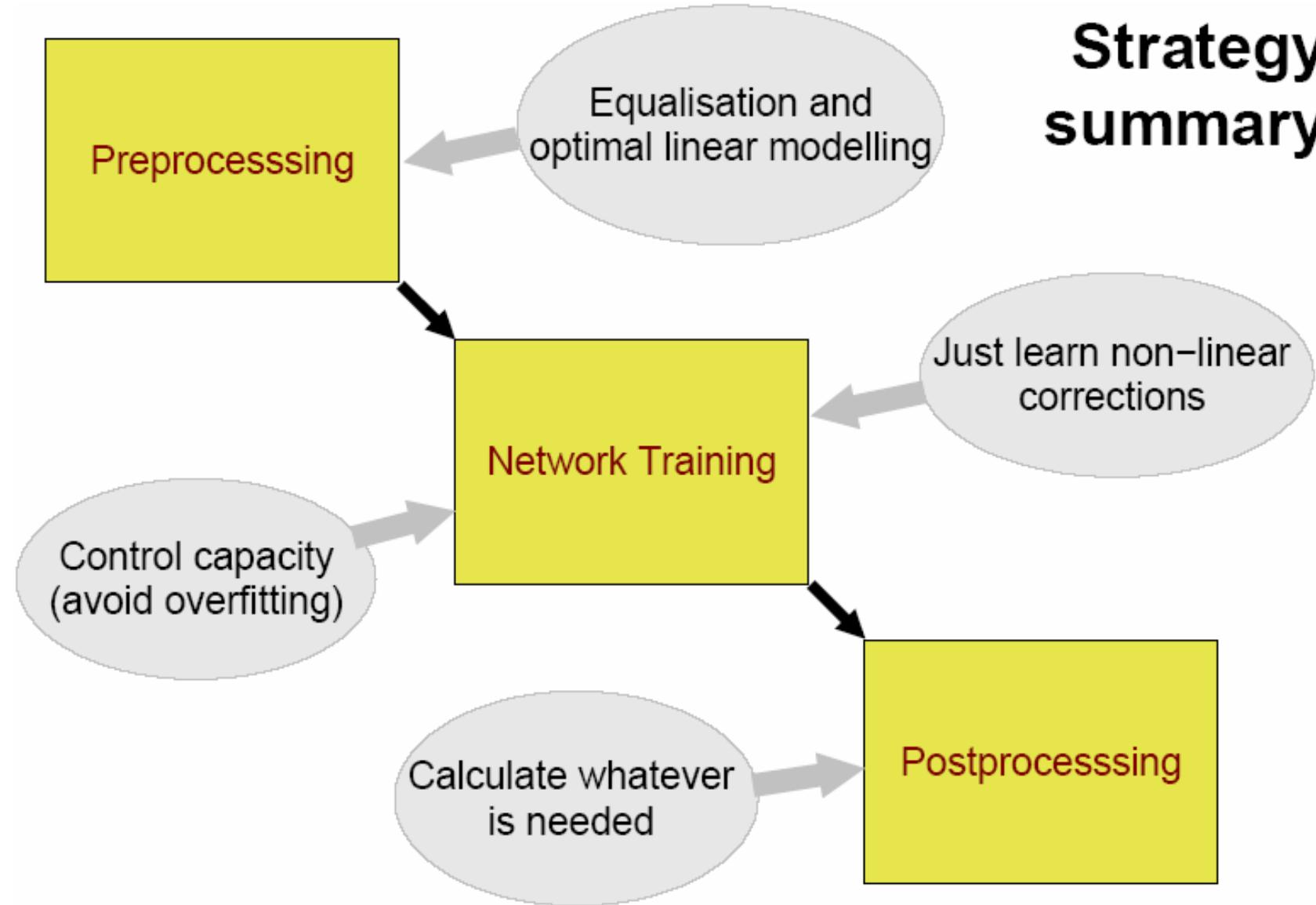
# NeuroBayes Network architecture: Teacher

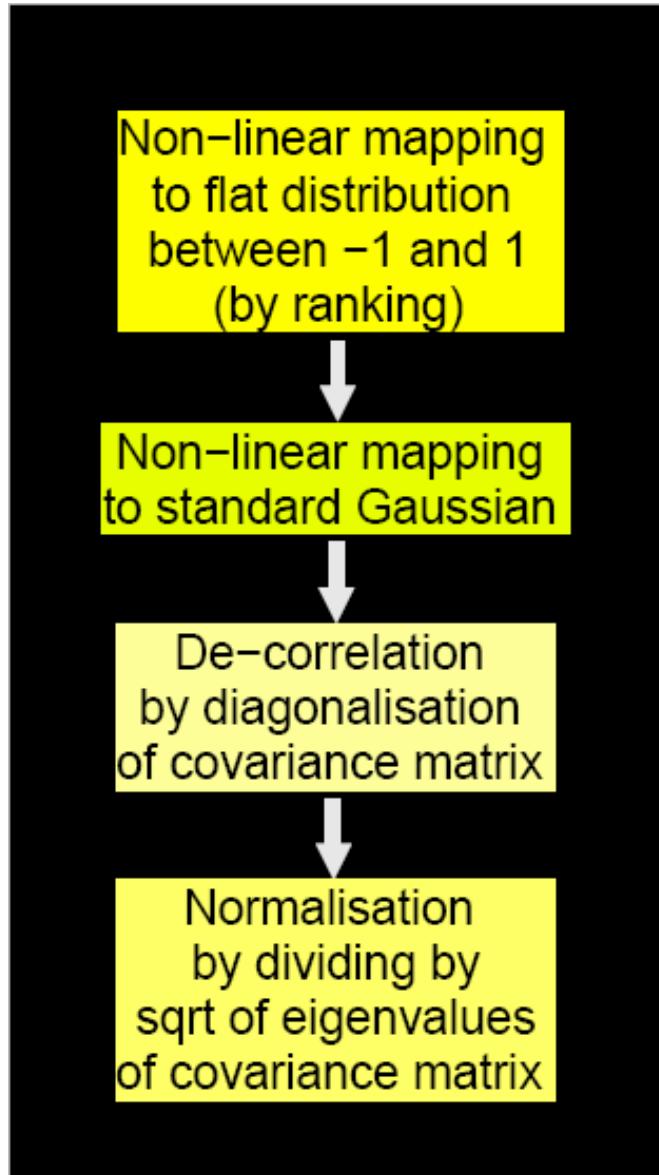


# NeuroBayes network architecture: Expert



## Strategy summary





# Preprocessing of input variables I

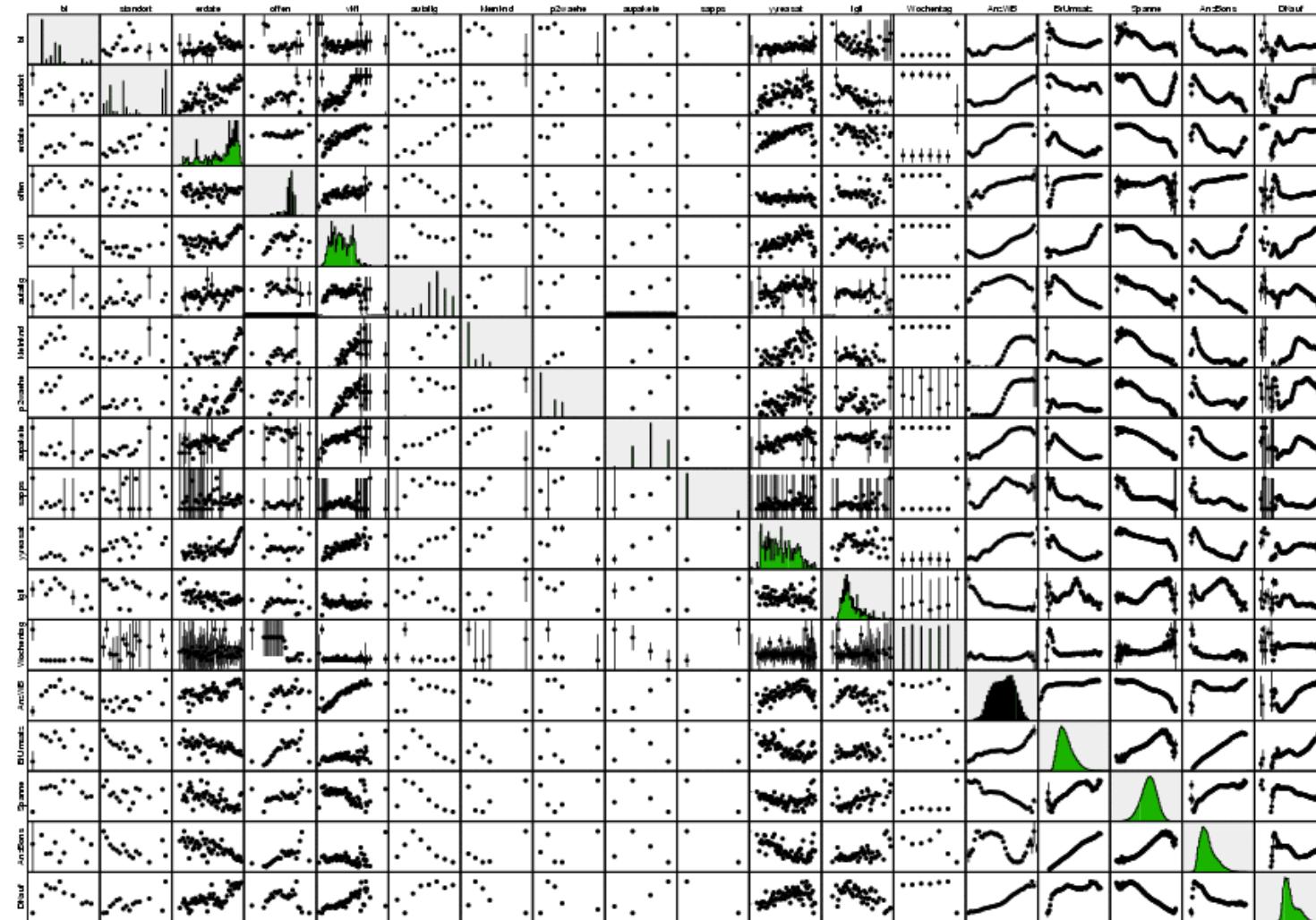
At this stage all input variables are

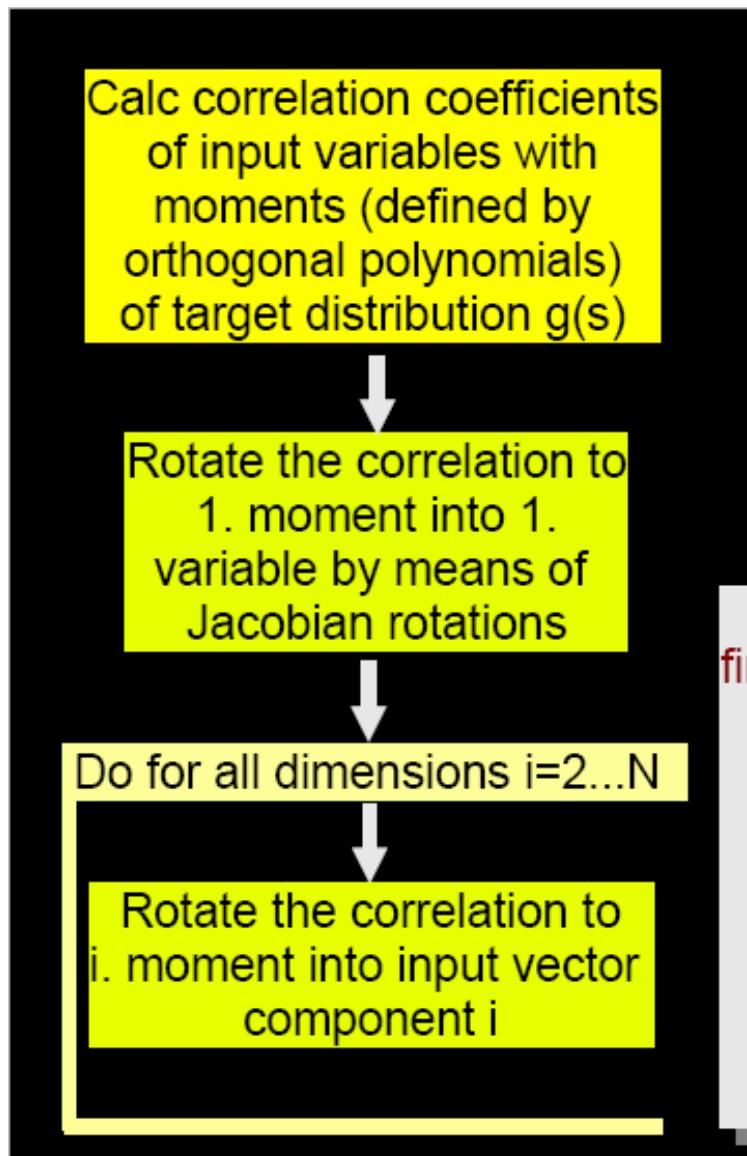
- independent,
- centered at 0
- $\sigma=1$
- Gaussian (if no discrete input variables)

New covariance matrix is unit matrix

completely automatic  
absolutely robust

# Ramler-Plot (erweiterte Korrelationsmatrix)





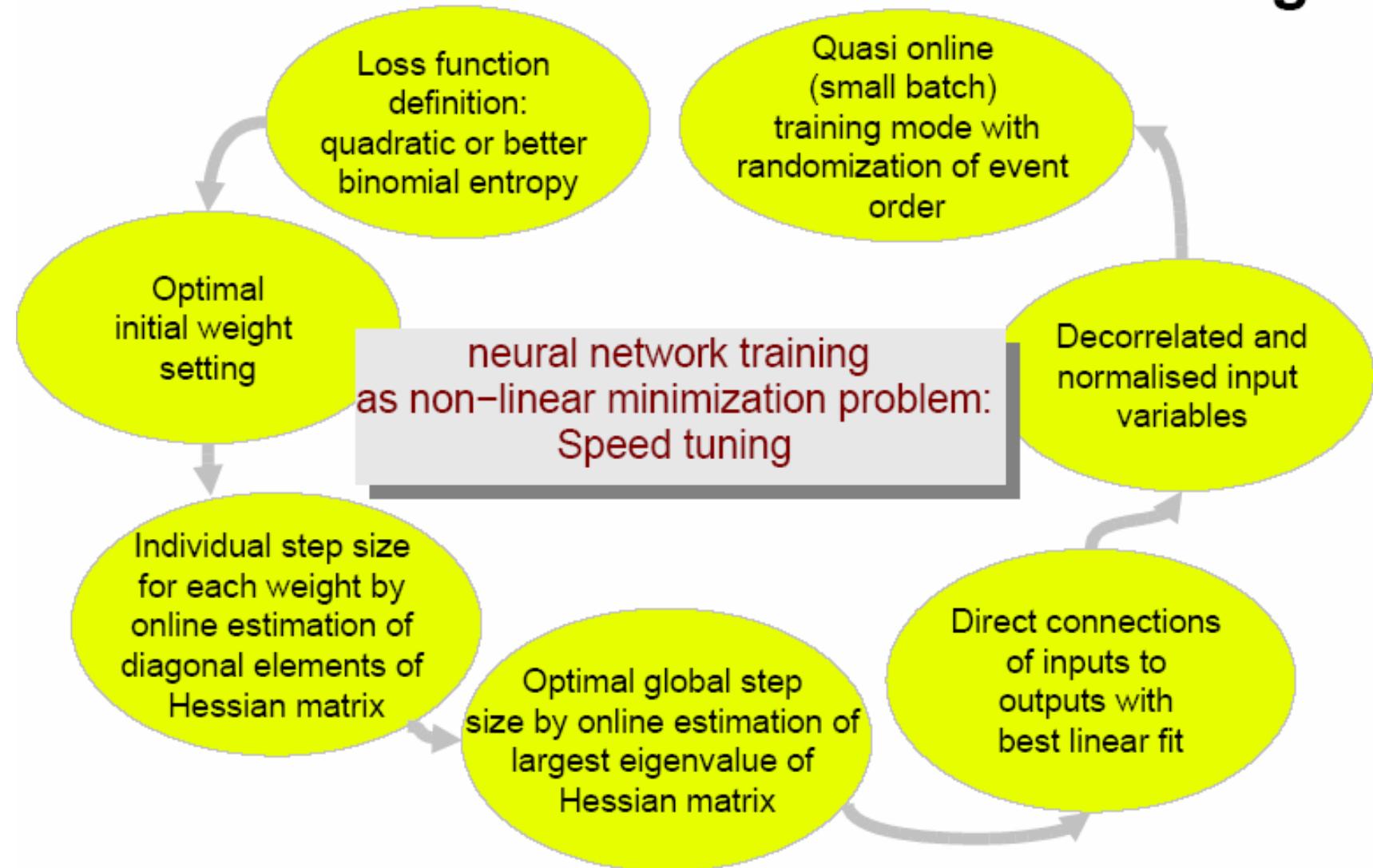
## Preprocessing of input variables II

Degeneracy of new covariance matrix allows arbitrary rotation of  $n$ -dimensional basis

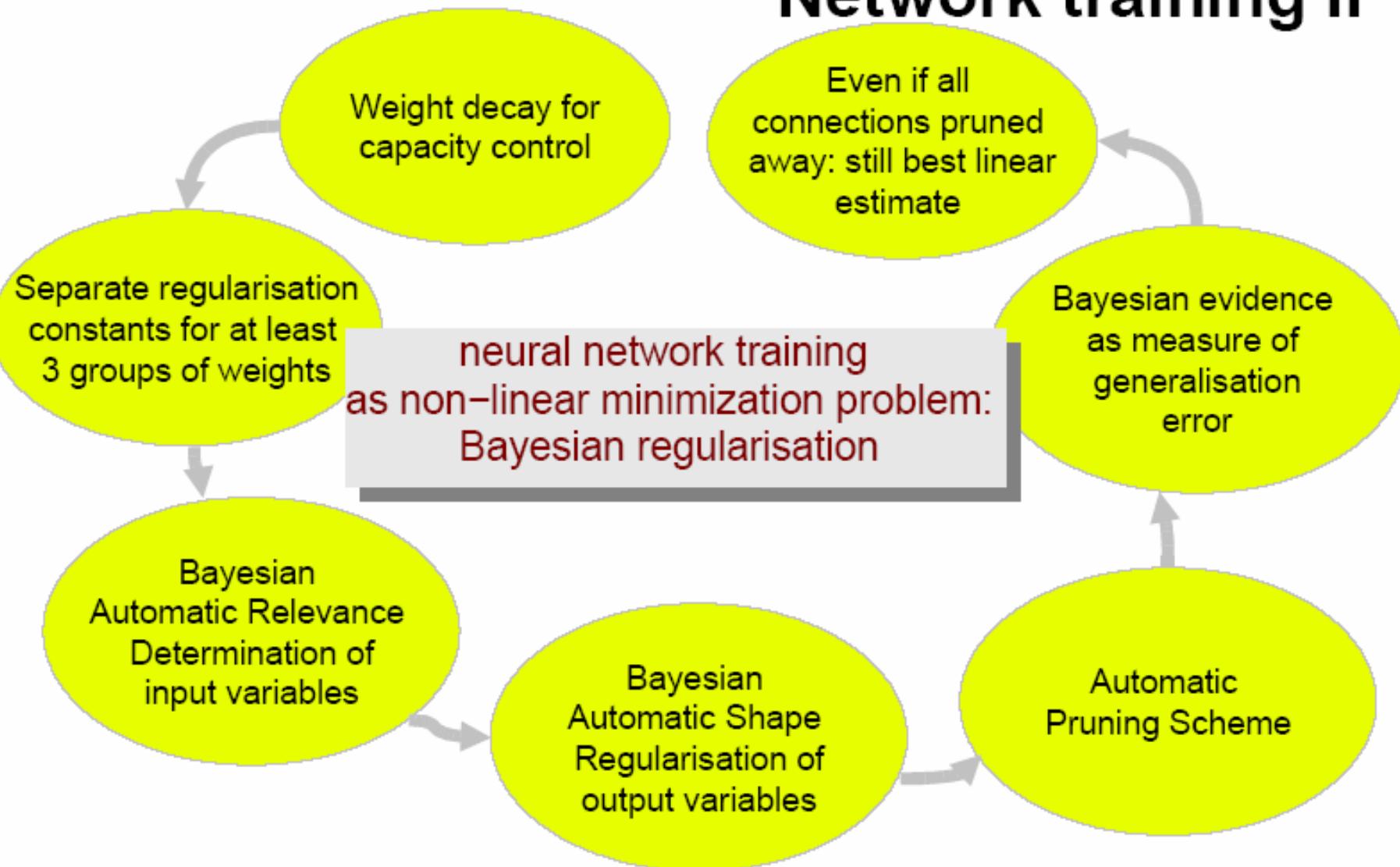
Rotate such that first variable contains all linear information on the mean value, second all linear information on width,  $i$ -th variable on  $i$ -th moment.

The larger  $i$ , the more statistical uncertainty: high frequency oscillations in solution (ill-posed problem)

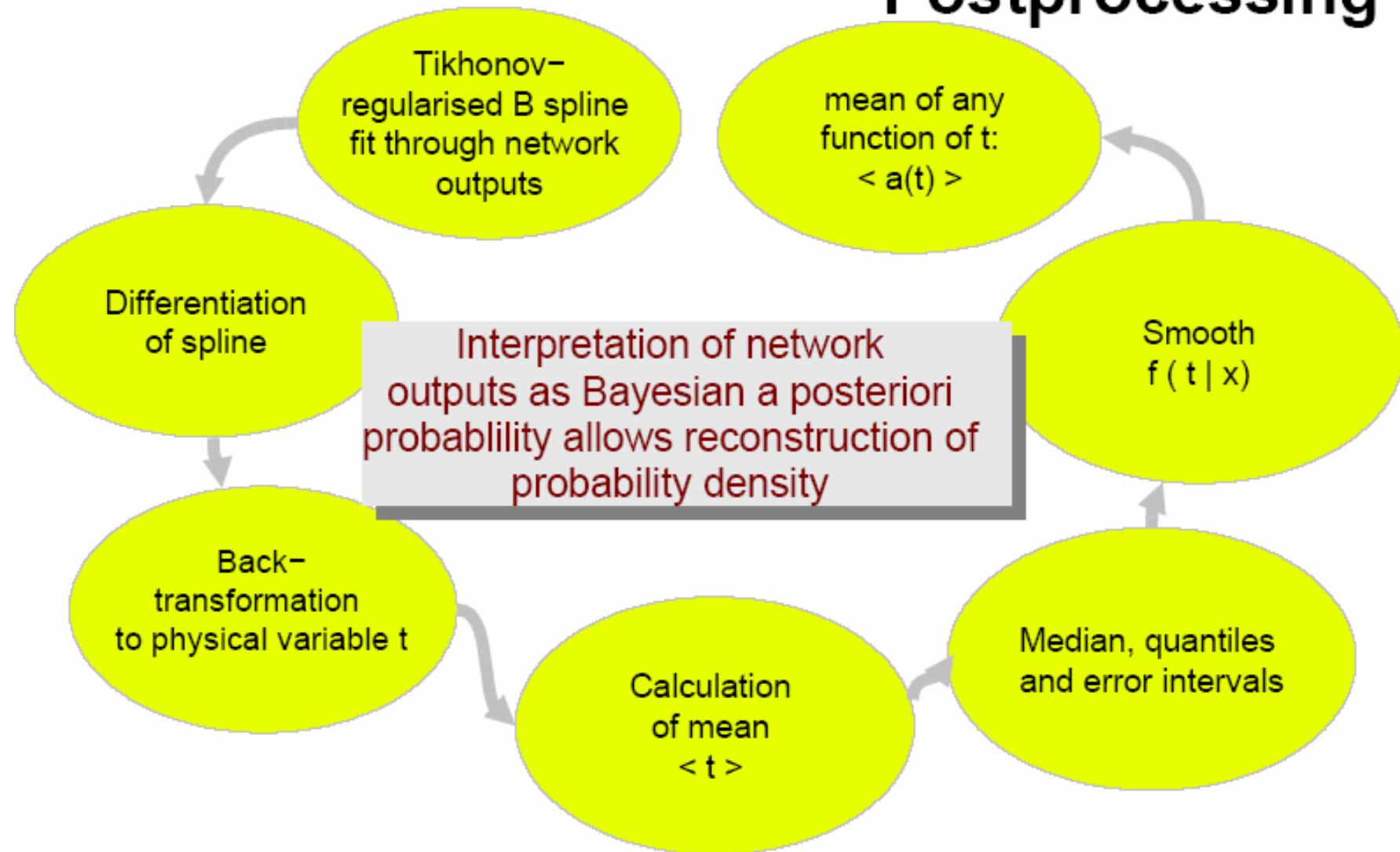
# Network training I

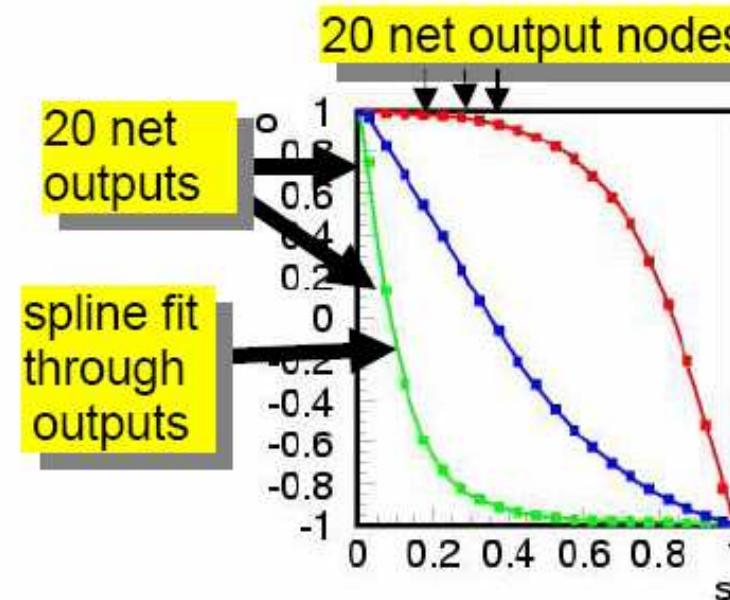


## Network training II

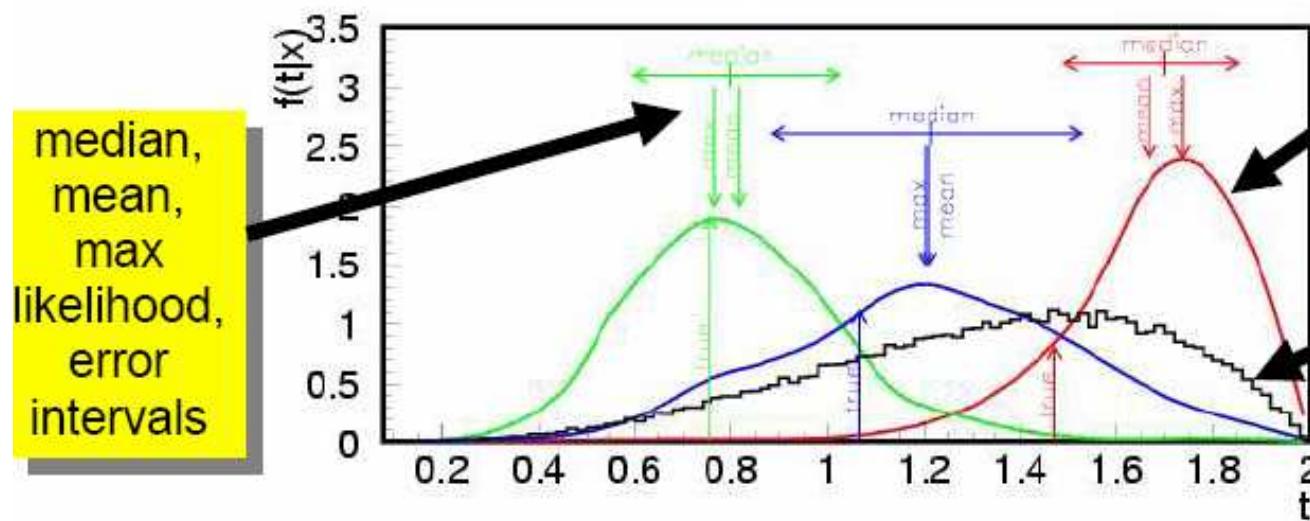
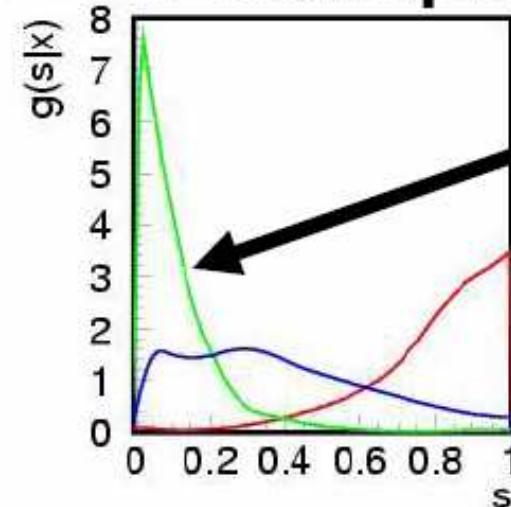


# Postprocessing





## 3 example events



# automatic error propagation

toy experiment :

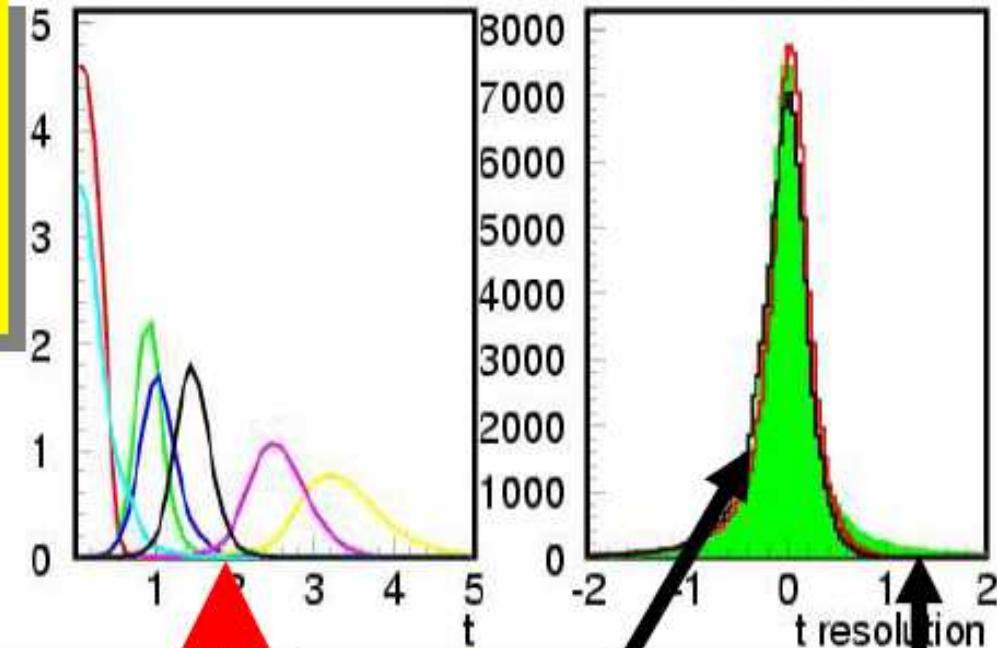
measure (with errors)

- decay length d

- momentum p

and train for proper time t:

$$t = \frac{m}{c} \cdot \frac{d}{p}$$



Result :

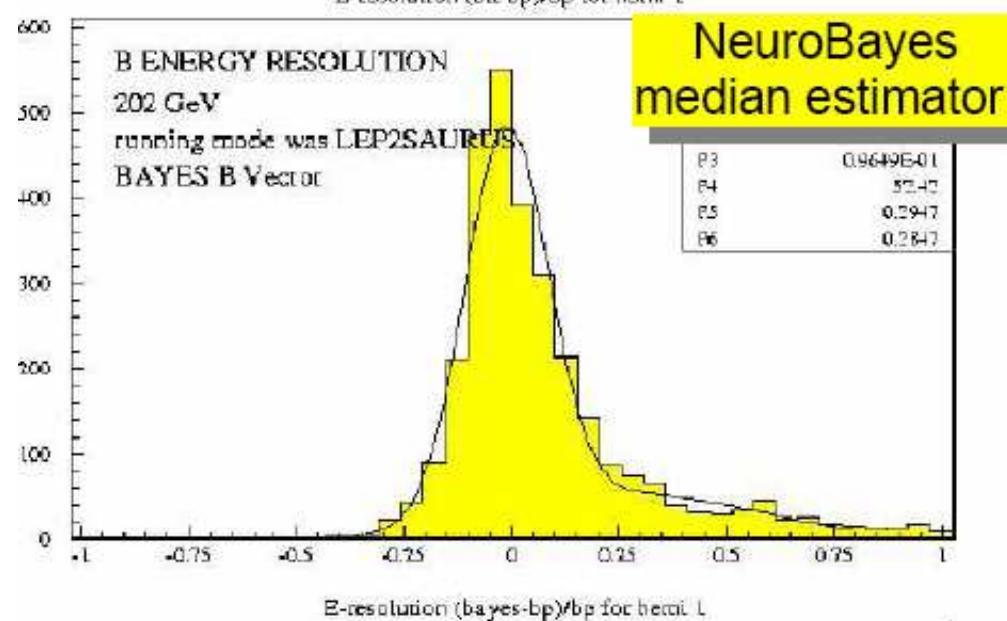
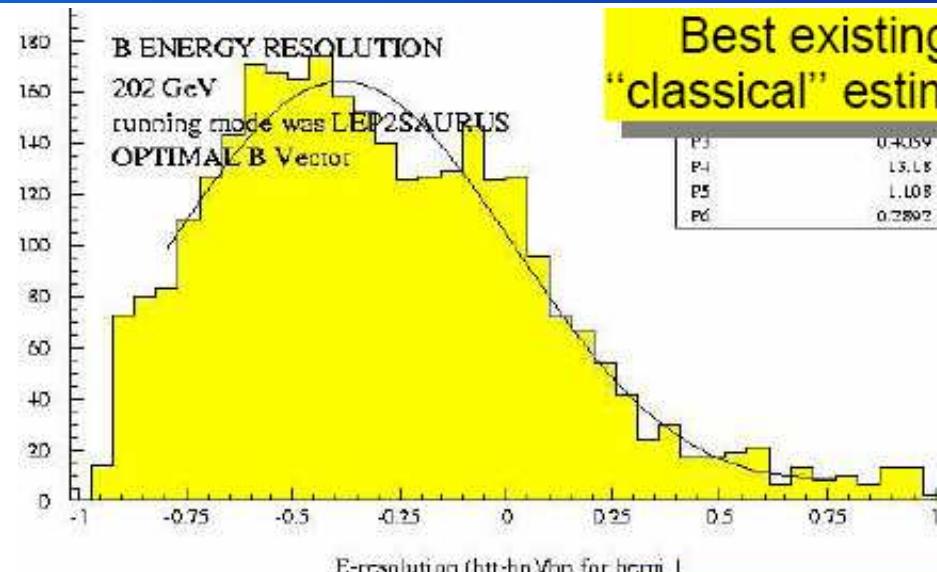
networks learns automatically from data

- that it should divide d by p
- how it should propagate errors
- true lifetimes are never negative (although both measured d and p can be)

Max likelihood estimate  
median estimate

classical approach:  
tail from negative  
lifetimes

# Recent real world application

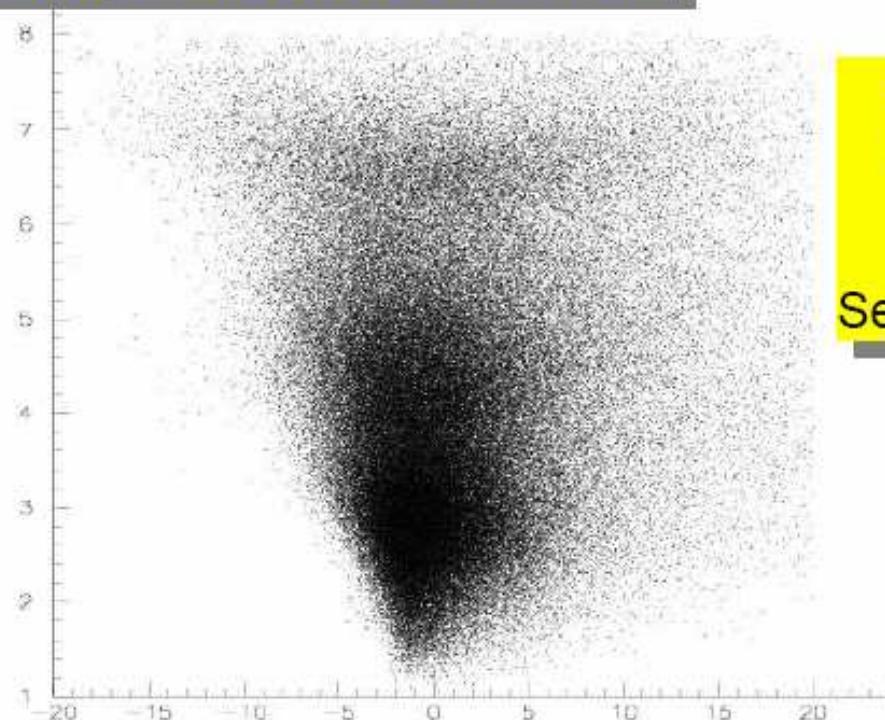


Relative resolution of reconstructed B hadron energy in DELPHI at LEP II at 202 GeV energy  
(completely inclusive)

core resolution  
40% → 10%



## NeuroBayes mean error estimator



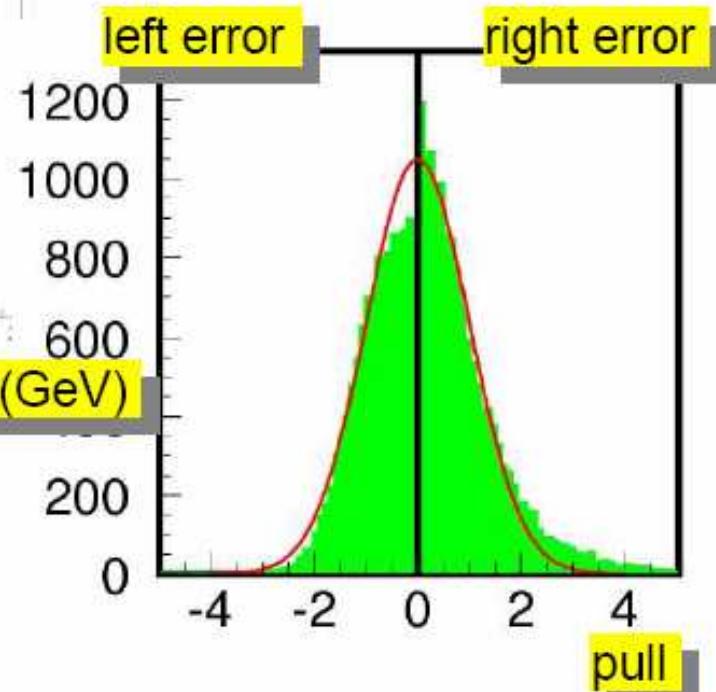
## NeuroBayes median estimator – true energy (GeV)

Resolution of reconstructed  
B hadron energy in DELPHI at LEP I  
(completely inclusive)

## Error estimates

error estimates make sense!  
Pulls are almost Gaussians  
of width 1

Separate left and right uncertainty



# Richtung von B-Mesonen (DELPHI)



# Anwendungen von NeuroBayes in der Physik



## DELPHI:

Auflösungsoptimierung inklusive B-, E, phi, Theta, Q-Wert  
B\*\* Anreicherung  
B Fragmentationsfunktion  
Limit auf B\_s-Mixing  
B0-Mixing  
B- F/B-Asymmetrie  
B-> wrong sign charm

## CDF: (alles Work in Progress)

Elektron ID  
Kaon/Proton ID  
Resonanzoptimierung  
B-Tagging  
F-Flavour Tagging

## Erkenntnis

Diese Methoden sind nicht  
nur in der Physik  
anwendbar

<phi-t>: Ausgründung aus der Universität Karlsruhe,  
gefördert durch das exist-seed-Programm des  
Bundesministeriums für Bildung und Forschung

# Historie

- 2000-2002 NeuroBayes®-Spezialisierung  
für die Wirtschaft in Universität  
Karlsruhe
- Okt. 2002: GmbH gegründet,  
erste bezahlte Aufträge
- Juni 2003: Umzug in neues Büro  
199 qm IT-Portal Karlsruhe

Exklusivrechte für NeuroBayes®  
Juli 2004: Partnerschaftsvertrag mit  
2000-Mann-Firma msg Systems AG

Personal September 2004:  
3 hauptberufliche und  
6 nebenberufliche Mitarbeiter,  
Prof. Beratung z.B. durch Volker Blobel,  
Wirtschaftliche/rechtliche/Marketing- Expertise durch stille Beteiligte,  
Pool freier Mitarbeiter.



# Anwendungen von NeuroBayes in der Wirtschaft

- > Industrielle Forschung
    - z.B. Qualitätsklassifizierung, Fertigungskontrolle
  - > Medizin- und Pharma-Forschung
    - z.B. Wirkungen, Nebenwirkungen, Wechselwirkungen von Medikamenten, Tumorfrüherkennung
  - > Banken
    - z.B. Kredit-Scoring (Basel II), Kursvorhersage, Bewertung von Derivaten, risikominimierte Handelsstrategien, Kundenbewertung, Kündigungsfrüherkennung
  - > Versicherungen
    - z.B. Risikovorhersage, Kündigungswahrscheinlichkeit
  - > Handelsketten: Umsatzprognosen
- Voraussetzung:  
es müssen (historische oder simulierte) Daten vorhanden sein.

# NeuroBayes® Klassifikationen

**Klassifikationen:** Ja/Nein-Entscheidungen.  
Ausgabewert ist die Wahrscheinlichkeit, dass die Entscheidung richtig ist.

**Beispiele:**

- > Dieses Elementarteilchen ist ein K-Meson.
- > Deutschland wird bei der WM 2006 Weltmeister.
  
- > Kunde Meier wird innerhalb des nächsten Jahres kündigen.
- > Kunde Schmidt wird innerhalb des nächsten Jahres einen Schaden melden.
- > Der Ölpreis steigt.
- > Der Patient wird bei Einnahme des Medikaments die Nebenwirkung erleiden

# NeuroBayes®

## Wahrscheinlichkeitsdichten

### Wahrscheinlichkeitsdichte:

Für jeden möglichen Wert wird eine Wahrscheinlichkeit angegeben. Daraus können Mittelwert, Median, Modus, Standardabweichung, Percentile etc. abgeleitet werden.

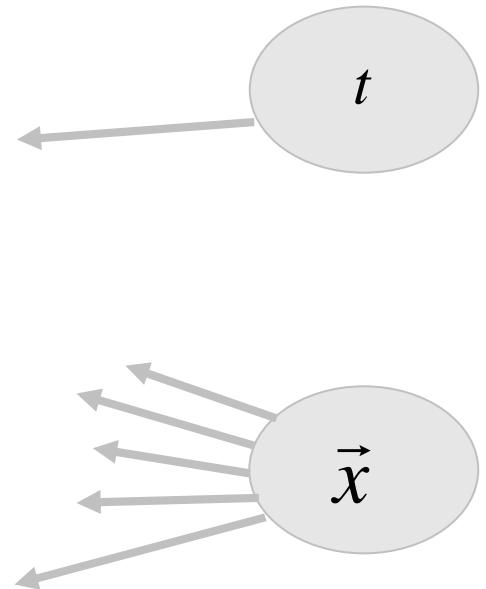
### Beispiele:

- > Energie eines Elementarteilchens
- > Änderung des Preises einer Aktie oder Option
- > Umsatz einer Filiale in der Zukunft
- > Lebenserwartung
- > Schadenssumme eines bestimmten Versicherungsnehmers

# Bedingte Wahrscheinlichkeitsdichten in der Teilchenphysik

Was ist die Wahrscheinlichkeitsdichteverteilung von der wahren B-Meson-Energie  $E$  in diesem Ereignis, das mit dem DELPHI-Detektor bei LEP II bei dieser Strahlenergie, mit dieser effektiven Schwerpunktenergie  $s'$ , diesen  $n$  Spuren mit jenen Impulsen und Rapiditäten in der Hemisphäre, die diesen Vertex bilden, mit dieser Zerfallslänge und Wahrscheinlichkeit, dieser Anzahl nicht gut rekonstruierter Spuren etc. pp. aufgenommen wurde?

$$f(t | \vec{x})$$



# Bedingte Wahrscheinlichkeitsdichten im Investment-Banking

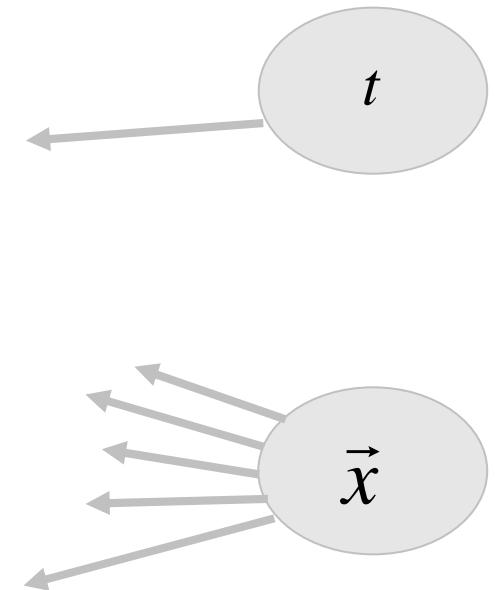
Was ist die WahrscheinlichkeitsdichteVerteilung  
für die Preisänderung in 10 Tagen, der Aktie der  
Firma A AG...

die in den letzten Tagen und Wochen diese und  
jene Kursbewegung vollführt hat...

soundsoviel teurer ist als das n-Tage moving  
average...

aber soundsoviel Prozent unter dem absoluten  
Maximalpreis liegt...

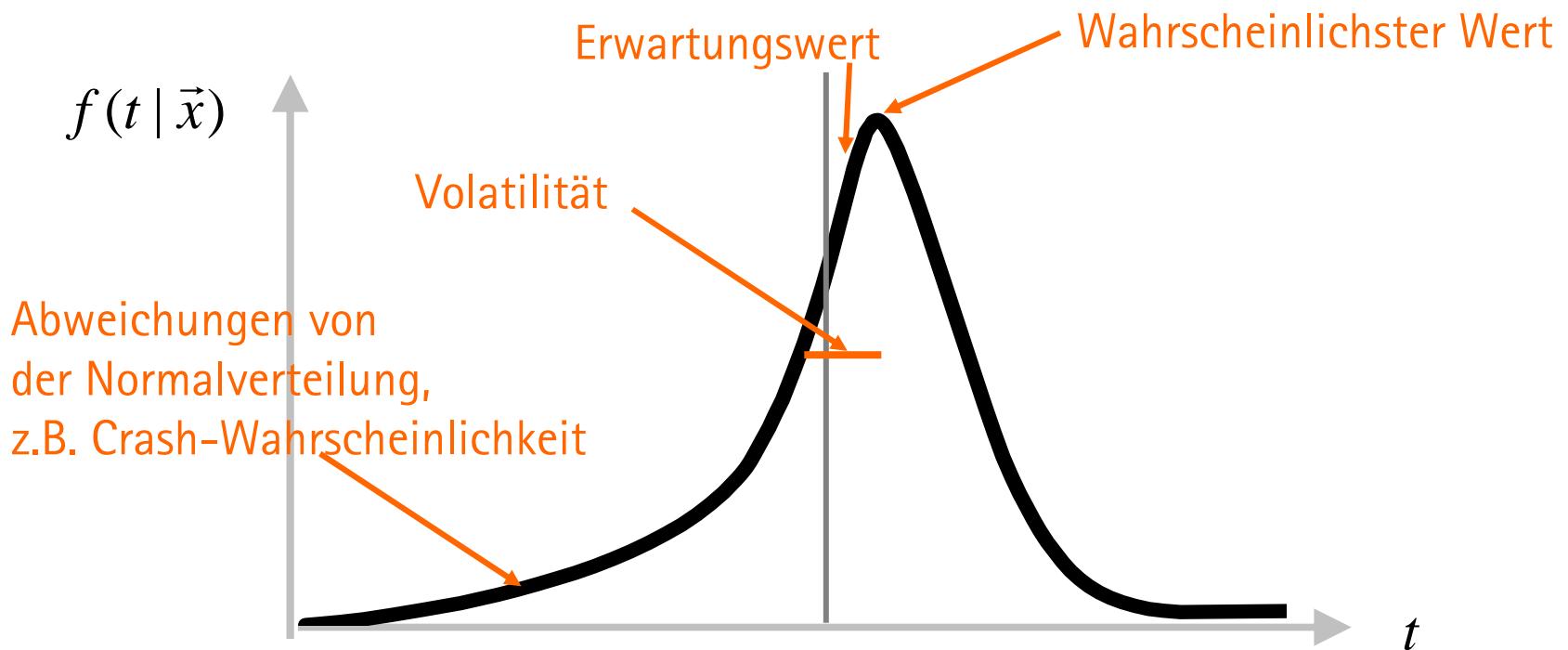
diese Korrelation die den Index...und dem Ölpreis  
aufweist..., der wiederum....etc pp...



$$f(t | \vec{x})$$

# Vorhersage der kompletten Wahrscheinlichkeitsverteilung

mit dem  $\langle \phi-t \rangle$  NeuroBayes neuronalen Netzwerk



## <phi-t>finance Optionspreis

- > Durch Vorhersage der kompletten Wahrscheinlichkeitsdichte ist eine Berechnung von fairen Optionspreisen möglich.
- > Besser als mit Nobelpreis ausgezeichnete Black-Scholes Formel, weil keine Annahmen über Normalverteilung nötig sind.
- > Nicht nur eine Markttendenz, sondern auch die Volatilität wird prognostiziert.
- > Wahrscheinlichkeit und Ausmaß eines Marktcrashes so wie in historischen Daten.

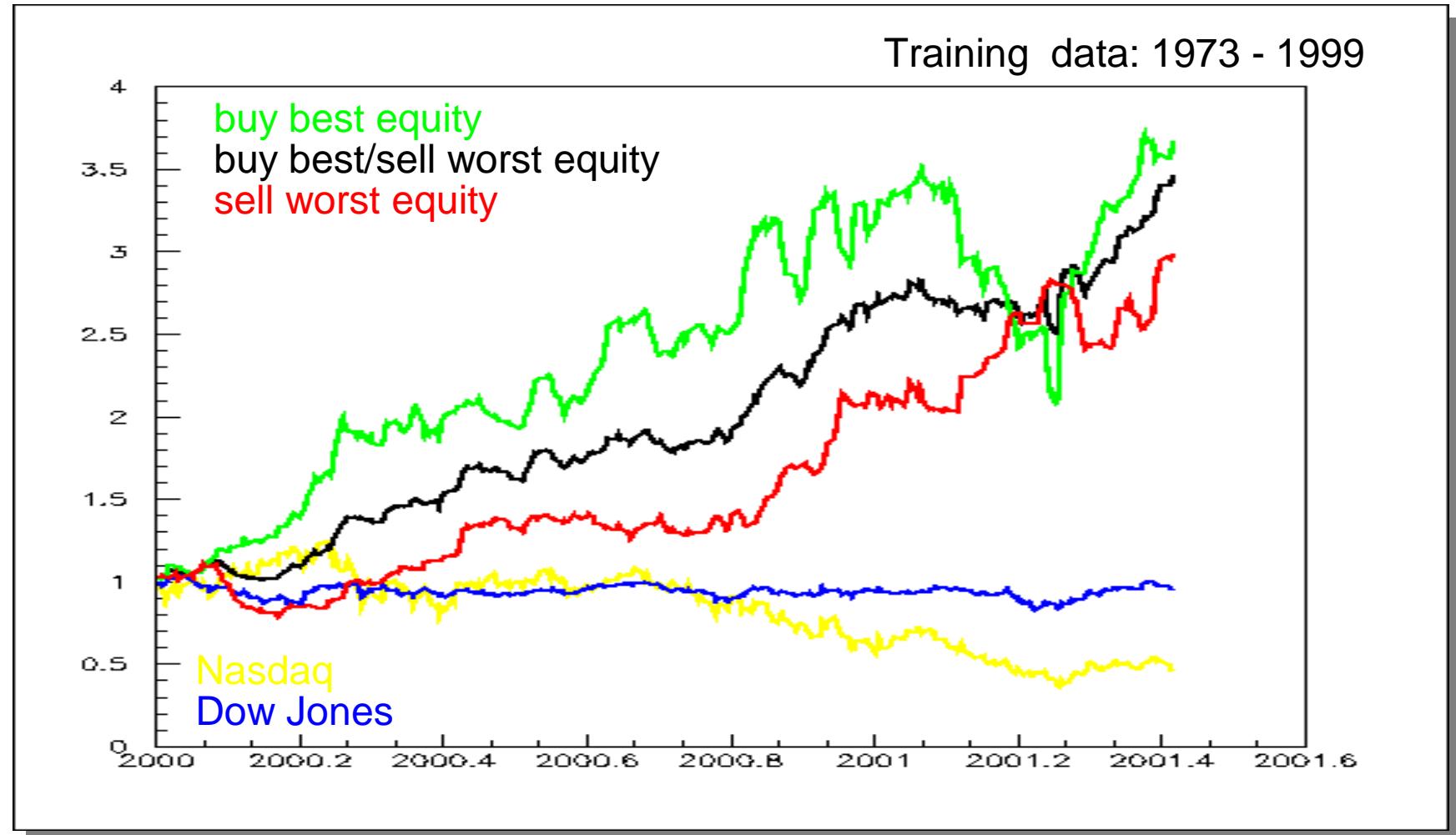
# Input-Variablen

Technische Indikatoren aus der Literatur  
(RSI, moving averages, MACD, Wilders Volatilität etc.),  
deren Relevanz durch eigene statistische Untersuchungen belegt  
wurde.

Neu entwickelte mathematische Kombinationen  
von Kursverläufen und Volumina, basierend auf

- Fouriertransformationen
- Wavelet-Entwicklungen
- Maximum Entropie-Filter
- Perzeptrons
- neuronalen Netzwerken
- Bayes'scher Statistik

# Aktienkursvorhersage



# Ameritrade Börsenspiel

Netscape: Investors Cup

File Edit View Go Communicator Help

- Transactions  
Player Performance  
- Top 20 Weekly  
...Gainers  
- Top 20 Daily  
...Gainers  
- Relative  
...Performance  
Education  
- Education Center  
Personal Finance 

Contact Us  
Register  
Official Rules

Your Portfolio at a Glance  
Net Account Value  
\$61,304.84  
Buying Power  
\$11,736.09  
State Rank Rank  
1 356

  
25 FREE TRADES OPEN YOUR ACCOUNT NOW! GO!

**Ameritrade Investors Cup**

Rank	Name	Percentile Rank	Account Value	Percentage Gain
1	JEN	100.00%	\$124,905.36	149.81%
2	fatcat1	99.99%	\$113,448.24	126.90%
3	BigAI	99.99%	\$101,452.42	102.90%
4	bags	99.98%	\$101,019.72	102.04%
5	bowlingout	99.97%	\$100,638.61	101.28%
351	3base4eva	97.63%	\$61,445.48	22.89%
352	Josevi	97.63%	\$61,392.65	22.79%
353	AlainFlores	97.62%	\$61,375.48	22.75%
354	Illwah	97.61%	\$61,346.37	22.69%
355	KaoLee	97.61%	\$61,320.34	22.64%
356	phi-t	97.60%	\$61,304.84	22.61%
357	Buffdaddy17	97.59%	\$61,297.63	22.60%
358	pishi	97.58%	\$61,296.77	22.59%
359	idwise	97.58%	\$61,287.46	22.57%
360	Unclebubba63	97.57%	\$61,285.48	22.57%
361	ADFUTURES	97.56%	\$61,270.95	22.54%
14778	Steve007	0.03%	\$11,024.97	-77.95%
14779	kamil z. madi	0.03%	\$10,348.72	-79.30%
14780	Buster Jigs	0.02%	\$8,736.80	-82.53%
14781	ZIO	0.01%	\$5,894.09	-88.21%
14782	bongowally	0.01%	\$5,566.66	-88.87%

SnapQuote   For Options, use "OP"(+BBMAT). For an Index, use "SI"(\$INDU).  
Quotes provided by S&P Comstock.



Jan-Mar 2002:  
Performance +22.61%  
Platz 356 von 14782  
(beste 2.5%)

# Ergebnisse des Projekts Risiko-Analyse für BGV



Resultate für die Badischen Gemeinde-Versicherungen:

seit Mai 2003: **radikal neuer Tarif für junge Fahrer !**

Neue Variablen zur Prämienberechnung  
hinzugefügt:

-Fahrzeugtyp

-Fahrzeugalter

viele neue Erkenntnisse

u.a. Risiko und Prämie teilweise um Faktor **3** auseinander!

seit Juli 2003: **Kooperationsvertrag zw. <phi-t> und BGV**

# Stornovorhersage:

Stornowahrscheinlichkeit für Versicherungsverträge

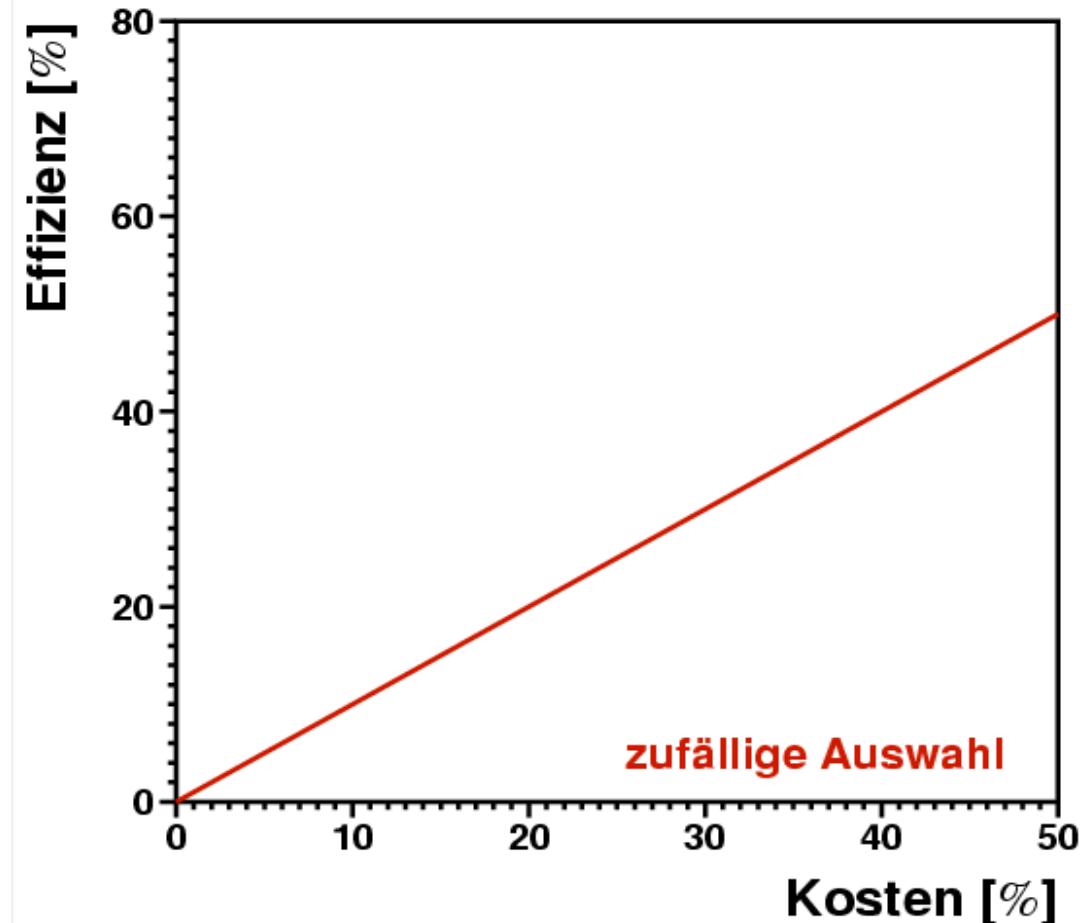
## Problem und Zielsetzung:

1. Hohe Stornorate führt zu Umsatzrückgang und Verwaltungsaufwand.
2. Früherkennung von potenziellen Kündigern erlaubt gezielte Kundenbindungsmaßnahmen.
3. Vorhersage der individuellen Kündigungswahrscheinlichkeit für jeden Kunden aus allen bekannten Informationen ist mit unseren modernen Methoden erstaunlich gut möglich!

# Vertrags – Storno einer Individual-Unfallversicherung

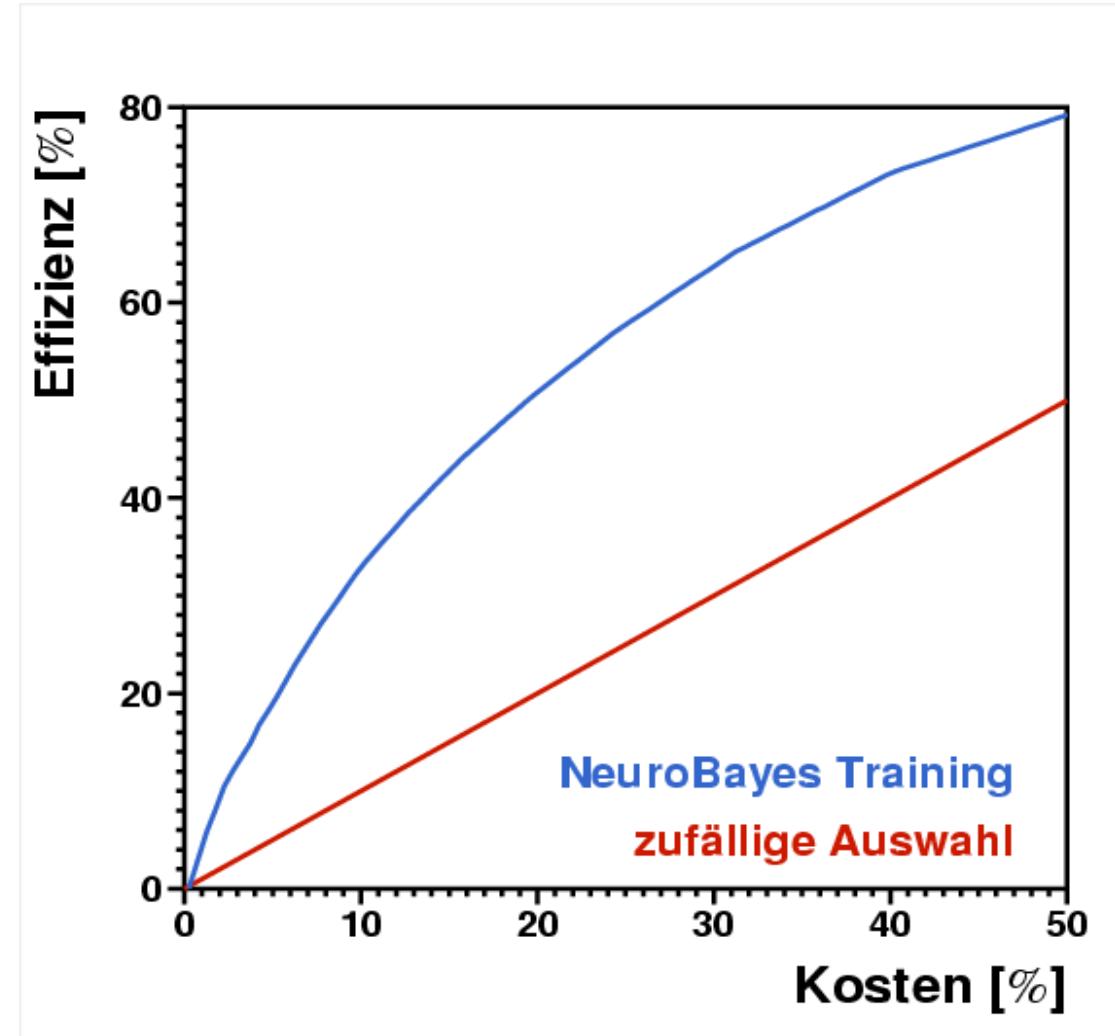
Effizienz:  
Prozentsatz der  
selektierten an allen  
Kündigern

Kosten:  
Prozentsatz der  
selektierten an allen  
Kunden



# Vertrags – Storno

Mit NeuroBayes®  
Können Kunden  
mit erhöhtem  
Kündigungs-  
potenzial  
Identifiziert  
werden

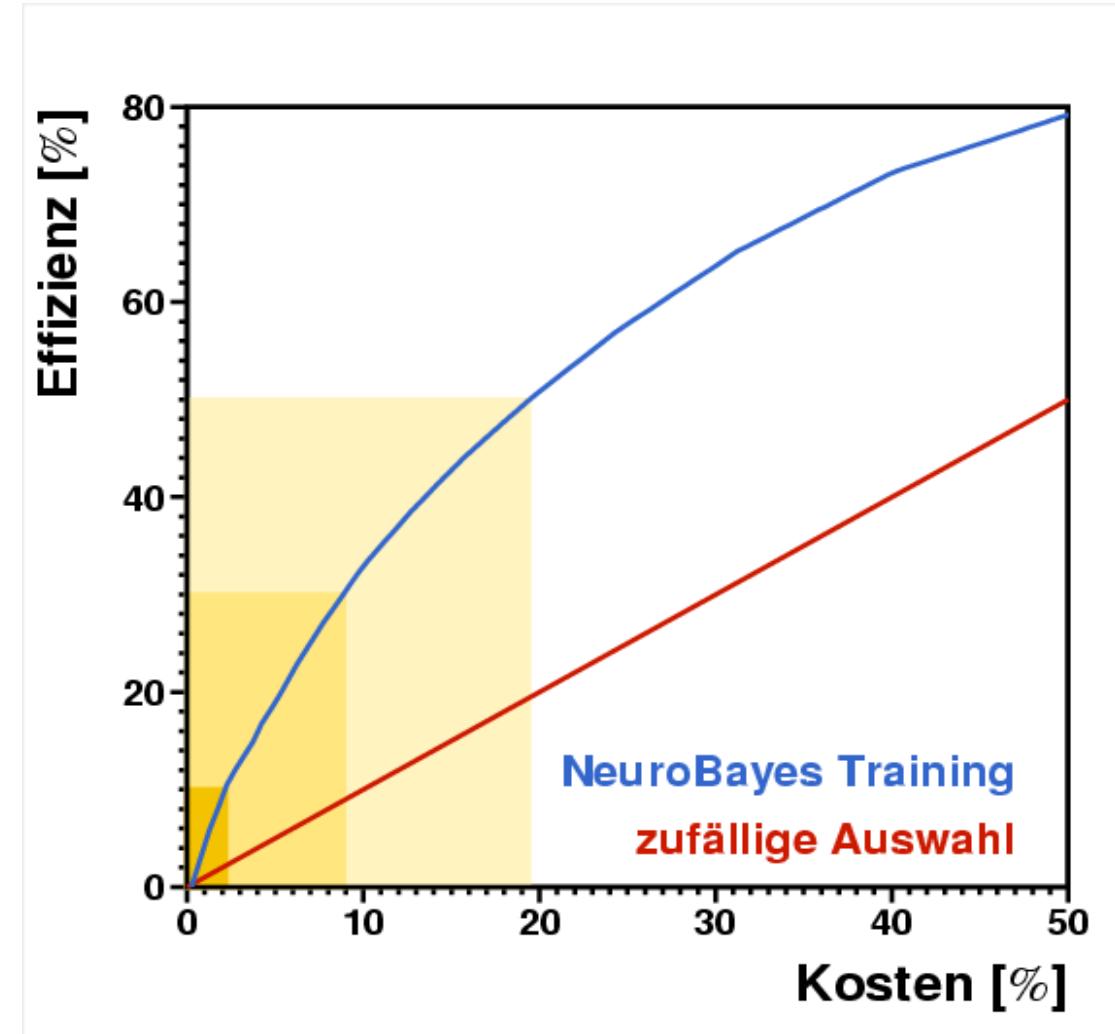


# Vertrags – Storno

Wir können 10% aller Kündiger in 2,2% aller Kunden identifizieren!

Oder 30% in 9%.

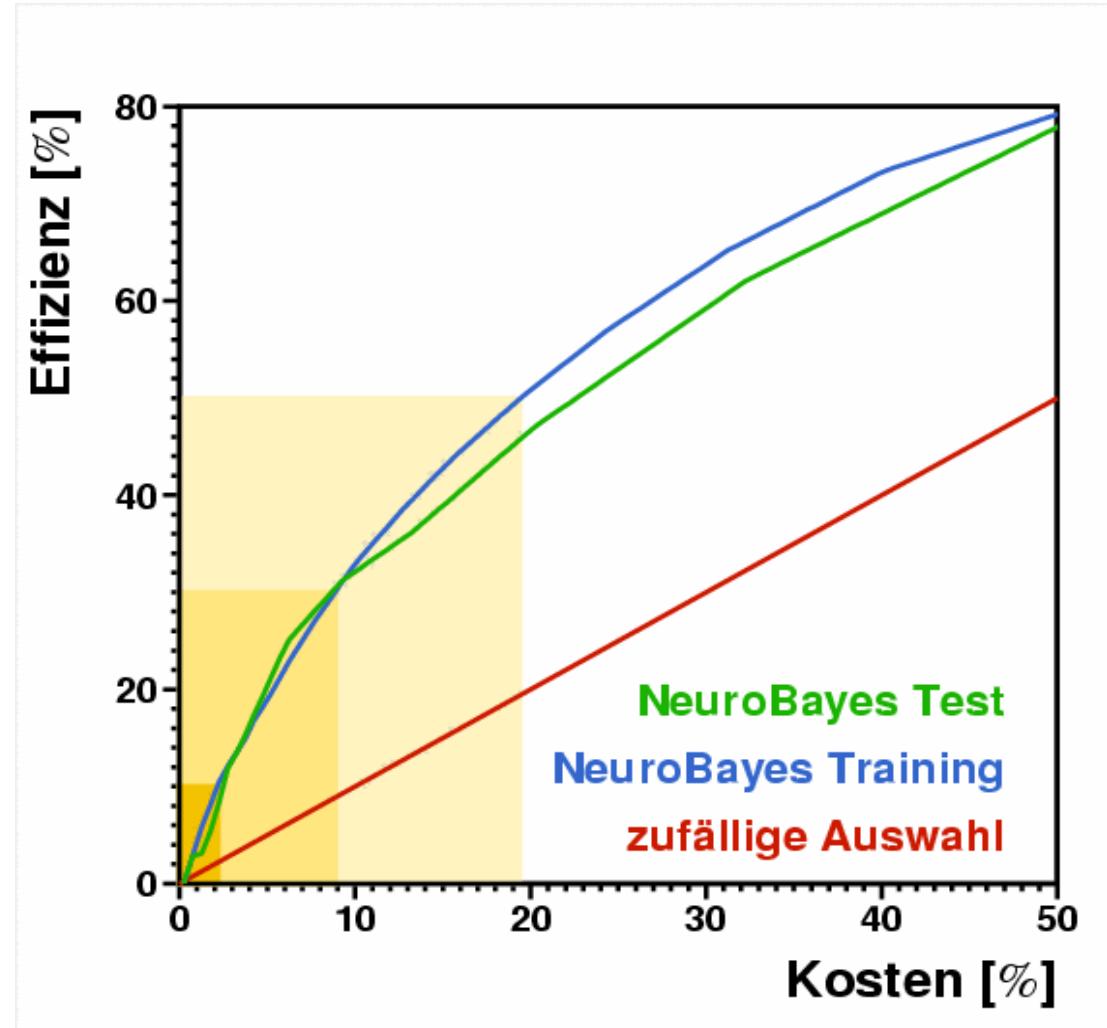
Oder 50% in 19%.



# Vertrags – Storno

Und unsere  
Vorhersage  
trifft auch ein:

Test auf dem  
Training unbekannten  
Daten eines anderen  
Jahrgangs



# Ergebnisse des Projekts Risiko-Analyse für BGV



Resultate für die Badischen Gemeinde-Versicherungen:

seit Mai 2003: **radikal neuer Tarif für junge Fahrer !**

viele neue Erkenntnisse  
u.a. Risiko und Prämie teilweise um Faktor **3** auseinander!

seit Juli 2003: **Kooperationsvertrag zw. <phi-t> und BGV**

Risikoprognose  
Schadenhöhenprognose Haftpflicht, Kasko  
Kündigungsprognose

# Mögliche Eingabedaten am Beispiel einer Kfz-Versicherung

## persönliche Daten:

Alter, Geschlecht, Beruf, Wohnort, jährliche Kilometerleistung, Benutzungsbeschränkungen, Mitfahrer

## Fahrzeugdaten:

Fahrzeugtyp, Motorleistung, Baujahr, Farbe, Garagenwagen, Tuning-Extras,...

## Versicherungsdaten:

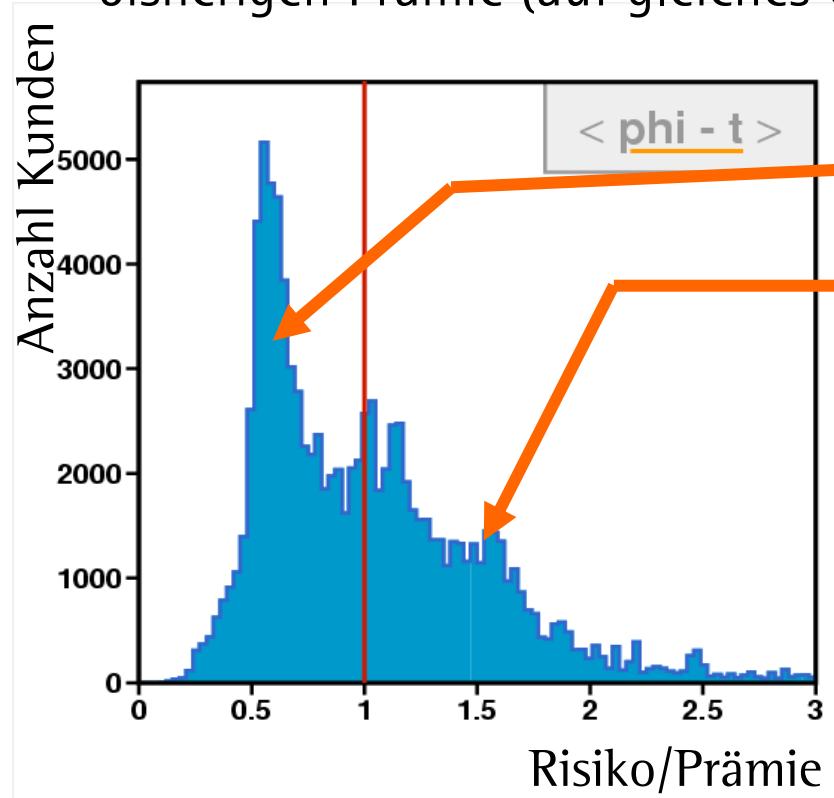
letzte Schadenssumme, Zeit seit letzter Schadensmeldung, Schadenfreiheitsrabatt, Art der letzten Versicherung (Haftpflicht, Kasko etc.)

## Vorhandene Expertenmodelle:

Prämie nach bisherigem Modell, Prämie nach Alternativmodell

# Die „Ungerechtigkeit“ der Prämie

Verhältnis des mit NeuroBayes® errechneten Risikos zur bisherigen Prämie (auf gleiches Gesamtprämienvolumen normiert):



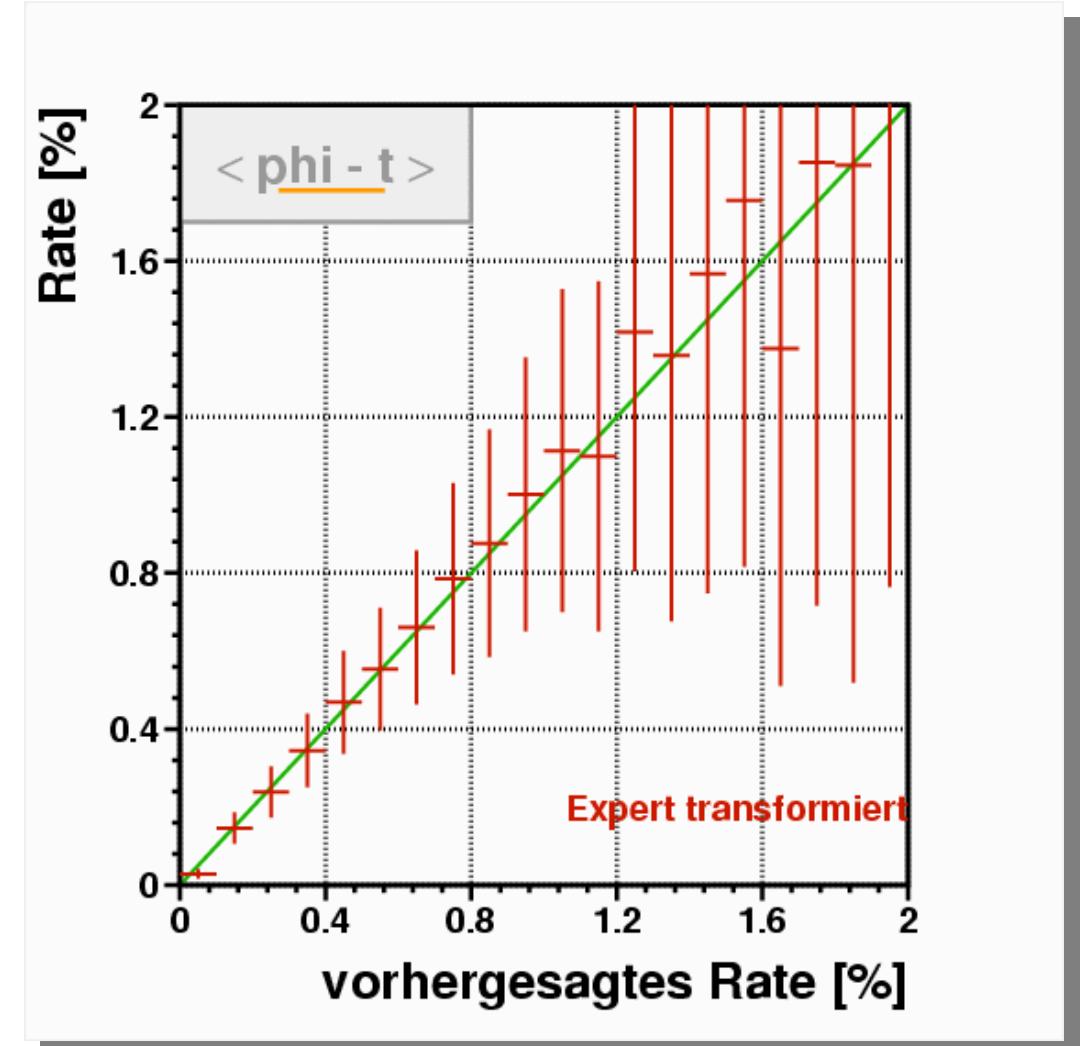
Die Mehrheit der (eher wenig gefährdeten) Kunden zahlt zu viel.

Weniger als die Hälfte der Kunden (die eher unfallgefährdet sind) zahlen zu wenig, zum Teil viel zu wenig. Diese werden z.Z. von den weniger gefährdeten „subventioniert“.

# Bausparkasse Schwäbisch Hall

Prognose vorzeitiger  
Kündigungen

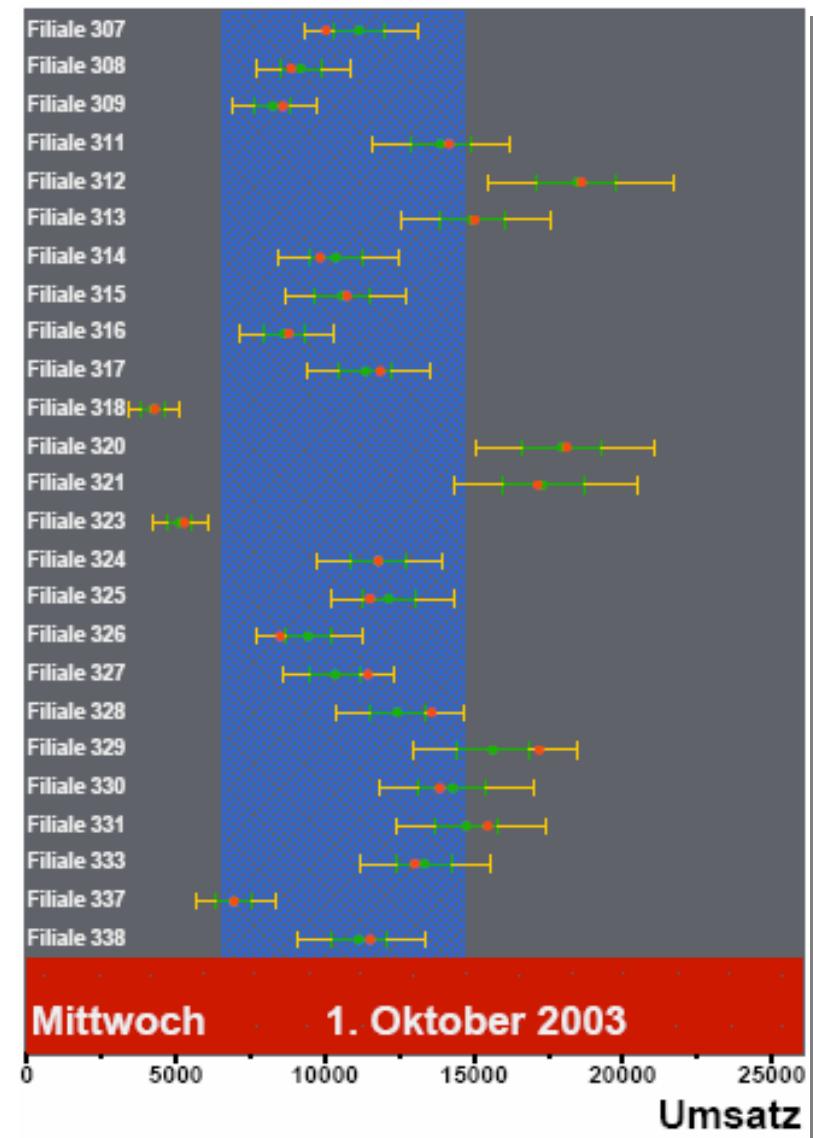
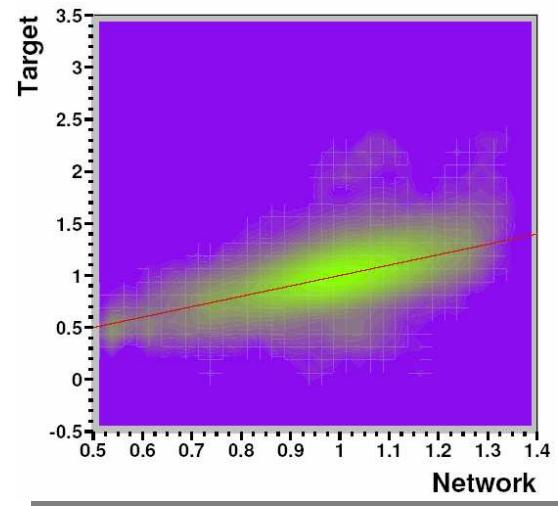
Massive Anreicherung  
möglich.  
CRM-Maßnahmen  
Eingeleitet.



# dm drogerie-markt

Umsatzprognosen für jede Filiale  
in der Zukunft:  
(1 Tag, 2 Tage, 1 Woche, 1 Monat)

1. Zeitreihen-Modellbildung
2. Korrektur und Fehlerabschätzung  
durch NeuroBayes



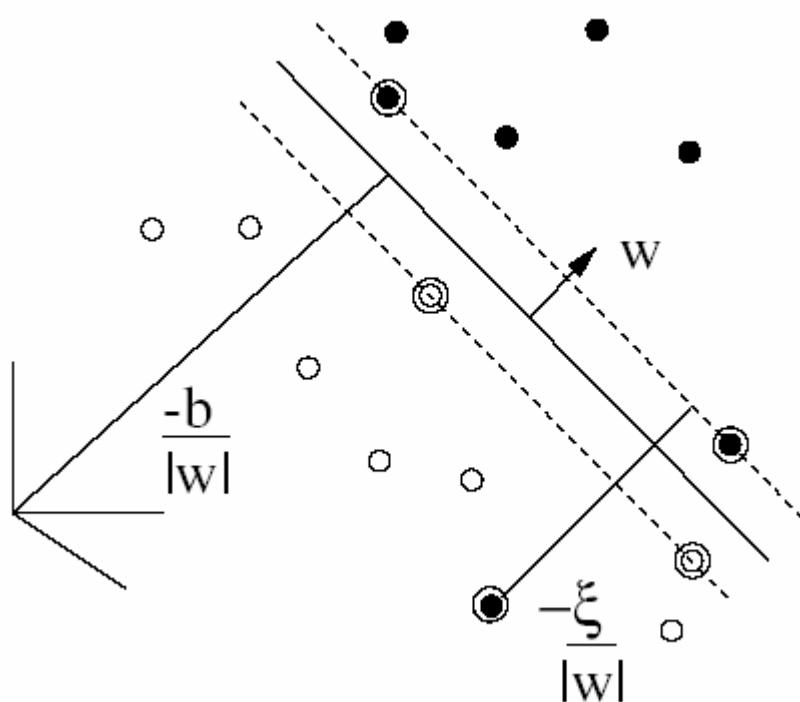
# Weiterer Pläne

Anwendungen in der Medizin

Anwendungen in der Pharmaforschung

# Support-Vektor-Maschinen

Trennbare Probleme in vielen Dimensionen mit wenig Trainingsereignissen (z.B. Spamfilter)



Finde Hyperebene, die zum minimalen Abstand zwischen irgendwelchen Punkten und der Hyperebene führt.

- Die Punkte, die das Minimum-Kriterium erfüllen, heißen Support-Vektoren.  
(Auch nichtlineare Trennflächen möglich)

# Support-Vektor-Maschinen

*vector  $\mathbf{x}_i$  belongs to*

$\mathbf{x}_i \cdot \mathbf{w} + b \geq +1 \quad \text{for } y_i = +1 \quad (\text{class1, signal})$

$\mathbf{x}_i \cdot \mathbf{w} + b \leq -1 \quad \text{for } y_i = -1 \quad (\text{class2, background})$

