Karlsruhe Institute of Technology

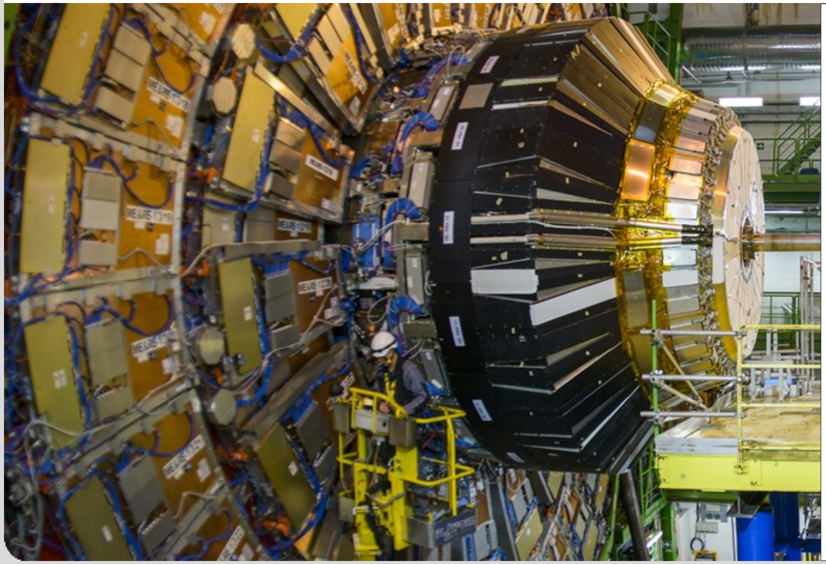
Rechnernutzung in der Physik

Teil 3 – Statistische Methoden in der Datenanalyse

Roger Wolf

19. Januar 2015

INSTITUTE OF EXPERIMENTAL PARTICLE PHYSICS (IEKP) – PHYSICS FACULTY



- Grundlagen der Wahrscheinlichkeitstheorie, Werkzeuge zur statistischen Datenanalyse
- Gängige Wahrscheinlichkeitsverteilungen
- Monte-Carlo Methoden
- Parameterschätzung
- Hypothesentests

Kapiel 3.6:

Parameterschätzung

- Einführung und Grundlagen der Statistik (→ Stichprobe, Schätzfunktion, Stichprobenverteilung, Verzerrung).
- Charakterisierung von Stichproben (→ Erwartungswert, Varianz, Korrelationskoeffizient).

Kapiel 3.7:

Parameterschätzung mit Hilfe der Maximum Likelihood Methode

- Stochastische Modelle, Hypothesen, Teststatistik, Likelihood Funktion.
- Maximum Likelihood Prinzip.
- Maximum Likelihood Schätzfunktion und Unsicherheit am Beispiel des radioaktiven Zerfalls.
- Abschätzung der Unsicherheit der ML Schätzfunktion, analytisch, durch Monte Carlo Methoden, durch die RCF Ungleichung, graphisch.

Kapiel 3.8:

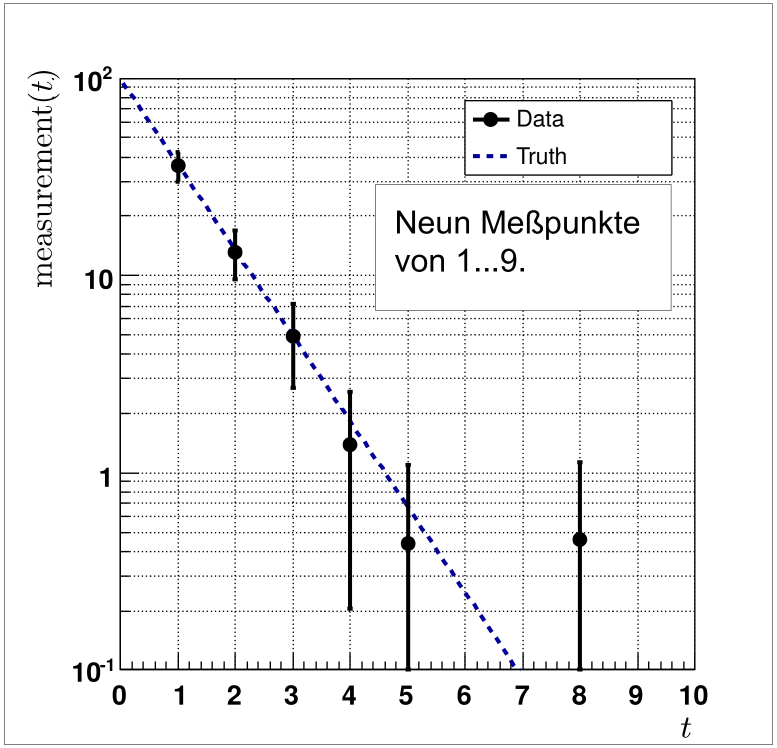
Parameterschätzung mit Hilfe der Least Square Methode

Gaußverteilte Einzelmessungen

- Stellen Sie sich vor Sie hätten eine Meßreihe aus 9 unabhängigen Einzelmessungen $\{y_i\}$ die alle nach Gauß verteilt sind. Es kann sich dabei z.B. um die Zerfallsrate eines radioaktiven Präparats zu bestimmten Zeitpunkten $\{t_i\}$ handeln.[(1)]

- Nehmen Sie weiter an, daß jede Einzelmessung y_i einen anderen unbekannten Erwartungswert $\mu_i(t_i)$ und eine bekannte Varianz σ_i^2 besitzt. In unserem Beispiel abgeschätzt durch $\sqrt{N_i}$ bei N_i registrierten Zerfällen.

- Die μ_i können außerdem mehreren weiteren Parametern θ_k abhängen. In unserem Beispiel entspricht ein enzelner Parameter θ der Lebensdauer des Isotops.



(1) Die Zeitpunkte der Messungen setzen Sie in diesem Fall als exakt bekannt vorraus.

ML-Schätzfunktion für Gaußverteilte Messungen

- Sie können wieder die gesamte Meßreihe aus 9 Einzelmessungen als einzelnes Zufallexperiment betrachen (vgl VL-11). In einer Maximum Likelihood Analyse, mit $\mu_i = \mu(t_i, \theta)$ würden Sie die Likelihood Funktion wie folgt ansetzen:

$$\mathcal{L}(\{y_i\}, \{\mu_i(\theta)\}, \{\sigma_i\}) = \prod_{i=1}^{9} \frac{1}{\sqrt{2\pi\sigma_i^2}} \, e^{-(y_i - \mu_i(\theta))^2 / 2\sigma_i^2}$$

$$\ln(\mathcal{L}(\theta)) = -\frac{1}{2} \sum_{i=1}^{9} \frac{(y_i - \mu_i(\theta))^2}{\sigma_i^2}$$

- Die Likelihood Funktion wird maximal wenn die Funktion $\chi^2(\theta) = \sum_{i=1}^{9} \frac{(y_i - \mu_i(\theta))^2}{\sigma_i^2}$ minimal wird.

- Für nach Gauß verteilte Zufallsgrößen gilt: $\chi^2(\theta) = -2\ln(\mathcal{L}(\theta))$

Least Square (LS-)Schätzfunktion

Seien $\{y_i\}$ Einzelmessungen einer Meßreihe, deren Erwartungswerte $\{\mu_i\}$ unbekannt, deren Varianzen $\{\sigma_i^2\}$ jedoch bekannt sind. Diese Einzelmessungen müssen nicht unabhängig sein sondern können über eine bekannte Matrix $\{V_{ij}\}$ korreliert sein. Die Funktion

$$\chi^2(\theta) = \sum_{i,j=1}^{n} (y_i - \mu_i(\theta))^\mathsf{T} \, V_{ij}^{-1} \, (y_j - \mu_j(\theta))$$

heißt **Schätzfunktion der kleinsten Quadrate** (Least Square, LS-Schätzfunktion). Die Parameter $\hat{\theta}_{LS}$, die $\chi^2(\theta)$ minimieren heissen **Schätzwerte der kleinsten Quadrate** (Least Square, LS-Schätzwerte).

Eigenschaften der LS-Schätzfunktion

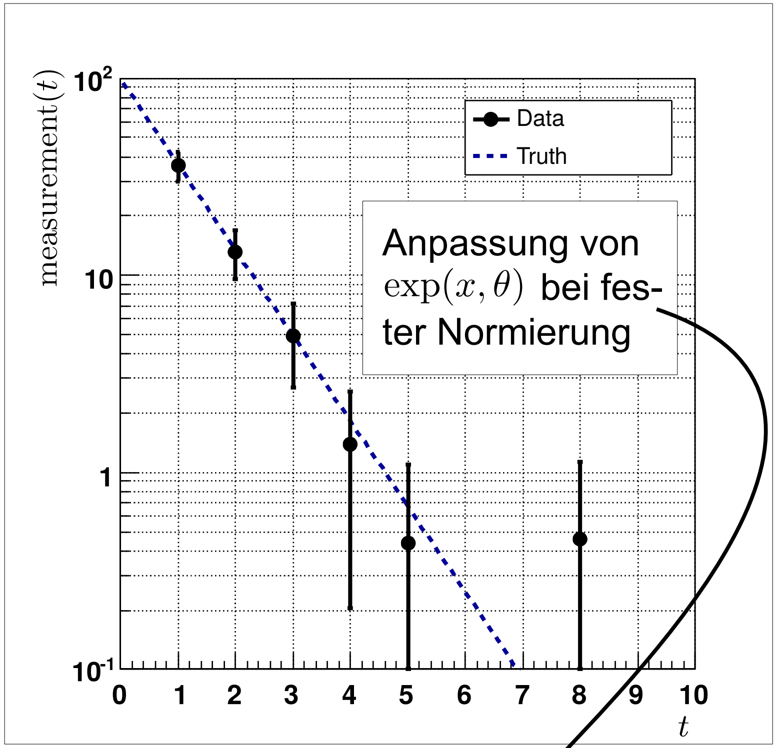
- Die Anpassungsmethode der kleinsten Quadrate ist für nach Gauß verteilte Zufallszahlen ein Spezialfall der Maximum Likelihood Methode.

- Wie die $\{y_i\}$ nach t verteilt sind kommt in der LS-Schätzfunktion nicht vor. Man sagt die LS-Schätzfunktion ist verteilungsfrei.

- Sind die Einzelmessungen $\{y_i\}$ nach Gauß verteilt, dann folgt die LS-Schätz-funktion Funktion der χ^2 Funktion aus VL-09 (slide 41) für $(n-k)$ Freiheitsgrade.

$$\chi^2(\theta) = \sum_{i=1}^{n} \frac{(y_i - \mu_i(\theta))^2}{\sigma_i^2}$$

dabei ist n die Anzahl der Messungen und k die Anzahl der freien Parameter des Modells.



9 Meßpunkte minus 1 freier Parameter → 8 Freiheitgrade.

Eigenschaften der LS-Schätzfunktion

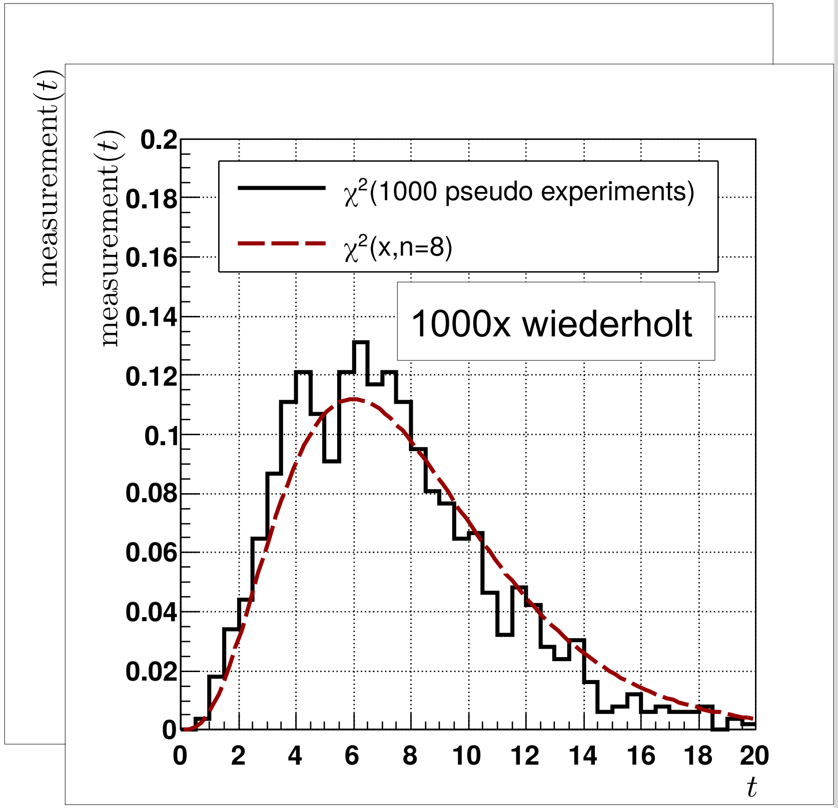
- Die Anpassungsmethode der kleinsten Quadrate ist für nach Gauß verteilte Zufallszahlen ein Spezialfall der Maximum Likelihood Methode.

- Wie die $\{y_i\}$ nach t verteilt sind kommt in der LS-Schätzfunktion nicht vor. Man sagt die LS-Schätzfunktion ist verteilungsfrei.

- Sind die Einzelmessungen $\{y_i\}$ nach Gauß verteilt, dann folgt die LS-Schätzfunktion Funktion der χ^2 Funktion aus VL-09 (slide 41) für $(n-k)$ Freiheitsgrade.

$$\chi^2(\theta) = \sum_{i=1}^{n} \frac{(y_i - \mu_i(\theta))^2}{\sigma_i^2}$$

dabei ist n die Anzahl der Messungen und k die Anzahl der freien Parameter des Modells.



9 Meßpunkte minus 1 freier Parameter → 8 Freiheitgrade.

Die Linear Least Squares Anpassung

- Die LS-Schätzfunktion kann numerisch minimiert werden. Wenn die Schätzfunktion linear von den Parametern θ_k abhängt ist die jedoch auch analytisch lösbar.

$$\mu(k_i, \theta_k) = \sum_{k=1}^{m} a_k(y_i)\theta_k = \sum_{k=1}^{m} A_{ik}\theta_k$$

in Matrixschreibweise schreibt sich χ^2 als:

$$\chi^2 = (\vec{y} - \vec{\mu})^\intercal V^{-1}(\vec{y} - \vec{\mu}) = (\vec{y} - A\vec{\theta})^\intercal V^{-1}(\vec{y} - A\vec{\theta})$$

Um das Minimum zu finden setzen wir:

$$\vec{\partial}_\theta \chi^2 = -2\left(A^\intercal V^{-1}\vec{y} - A^\intercal V^{-1}A\,\vec{\theta}\right) \stackrel{!}{=} 0$$

$$\boxed{\hat{\vec{\theta}}_{LS} = \underbrace{\left(A^\intercal V^{-1}A\right)^{-1} A^\intercal V^{-1}}_{\equiv B}\vec{y}}$$

d.h. die LS-Schätzwerte sind Linearkombinationen der ursprünglichen Messungen:

$$\hat{\theta}_{LS,k} = \sum_{i=1}^{n} B_{ki} y_i$$

Varianz der $\hat{\theta}_{LS}$

- Es gibt zwei offensichtliche Methoden die Varianz der $\hat{\theta}_{LS}$ auszurechnen:
- **Methode-1:** aus Fehlerfortpflanzung (=Variablentransformation) von V.

$$\mathrm{cov}(\hat{\theta}_i, \hat{\theta}_j) = BVB^\intercal = (A^\intercal V^{-1} A)^{-1}$$

$$\left((A^\intercal V^{-1} A)^{-1} A^\intercal V^{-1}\right) V \left((A^\intercal V^{-1} A)^{-1} A^\intercal V^{-1}\right)^\intercal$$

$$(A^\intercal)^{-1} V A^{-1} A^\intercal \underbrace{V^{-1}\ V}_{\equiv 1} (V^{-1})^\intercal A (A^{-1})^\intercal V^\intercal A^{-1}$$

$$(A^\intercal)^{-1} V \underbrace{\underbrace{A^{-1} A^\intercal (V^{-1})^\intercal}_{X^{-1}} \underbrace{A (A^{-1})^\intercal V^\intercal}_{\equiv X}}_{\equiv 1} A^{-1} = (A^\intercal)^{-1} V A^{-1} = (A^\intercal V^{-1} A)^{-1}$$

Unter Verwendung von:

$$(AB)^\intercal = B^\intercal A^\intercal$$

$$(A^{-1})^\intercal = (A^\intercal)^{-1}$$

Varianz der $\hat{\theta}_{LS}$

- Es gibt zwei offensichtliche Methoden die Varianz der $\hat{\theta}_{LS}$ auszurechnen:

- **Methode-1:** aus Fehlerfortpflanzung (=Variablentransformation) von V.

$$\mathrm{cov}(\hat{\theta}_i, \hat{\theta}_j) = BVB^\intercal = (A^\intercal V^{-1} A)^{-1}$$

$$\ldots VA^{-1} = (A^\intercal V^{-1} A)^{-1}$$

Ein Vergleich mit der RCF-Ungleichung ergibt:

$$\frac{\partial^2}{\partial\theta_i \partial\theta_j}\chi^2(\theta) = 2A^\intercal V^{-1} A$$

$$\mathrm{cov}[\hat{\theta}_i, \hat{\theta}_j] \geq \left[-\frac{\partial^2}{\partial\theta_i \partial\theta_j} \underbrace{\ln \mathcal{L}(\theta)}_{-\frac{1}{2}\chi^2(\theta)}\right]^{-1}_{\substack{\theta_i = \hat{\theta}_i \\ \theta_j = \hat{\theta}_j}} = (A^\intercal V^{-1} A)^{-1}$$

Für normalverteilte Zufallsgrößen

D.h. für nach Gauß verteilte Zufallsgrößen ist die LS-Schätzfunktion effizient![(2)]

Unter Verwendung von:

$$(AB)^\intercal = B^\intercal A^\intercal$$

$$(A^{-1})^\intercal = (A^\intercal)^{-1}$$

(2)
cf. VL-11 slide 32. Dieser Ausgang war allerdings bereits klar, weil es sich dann um eine ML-Schätzfunktion handelt (siehe slide 33 der selben VL).

Varianz der $\hat{\theta}_{LS}$

- Es gibt zwei offensichtliche Methoden die Varianz der $\hat{\theta}_{LS}$ auszurechnen:
- **Methode-2:** graphisch aus Abweichung von $\Delta\chi^2$ vom Wert am Minimum $\chi^2_{\min}$.

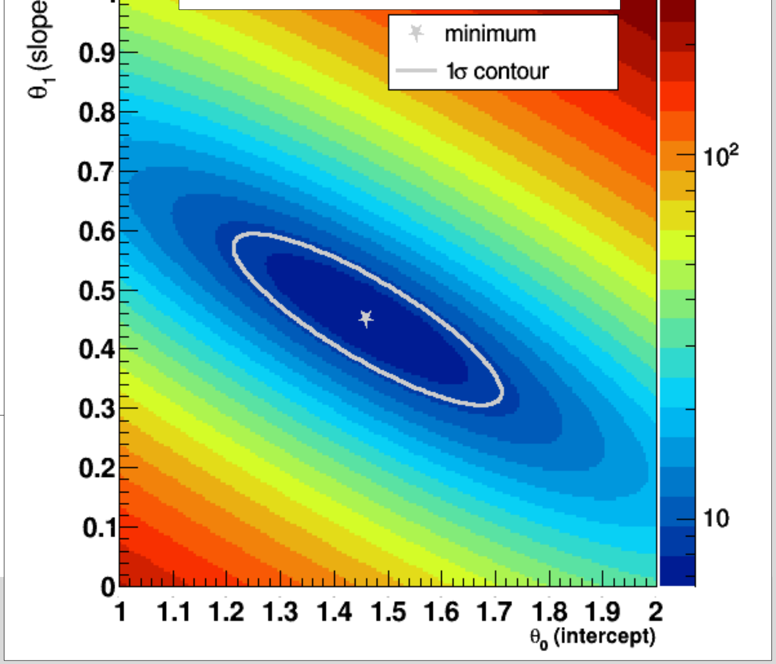
Aus der **Taylor-Entwicklung der Likelihood** Funktion im Maximum ergibt sich:

$$\chi^2(\theta_i) \approx \underbrace{\chi^2(\{\hat{\theta}_{LS,i}\})}_{\equiv \chi^2_{\max}} + \underbrace{\frac{1}{2!}\sum\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\chi^2\right]}_{\equiv \hat{\mathrm{cov}}^{-1}_{ij}}{}_{\theta=\hat{\theta}_{LS}}\left(\theta_i - \hat{\theta}_{LS,i}\right)\left(\theta_j - \hat{\theta}_{LS,j}\right) + \ldots$$

$$\chi^2(\vec{\hat{\theta}}_{LS} \pm \hat{\mathrm{cov}}(\{\hat{\theta}_{LS,i}\}) \cdot \vec{\hat{\theta}}_{LS}) = \chi^2_{\min} + 1$$

d.h. die Variation aus dem Minimum um $\pm\hat{\sigma}$ bewirkt die Erhöhung von $\chi^2(\{\theta_i\})$ um den Wert 1![3]

Aus dem folgenden Beispiel
(Anpassung einer Geraden)



(3)
cf. VL-11 slide 33. Der Faktor 2 ergibt sich aus der Definition von $\chi^2 = -2\ln\mathcal{L}$ (s.a. slide 6 dieser VL).

Beispiel: Anpassung eines Polynoms

- Fallbeispiel: Anpassung eines Polynoms n-ten Grades:

$$\mu(y_i, \theta_k) = \sum_{k=1}^{m} a_k(y_i)\theta_k = \sum_{k=1}^{m} A_{ik}\theta_k$$

$$A_{ik} = a_k(y_i) = y_i^k$$

Beispiel: Anpassung eines Polynoms

- Fallbeispiel: Anpassung eines Polynoms n-ten Grades:

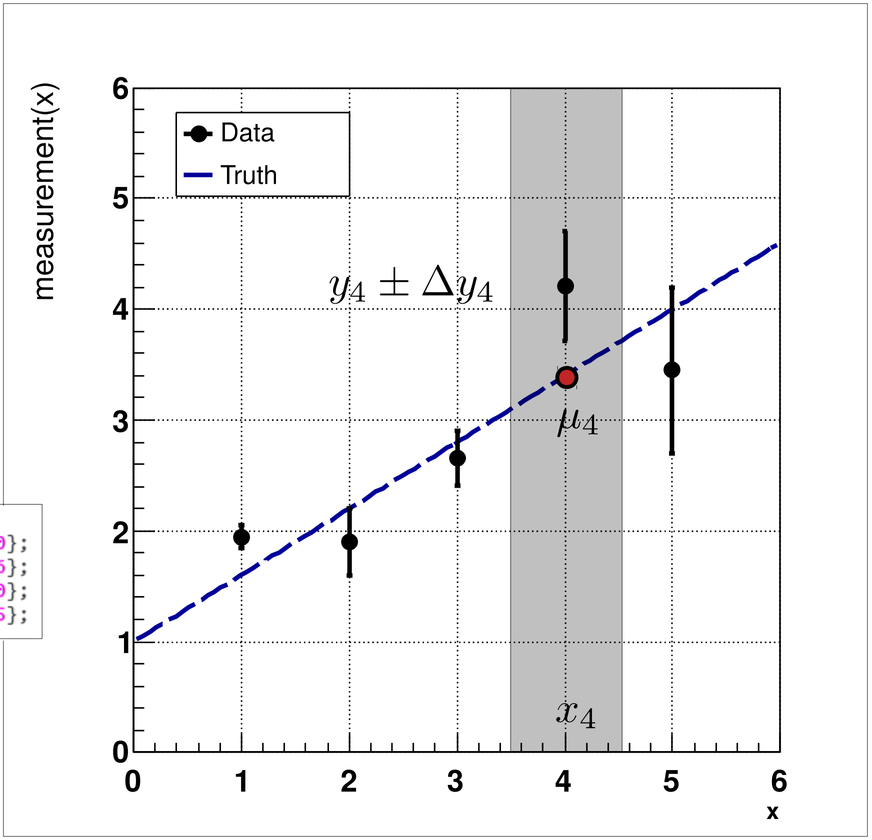
$$\mu(y_i, \theta_k) = \sum_{k=1}^{m} a_k(y_i)\theta_k = \sum_{k=1}^{m} A_{ik}\theta_k$$

$$A_{ik} = a_k(y_i) = y_i^k$$

- Fünf unkorrelierte Datenpunkte (gaußisch nach einer unbekannten linearen Funktion verteilt, $m = 5$):

```
static const int LENGTH = 5;
static float XVALUES[] = {   1.00,     2.00,     3.00,     4.00,     5.00};
static float YVALUES[] = {1.94759, 1.90523, 2.65621, 4.20916, 3.44776};
static float XERRORS[] = {   0.00,     0.00,     0.00,     0.00,     0.00};
static float YERRORS[] = {   0.10,     0.30,     0.25,     0.50,     0.75};
```

- Anpassung einer Geraden ($k = 2$) mit θ_0 (y-Achsenabschnitt) und θ_1 (Steigung).



Beispiel: Anpassung eines Polynoms

$$A^\intercal V^{-1} A = \begin{pmatrix} \sum 1/\sigma_i^2 & \sum x_i/\sigma_i^2 \\ \sum x_i/\sigma_i^2 & \sum x_i^2/\sigma_i^2 \end{pmatrix}$$

$$\left(A^\intercal V^{-1} A\right)^{-1} = \frac{1}{\det(A^\intercal V^{-1} A)} \begin{pmatrix} \sum x_i^2/\sigma_i^2 & -\sum x_i/\sigma_i^2 \\ -\sum x_i/\sigma_i^2 & 1/\sigma_i^2 \end{pmatrix}^{(4)}$$

$$A^\intercal V^{-1} \vec{x} = \begin{pmatrix} \sum \mu_i/\sigma_i^2 \\ \sum \mu_i x_i/\sigma_i^2 \end{pmatrix}$$

$$\begin{pmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \end{pmatrix} = \frac{1}{\det(A^\intercal V^{-1} A)} \begin{pmatrix} (\sum x_i^2/\sigma_i^2)(\sum y_i/\sigma_i^2) - (\sum x_i/\sigma_i^2)(\sum y_i x_i/\sigma_i^2 \\ (\sum 1/\sigma_i^2)(\sum y_i x_i/\sigma_i^2) - (\sum x_i/\sigma_i)(\sum y_i/\sigma_i^2 \end{pmatrix}$$

$$\hat{\vec{\theta}}_{LS} = \underbrace{\left(A^\intercal V^{-1} A\right)^{-1}}_{\equiv B} A^\intercal V^{-1} \vec{x}$$

$$A = \begin{pmatrix} 1 & x_0 \\ 1 & x_1 \\ \vdots & \vdots \\ 1 & x_5 \end{pmatrix} \qquad \vec{x} = \begin{pmatrix} 1 \\ 2 \\ \vdots \\ 5 \end{pmatrix}$$

$$V^{-1} = \begin{pmatrix} 1/\sigma_1^2 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2^2 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \ldots & 1/\sigma_5^2 \end{pmatrix}$$

(4) $\det(A^\intercal V^{-1} A) = (\sum 1/\sigma_i^2)(\sum x_i^2/\sigma_i^2) - (\sum x_i/\sigma_i^2)^2$

Beispiel: Anpassung eines Polynoms

$$A^{\intercal}V^{-1}A = \begin{pmatrix} \sum 1/\sigma_i^2 & \sum x_i/\sigma_i^2 \\ \sum x_i/\sigma_i^2 & \sum x_i^2/\sigma_i^2 \end{pmatrix}$$

$$\boxed{\left(A^{\intercal}V^{-1}A\right)^{-1} = \frac{1}{\det(A^{\intercal}V^{-1}A)} \begin{pmatrix} \sum x_i^2/\sigma_i^2 & -\sum x_i/\sigma_i^2 \\ -\sum x_i/\sigma_i^2 & 1/\sigma_i^2 \end{pmatrix}^{(4)}}$$

Steigung ($\hat{\theta}_1$) und y-Achsenabschnitt ($\hat{\theta}_0$) antikorreliert.

(Korrelationsmatrix)

$$A^{\intercal}V^{-1}\vec{x} = \begin{pmatrix} \sum \mu_i/\sigma_i^2 \\ \sum \mu_i x_i/\sigma_i^2 \end{pmatrix}$$

$$\begin{pmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \end{pmatrix} = \frac{1}{\det(A^{\intercal}V^{-1}A)} \begin{pmatrix} (\sum x_i^2/\sigma_i^2)(\sum y_i/\sigma_i^2) - (\sum x_i/\sigma_i^2)(\sum y_i x_i/\sigma_i^2 \\ (\sum 1/\sigma_i^2)(\sum y_i x_i/\sigma_i^2) - (\sum x_i/\sigma_i)(\sum y_i/\sigma_i^2 \end{pmatrix}$$

$$\boxed{\vec{\hat{\theta}}_{LS} = \underbrace{\left(A^{\intercal}V^{-1}A\right)^{-1} A^{\intercal}V^{-1}}_{\equiv B}\vec{x}}$$

$$A = \begin{pmatrix} 1 & x_0 \\ 1 & x_1 \\ \vdots & \vdots \\ 1 & x_5 \end{pmatrix} \qquad \vec{x} = \begin{pmatrix} 1 \\ 2 \\ \vdots \\ 5 \end{pmatrix}$$

$$V^{-1} = \begin{pmatrix} 1/\sigma_1^2 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2^2 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \ldots & 1/\sigma_5^2 \end{pmatrix}$$

(4) $\det(A^{\intercal}V^{-1}A) = (\sum 1/\sigma_i^2)(\sum x_i^2/\sigma_i^2) - (\sum x_i/\sigma_i^2)^2$

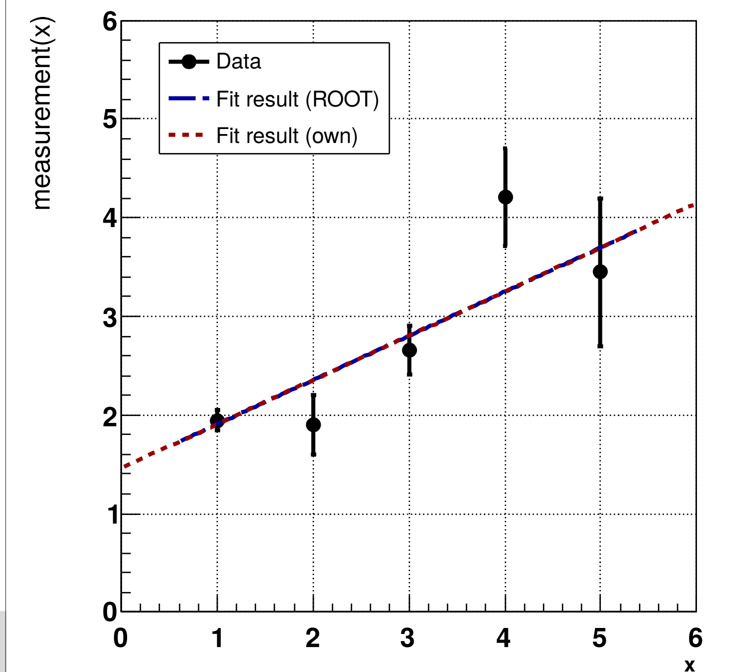
Beispiel: Anpassung eines Polynoms

```
std::pair<TMatrixD,TMatrixD> LLS(TMatrixD& A, TMatrixD& V, TMatrixD& X){
  TMatrixD U=TMatrixD(A,TMatrixD::kTransposeMult,TMatrixD(V,TMatrixD::kMult,A)).Invert();  /* (A^t*V*A)^(-1)       */
  TMatrixD B=TMatrixD(U,TMatrixD::kMult,TMatrixD(A,TMatrixD::kTransposeMult,V));           /* (A^t*V*A)^(-1)*A^t*V */
  TMatrixD T=TMatrixD(B,TMatrixD::kMult,X);                                                 /* fit result           */
  return std::make_pair(T,U);
}

/* -------------------------------------------------------------------------------
 *  Anpassung einer Geraden (f(x)=a[0]+a[1]*x)
 *-------------------------------------------------------------------------------*/
const int NPARAM=2;                                /* Koeffizientenmatrix            */
TMatrixD A(LENGTH,NPARAM);                         /*                                */
for(int i=0;i<LENGTH;++i){                         /*                                */
  for(int j=0;j<NPARAM;++j){                       /*                                */
    A(i,j)=(j==0?1.0:XVALUES[i]);                  /*                                */
  }
}
// std::cout << "PRINTING A" << std::endl; print(A);
// inverse Korrelationsmatrix
TMatrixD V(LENGTH,LENGTH);                         /* inverse Korrelationsmatrix     */
for(int i=0;i<LENGTH;++i){                         /*                                */
  for(int j=0;j<LENGTH;++j){                       /*                                */
    V(i,j)=(i==j?1./(YERRORS[i]*YERRORS[i]):0.);   /*                                */
  }
}
// std::cout << "PRINTING V" << std::endl; print(V);
TMatrixD Y(LENGTH,1);                              /* Vektor der Messwerte           */
for(int i=0;i<LENGTH;++i){                         /*                                */
  Y(i,0)=YVALUES[i];                               /*                                */
}
// std::cout << "PRINTING X" << std::endl; print(X);
std::pair<TMatrixD,TMatrixD> fit_result = LLS(A,V,Y);
TMatrixD T = fit_result.first;
TMatrixD U = fit_result.second;
```

$$\vec{\hat{\theta}}_{LS} = \underbrace{\left(A^{\intercal}V^{-1}A\right)^{-1} A^{\intercal}V^{-1}}_{\equiv B}\vec{x}$$

LS-Schätzwerte:	Wahrheit:
$\hat{\theta}_0 = 1.46 \pm 0.16$	$\theta_0 = 1.0$
$\hat{\theta}_1 = 0.45 \pm 0.09$	$\theta_1 = 0.6$
$\rho(\hat{\theta}_0, \hat{\theta}_1) = -0.85$	



Beispiel: Anpassung eines Polynoms

```
std::pair<TMatrixD,TMatrixD> LLS(TMatrixD& A, TMatrixD& V, TMatrixD& X){
  TMatrixD U=TMatrixD(A,TMatrixD::kTransposeMult,TMatrixD(V,TMatrixD::kMult,A)).Invert();   /* (A^t*V*A)^(-1)       */
  TMatrixD B=TMatrixD(U,TMatrixD::kMult,TMatrixD(A,TMatrixD::kTransposeMult,V));            /* (A^t*V*A)^(-1)*A^t*V */
  TMatrixD T=TMatrixD(B,TMatrixD::kMult,X);                                                 /* fit result           */
  return std::make_pair(T,U);
}

/* ---------------------------------------------------------------------------
 *  Anpassung einer Geraden (f(x)=a[0]+a[1]*x)
 *---------------------------------------------------------------------------*/
const int NPARAM=2;                             /* Koeffizientenmatrix        */
TMatrixD A(LENGTH,NPARAM);                      /*                            */
for(int i=0;i<LENGTH;++i){                      /*                            */
  for(int j=0;j<NPARAM;++j){                    /*                            */
    A(i,j)=(j==0?1.0:XVALUES[i]);               /*                            */
  }
}
// std::cout << "PRINTING A" <<
// inverse Korrelationsmatrix
TMatrixD V(LENGTH,LENGTH);
for(int i=0;i<LENGTH;++i){
  for(int j=0;j<LENGTH;++j){
    V(i,j)=(i==j?1./(YERRORS[i]*
  }
}
// std::cout << "PRINTING V" <<
TMatrixD Y(LENGTH,1);
for(int i=0;i<LENGTH;++i){
  Y(i,0)=YVALUES[i];
}
// std::cout << "PRINTING X" <<
std::pair<TMatrixD,TMatrixD> fit
TMatrixD T = fit_result.first;
TMatrixD U = fit_result.second;
```

Durch geschickte Wahl können die Parameter dekorreliert werden:

$$y(x) = \theta_0 + \theta_1 (x - \bar{x})$$

mit $\bar{x} = \frac{\sum x_i/\sigma_i^2}{\sum 1/\sigma_i^2} = 1.47$

führt auf: $\rho(\hat{\theta}_0, \hat{\theta}_1) = 10^{-8}$

$$\hat{\vec{\theta}}_{LS} = \underbrace{\left(A^\intercal V^{-1} A\right)^{-1} A^\intercal V^{-1}}_{\equiv B} \vec{x}$$

LS-Schätzwerte:

$\hat{\theta}_0 = 1.46 \pm 0.16$

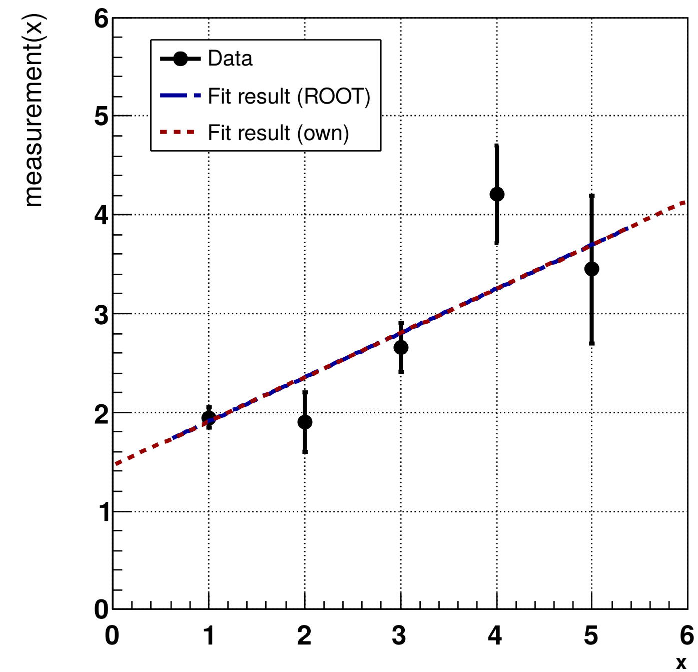
$\hat{\theta}_1 = 0.45 \pm 0.09$

$\rho(\hat{\theta}_0, \hat{\theta}_1) = -0.85$

Wahrheit:

$\theta_0 = 1.0$

$\theta_1 = 0.6$



Kapiel 3.8:

Parameterschätzung mit Hilfe der Least Square Methode

- Least Squares Anpassung als Spezialfall der Maximum Likelihood Methode.
- Varianz des LS-Schätzwertes und Vergleich zur Maximum Likelihood Methode.
- Analytische Lösung der Linear Least Squares Anpassung am Beispiel der Anpassung einer Geraden.

Kapiel 3.9:

Hypothesentests

Hypothesentests

- **Aufgabenstellung:** Bewertung von Hypothesen H_i im Zusammenhang mit einer Meßreihe (Stichprobe). Die Hypothesen entsprechen den Aussagen stochastischer Modelle.

- Oft geht es um die Bewertung einer ausgezeichneten Hypothese H_0 (z.B. Standardmodell der Teilchenphysik, Existenz oder Nichtexistenz eines neuen Teilchens, …).

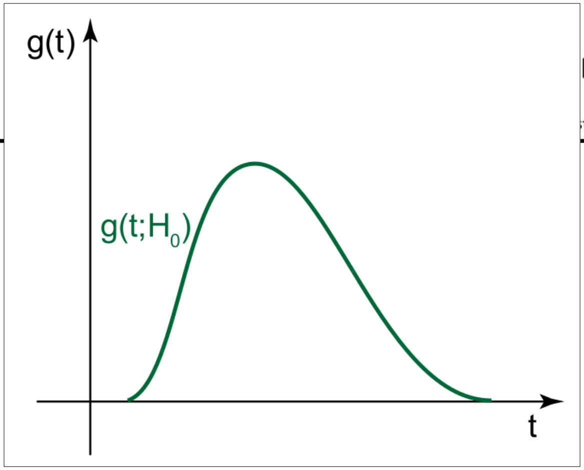
- Quantitative, objektive Verfahren zur Beantwortung der Fragen:

Kann/sollte H_0 aufgrund der vorliegenden Meßreihe verworfen werden?

Wie sicher ist es H_0 aufgrund der vorliegenden Meßreihe zu verwerfen?

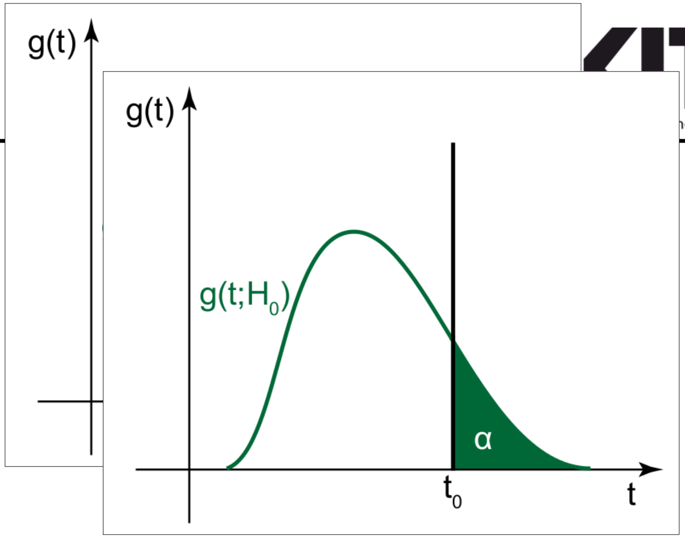
Vorgehen beim Hypothesentest

- Meßreihe: $\vec{x} = (x_1, x_2, \ldots, x_n)$
- Definition einer Teststatistik: $t(\{x_i\})$ [5] Die Teststatistik ist selbst zufallsverteilt nach der Stichprobenverteilung $g_{H_0}(t(\{x_i\}))$.
- Hier spezifizieren wir die Stichprobenverteilung als bedingte Wahrscheinlichkeitsverteilung unter der Bedingung H_0.



(5) cf. VL-11 slide 7.

Vorgehen beim Hypothesentest



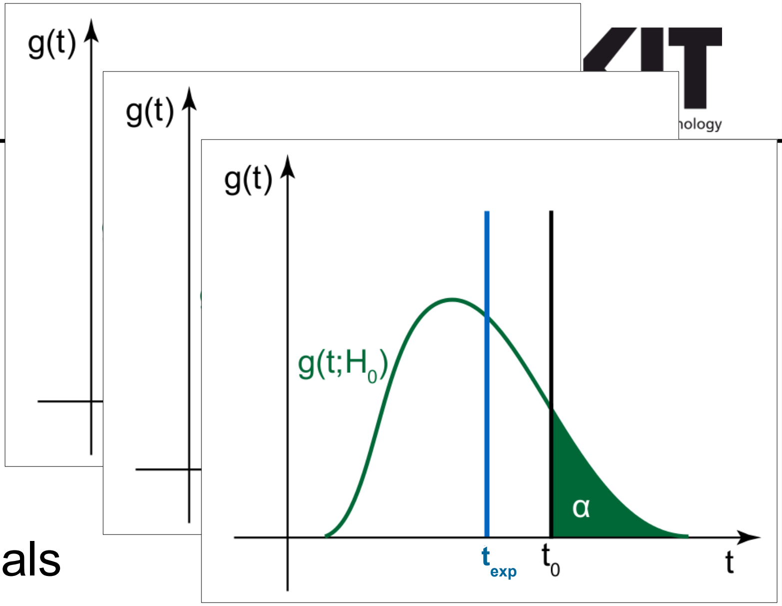
Angenommen H_0 ist wahr, dann ist in einem Bruchteil von α Fällen $t \geq t_0$.

- Meßreihe: $\vec{x} = (x_1, x_2, \ldots, x_n)$
- Definition einer Teststatistik: $t(\{x_i\})$ (5) Die Teststatistik ist selbst zufallsverteilt nach der Stichprobenverteilung $g_{H_0}(t(\{x_i\}))$.
- Hier spezifizieren wir die Stichprobenverteilung als bedingte Wahrscheinlichkeitsverteilung unter der Bedingung H_0.
- Festlegen eines Kriteriums um H_0 zu verwerfen (vor der Messung!):

$$\alpha = \int_{t_0}^{\infty} g_{H_0}(t)\mathrm{d}t$$ **(Signifikanzniveau)**

(5) cf. VL-11 slide 7.

Vorgehen beim Hypothesentest



- Meßreihe: $\vec{x} = (x_1, x_2, \ldots, x_n)$

- Definition einer Teststatistik: $t(\{x_i\})$ [5] Die Teststatistik ist selbst zufallsverteilt nach der Stichprobenverteilung $g_{H_0}(t(\{x_i\}))$.

- Hier spezifizieren wir die Stichprobenverteilung als bedingte Wahrscheinlichkeitsverteilung unter der Bedingung H_0.

- Festlegen eines Kriteriums um H_0 zu verwerfen (vor der Messung!):

$$\alpha = \int_{t_0}^{\infty} g_{H_0}(t)\mathrm{d}t \quad \textbf{(Signifikanzniveau)}$$

- Auswertung des Experiments (Meßreihe):

$$p = P_{H_0}(t \geq t_{exp}) = \int_{t_{exp}}^{\infty} g_{H_0}(t)\mathrm{d}t \quad \textbf{(\textit{p}-Wert)}$$

$t \leq t_0 \rightarrow H_0$ nicht verwerfen.

$t > t_0 \rightarrow H_0$ verwerfen.

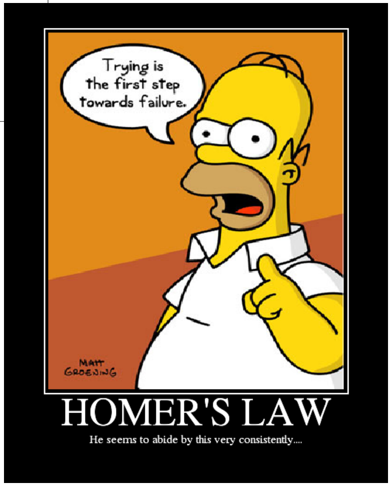
(5) cf. VL-11 slide 7.

Fehler 1. & 2. Art

Fehler erster Art:
(„false positive“)

H_0 mit Wahrscheinlichkeit α verworfen obwohl sie wahr ist.

- Patient krank diagnostiziert obwohl er gesund ist.
- Fälschliche entdeckung eines neuen Teilchens.



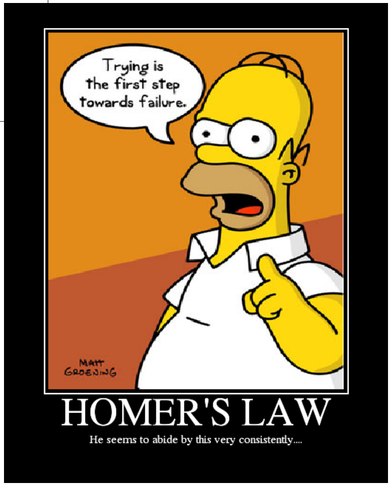
Fehler 1. & 2. Art

Karlsruhe Institute of Technology

Fehler erster Art:
(„false positive“)

H_0 mit Wahrscheinlichkeit α verworfen obwohl sie wahr ist.

- Patient krank diagnostiziert obwohl er gesund ist.
- Fälschliche entdeckung eines neuen Teilchens.



Fehler zweiter Art:
(„false negative“)

H_0 nicht verworfen obwohl falsch.

- Patient gesund diagnostiziert obwohl er krank ist.
- Neues Teilchen nicht entdeckt, obwohl es da ist.
- Von Relevanz beim Vergleich von Hypothesen.

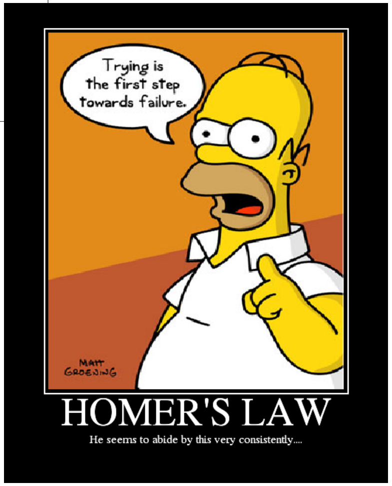
Fehler 1. & 2. Art

Karlsruhe Institute of Technology

Fehler erster Art:
(„false positive“)

H_0 mit Wahrscheinlichkeit α verworfen obwohl sie wahr ist.

- Patient krank diagnostiziert obwohl er gesund ist.
- Fälschliche entdeckung eines neuen Teilchens.



Fehler zweiter Art:
(„false negative“)

H_0 nicht verworfen obwohl falsch.

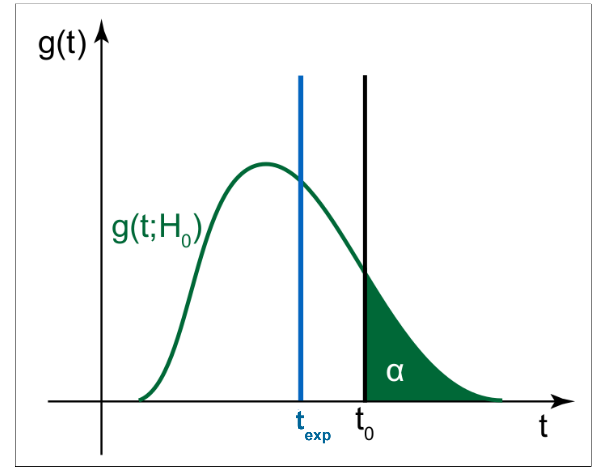
- Patient gesund diagnostiziert obwohl er krank ist.
- Neues Teilchen nicht entdeckt, obwohl es da ist.
- Von Relevanz beim Vergleich von Hypothesen.

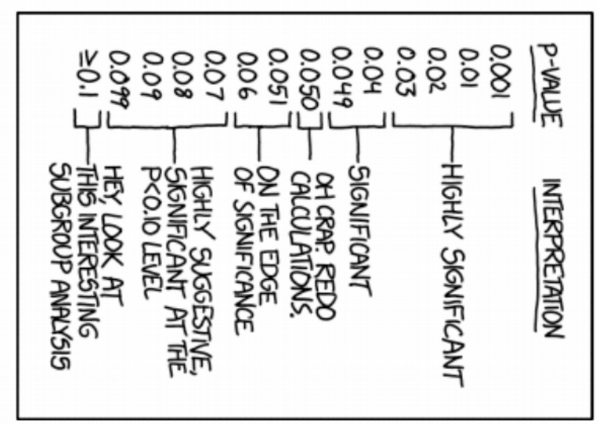
- Wahrscheinlichkeiten für den Ausgang der Messung. Eigenschaft der Messung. Festlegung von α erfolgt (unbeeinflußt vom Ausgang des Experiments!) vor der Auswertung der Meßreihe.

p-Wert

- Der *p*-Wert ist eine a posteriori Wahrscheinlichkeit bestimmt nach(!) Auswertung des Experiments:
- Wahrscheinlichkeit für $t > t_{exp}$ wenn H_0 wahr ist wenn die Meßreihe beliebig oft wiederholt würde.
- Achten Sie auf die Interpretation ihrer Ergebnisse. Die Mathematik versort Sie mit nackten Zahlen. Die Deutungshoheit liegt bei Ihnen.

Etwas unlautere Interpretation des *p*-Wertes:





[xkcd.com]

Einfacher Test der H_0 Hypothese

- Oft ist man in der Situation, beurteilen zu müssen, wie kompatibel ein vorgegebenes Modell (H_0) mit einer vorliegenden Meßreihe ist.

- Die Kompatibilität kann mit Hilfe des *p*-Wertes quantifiziert werden. Solche Tests bezeichnet man gemeinhin als **goodness-of-fit** Tests.

Beispiel Münzwurf

- Stellen Sie sich vor, Sie haben nach 20-maligem Wurf einer Münze 17 mal Kopf und 3 mal Zahl erhalten. Wie kompatibel ist die Hypothese $p(\mathrm{Kopf}) = 0.5$ mit diesem Ausgang der Meßreihe?

$$B(p,n,k) = \binom{n}{k} p^k \, (1-p)^{n-k}$$

$$\mu = np = 10 \quad \sigma^2 = np(1-p) = 5$$ (Abweichung von 3.13σ)

Sie berechnen den *p*-Wert als Summe der Wahrscheinlichkeiten für das Auftreten einer Abweichung vom Erwartungswert, die größer oder gleich der beobachteten ist:

$$p = \sum_{k \in \{0,1,2,3,17,18,19,20\}} B(p = 1/2, n = 20, k) = 0.0026$$

Achtung:
Bei einer Münze ist $H_0 : p(\mathrm{Kopf}) = 0.5$ eine sehr valide Annahme. Überlegen Sie sich genau, wann sie diese Hypothese ins Wanken bringen möchten.

Beispiel: Zählexperiment

Karlsruhe Institute of Technology

- Stellen Sie sich vor, sie haben in einem Zählexperiment (z.B. für den Nachweis eines sterilen Neutrinos) 4 Ereignisse nachgewiesen. Sie erwarten aus Ihrer Untergrundhypothese 1 Ereignis. Sollten Sie das Nobelpreiskommitee informieren?

$$P(\mu, k) = \frac{\mu^k}{k!} e^{-\mu}$$

$\mu = np = 1 \qquad \sigma^2 = np = 1$ (Abweichung von 3.0σ)

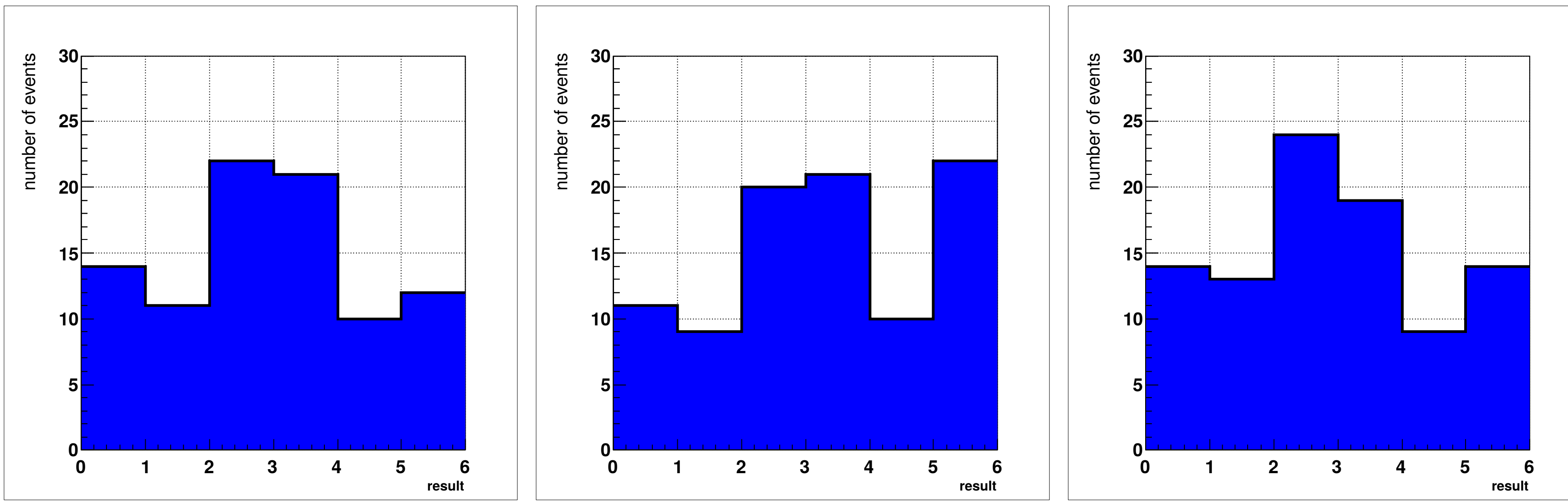
Sie berechnen den *p*-Wert als Summe der Wahrscheinlichkeiten für das Auftreten einer Abweichung vom Erwartungswert, die größer oder gleich der beobachteten ist:

$$p = 1 - (1 + \mu + \mu^2/2 + \mu^3/6)e^{-1} = 0.019$$

0 Ereignis(se)
1 Ereignis(se)
2 Ereignis(se)
3 Ereignis(se)

Beispiel: Least Square Teststatistik

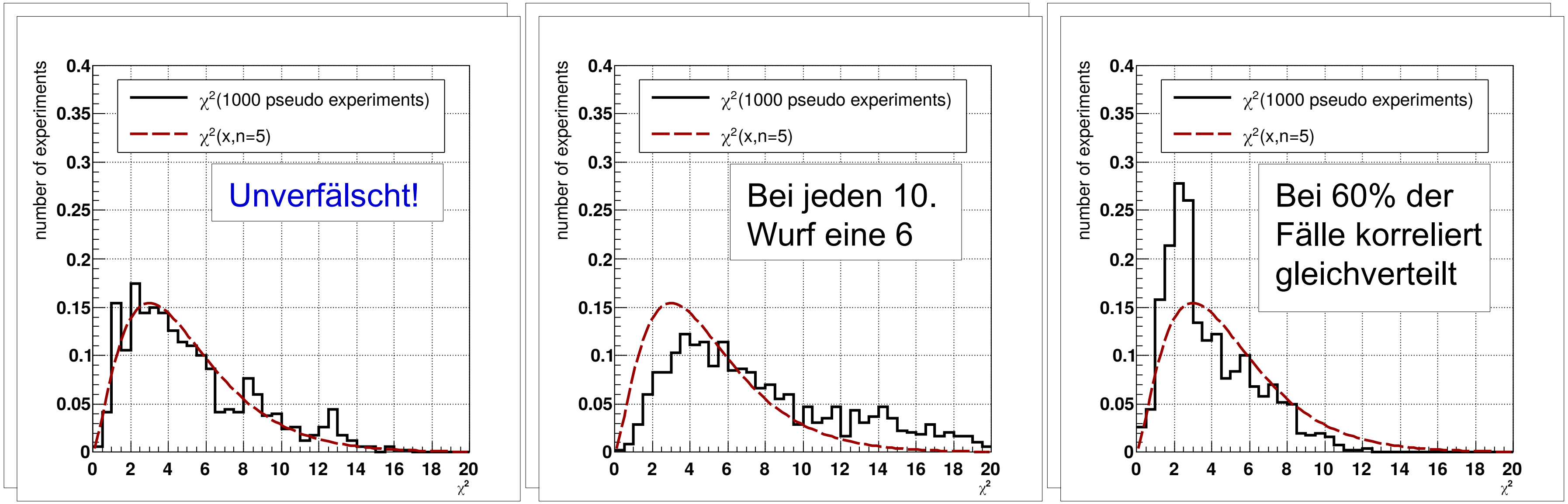
- Sie können auch die Least Square Teststatistik, $\hat{\theta}_{LS}$ für einen goodness-of-fit Test verwenden. Stellen Sie sich vor, Sie möchten einen Würfel auf seine Unverfälschtheit überprüfen. Dafür werfen Sie den Würfen 100 mal und histogrammieren wie oft die Zahlen 1 bis 6 auftauchen. Wir geben drei Beispiele von Meßreihen vor:



- Welcher ist der unverfälschte Würfel?

Beispiel: Least Square Teststatistik

- Antwort können sie z.B. nach 1000 maliger Wiederholung der Meßreihe erhalten, wenn $\hat{\theta}_{LS}$ einer $\chi^2(x,n)$ für 5 Freiheitsgrade folgt. Warum gerade 5?

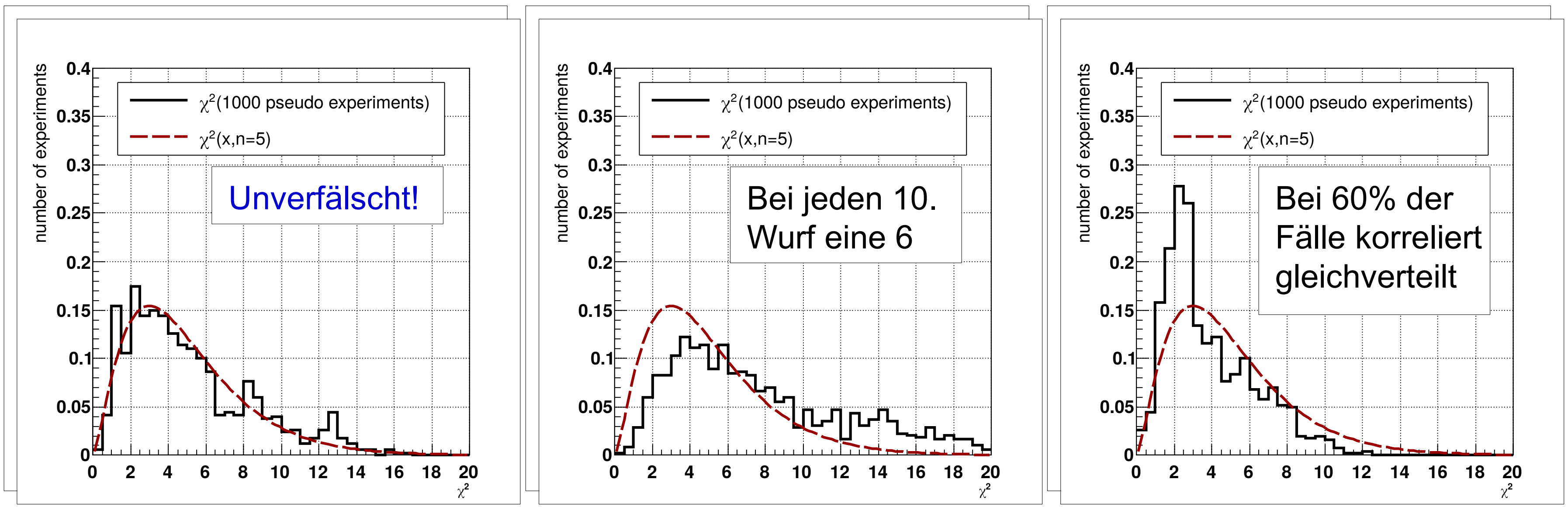


- Welcher ist der unverfälschte Würfel?

Beispiel: Least Square Teststatistik

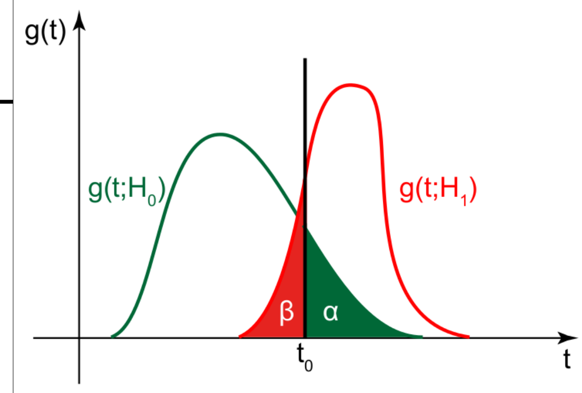
- Antwort können sie z.B. nach 1000 maliger Wiederholung der Meßreihe erhalten, wenn $\hat{\theta}_{LS}$ einer $\chi^2(x,n)$ für 5 Freiheitsgrade folgt. Warum gerade 5?

Sie haben 6 Meßpunkte 1...6 minus 1 Parameter im Modell → 5 Freiheitsgrade.



- Welcher ist der unverfälschte Würfel?

Test alternativer Hypothesen



Je geringer der Überlapp zwischen H_0 und H_1 desto höher die Trennschärfe (Mächtigkeit).

- Beim Test alternativer Hypothesen stellt sich die Frage wie gut sich bei vorgegebenem Signifikanzniveau, α, die alternative Hypothese, H_1, von der Nullhypothese H_0 unterscheiden läßt.

$$\beta = \int_{-\infty}^{t_0} g_{H_1}(t)\mathrm{d}t$$ **(Fehler zweiter Art)**

- Je geringer der Überlapp der beiden Wahrscheinlichkeitsdichten $g_{H_0}(t)$ und $g_{H_1}(t)$ in t, desto besser ist das Experiment dazu in der Lage die beiden Hypothesen zu unterscheiden.

- Man bezeichnet $1-\beta$ als **Trennschärfe** oder Mächtigkeit des Tests. Die Trennschärfe ist eine Eigenschaft des Experiments (und der Wahl der Teststatistik).

Wahl der Teststatistik

- Zum Glück gibt es bei der Wahl der Teststatistik einen eindeutigen Vorzug: Nach dem Lemma von Neyman und Pearson ist die Statistik mit der größten erreichbaren Trennschärfe der **Likelihood Ratio**:

$$Q = \frac{\mathcal{L}(x,\theta,H_1)}{\mathcal{L}(x,\theta,H_0)}$$ (**Likelihod Ratio**)

- Oft auch als Log-Likelihood Ratio (LLR) angegeben:

$$q = -2\ln\left(\frac{\mathcal{L}(x,\theta,H_1)}{\mathcal{L}(x,\theta,H_0)}\right)$$ (**Log-Likelihod Ratio**)

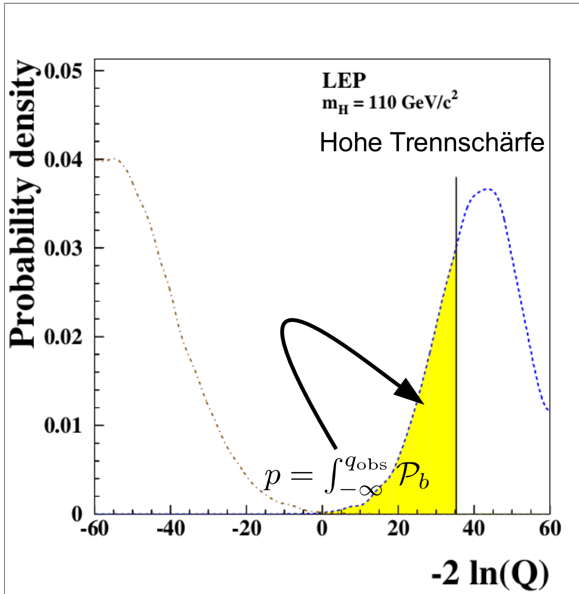
- **NB:**
  Der Faktor -2 wird oft verwendet und ergibt sich wieder aus der Definition von $\chi^2 = -2\ln\mathcal{L}$. Beachten Sie, sobald Sie annehmen können, daß die Werte der Einzelmessungen nach Gauß verteilt sind, folgt die Likelihood Funktion (und unter bestimmten Bedingungen auch der Likelihood Ratio) einer χ^2-Funktion. Das Erlaubt den Einsatz asymptotischer Formeln zur Berechnung der Varianz der Schätzwertes statt mühseliger Pseudo-Experimente.

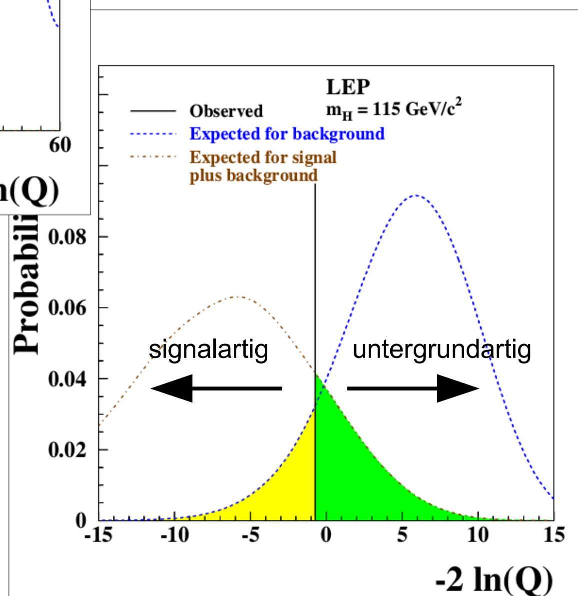
Beispiele aus der Teilchenphysik

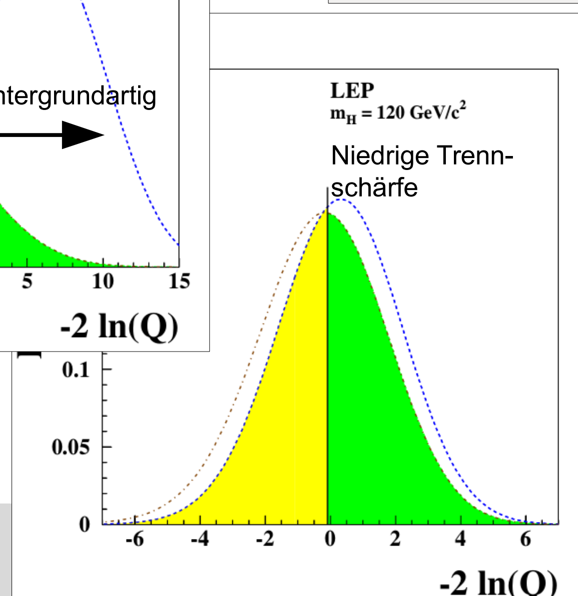
$$q_\mu = -2\ln\left(\frac{\mathcal{L}(n|\mu s+b)}{\mathcal{L}(n|b)}\right), \quad 0 \le \mu$$

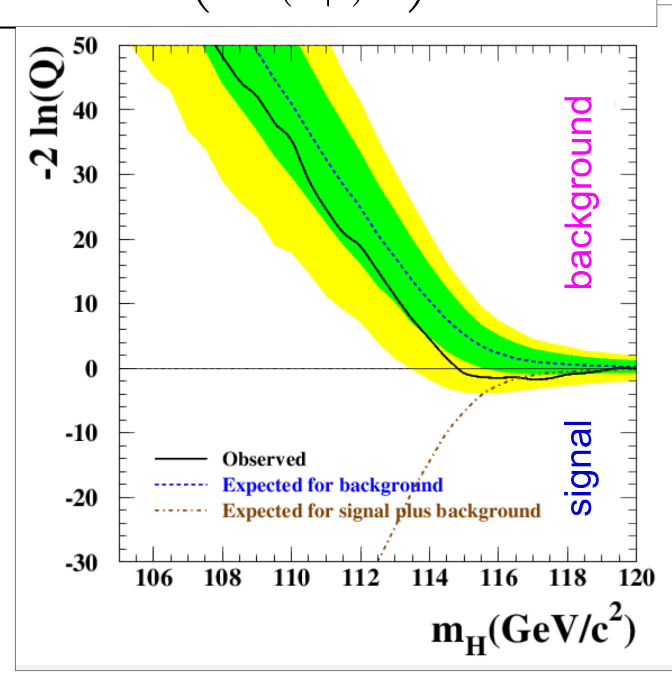
Suche nach dem Higgs Boson bei LEP

Hier ist H_0 die Hypothese: „kein Higgs Boson entdeckt".





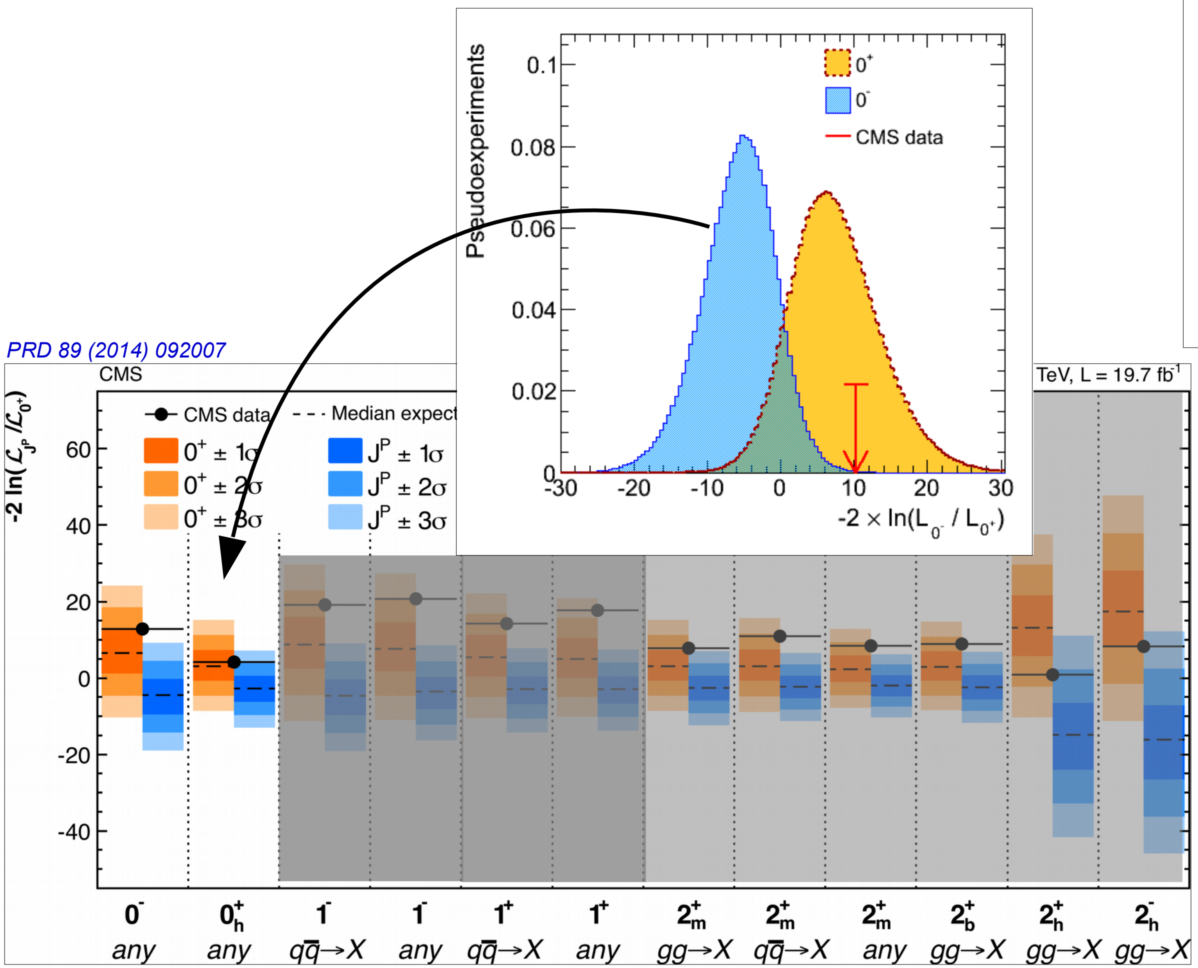


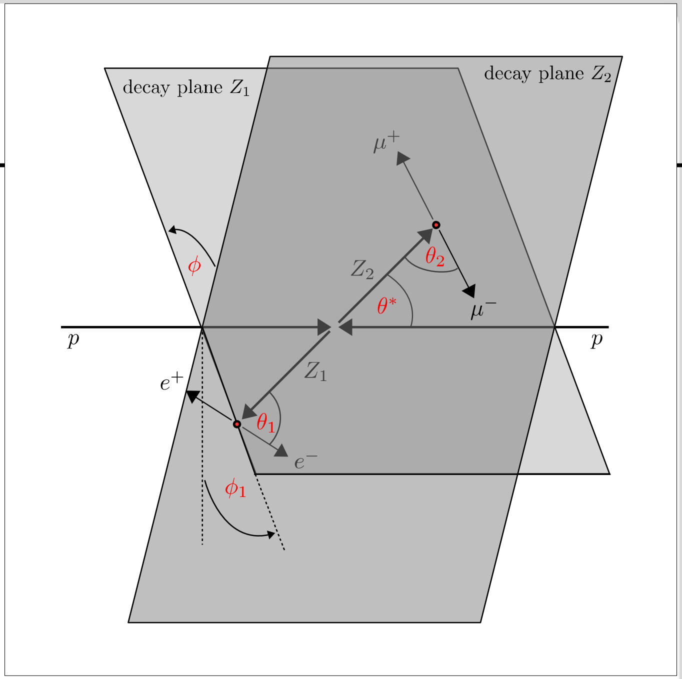


arXiv:hep-ex/030633 (im übrigen ein sehr lesenswertes Papier, dessen Anwendung statistischer Methoden zur Signalbestimmung Sie mit Abschluß dieser VL vollständig verstehen können. Dieses Papier hat das mathematische Fundament für alle weiteren Higgs Boson Suchen bis 2012 gelegt!)

Beispiele aus der Teilchenphysik

Spin und CP Hypothesentest für das Higgs Boson bei LHC





NB: wie Sie sehen ist die Trennschärfe dieses Tests ist nicht ausgesprochen hoch. Das liegt daran, daß der Test auf ~20 ausgewählten Ereignissen aus unzähligen am LHC aufgezeichneten Ereignissen besteht, die mit hoher Wahrscheinlichkeit Higgs Bosonen enthalten. Basierend auf mehr Ereignissen (wie sie derzeit am LHC weiter aufgezeichnet werden) werden die Stichprobenverteilungen der beiden Hypothesen immer schmaler werden, wodurch sich der Überlapp reduziert, gleichbedeutend mit einer höheren Trennschärfe.

Kapiel 3.9:

Hypothesentests

- Einfache und vergleichende Hypothesentests, χ^2-Test, Likelihood Ratio.
- Fehler erster und zweiter Art, *p*-Wert.
- Einige einfache Beispiele und einige Beispiele aus dem täglichen Leben.

Famous Last Words (FLW)

- Das war die letzte Vorlesung aus dem Block Statistik in der Rechnernutzung in der Physik.

- Was Sie mitnehmen sollten: **alles**..., kehren Sie von Zeit zu Zeit zu diesen Folien zurück, sie werden auf dem web-Server verbleiben und können als Referenz dienen.

- Falls Sie Fragen haben, kommen Sie gerne bei mir vorbei: Physikhochhaus 9.20 (roger.wolf@cern.ch).

- Rechnen Sie die Übungen! Davon lernen Sie am nachhaltigsten.

- Und bleiben Sie neugierig...