

# Moderne Methoden der Datenanalyse – Parameterschätzung –

**Roger Wolf**  
21. Mai 2020

# Inhalt der nächsten (drei) Vorlesungen

---

- Grundlagen zur Parameterschätzung und Einführung von Begriffen und Nomenklatur.
- Parameterschätzung mit Hilfe der Maximum Likelihood Methode.
- Parameterschätzung mit und Besonderheiten der  $\chi^2$ -Methode.
- Parameterisierungs- und Minimierungsmethoden (evtl. nur zum Lesen).

# Erweiterte Likelihood

---

- Gerade in der Teilchenphysik kann es vorkommen, dass nicht nur die Eigenschaften/Lage der  $\{x_i\}$  von Relevanz ist, sondern auch die Menge der Beobachtungen  $n$ .
- Die sog. erweiterte Likelihood nimmt dann die folgende Form an:

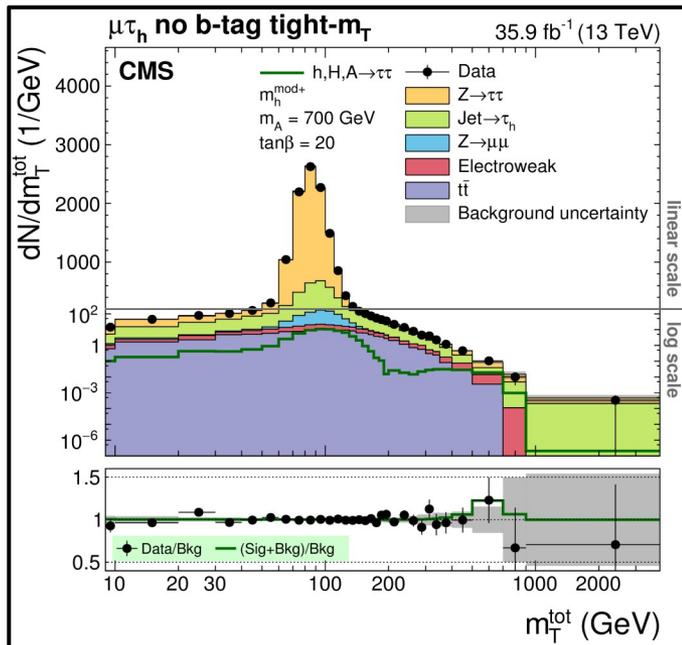
$$\mathcal{L}(\vec{x}, \vec{\theta}, \nu) = \frac{\nu^n}{n!} e^{-\nu} \prod_{i \leq n} p(x_i, \vec{\theta})$$

$$\ln \left( \mathcal{L}(\vec{x}, \vec{\theta}, \nu) \right) = -\nu + \sum_{i \leq n} \ln \left( \nu p(x_i, \vec{\theta}) \right) \quad (*)$$

- D.h. die Likelihood wird um die Wahrscheinlichkeitsdichte einer Poissonverteilung erweitert. Die Poissonverteilung ist intrinsisch die zugrunde liegende wahre Wahrscheinlichkeitsverteilung beim zählen von Häufigkeiten.
- Man unterscheidet zwei Fälle: (i) ist  $\nu$  unabhängig von den  $\vec{\theta}$  ist es einfach ein weiterer Parameter des Modells; (ii) ist  $\nu = \nu(\vec{\theta})$  trägt die Normierung der Stichprobe zur Reduktion der Varianz der Stichprobe bei.

# Binned Likelihood

- In Fällen, in denen die Länge der Stichprobe sehr groß wird kann es sich als aufwändig erweisen die vollständige Likelihood zu handhaben.
- In diesem Fall läßt sich der Informationsgehalt auf ein handbarbares Maß reduzieren, indem man die Skala von  $x$  in  $N$  bins aufteilt, und die beobachtete Häufigkeit pro bin mit der Erwartung vergleicht ( $\rightarrow$  Histogramm).
- Ein Beispiel aus der Teilchenphysik:



**NB:** Dieses Histogramm enthält einige 10k Einträge, aber nur 30 bins. Die Likelihood wäre also ein Produkt aus 30 statt 10k Einzelwahrscheinlichkeiten.

# Binned Likelihood

---

- Auch hier unterscheidet man zwei Fälle von unterschiedlicher Relevanz:
- **(i)** Normierung des Histogramms bekannt, d.h.  $\sum_{k \leq N} \nu_k = n_{\text{tot}}$  von aussen vorgegeben:

$$n_{\text{tot}} = \sum_{k \leq N} n_k \quad (\text{Beobachtete Häufigkeit})$$

$$\nu_k(\vec{\theta}) = n_{\text{tot}} \int_{x_k^{\min}}^{x_k^{\max}} p(x, \vec{\theta}) dx \quad (\text{Erwartete Häufigkeit/bin})$$

$$\mathcal{L}(\vec{n} | \vec{\nu}) = \frac{n_{\text{tot}}!}{n_1! \dots n_N!} \left( \frac{\nu_1}{n_{\text{tot}}} \right)^{n_1} \dots \left( \frac{\nu_N}{n_{\text{tot}}} \right)^{n_N} \quad (\text{Multinomialverteilung})$$

$$\ln(\mathcal{L}(\vec{n} | \vec{\nu})) = \sum_{k \leq N} n_k \ln(\nu_k(\vec{\theta})) \quad (*)$$

# Binned Likelihood

- Auch hier unterscheidet man zwei Fälle von unterschiedlicher Relevanz:
- **(ii)** Normierung des Histogramms nicht bekannt, d.h.  $\sum_{k \leq N} \nu_k = \nu_{\text{tot}}$  nicht vorgegeben:

$$n_{\text{tot}} = \sum_{k \leq N} n_k \quad (\text{Beobachtete Häufigkeit})$$

$$\nu_{\text{tot}} = \sum_{k \leq N} \nu_k \quad (\text{Erwartete Häufigkeit})$$

$$\mathcal{L}(\vec{n} | \vec{\nu}) = \frac{\nu_{\text{tot}}^{n_{\text{tot}}}}{n_{\text{tot}}!} e^{-\nu_{\text{tot}}} \frac{n_{\text{tot}}!}{n_1! \dots n_N!} \left( \frac{\nu_1}{\nu_{\text{tot}}} \right)^{n_1} \dots \left( \frac{\nu_N}{\nu_{\text{tot}}} \right)^{n_N} = \prod_{k \leq N} \frac{\nu_k^{n_k}}{n_k!} e^{-\nu_k}$$

$$\ln(\mathcal{L}(\vec{n} | \vec{\nu})) = \sum_{k \leq N} \left( -\nu_k + n_k \ln \left( \nu_k(\vec{\theta}) \right) \right) = -\nu_{\text{tot}} + \sum_{k \leq N} \left( n_k \ln \left( \nu_k(\vec{\theta}) \right) \right) \quad (*)$$

- **(ii)** ist die erweiterte binned Likelihood. Sie ist ein Arbeitspferd der Teilchenphysik. Wir werden im folgenden einige Beispiele betrachten.
- Machen Sie sich den Übergang von **(i)** nach **(ii)** anschaulich klar.

# Binned Likelihood

---

- **Anm.:** Für  $N \rightarrow \infty$ , d.h. sehr viele schmale bins geht die binned Likelihood in die unbinned Likelihood über, d.h. (im Gegensatz zur  $\chi^2$ -Statistik, die wir in dieser VL noch diskutieren werden) sind leere oder schwach populierte bins bei einer ML Anpassung kein Problem. Machen Sie sich diesen Übergang nochmal gedanklich klar.

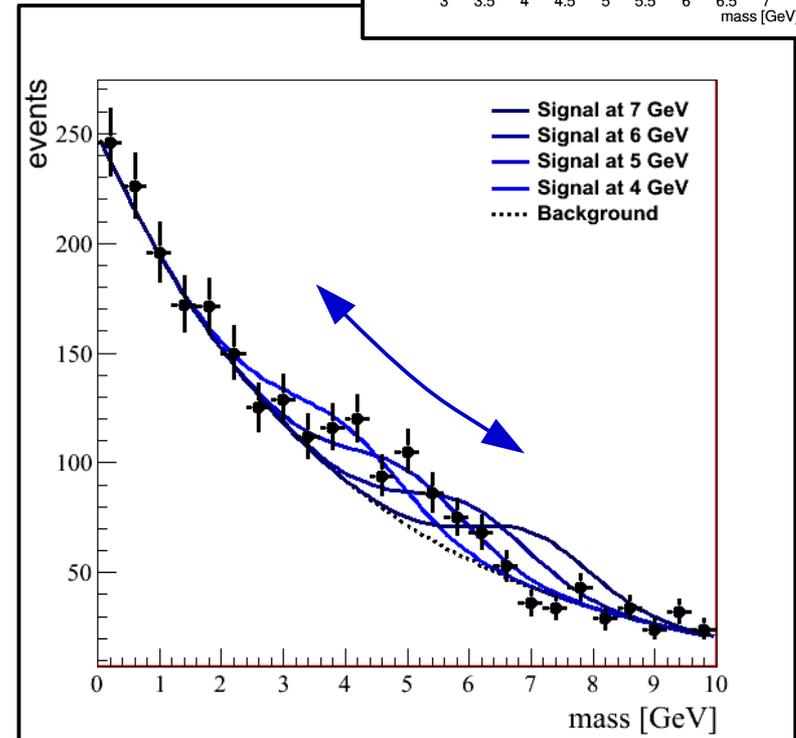
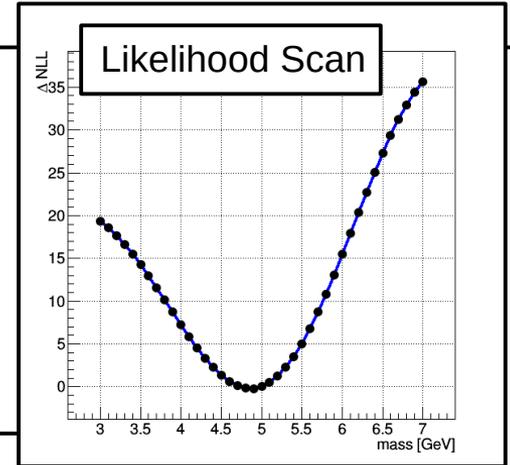
# Beispiel – 1

- Signal an unbekannter Stelle über bekanntem Untergrund:
  - Extended Binned Likelihood, unbekannte Normierung.
  - Physikalisches Modell mit vier freien Parametern.
  - *Parameter of interest* (POI): Lage des peaks ( $\theta_3$ ).

$$\mathcal{L}(\{x_k\}, \{\theta_j\}) = \prod_{k \leq 25} \mathcal{P}(\{x_k\}, \nu_k(\{\theta_j\}))$$

Produkt Poissondichte  
für 25 bins.

$$\nu_k(\{\theta_j\}) = \underbrace{\theta_0 e^{-\theta_1 x_k}}_{\text{Untergrund}} + \underbrace{\theta_2 e^{-(\theta_3 - x_k)^2}}_{\text{Signal}}$$



# Beispiel – 2

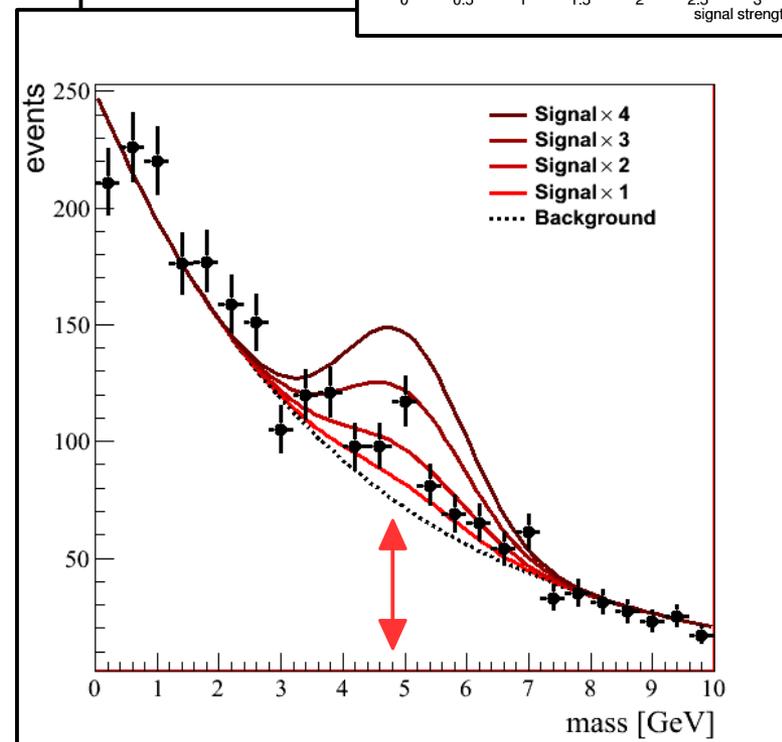
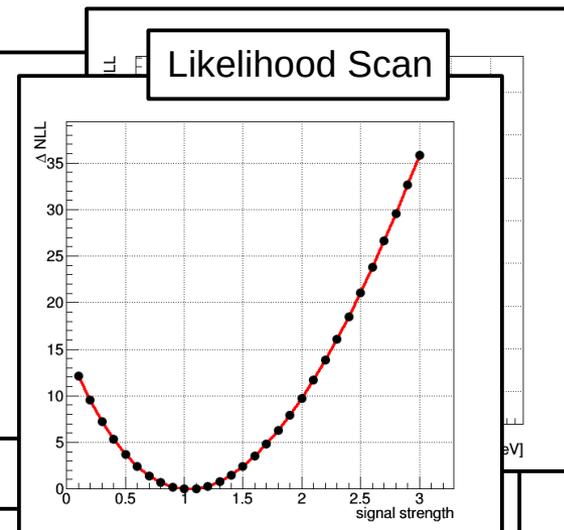
- Signal an unbekannter Stärke über bekanntem Untergrund:
  - Extended Binned Likelihood, unbekannte Normierung.
  - Physikalisches Modell mit vier freien Parametern.
  - *Parameter of interest (POI): Höhe des peaks* bei bekannter Lage ( $\theta_2$ ).

$$\mathcal{L}(\{x_k\}, \{\theta_j\}) = \prod_{k \leq 25} \mathcal{P}(\{x_k\}, \nu_k(\{\theta_j\}))$$

Produkt Poissondichte für 25 bins.

$$\nu_k(\{\theta_j\}) = \underbrace{\theta_0 e^{-\theta_1 x_k}}_{\text{Untergrund}} + \underbrace{\theta_2 e^{-(\theta_3 - x_k)^2}}_{\text{Signal}}$$

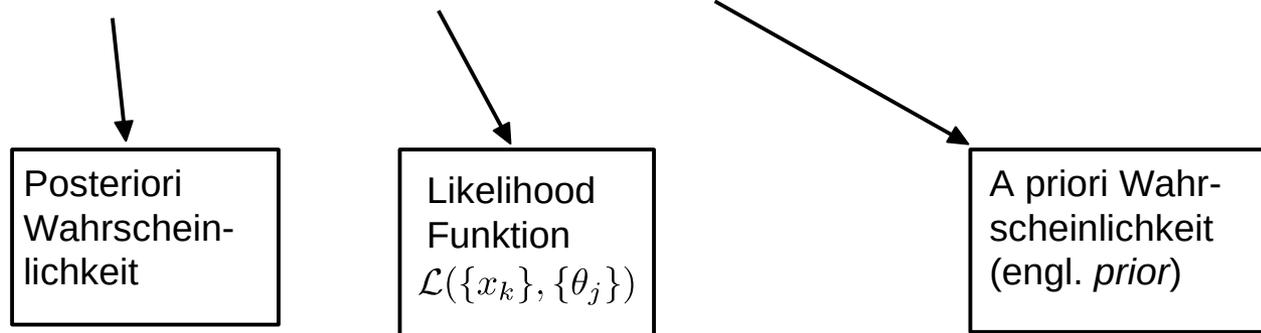
Ein lauffähiges RooT macro finden Sie [hier](#).



# Bayesianische Statistik vs. Likelihood

- Wir kehren an dieser Stelle nochmal zum Zusammenhang zwischen Bayesianischer Statistik und Likelihood zurück:

$$P_{\text{Daten}}(\text{Modell}) \propto P_{\text{Modell}}(\text{Daten}) \cdot P(\text{Modell})$$



- Wir erhalten die bedingte Bayesianische Wahrscheinlichkeit für das *Fürwahrhalten* des Modells (mit Parametern  $\{\theta_j\}$ ) bei Vorliegen der Daten  $\{x_k\}$  aus:

$$P_{\{x_k\}}(\{\theta_j\}) = \frac{\mathcal{L}(\{x_k\}, \{\theta_j\}) \pi(\{\theta_j\})}{\int \mathcal{L}(\{x_k\}, \{\theta'_j\}) \pi(\{\theta'_j\}) d\theta'}$$

$$\mathcal{P}_{\text{Daten}}(\text{Modell}) \equiv P_{\{x_k\}}(\{\theta_j\}) \quad (\text{Posterior})$$

$$\mathcal{P}_{\text{Modell}}(\text{Daten}) \equiv \mathcal{L}(\{x_k\}, \{\theta_j\}) \quad (\text{Likelihood})$$

$$\mathcal{P}(\text{Modell}) \equiv \pi(\{\theta_j\}) \quad (\text{Prior})$$

**Anm:**  $\mathcal{L}(\{x_k\}, \{\theta_j\})$  allein ist als Funktion von  $\{\theta_j\}$  noch keine Wahrscheinlichkeitsdichte.

# Bayesianische Statistik vs. Likelihood

---

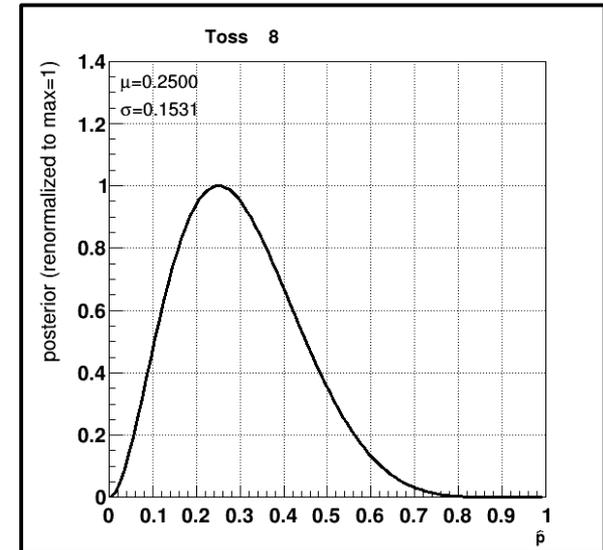
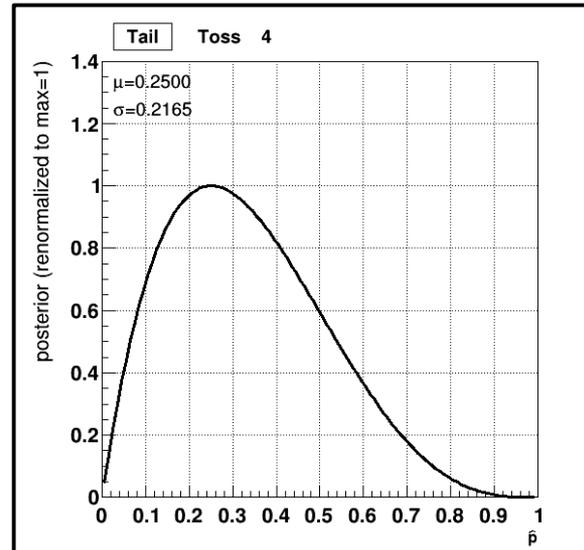
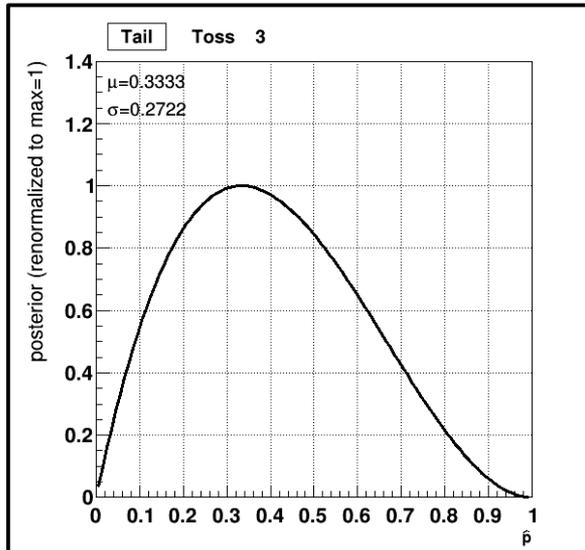
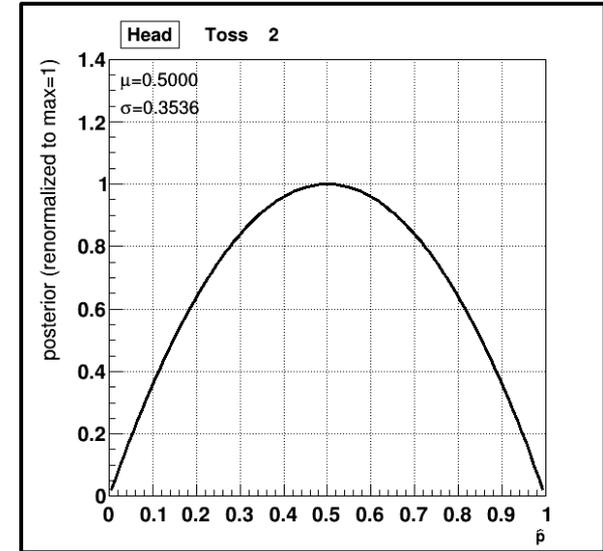
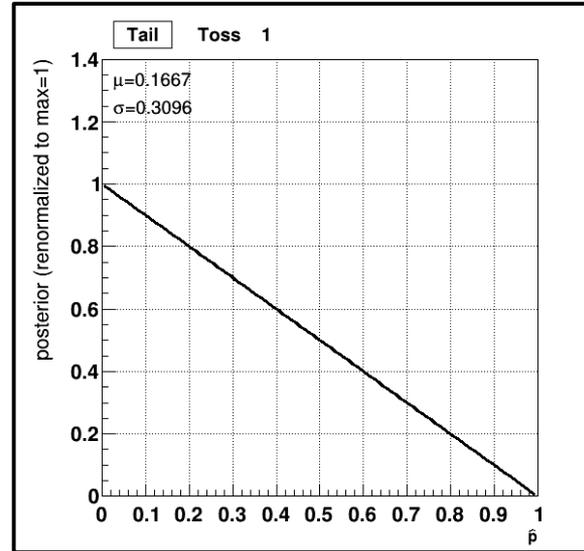
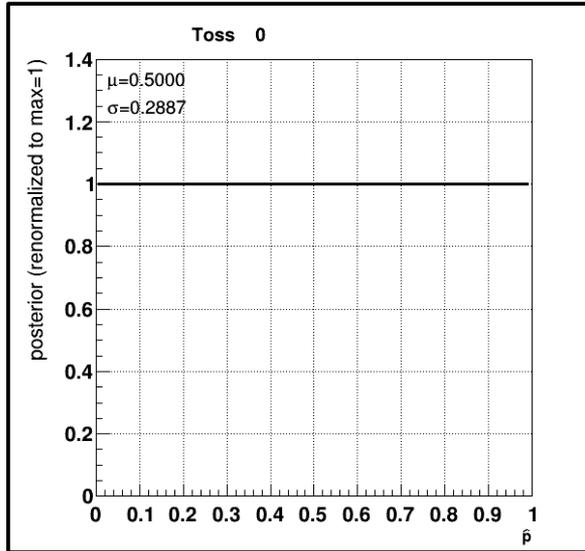
- Der Posterior  $P_{\{x_k\}}(\{\theta_j\})$  enthält im Bayesianischen Paradigma die volle Information über das Modell, nach Auswertung der Daten.
- Für  $\pi(\{\theta_j\}) = \text{const}$  (*Bayesianischer Prior* genannt) ist die ML Abschätzung zur Abschätzung durch den Bayesianischen *Posterior Modus* äquivalent.

# Bayesianische Statistik vs. Likelihood

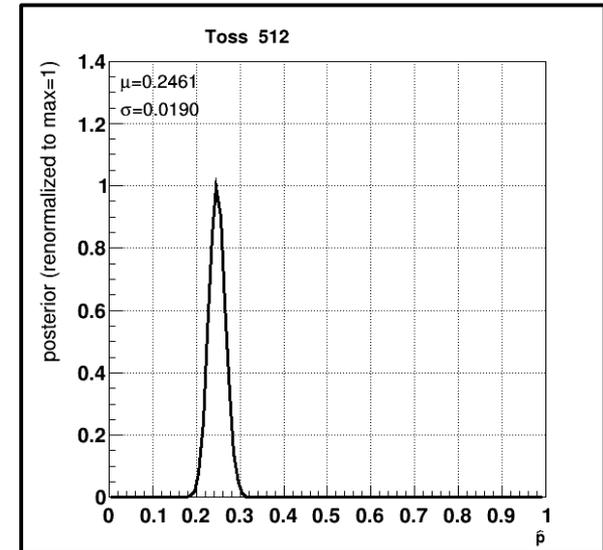
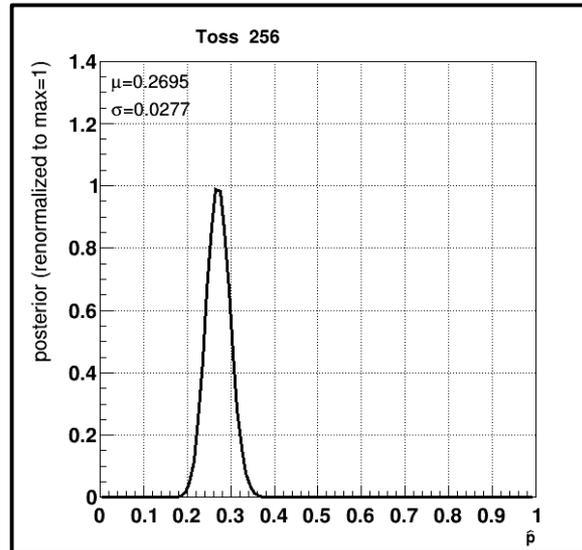
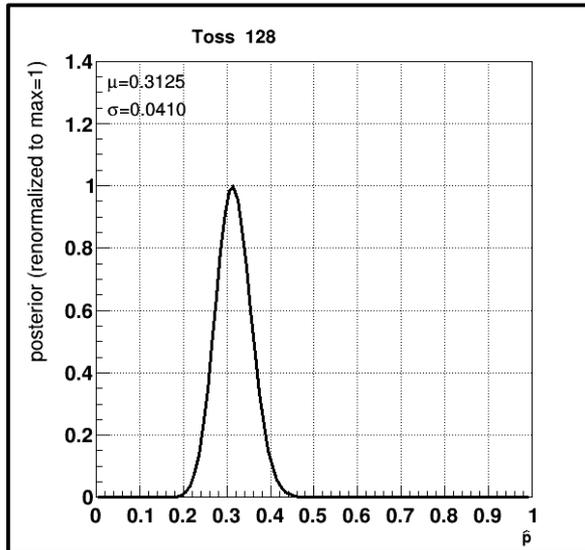
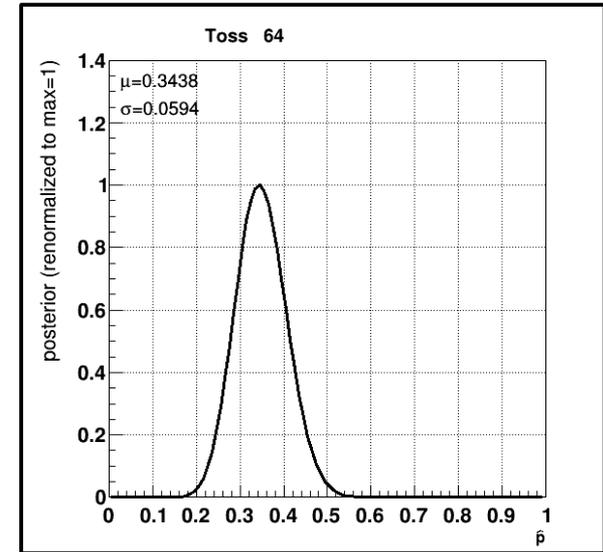
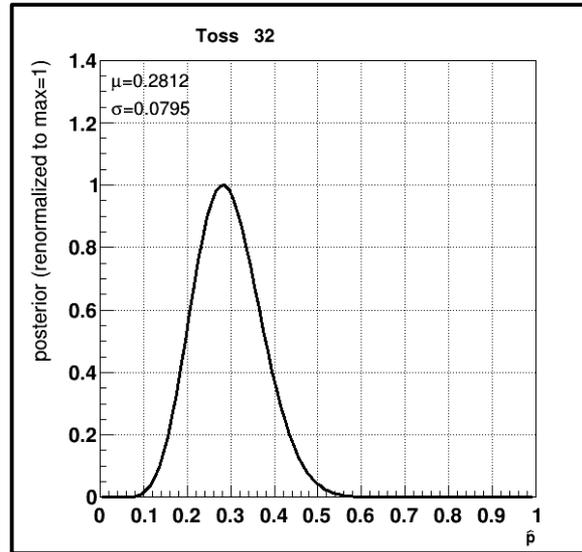
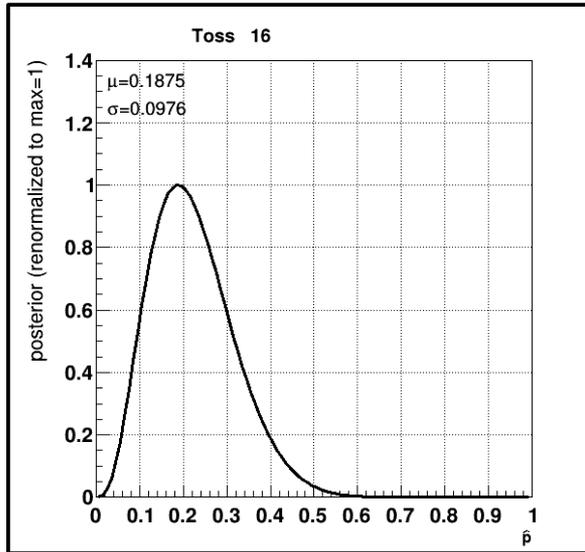
---

- Der Posterior  $P_{\{x_k\}}(\{\theta_j\})$  enthält im Bayesianischen Paradigma die volle Information über das Modell, nach Auswertung der Daten.
- Für  $\pi(\{\theta_j\}) = \text{const}$  (*Bayesianischer Prior* genannt) ist die ML Abschätzung zur Abschätzung durch den Bayesianischen *Posterior Modus* äquivalent.
- Wir geben zum Abschluss ein Beispiel wie man im Bayesianischen Paradigma den Posterior nach Auswertung der Daten sukzessive aktualisieren kann. Wir wählen hierzu die folgende Fragestellung:
  - $n$ -facher Münzwurf.
  - Wie hoch ist die Wahrscheinlichkeit  $\theta = p$  bei der Münze „Kopf“ zu erhalten?
  - Das zugrunde liegende Modell ist die Binomialverteilung  $B(p, n, k)$  nach  $n$ -fachem Wurf.
  - Vor dem ersten Wurf ist  $p$  gleichverteilt (Bayesianischer Prior).
  - Der Prior wird nach jedem Münzwurf aktualisiert,  $\hat{p} = \frac{k}{n}$ . Die Daten tragen so zur sukzessiven Verfeinerung des Modells bei.

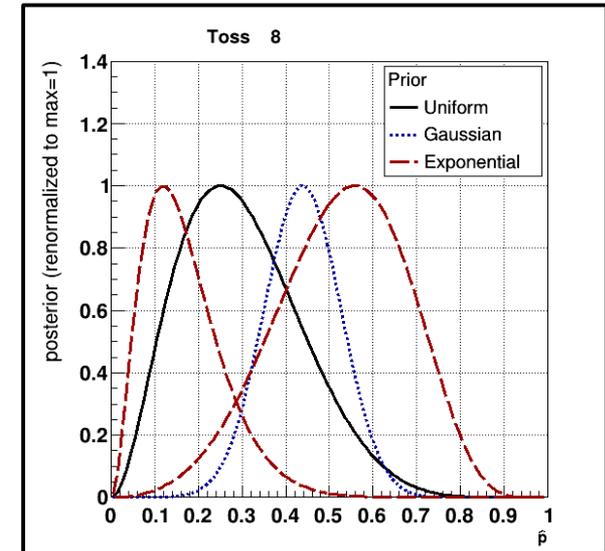
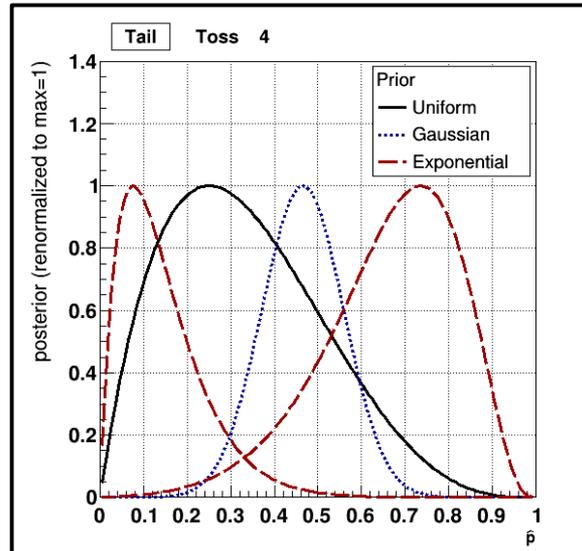
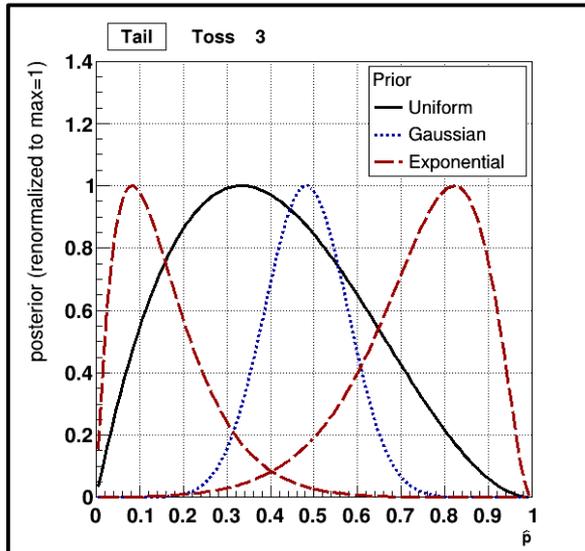
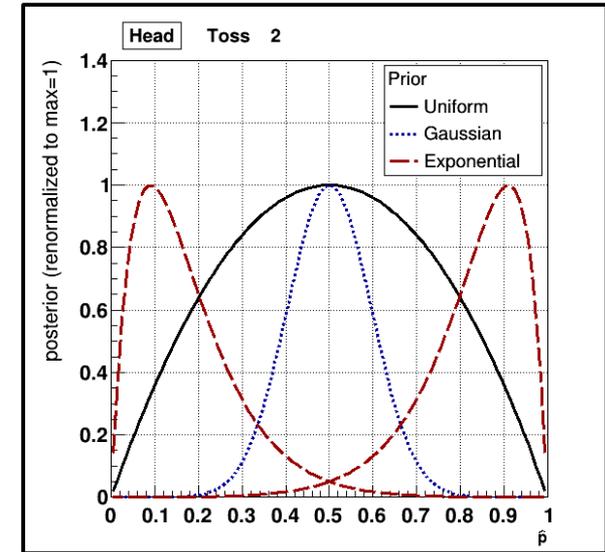
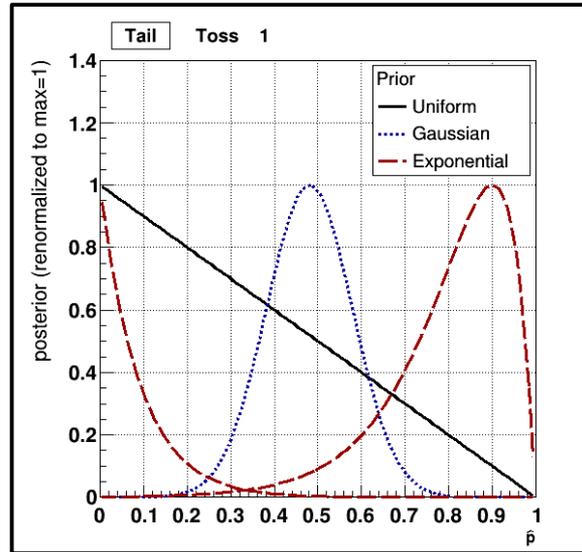
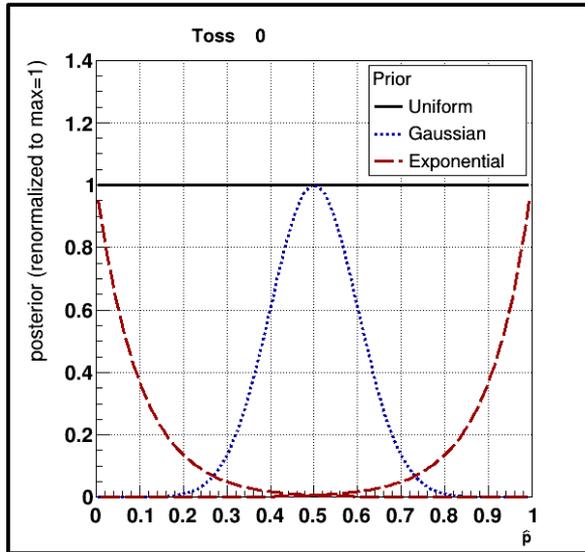
# Bayesianischer Posterior bei n-fachem Münzwurf



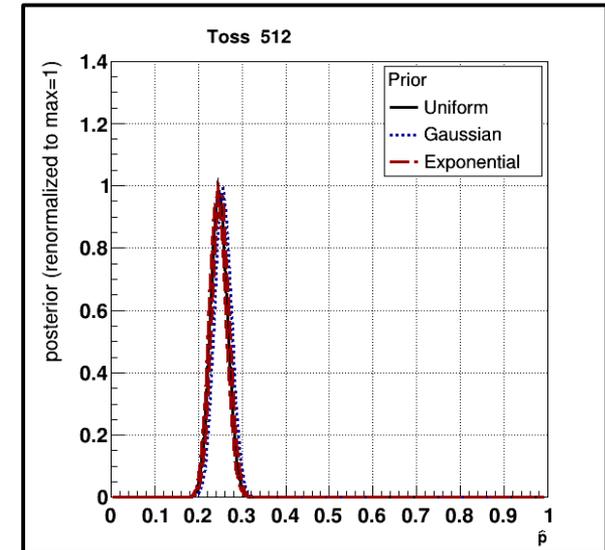
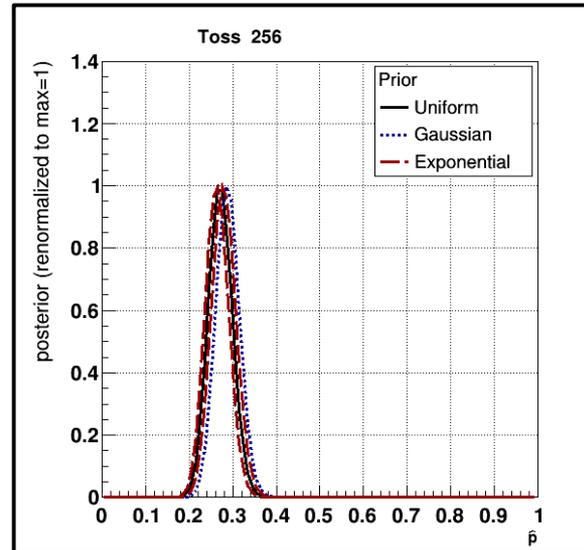
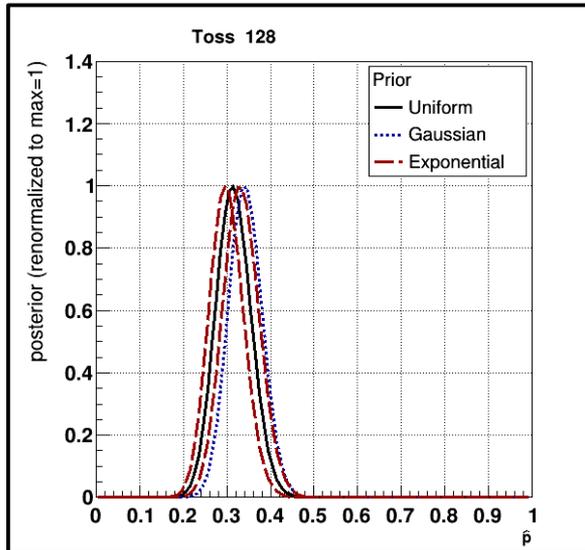
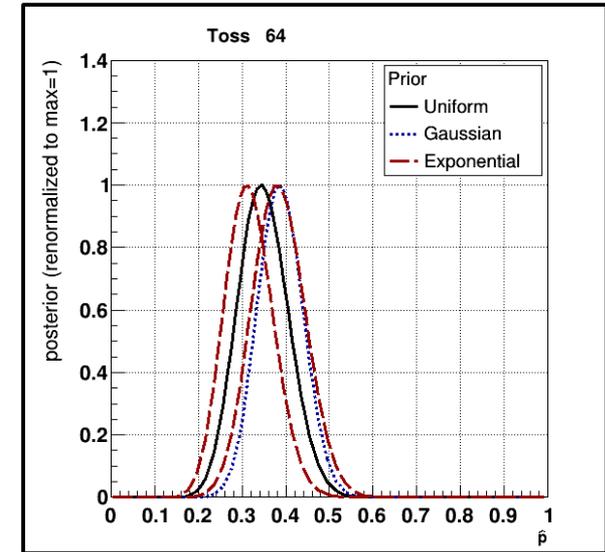
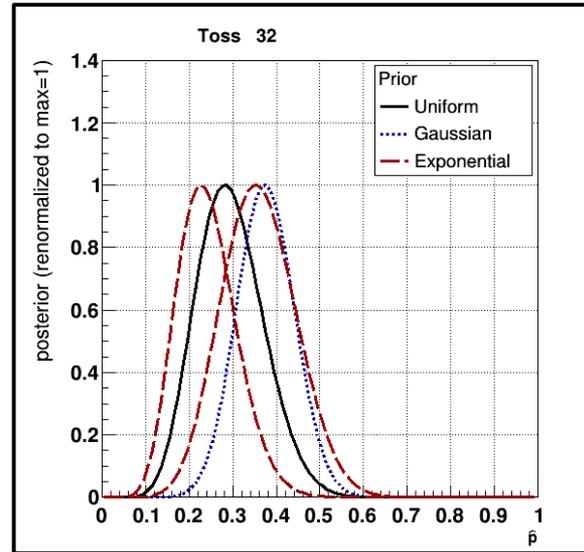
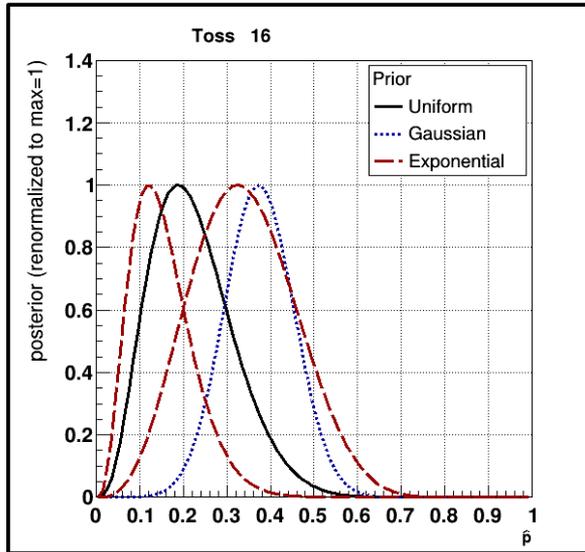
# Bayesianischer Posterior bei n-fachem Münzwurf



# Bayesianischer Posterior bei n-fachem Münzwurf



# Bayesianischer Posterior bei n-fachem Münzwurf

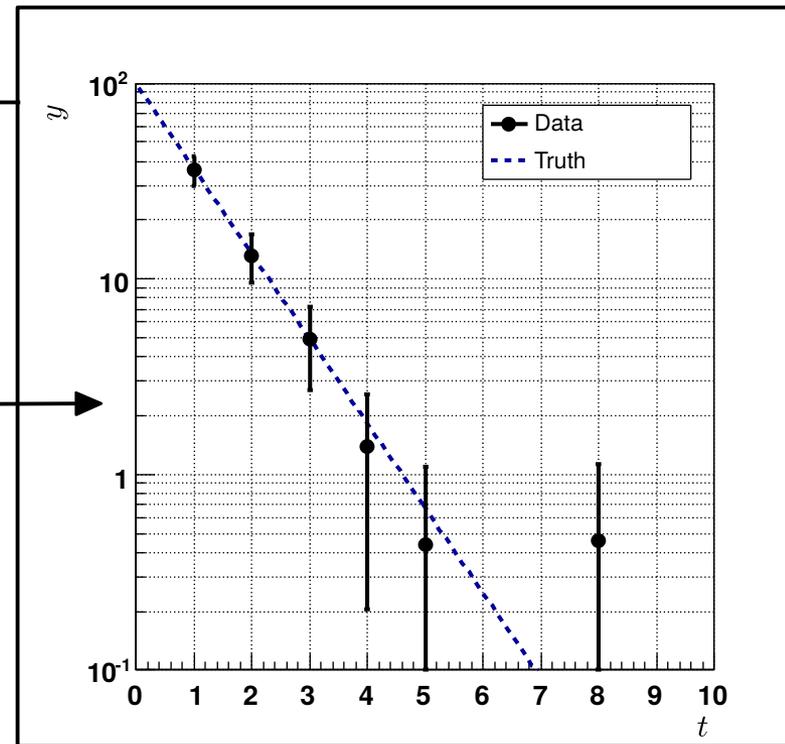


# Parameterschätzung mit Hilfe der $\chi^2$ Methode

# Likelihood $\rightarrow \chi^2$

- Für die weitere Diskussion nutzen wir das folgende Beispiel:

9 Messpunkte  $\{y_i\}$  mit Gaußischer Wahrscheinlichkeitsdichte mit  $\{(\mu_i, \sigma_i)\}$ , die einer exponentiellen Verteilung mit fester Normierung folgen (siehe Bild rechts).



- Likelihood Funktion:**

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

$$\ln(\mathcal{L}(\{y_i\}, \theta)) = - \underbrace{\sum_{i \leq 9} \frac{(y_i - \mu_i(\theta))^2}{2\sigma_i^2}}_{\equiv \chi^2(\theta)} + \underbrace{\sum_{i \leq 9} \frac{1}{2} \ln(2\pi\sigma_i^2)}_{\text{const.}}$$

- Der quadratische Abstand der Messpunkte von der wahren Verteilung,  $\chi^2(\theta)$ , folgt der Verteilung,  $\chi^2(x, n)|_{n=9}$ , in unserem Fall mit  $n = 9$  Freiheitsgraden.
- Die Likelihood wird maximal, wenn  $\chi^2(\theta)$  minimal wird.

# Erinnerung $\chi^2$ Funktion

$$\chi^2(x, n) = \frac{1}{2^{n/2} \Gamma(n/2)} x^{n/2-1} e^{-x/2}$$

$$\Gamma(x) = \int e^{-t} t^{x-1} dt \quad (\text{Gammafunktion})$$

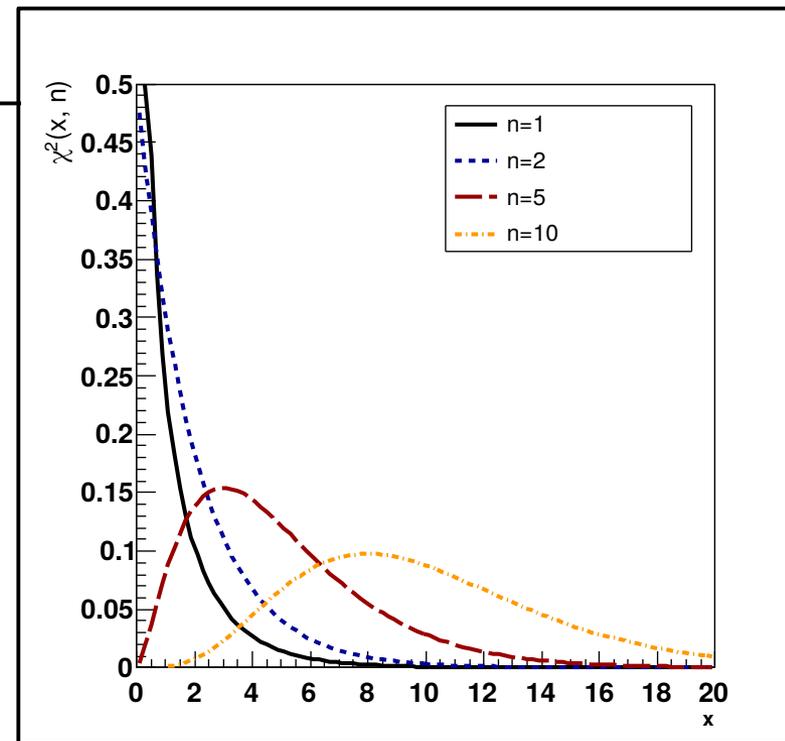
$$E[x] = n \quad (\text{Erwartungswert})$$

$$\text{var}[x] = 2n \quad (\text{Varianz})$$

- Die Summe der Quadrate von  $n$  normalverteilten Zufallsgrößen  $\{x_i\}$

$$\sum_{i \leq n} \frac{(x_i - \mu_i)^2}{\sigma_i^2} \quad \text{ist nach der Funktion } \chi^2(x, n) \text{ verteilt.}$$

- Die Zahl  $n$  heißt Freiheitsgrad. Sie entspricht der Anzahl unabhängiger Normalverteilungen in der Summe.
- Für große  $n \gg 1$  geht  $\chi^2(x, n)$  in eine Normalverteilung  $\varphi(x, n, \sqrt{2n})$  über.



# $\chi^2$ -Schätzfunktion

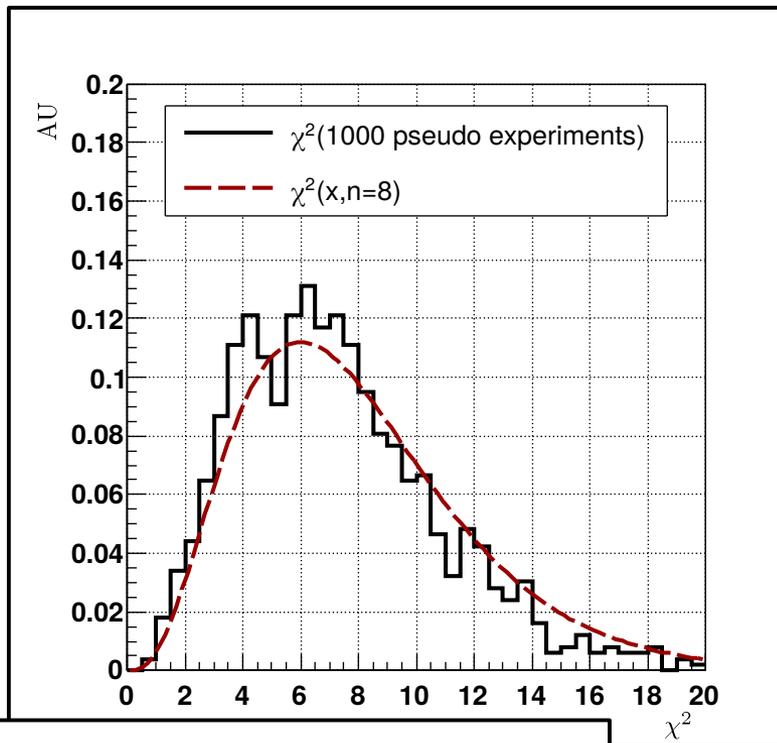
Seien  $\{y_i\}$  Einzelmessungen einer Messreihe, deren Erwartungswerte  $\{\mu_i\}$  unbekannt, deren Varianzen  $\{\sigma_i\}$  jedoch bekannt sind. Diese Einzelmessungen müssen nicht unabhängig sein, sondern können über eine bekannte Matrix  $V_{ij}$  korreliert sein. Die Funktion

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

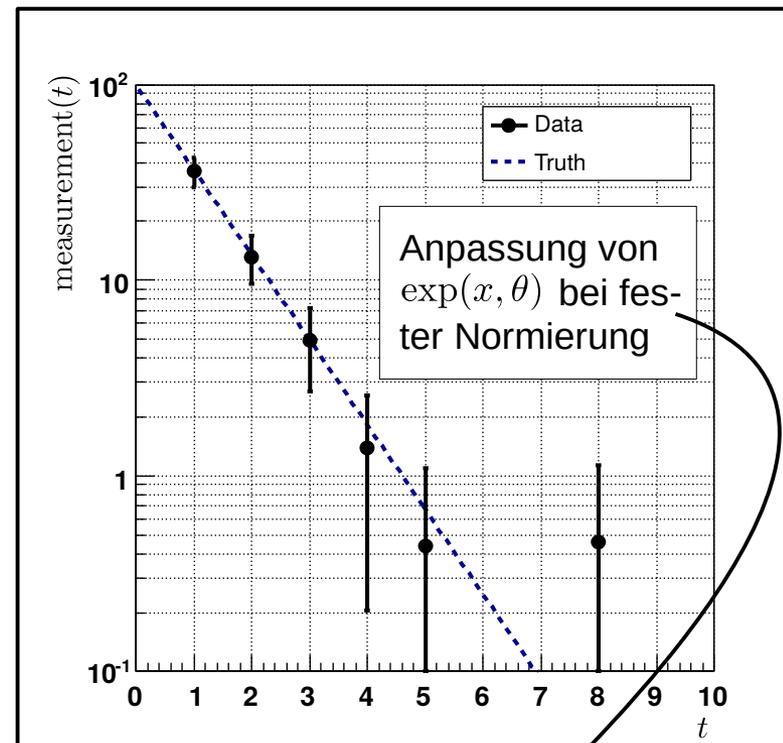
heißt Schätzfunktion der kleinsten Quadrate (Least Square, LS Schätzfunktion). Die Parameter  $\hat{\theta}_{\text{LS}}$ , die  $\chi^2(\theta)$  minimieren heißen Schätzwerte der kleinsten Quadrate (LS Schätzwerte).

- Sind die Messwerte  $\{y_i\}$  normalverteilt, dann ist die LS Schätzfunktion zur ML Schätzfunktion äquivalent.

# $\chi^2$ -Schätzfunktion



**NB:** Für dieses Histogramm habe ich die 9 Datenpunkte 1000 mal mit einer zugrunde liegenden Normalverteilung mit entsprechender Standardabweichung neu generiert. Danach habe ich die Summe der quadratischen Abstände der Messwerte von der angepassten Verteilung bestimmt und in dieses Histogramm abgetragen.



9 Messpunkte minus 1 freier Parameter  $\rightarrow$  8 Freiheitsgrade.

## Bei der Anpassung ist folgendes zu beachten:

Der quadratische Abstand der Messpunkte von der angepassten Verteilung,  $\chi^2(\hat{\theta})$ , folgt  $\chi^2(x, n - k)$ , wobei  $k$  der Anzahl der anzupassenden Parameter entspricht. Ein Messwert legt jeweils einen Parameter der anzupassenden Funktion fest und ist daher nicht mehr von den anderen Messwerten unabhängig.

# Lineare LS Schätzfunktion

- Die LS Schätzfunktion kann numerisch minimiert werden. Wenn sie linear von den Parametern  $\{\theta_j\}$  abhängt, d.h. wenn

$$\mu_i(\{\theta_j\}) = \sum_{j \leq n} A_{ij} \theta_j$$

ist sie auch analytisch lösbar. Die  $\{\mu_i\}$  hängen hier über die Matrix  $A$  linear von den  $\{\theta_j\}$  ab. In Matrixschreibweise schreibt sich  $\chi^2$  als:

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

- Um das Minimum zu finden setzen wir:

$$\vec{\nabla}_{\theta} \chi^2(\vec{\theta}) = -2 \left( A^\top V^{-1} \vec{y} - A^\top V^{-1} A \vec{\theta} \right) = 0$$

$$\vec{\hat{\theta}}_{\text{LS}} = \underbrace{(A^\top V^{-1} A)^{-1} A^\top V^{-1}}_{\equiv B} \vec{y}$$

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

$$\hat{\theta}_{\text{LS},k} = \sum_{i \leq n} B_{ik} y_i$$

# Varianz von $\theta_{LS}$

- Es gibt zwei offensichtliche Methoden die Varianz von  $\theta_{LS}$  auszurechnen:
- **Methode 1:** aus Fehlerfortpflanzung (d.h. Variablentransformation) von  $V$ .

$$\text{cov}(\theta_i, \theta_j) = BVB^T \equiv (A^T V^{-1} A)^{-1}$$

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

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

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

$$\underbrace{\hspace{10em}}_{\equiv 1}$$

Unter Verwendung von:

$$(AB)^T = B^T A^T$$

$$(A^{-1})^T = (A^T)^{-1}$$

# Vergleich mit der RCF-Ungleichung

- Für normalverteilte Zufallsgrößen gilt:

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

vgl. mit [Folie 22](#).

für normalverteilte  
Zufallsgrößen

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

vgl. mit [Folie 21](#).

- Für normalverteilte Zufallsgrößen ist die LS Abschätzung effizient.

# Varianz von $\theta_{LS}$

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

$$\chi^2(\theta) \approx \underbrace{\chi^2(\hat{\theta}_{LS})}_{\equiv \chi_{\min}^2} + \underbrace{\frac{1}{2} \left[ \frac{\partial^2}{\partial \theta^2} \chi^2(\theta) \right]_{\theta=\hat{\theta}_{LS}}}_{\hat{\sigma}^{-2}} (\theta - \hat{\theta}_{LS})^2$$

$$\chi^2(\hat{\theta}_{LS} \pm \hat{\sigma}) = \chi_{\min}^2 + 1$$

d.h. die Variation aus dem Minimum um  $\pm \hat{\sigma}$  bewirkt die Erhöhung von  $\chi^2(\hat{\theta}_{ML})$  um 1.

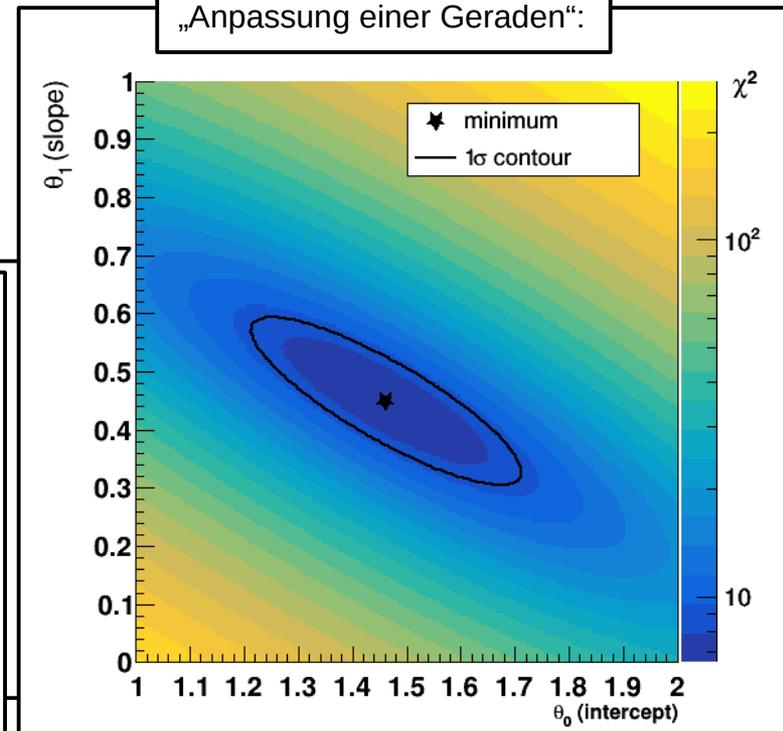
Beispiel in 1-dim.

**NB:** für n unabhängige Parameter müssen Sie die Differenz  $\Delta\chi^2$  zu  $\chi_{\min}^2$  als 68% Quantile einer  $\chi^2(x, n)$ -Funktion mit n Freiheitsgraden bestimmen. Das können Sie z.B. unter diesem [link](#) tun. Im folgenden sind einige Werte angegeben:

Dim.	$\Delta\chi^2(68\% \text{ CL})$	$\Delta\chi^2(95\% \text{ CL})$
1	1.0	3.8
2	2.3	6.0
3	3.5	7.8

Beispiel in 2-dim.

„Anpassung einer Geraden“:



Ein lauffähiges ROOT macro finden Sie [hier](#).

# Beispiel: Anpassung einer Gerade an 5 Messpunkte

$$\mu_i(\{\theta_j\}) = \sum_{j \leq n} A_{ij} \theta_j$$

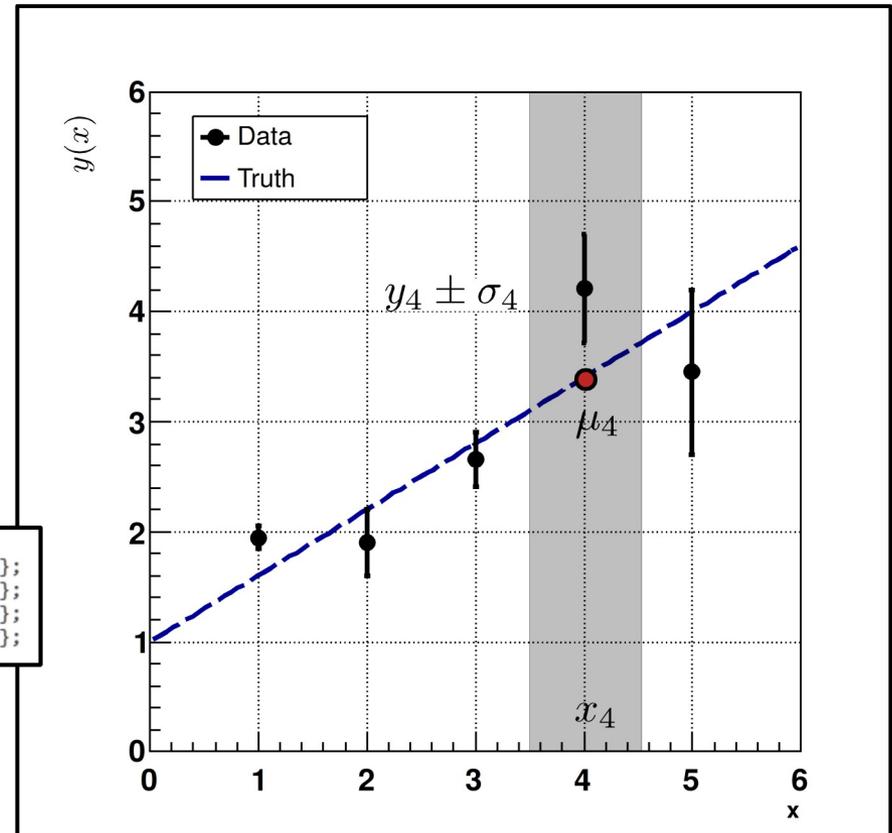
mit:

$$i = 1 \dots 5; \quad j = 0, 1; \quad A_{ij} = y_i^j$$

- Fünf unkorrelierte Datenpunkte  $\{y_i\}$  nach einer unbekanntem linearen Funktion normalverteilt:

```
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 mit y-Achsenabschnitt  $\theta_0$  und Steigung  $\theta_1$ .



Ein lauffähiges Root macro finden Sie [hier](#).

# Beispiel: Anpassung einer Geraden an

$$\underbrace{\begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_4 \end{pmatrix}}_{\vec{y}} = \underbrace{\begin{pmatrix} 1 & x_0 \\ 1 & x_1 \\ \vdots & \vdots \\ 1 & x_5 \end{pmatrix}}_A \cdot \underbrace{\begin{pmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \end{pmatrix}}_{\vec{\theta}}$$

$$V^{-1} = \begin{pmatrix} 1/\hat{\sigma}_0^2 & 0 & \dots & 0 \\ 0 & 1/\hat{\sigma}_1^2 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \dots & 1/\hat{\sigma}_4^2 \end{pmatrix}$$

$$\vec{\hat{\theta}}_{\text{LS}} = \underbrace{(A^T V^{-1} A)^{-1} A^T V^{-1} \vec{y}}_{\equiv B}$$

$$A^T V^{-1} A = \begin{pmatrix} \sum 1/\hat{\sigma}_i^2 & \sum x_i/\hat{\sigma}_i^2 \\ \sum x_i/\hat{\sigma}_i^2 & \sum x_i^2/\hat{\sigma}_i^2 \end{pmatrix}$$

$$(A^T V^{-1} A)^{-1} \stackrel{(4)}{=} \frac{1}{\det(A^T V^{-1} A)} \begin{pmatrix} \sum x_i^2/\hat{\sigma}_i^2 & -\sum x_i/\hat{\sigma}_i^2 \\ -\sum x_i/\hat{\sigma}_i^2 & \sum 1/\hat{\sigma}_i^2 \end{pmatrix}$$

$$A^T V^{-1} \vec{y} \stackrel{(4)}{=} \begin{pmatrix} \sum \mu_i/\hat{\sigma}_i^2 \\ \sum \mu_i x_i/\hat{\sigma}_i^2 \end{pmatrix}$$

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

# Beispiel: Anpassung einer Geraden an

$$\underbrace{\begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_4 \end{pmatrix}}_{\vec{y}} = \underbrace{\begin{pmatrix} 1 & x_0 \\ 1 & x_1 \\ \vdots & \vdots \\ 1 & x_5 \end{pmatrix}}_A \cdot \underbrace{\begin{pmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \end{pmatrix}}_{\vec{\theta}}$$

$$V^{-1} = \begin{pmatrix} 1/\hat{\sigma}_0^2 & 0 & \dots & 0 \\ 0 & 1/\hat{\sigma}_1^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/\hat{\sigma}_4^2 \end{pmatrix}$$

$$\vec{\hat{\theta}}_{\text{LS}} = \underbrace{(A^T V^{-1} A)^{-1} A^T V^{-1}}_{\equiv B} \vec{y}$$

$$A^T V^{-1} A = \begin{pmatrix} \sum 1/\hat{\sigma}_i^2 & \sum x_i/\hat{\sigma}_i^2 \\ \sum x_i/\hat{\sigma}_i^2 & \sum x_i^2/\hat{\sigma}_i^2 \end{pmatrix}$$

Korrelationsmatrix:

$$(A^T V^{-1} A)^{-1} \stackrel{(4)}{=} \frac{1}{\det(A^T V^{-1} A)} \begin{pmatrix} \sum x_i^2/\hat{\sigma}_i^2 & -\sum x_i/\hat{\sigma}_i^2 \\ -\sum x_i/\hat{\sigma}_i^2 & 1/\hat{\sigma}_i^2 \end{pmatrix}$$

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

$$A^T V^{-1} \vec{y} \stackrel{(4)}{=} \begin{pmatrix} \sum \mu_i/\hat{\sigma}_i^2 \\ \sum \mu_i x_i/\hat{\sigma}_i^2 \end{pmatrix}$$

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

# Beispiel: Anpassung einer Gerad an 5 Messpunkte

Ein lauffähiges ROOT macro finden Sie [hier](#).

```
std::pair<TMatrixD,TMatrixD> LLS(TMatrixD& A, TMatrixD& V, TMatrixD& X){
  TMatrixD U=TMMatrixD(A, TMMatrixD::kTransposeMult, TMMatrixD(V, TMMatrixD::kMult, A)).Invert(); /* (A^t*V*A)^(-1) */
  TMatrixD B=TMMatrixD(U, TMMatrixD::kMult, TMMatrixD(A, TMMatrixD::kTransposeMult, V)); /* (A^t*V*A)^(-1)*A^t*V */
  TMatrixD T=TMMatrixD(B, TMMatrixD::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[j]):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;
```

LS Schätzwerte:

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

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

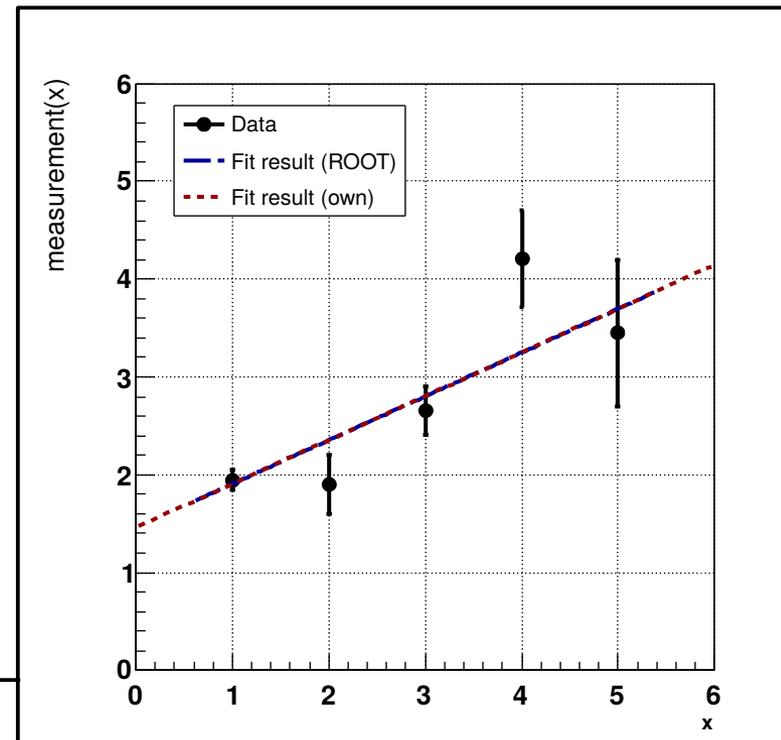
$$\rho(\hat{\theta}_0, \hat{\theta}_1) = -0.85$$

Wahrheit:

$$\theta_0 = 1.0$$

$$\theta_1 = 0.6$$

$$\vec{\hat{\theta}}_{\text{LS}} = \underbrace{(A^T V^{-1} A)^{-1} A^T V^{-1}}_{\equiv B} \vec{y}$$



# Beispiel: Anpassung einer Gerad an 5 Messpunkte

Ein lauffähiges Root macro finden Sie [hier](#).

```
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
* -----
const int NPARAM=2;
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:XVALUE
}
}
// std::cout << "PRINTING A
// inverse Korrelationsmatr
TMatrixD V(LENGTH,LENGTH);
for(int i=0;i<LENGTH;++i){
  for(int j=0;j<LENGTH;++j)
    V(i,j)=(i==j?1./(YERROR
}
}
// 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>
TMatrixD T = fit_result.first;
TMatrixD U = fit_result.second;
```

Durch geschickte Wahl können die Parameter  $\{\theta_j\}$  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

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

LS Schätzwerte:

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

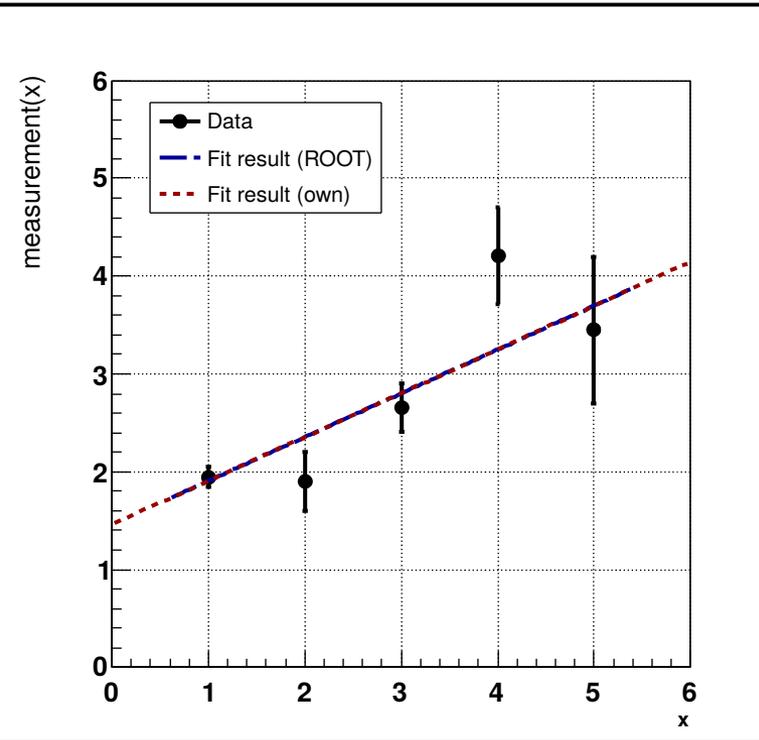
$$\hat{\theta}_1 = 0.45 \pm 0.09$$

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

Wahrheit:

$$\theta_0 = 1.0$$

$$\theta_1 = 0.6$$



$$\vec{\hat{\theta}}_{\text{LS}} = \underbrace{(A^T V^{-1} A)^{-1} A^T V^{-1}}_{\equiv B} \vec{y}$$

# Zusammenfassung

---

- Erweiterte/binning Likelihood.
- Zusammenhang zwischen Likelihood und Bayesianischen Parameterschätzungen.
- Parameterschätzung mit Hilfe der Least Squares Methode:
  - Übergang Likelihood → Least Squares (LS) Abschätzung.
  - Definition LS Abschätzung.
  - Lineare LS Abschätzung.
  - Varianz der LS Abschätzung, Auswertung der RCF-Ungleichung.
  - Graphische Abschätzung der Varianz.
  - Konkretes Beispiel: Anpassung einer Geraden an 5 Messpunkte.