Karlsruhe Institute of Technology

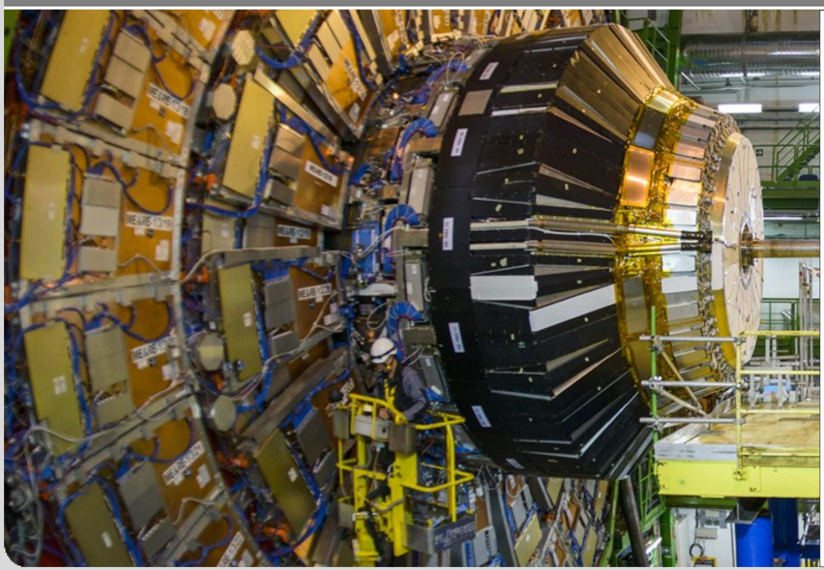
Rechnernutzung in der Physik

Teil 3 – Statistische Methoden in der Datenanalyse

Roger Wolf

12. Januar 2015

INSTITUTE OF EXPERIMENTAL PARTICLE PHYSICS (IEKP) – PHYSICS FACULTY



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

Zusammenfassung

Kapiel 3.5:

Monte Carlo Methoden

- Generatoren von (Pseudo-)Zufallszahlen (→ wichtige Eigenschaften, LCG).
- Transformation gleichverteilter Zufallszahlen auf beliebig verteilte Zufallszahlen (→ analytisch, rejection sampling, verbesserte Konvergenz).
- Monte Carlo als Integrationsmethode.
- Bedeutung in der (Teilchen-)Physik.

Kapiel 3.6:

Parameterschätzung

Stochastik vs Statistik

- Bisher haben wir uns mit der Stochastik beschäftigt: d.h. wir haben die wahre Wahrscheinlichkeitsdichte als bekannt vorrausgesetzt.

- In der Realität ist der Wahrscheinlichkeitsraum nicht (vollständig) bekannt!

- Charakterisierung muß aus begrenzten Stichproben abgeschätzt werden.

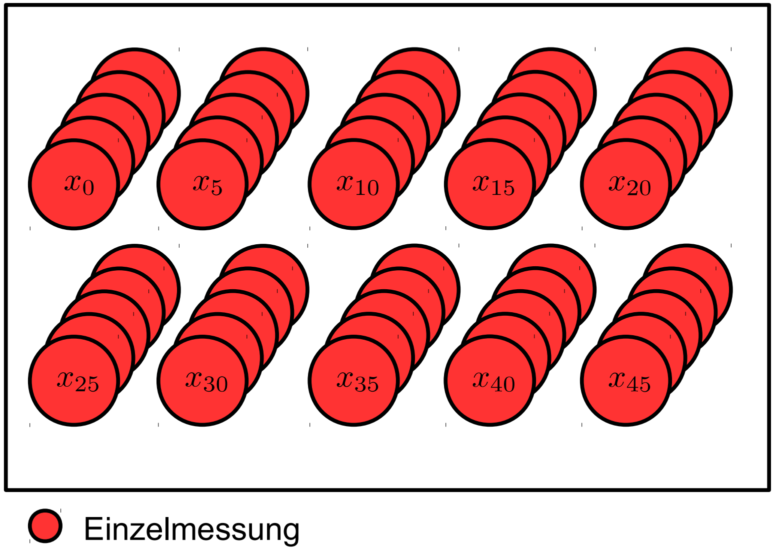
Sei x eine Zufallsvariable, die nach der Wahrscheinlichkeitsdichte $p(x)$ verteilt ist und A eine Menge von n unabhängigen Zufallsexperimenten. Die n Werte x_i können als n-dimensionaler Vektor $\vec{x} = (x_1, x_2, \ldots, x_n)$ zu einem neuen Ergebnisraum zusammengefaßt werden. Da die Einzelergebnisse x_i (stochastisch) unabhängig sind ist die Wahrscheinlichkeitsverteilung über dem neuen Ergebnisraum gegeben durch:

$$P_{\exp}(x_1, x_2, \ldots, x_n) = \prod_{i \leq n} p(x_i)$$

- $\vec{x} = (x_1, x_2, \ldots, x_n)$ kann als Ausgang eines Experiments/einer Meßreihe betrachtet werden.

Einzelmessung vs Meßreihe

Meßreihe aus 50 Einzelmessungen:



$$P_{\text{exp}}(x_1, x_2, \ldots, x_n) = \prod_{i \leq n} p(x_i)$$

- Eines der Grundprobleme der Statistik ist es aus den gemessenen Werten x_i auf die Eigenschaften von $p(x)$ zu schließen, wenn man $p(x)$ nicht kennt.

(Stochastisches) **Modell**

- Man konstruiert i.a. Modelle von $p(x, \vec{\theta})$ mit zusätzlichen unbekannten Parametern $\vec{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)$, die durch das Experiment abgeschätzt werden sollen (cf VL-08 slide 24).

Realität:

Mit welcher Wahrscheinlichkeit tritt ein reales Ereignis A_r ein?

Konstruktion des Modells.

Übertragung des mathematischen Resultats in die Realität.

Modell:

Wahrscheinlichkeitsraum $(\Omega, \mathcal{P})$, Modellereignis A.

Teststatistik

Eine Funktion einer beliebigen Anzahl beobachteter Einzelergebnisse $\{x_i\}$ heißt **Teststatistik**.

- Eine Teststatistik kann beliebig viele freie Parameter besitzen, sie kann ein mehrdimensionaler Vektor $\vec{t} = (t_1, t_2, \ldots, t_m)$ (für $n \geq m$ Einzelmessungen $\{x_i\}$) sein.[(1)]
- Die Teststatistik ist eine Funktion der zufallsverteilten $\{x_i\}$ und folgt damit selbst einer Wahrscheinlichkeitsverteilung $g(t(\{x_i\}))$.

(1)
Im Extremfall kann sie den Einzelmessungen selbst entprechen. Üblicherlweise ist t aber eine einfache skalare Funktion. Die Reduktion der Dimension erfolgt im allgemeinen aufgrund der besseren Handhabarkeit.

Schätzfunktion

Eine Statistik, zur Bestimmung der Eigenschaften einer Wahrscheinlichkeits-dichte heißt **Schätzfunktion** (Abschätzung), zur Bestimmung eines **Schätzwer-tes**.

- Die Schätzfunktion $\hat{\theta}$ einer Eigenschaft einer Wahrscheinlichkeitsdichte wird oft mit einem $\hat{.}$ bezeichnet, um sie vom wahren Wert θ zu unterscheiden.
- Wenn $\lim_{n \to \infty} \mathcal{P}(|\hat{\theta} - \theta| > \epsilon) = 0$, dann heißt $\hat{\theta}$ **konsistent**.
- Dabei ist n die Länge der Stichprobe/Meßreihe.

Schätzfunktion als Zufallsvariable

- Da die Schätzfunktion $\hat{\theta}(\{x_i\})$ eine Funktion der $\{x_i\}$ ist ist sie selbst eine Zufallsvariable. D.h. nach vielfacher Wiederholung des Experiments folgt der Ausgang der Experimente selbst einer Wahrscheinlichkeitverteilung $g(\hat{\theta}, \theta)$.

- **NB:** $g(\hat{\theta}, \theta)$ hängt im allgemeinen auch von den wahren Werten θ ab.

$g(\hat{\theta}, \theta)$ heißt **Stichprobenverteilung**. Der **Erwartungswert der Schätzfunktion** $\hat{\theta}$ ist definiert als:

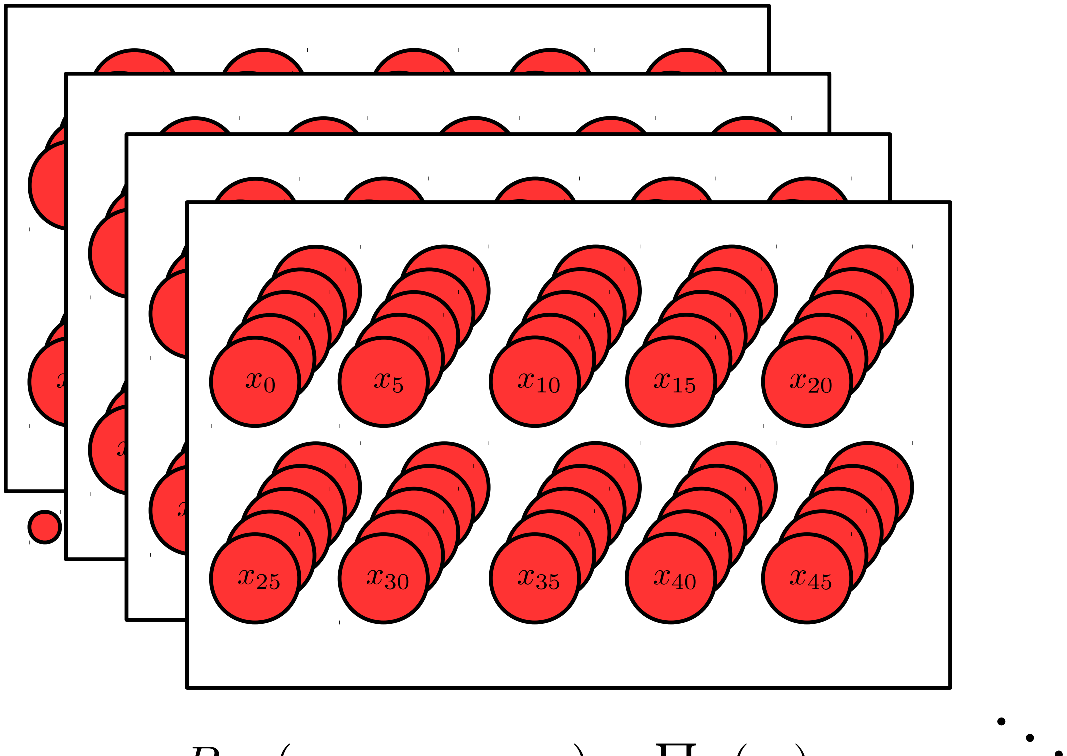
$$E[\hat{\theta}(\vec{x})] = \int \hat{\theta} \cdot g(\hat{\theta}, \theta) \mathrm{d}\hat{\theta}$$

$$= \int \cdots \int \hat{\theta}(\{x_i\}) \cdot \prod_{i \leq n} p(x_i, \theta) \mathrm{d}x_i$$

Der Erwartungswert ist definiert für eine unendliche große Stichprobe von Experimenten der Länge n.

Vielfache Wiederholung des Experiments

Experiment aus 50 Einzelmessungen:



$$P_{\exp}(x_1, x_2, \ldots, x_n) = \prod_{i \leq n} p(x_i)$$

Auf diesem Raum schätzen wir Eigenschaften von $P_{\exp}(\{x_i\})$ ab, selbst wenn die eigentliche Form unbekannt bleibt.

Verzerrung (=bias)

Die Größe

$$b = E[\hat{\theta}] - \theta$$

heißt **Verzerrung** (***bias***) der Schätzfunktion.

- **NB:** der *bias* der Schätzfunktion hängt nicht von den Einzelmessungen $\{x_i\}$ ab sondern von der Stichprobenlänge der Meßreihe, der funktionalen Form der Schätzfunktion, den Eigenschaften von $p(x)$ und den wahren Werten $\vec{\theta}$.
- Ein Parameter für den der *bias* unabhängig von der Stichprobenlänge der Meßreihe 0 ist heißt **erwartungtreu**, ein Parameter für den $b = 0$ für $n \to \infty$ heißt asymptotisch erwart-ungstreu.
- **NB:** ein Parameter kann verzerrt sein, selbst wenn er konsistent ist.
- **Problem:** wie kann ich den *bias* bestimmen, wenn mir der wahre Wert des Parameters θ nicht bekannt ist?

Verzerrung (=bias)

Die Größe

$$b = E[\hat{\theta}] - \theta$$

heißt **Verzerrung** (***bias***) der Schätzfunktion.

- **NB:** der *bias* der Schätzfunktion hängt nicht von den Einzelmessungen $\{x_i\}$ ab sondern von der Stichprobenlänge der Meßreihe, der funktionalen Form der Schätzfunktion, den Eigenschaften von $p(x)$ und den wahren Werten $\vec{\theta}$.
- Ein Parameter für den der *bias* unabhängig von der Stichprobenlänge der Meßreihe 0 ist heißt **erwartungtreu**, ein Parameter für den $b = 0$ für $n \to \infty$ heißt asymptotisch erwartungstreu.
- **NB:** ein Parameter kann verzerrt sein, selbst wenn er konsistent ist.
- **Problem:** wie kann ich den *bias* bestimmen, wenn mir der wahre Wert des Parameters θ nicht bekannt ist? Anwort von Fall zu Fall verschieden (z.B. Erwartungstreue in den Schätzfunktionen, die im folgenden diskutiert werden allgemein beweisbar, in anderen Fällen in der Praxis Überprüfung mit Hilfe von Modellen und Simulation)

Statistische und Systematische Unsicherheiten

Die Größe

$$\mathrm{MSE} = E[(\hat{\theta}-\theta)^2] = E[(\hat{\theta}-E[\hat{\theta}])^2] + (E[\hat{\theta}-\theta])^2 = \mathrm{var}[\hat{\theta}] + b^2 = \sigma^2_{\hat{\theta}} + b^2_{\hat{\theta}}$$

heißt **Genaugkeit** (***mean squared error***) der Schätzfunktion. In einer physikalischen Messung beschreibt $\mathrm{var}[\hat{\theta}]$ die statistische Unsicherheit der Messung und b die systematische Unsicherheit.

- Wir weisen die angegebene Beziehung nach:

LHS: $E[(\hat{\theta}-\theta)^2] \stackrel{(2)}{=} E[\hat{\theta^2}] - 2\theta E[\hat{\theta}] + \theta^2$

RHS: $E[(\hat{\theta}-E[\hat{\theta}])^2] = E[\hat{\theta^2}] - E[\hat{\theta}]^2$

$(E[\hat{\theta}-\theta])^2 = E[\hat{\theta}]^2 - 2\theta E[\hat{\theta}] + \theta^2$

$E[(\hat{\theta}-E[\theta])^2] + (E[\hat{\theta}-\theta]])^2 = E[\hat{\theta^2}] - E[\hat{\theta}]^2 + E[\hat{\theta}]^2 - 2\theta E[\hat{\theta}] + \theta^2 = E[\hat{\theta^2}] - 2\theta E[\hat{\theta}] + \theta^2$ qed

- Abbildung in die physikalische Messung (siehe VL-08 slide 6) durch statistische und systematische Unsicherheit.

(2) $E[\theta] = \int \theta g(\hat{\theta},\theta)\mathrm{d}\hat{\theta} = \theta \cdot \int g(\hat{\theta},\theta)\mathrm{d}\hat{\theta} = \theta$

Mittelwert der Stichprobe (Meßreihe)

Die Größe $\overline{x} = \frac{1}{n}\sum\limits_{i \leq n} x_i$ heißt **Mittelwert der Stichprobe**. Sie hat den Erwartungswert

$$E[\overline{x}] = E[\tfrac{1}{n}\sum_{i \leq n} x_i] = \tfrac{1}{n}\sum_{i \leq n} E[x_i] = \tfrac{1}{n}\sum_{i \leq n} \mu = \mu$$

und die Varianz

$$\mathrm{var}[\overline{x}] = E[\overline{x}^2] - E[\overline{x}]^2 = E\left[\left(\tfrac{1}{n}\sum_{i \leq n} x_i\right)\left(\tfrac{1}{n}\sum_{j \leq n} x_j\right)\right] - \mu^2 = \tfrac{1}{n^2}\sum_{i,j \leq n} E[x_i x_j] - \mu^2$$

$$= \tfrac{1}{n^2}\left[\underbrace{(n^2 - n)\mu^2} + \underbrace{n(\mu^2 + \sigma^2)}\right] - \mu^2 = \sigma^2/n$$

$n(n-1)$ off-diagonale Elemente ($i \neq j$) mit:

$E[x_i x_j] = E[x_i]E[x_j]$
$= \mu^2$

n diagonale Elemente ($i = j$) mit:

$E[x^2] = \sigma^2 + \mu^2$

Der Mittelwert der Stichprobe ist erwartungstreu unabhängig von $p(x)$.

Varianz der Stichprobe (Meßreihe)

Die Größe $s^2 = \frac{1}{n-1}\sum_{i \leq n}(x_i - \overline{x})^2 = \frac{n}{n-1}(\overline{x^2} - \overline{x}^2)$ heißt **Varianz der Stichprobe**.

Sie hat den Erwartungswert σ^2 und die Varianz

$$\mathrm{var}[s^2] = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-4}\mu_2^2\right)$$

wobei μ_k das k-te algebraische Moment um μ ist.[(3)]

- Die μ_k können durch $m_k = \frac{1}{n-1}\sum(x_i - \overline{x})^k$ abgeschätzt werden.

Die Varianz der Stichprobe ist erwartungstreu unabhängig von $p(x)$.

(3) cf. VL-09 slide 16.

Korrelationskoeffizient der Stichprobe (Meßreihe)

Die Größe

$$r = \frac{\hat{V}_{xy}}{s_x s_y} = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sum(x_j - \overline{x})^2 \sum(y_k - \overline{y})^2} = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\sqrt{(\overline{x^2} - \overline{x}^2)(\overline{y^2} - \overline{y}^2)}}$$

ist eine **Schätzfunktion für den Korrelationskoeffizienten** zweier Einzelvariablen x und y. Sie hat den Erwartungswert

$$E[r] = r - \frac{r(1-r^2)}{2n} + \mathcal{O}(1/n^2)$$

und die Varianz

$$V[r] = \frac{1}{n}(1 - r^2)^2 + \mathcal{O}(1/n^2)$$

- Die Schätzfunktion r ist nur asymptotisch erwartungstreu.
- Obwohl sowohl $\hat{V}_{xy}$, s_x als auch s_y erwartungstreue Schätzfunktionen sind ist die nichtlineare Funktion aus diesen drei Schätzfunktionen nicht erwartungstreu.

Kapiel 3.6:

Parameterschätzung

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

Kapiel 3.7:

Parameterschätzung mit Hilfe der Maximum Likelihood Methode

Präzisierung: Hypothese

- Im weiteren werden wir uns mit der Auswahl und Bewertung von Modellen in Form von Hypothesen beschäftigen (cf VL-08 slide 23).

> Eine (statistische) **Hypothese** ist eine Annahme, die mit Methoden der mathematischen Statistik auf Basis empirischer Daten überprüft werden kann.

- (Überprüfbare) Vorhersagen für die Wahrscheinlichkeit des Eintretens eines Ereignisses.

- Eine Hypothese heißt **einfach**, wenn sie direkt Wahrscheinlichkeitsaussagen zuläßt (→ Hypothesentests VL-12), oder **zusammengesetzt** wenn sie unbekannte (d.h. noch zu bestimmende) Parameter, θ_i , enthält (→diese Vorlesung).

Hypothese vs Bayesianische Statistik

- **Beachte:** Zusammenhang mit Bayesianischer Wahrscheinlichkeitsinterpretation („Fürwahrhalten einer Hypothese“, cf VL-08 slide 23):

$$\mathcal{P}_{\mathrm{Daten}}(\mathrm{Modell}) \propto \mathcal{P}_{\mathrm{Modell}}(\mathrm{Daten}) \cdot \mathcal{P}(\mathrm{Modell})$$

Posteriori Wahrscheinlichkeit

Likelihood Funktion

A priori Wahrscheinlichkeit (*prior*)

Wahrscheinlichkeitsraum Ω
$\rightarrow$ Hypothesenraum.

Likelihood Funktion

Sei x eine Zufallsvariable, die nach einer Wahrscheinlichkeitsdichte $p(x, \theta)$ verteilt ist. Dabei sei θ (mindestens) ein unbestimmter Parameter. Seien weiterhin $\{x_i\}, i = 1 \ldots, n$ n Einzelmessungen von x. Dann ist die Wahrscheinlichkeit dafür die Einzelmessung x_i im Interval $[x_i, x_i + \mathrm{d}x]$ zu finden gegeben durch

$$P(x_i, \theta) = p(x_i, \theta)\mathrm{d}x_i$$

Die Wahrscheinlichkeit für den Ausgang des Experimentes ist gegeben durch

$$P(\{x_i\}, \theta) = \prod_{i=1}^{n} p(x_i, \theta)\mathrm{d}x_i$$

Die Funktion

$$\mathcal{L}(\{x_i\}, \theta) = \prod_{i=1}^{n} p(x_i, \theta)$$

(als Funktion des unbestimmten Parameters θ) heißt **Likelihood Funktion**.

Maximum Likelihood Prinzip

- Die Wahrscheinlichkeitsdichte $p(x,\theta)$ ist durch die Hypothese vorgegeben. D.h. wir nehmen an, daß die parametrische Form von $p(x,\theta)$ bekannt und richtig ist.

- Wir können dann davon ausgehen, daß die Wahrscheinlichkeit, $P(\{x_i\},\theta)$ für das Vorliegen von $\{x_i\}$ Einzelmessungen für den korrekten Wert von θ höher ist, als für jeden anderen Wert.

- Da die $\mathrm{d}x_i$ nicht von θ abhängen darf die gleiche Annahme über $\mathcal{L}(\theta)$ gemacht werden.

Wir bezeichnen das Maximum von $\mathcal{L}(\theta)$ als **Maximum Likelihood Schätzfunktion** $\hat{\theta}_{ML}$ des Parameters θ.

Wenn $\mathcal{L}(\theta)$ (mindestens zwei mal) stetig differenzierbar ist ist $\hat{\theta}_{ML}$ gegeben durch $\frac{\partial}{\partial\theta}\mathcal{L}(\theta) \stackrel{!}{=} 0$ wenn $\frac{\partial^2}{(\partial\theta)^2}\mathcal{L}(\theta) < 0$.

Maximum Likelihood Prinzip (Log-Likelihood)

- $\mathcal{L}(\theta)$ ist selbst eine Wahrscheinlichkeitsdichte. Insbesondere beim Vorliegen vieler Einzelmessungen kann $\mathcal{L}(\theta)$ sehr kleine Werte annehmen.

- Es erweißt sich sehr oft als praktikabel statt $\mathcal{L}(\theta)$ den Logarithmus $\ln(\mathcal{L}(\theta))$ zu verwenden.(4)

Vorteile:	Zuordnung:
• Kleine Zahlen lassen sich besser darstellen und sind numerisch besser zu verarbeiten. • Produkte transformieren sich in Summen.	• θ korrekt: $\mathcal{L}(\theta)$ groß, $\ln(\mathcal{L}(\theta))$ groß. • θ falsch: $\mathcal{L}(\theta)$ klein, $\ln(\mathcal{L}(\theta))$ klein.

- In der Praxis sucht man oft nach dem Minimum der **Negative Log Likelihood** $-\ln(\mathcal{L}(\theta))$ (oder auch $-2\ln(\mathcal{L}(\theta))$, cf VL-12).

(4) Da der Logarithmus eine monotone (und damit ein-eindeutige) Funktion ist bleibt die Position des Maximums erhalten.

Maximum Likelihood Prinzip (Transformationsinvarianz)

Karlsruhe Institute of Technology

- Es kann sein, daß Sie sich nicht direkt für die ML Schätzfunktion für θ sondern für die ML Schätzfunktion einer Funktion $a(\theta)$ interessieren. In diesem Fall gilt:

$$\frac{\partial}{\partial\theta}\mathcal{L}(a(\theta)) = \frac{\partial}{\partial a}\mathcal{L}(a(\theta)) \cdot \frac{\partial}{\partial\theta}a(\theta) \stackrel{!}{=} 0$$

d.h. für $\frac{\partial}{\partial\theta}a(\theta) \neq 0$ gilt:

$$\frac{\partial}{\partial\theta}\mathcal{L}(a(\theta)) \stackrel{!}{=} 0 \qquad \Rightarrow \qquad \frac{\partial}{\partial a}\mathcal{L}(a'(\theta)) \stackrel{!}{=} 0$$

Man erhält also $\hat{a}_{ML}$ als $\hat{a}_{ML}(\theta) = a(\hat{\theta}_{ML})$. Man bezeichnet diese Eigenschaft als **Transformationsinvarianz**.

Beispiel: $\exp(x,\theta) = \frac{1}{\theta} e^{-x/\theta}$ [5]

- 50 Ereignisse verteilt nach $\exp(x,\theta)$ mit $\theta = 1$ (Bsp: radioaktiver Zerfall mit der Halbwertszeit θ).

- **Likelihood Funktion**:

$$\mathcal{L}(\{x_i\},\theta) = \prod_{i=1}^{50} \frac{1}{\theta} e^{-x_i/\theta}$$

$$\ln(\mathcal{L}(\{x_i\},\theta)) = \sum_{i=1}^{50} (-\ln(\theta) - {}^{x_i}/\theta)$$

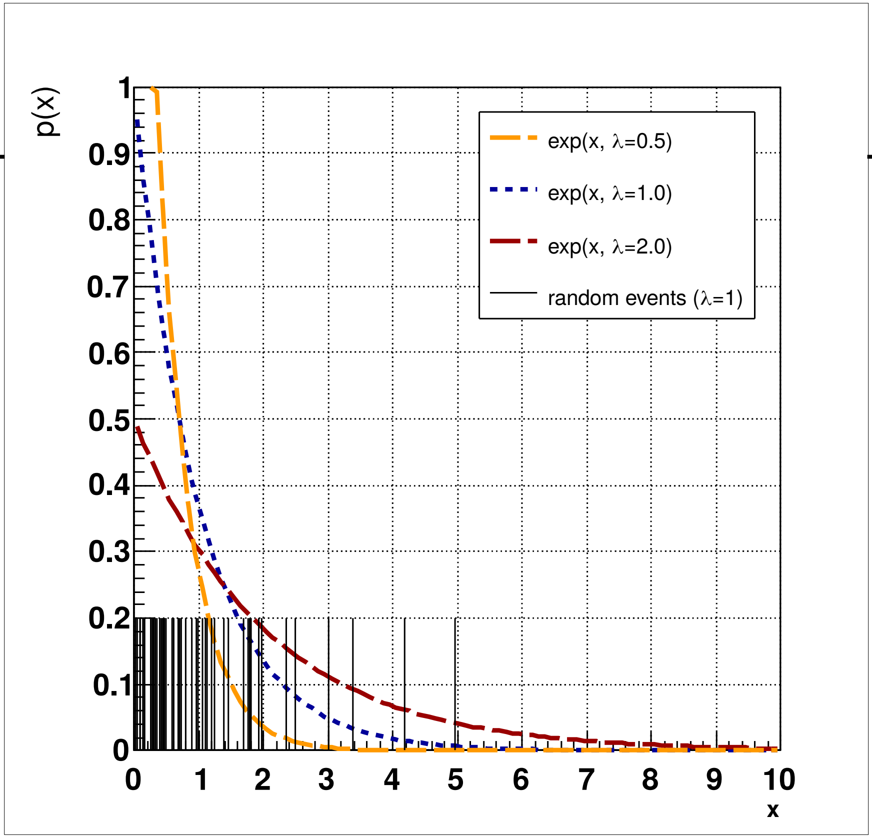
- **ML Schätzfunktion** $\hat{\theta}_{ML}$:

$$\frac{\partial}{\partial\theta} \ln(\mathcal{L}(\{x_i\},\theta)) = \sum_{i=1}^{50} (-1/\theta + {}^{x_i}/\theta^2)$$

$$= \sum_{i=1}^{50} \frac{-\theta + x_i}{\theta^2} \stackrel{!}{=} 0$$

$$\hat{\theta}_{ML} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

ML Schätzfunktion entspricht dem arithmetischen Mittel (in unserem Bsp für n=50).



$$\ln(\mathcal{L}(\theta))|_{\theta=1/2} = -67.5923$$

$$\ln(\mathcal{L}(\theta))|_{\theta=1} = -51.1248$$

$$\ln(\mathcal{L}(\theta))|_{\theta=2} = -60.2198$$

(5) cf. VL-09 slide 33.

Beispiel: $\exp(x,\theta) = \frac{1}{\theta} e^{-x/\theta}$ (5)

- **Erwartungswert** von $\hat{\theta}_{ML}$:

$$E\left[\hat{\theta}_{ML}(\{x_i\})\right] = \int \cdots \int \hat{\theta}_{ML}(\{x_i\}) \cdot \mathcal{L}(\{x_i\},\theta) \cdot \mathrm{d}x_1 \cdots \mathrm{d}x_n$$

$$= \int \cdots \int \frac{1}{n}\Big(\sum_{i=1}^{n} x_i\Big) \cdot \frac{1}{\theta} e^{-x_1/\theta} \cdots \frac{1}{\theta} e^{-x_n/\theta} \cdot \mathrm{d}x_1 \cdots \mathrm{d}x_n$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left(\underbrace{\int x_i \frac{1}{\theta} e^{-x_i/\theta} \mathrm{d}x_i}\, \prod_{j \neq i} \underbrace{\int \frac{1}{\theta} e^{-x_j/\theta} \mathrm{d}x_j} \right) = \frac{1}{n} \sum_{i=1}^{n} x_i = \theta$$

$$\int_0^\infty x/\theta \cdot e^{-x/\theta} \mathrm{d}x = \theta \int_0^\infty x' e^{-x'} \mathrm{d}x' = \theta$$

mit $x' = x/\theta \quad \mathrm{d}x = \theta \mathrm{d}x'$

$$\int_0^\infty x' e^{-x'} \mathrm{d}x' = \left[-x' e^{-x'}\right]_0^\infty + \int_0^\infty e^{-x'} \mathrm{d}x' = 1$$

$$= 1$$

$\hat{\theta}_{SM}$ ist konsistent & erwartungstreu!

(5) cf. VL-09 slide 33.

Beispiel: $\exp(x,\theta) = \frac{1}{\theta} e^{-x/\theta}$ [(5)]

- **Erwartungswert** von $\hat{\theta}_{ML}$:

$$E\left[\hat{\theta}_{ML}(\{x_i\})\right] = \int \cdots \int \hat{\theta}_{ML}(\{x_i\}) \cdot \mathcal{L}(\{x_i\},\theta) \cdot \mathrm{d}x_1 \cdots \mathrm{d}x_n$$

$$= \int \cdots \int \frac{1}{n} \left(\sum_{i=1}^{n} x_i\right) \cdot \frac{1}{\theta} e^{-x_1/\theta} \cdots \frac{1}{\theta} e^{-x_n/\theta} \cdot \mathrm{d}x_1 \cdots \mathrm{d}x_n$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left(\int x_i \frac{1}{\theta} e^{-x_i/\theta} \mathrm{d}x_i \prod_{j \neq i} \int \frac{1}{\theta} e^{-x_j/\theta} \mathrm{d}x_j \right) = \frac{1}{n} \sum_{i=1}^{n} x_i = \theta$$

Dieses Ergebnis war zu erwarten:

- $\hat{\theta}_{ML}$ erweißt sich als das arithmetische Mittel der Einzelmessungen $\{x_i\}$, die nach $p(x,\theta) = \exp(x,\theta)$ verteilt sind.
- Wir haben gelernt, daß das arithmetische Mittel der Einzelmessungen eine erwartungstreue Schätz-funktion für den Erwartungswert von $p(x,\theta)$ ist.
- θ ist der Erwartungswert von $\exp(x,\theta)$.[(5)]

$\hat{\theta}_{SM}$ ist konsistent & erwartungstreu!

(5) cf. VL-09 slide 33.

Beispiel: $\exp(x,\theta) = \frac{1}{\theta} e^{-x/\theta}$ [(5)]

- Die **statistische Unsicherheit** von $\hat{\theta}_{ML}$ ist durch die Varianz $\mathrm{var}[\hat{\theta}_{ML}]$ gegeben, die sich in Einzelfällen (wie diesem analystisch berechnen) läßt:

$$\mathrm{var}\left[\hat{\theta}_{ML}(\{x_i\})\right] = E\left[\hat{\theta}^2_{ML}(\{x_i\})\right] - E\left[\hat{\theta}_{ML}(\{x_i\})\right]^2$$

$$= \int \cdots \int \left(\hat{\theta}_{ML}(\{x_i\})\right)^2 \cdot \mathcal{L}(\{x_i\},\theta) \prod \mathrm{d}x_i$$

$$-\left(\int \cdots \int \hat{\theta}_{ML}(\{x_i\}) \cdot \mathcal{L}(\{x_i\},\theta) \prod \mathrm{d}x_i\right)^2 = \theta^2/n$$

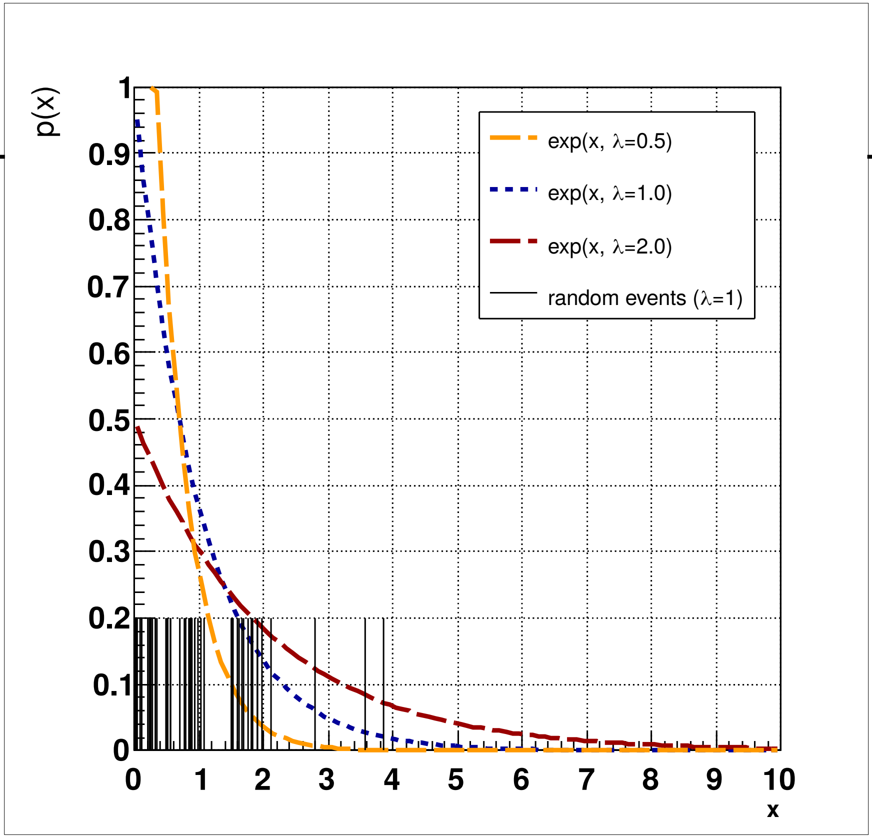
Dieses Ergebnis läßt sich leichter Ableiten:

- Wir haben gelernt, daß die Varianz des arithmetischen Mittels der Einzelmessungen σ^2/n ist, wobei σ^2 der Varianz von $p(x,\theta)$ und n der Länge der Stichprobe entspricht.

- θ^2 ist die Varianz von $\exp(x,\theta)$.[(5)]

(5) cf. VL-09 slide 33.

ML Schätzwert & Unsicherheit



- Nehmen Sie an, sie haben diese Reihe aus 50 Einzelmessungen der Lebensdauer θ eines radioaktiven Präparats vorgenommen. Das Meßergebnis für die Lebensdauer in unserem Beispiel ist:

$$\hat{\theta}_{ML} = 1.022 \pm 0.145\ (\mathrm{stat.})\ \mathrm{s}$$

Schätzwert aus Maximum Likelihood Analyse.

Standardabweichung des Schätzwertes.[6]

- **Interpretation:**

  Aus der Meßreihe die Sie vorgenommen haben haben sie den Wert $1.022\ \mathrm{s}$ für die Lebensdauer des Präparats erhalten. Wenn Sie die Meßreihe (aus 50 Einzelmessungen) sehr oft (=unendlich oft) wiederholen würden würde dieser Wert innerhalb von $\pm 0.144\ \mathrm{s}$ streuen.

(6)
Die Varianz enthält den wahren Wert von θ. Was machen wenn der wahre Wert eben nicht bekannt ist?!?
In diesem Fall ersetzen Sie den wahren Wert durch den Schätzwert.

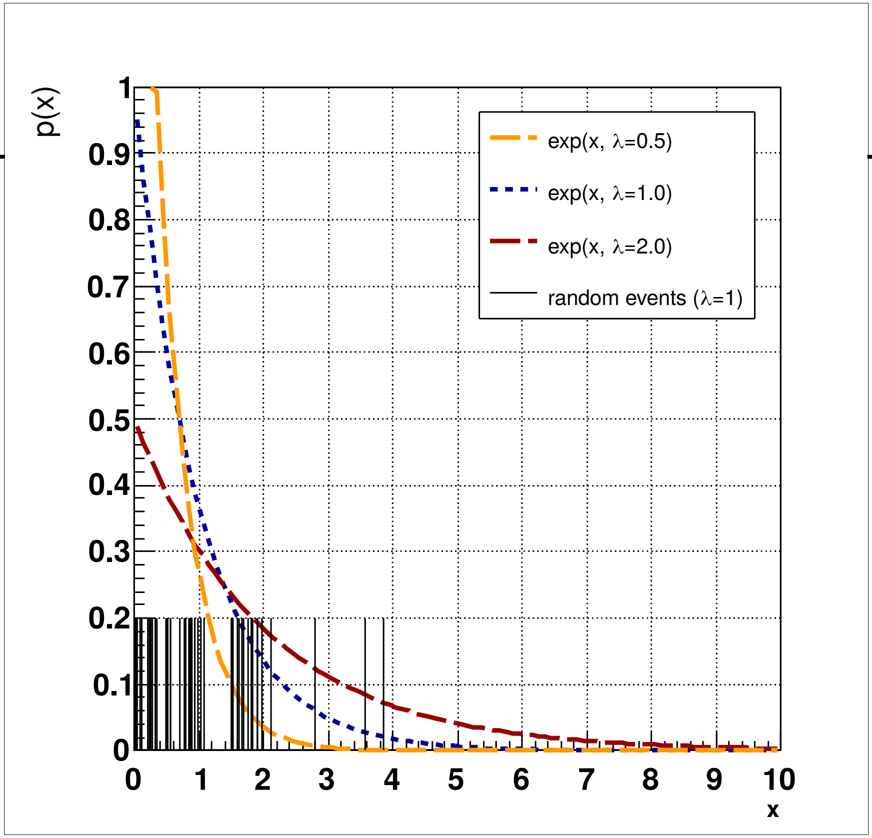
ML Schätzwert & Unsicherheit

- Nehmen Sie an, sie haben diese Reihe aus 50 Einzelmessungen der Lebensdauer θ eines radioaktiven Präparats vorgenommen. Das Meßergebnis für die Lebensdauer in unserem Beispiel ist:

$$\hat{\theta}_{ML} = 1.022 \pm 0.145\ (\mathrm{stat.})\ \mathrm{s}$$

Schätzwert aus Maximum Likelihood Analyse.

Standardabweichung des Schätzwertes [6]




```
double mean=0;
double nll1=0, nll2=0, nll3=0;
for(int i=0; i<50;++i){
  double x=unc1->GetRandom();
  nll1+=TMath::Log(unc1->Eval(x));
  nll2+=TMath::Log(unc2->Eval(x));
  nll3+=TMath::Log(unc3->Eval(x));
  line->DrawLine(x, 0., x, .2);
  mean+=1./(i+1)*(x-mean);
}
std::cout << "LogLikelihood(lambda=1.0)=" << nll1 << std::endl;
std::cout << "LogLikelihood(lambda=2.0)=" << nll2 << std::endl;
std::cout << "LogLikelihood(lambda=0.5)=" << nll3 << std::endl;
std::cout << "mean lifetime: t=" << mean << " \pm " << mean/TMath::Sqrt(50.) << std::endl;
```

Varianz durch Monte Carlo Methode

- In Fällen die zu kompliziert sind, um die Varianz der ML Schätzfunktion analytisch zu berechnen kann sie mit Hilfe der Monte Carlo Methode bestimmt werden.

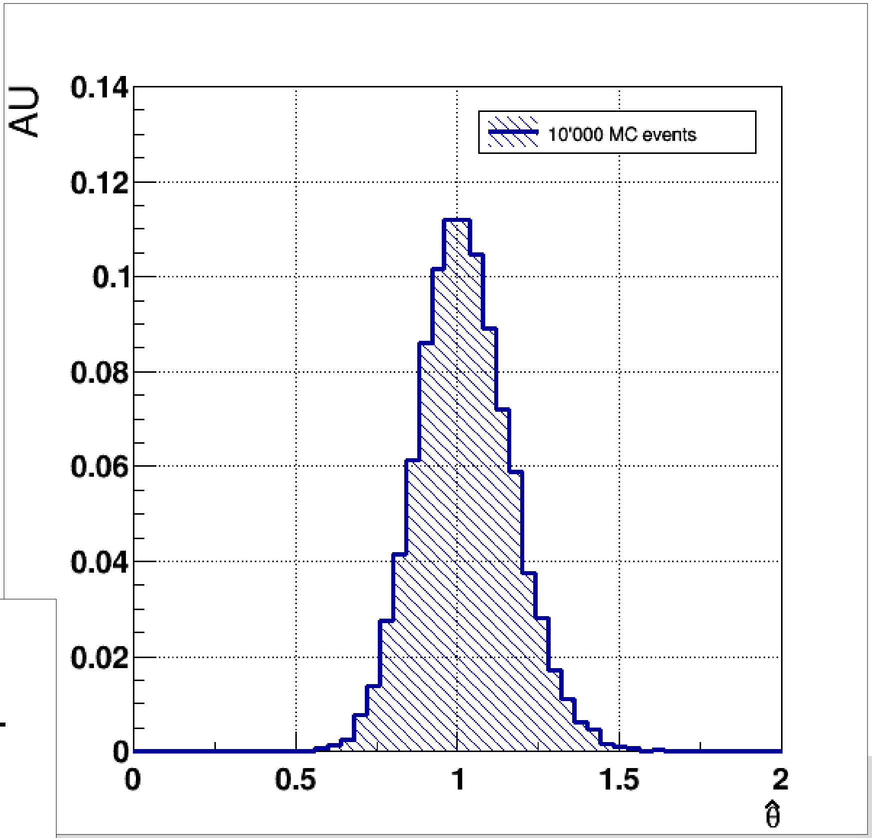
- Hierzu wiederholen Sie so viele Pseudo-Experimente (aus $n = 50$ Einzelmessungen) wie möglich.

- Für die wahre Lebensdauer setzen Sie den Schätzwert des Experiments $\hat{\theta}_{ML} = 1.022\ \mathrm{s}$.

- Ausgang für diesen Satz von 10'000 Pseudo-Experimenten:

  $$\langle \theta_{MC} \rangle = 1.022 \quad \sigma(\theta_{MC}) = 0.144$$

  d.h. deckungsgleich mit dem analytischen Ergebnis.



Mögliches Problem:

Die Erzeugung der Pseudo-Ereignisse kann aufwändig werden.

RCF-Ungleichung

- In Fällen in denen sowohl die analystische Berechnung, als auch eine Abschätzung durch Monte Carlo Methoden zu aufwändig ist, ist es möglich die Varianz durch die Rao-Cramér-Frechet (RCF) Ungleichung (aka Informationsungleichung, ohne Beweis) abzuschätzen:

$$\mathrm{var}[\hat{\theta}] \geq \frac{\left(1+\frac{\partial b}{\partial \theta}\right)^2}{E\left[-\frac{\partial^2 \ln \mathcal{L}}{\partial \theta^2}\right]}$$

Verzerrung.

Likelihood Funktion.

- Eine Schätzfunktion, die die untere Grenze der RCF-Ungleichung erreicht heißt **effizient**.

RCF-Ungleichung (für unser Beispiel)

- In unserem Beispiel läßt sich die untere Schwelle der RCF-Ungleichung explizit ausrechnen:

$$\mathrm{var}[\hat{\theta}] \geq \frac{\left(1+\frac{\partial b}{\partial \theta}\right)^2}{E\left[-\frac{\partial^2}{\partial \theta^2}\ln \mathcal{L}\right]}$$

$$\geq \frac{1}{E\left[-\frac{n}{\theta^2}\left(1-\frac{2\hat{\theta}}{\theta}\right)\right]} = \frac{1}{-\frac{n}{\theta^2}\left(1-\frac{2E[\hat{\theta}]}{\theta}\right)} = \theta^2/n$$

mit

$$b = 0 \text{ (const)} \quad E[\hat{\theta}] = \theta$$

$$\frac{\partial^2}{\partial \theta^2}\ln \mathcal{L} = \frac{n}{\theta^2}\left(1 - \frac{2}{\theta}\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{n}{\theta^2}\left(1 - \frac{2\hat{\theta}}{\theta}\right)$$

$\hat{\theta}_{ML} = \frac{1}{n}\sum x_i$ ist also eine effiziente Schätz-funktion.

- **NB:** Wenn es für ein Problem überhaupt effiziente Schätzfunktionen gibt, dann ist die ML Schätzfunktion effizient (ohne Beweis). Für eine ML Schätzfunktion wird die Ungleichung also zur Gleichung!

Graphische Abschätzung der Varianz

- Ist n „hinreichend“ groß läßt sich der Erwartungwert in der RCF-Ungleichung durch die Auswertung der zweiten Ableitung der Likelihood Funktion am Schätzwert für θ abschätzen:

$$\mathrm{var}[\hat{\theta}] = -\frac{1}{E\left[-\frac{\partial^2}{\partial\theta^2}\ln\mathcal{L}\right]} \xrightarrow[\text{large}]{n} \left[-\frac{1}{\frac{\partial^2}{\partial\theta^2}\ln\mathcal{L}}\right]_{\theta=\hat{\theta}}$$

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

$$\ln\mathcal{L}(\theta) = \underbrace{\ln\mathcal{L}(\hat{\theta}_{ML})}_{\equiv\,\mathcal{L}_{\max}} + \underbrace{\left[\frac{\partial}{\partial\theta}\ln\mathcal{L}\right]_{\theta=\hat{\theta}}}_{\equiv\,0}(\theta-\hat{\theta}_{ML}) + \frac{1}{2!}\underbrace{\left[\frac{\partial^2}{\partial\theta^2}\ln\mathcal{L}\right]_{\theta=\hat{\theta}}}_{\equiv\,-1/\hat{\sigma}^2}(\theta-\hat{\theta}_{ML})^2 + \cdots$$

$$\ln\mathcal{L}(\hat{\theta}_{ML}\pm\hat{\sigma}) = \mathcal{L}_{\max} - \tfrac{1}{2}$$

d.h. die Variation aus dem Maximum um $\pm\hat{\sigma}$ bewirkt die Reduktion von $\ln\mathcal{L}(\theta)$ um den Wert $1/2$!

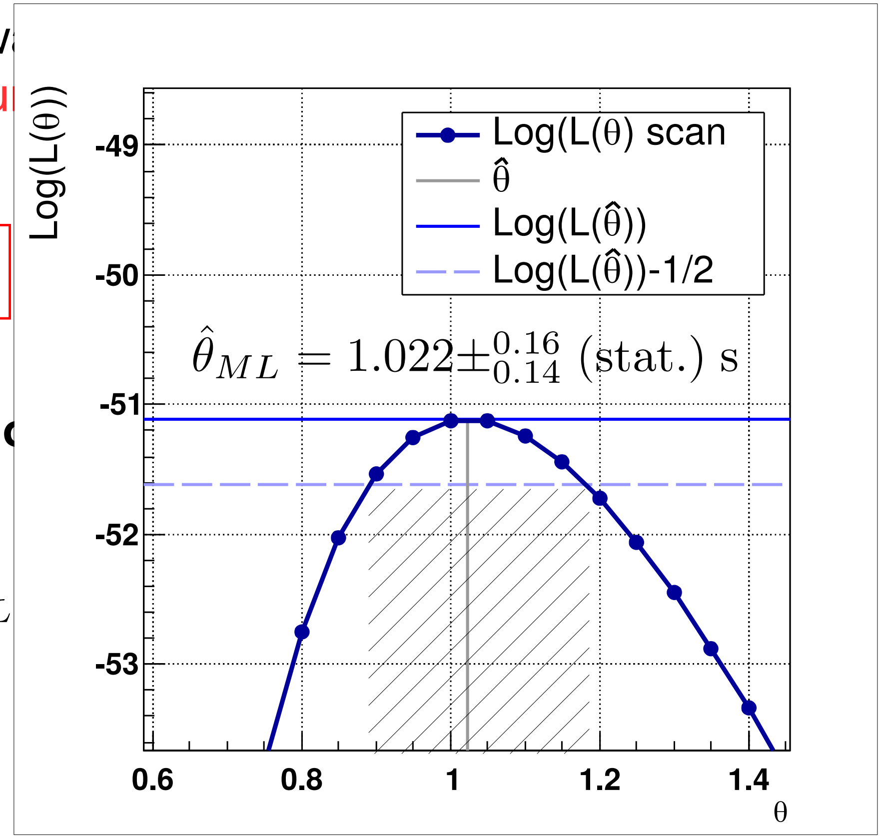
Graphische Abschätzung der Varianz

- Ist n „hinreichend“ groß läßt sich der Erw
  durch die Auswertung der zweiten Ableitu
  Schätzwert für θ abschätzen:

$$\mathrm{var}[\hat{\theta}] = -\frac{1}{E\left[-\frac{\partial^2}{\partial\theta^2}\ln\mathcal{L}\right]} \xrightarrow[\text{large}]{n} \left[-\frac{1}{\frac{\partial^2}{\partial\theta^2}\ln\mathcal{L}}\right]$$

- Aus der **Taylor-Entwicklung der Likelih**
  also:

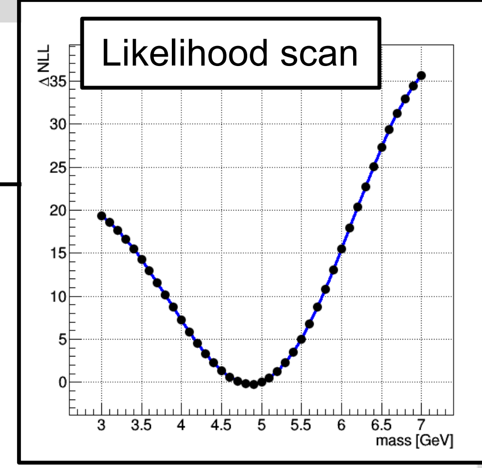
$$\ln\mathcal{L}(\theta) = \underbrace{\ln\mathcal{L}(\hat{\theta}_{ML})}_{\equiv\,\mathcal{L}_{\max}} + \underbrace{\left[\frac{\partial}{\partial\theta}\ln\mathcal{L}\right]_{\theta=\hat{\theta}}}_{\equiv\,0}(\theta-\hat{\theta}_{ML}$$



$$\ln\mathcal{L}(\hat{\theta}_{ML} \pm \hat{\sigma}) = \mathcal{L}_{\max} - \frac{1}{2}$$

d.h. die Variation aus dem Maximum um $\pm\hat{\sigma}$ bewirkt die Reduktion von $\ln\mathcal{L}(\hat{\theta}_{ML})$ um den Wert $1/2$!

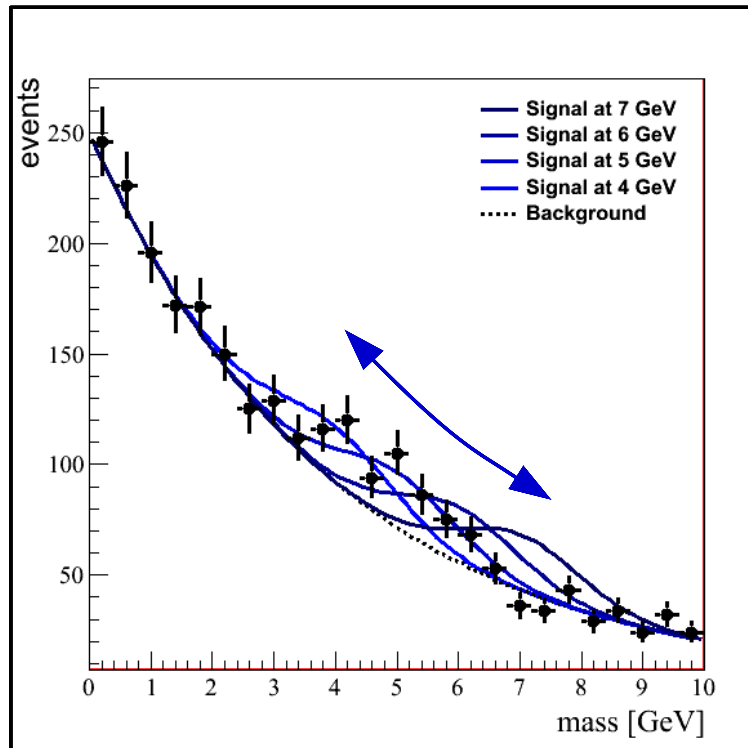
Praxisbezogene Beispiele



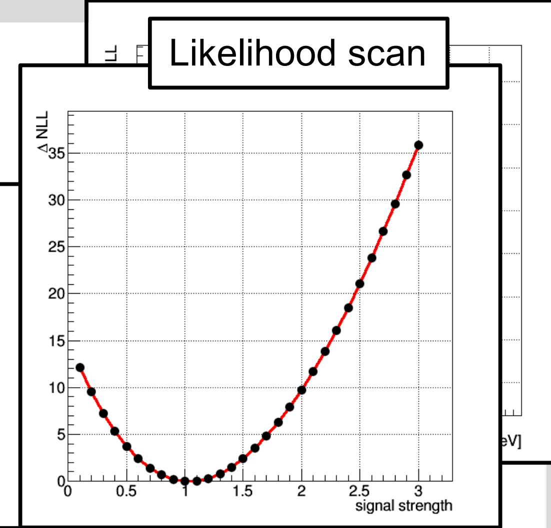
- Unbekanntes Signal auf einem bekannten Untergrund.
- Gebinnte Likelihood Funktion.
- Physikalisches Modell mit vier freien Parametern.
- *Parameter of interest*: Lage des massen peaks (θ_3).
- Scan der Negative Log-Likelihood.

$$\mathcal{L}(\{x_i\},\{\theta_j\}) = \prod_i \underbrace{\mathcal{P}(x_i,\mu_i(\theta_j))}_{\text{Product for each bin (Poisson).}}$$

$$\mu_i(\theta_j) = \underbrace{\theta_0 \cdot e^{-\theta_1 x_i}}_{\text{background}} + \underbrace{\theta_2 \cdot e^{-(\theta_3 - x_i)^2}}_{\text{signal}}$$



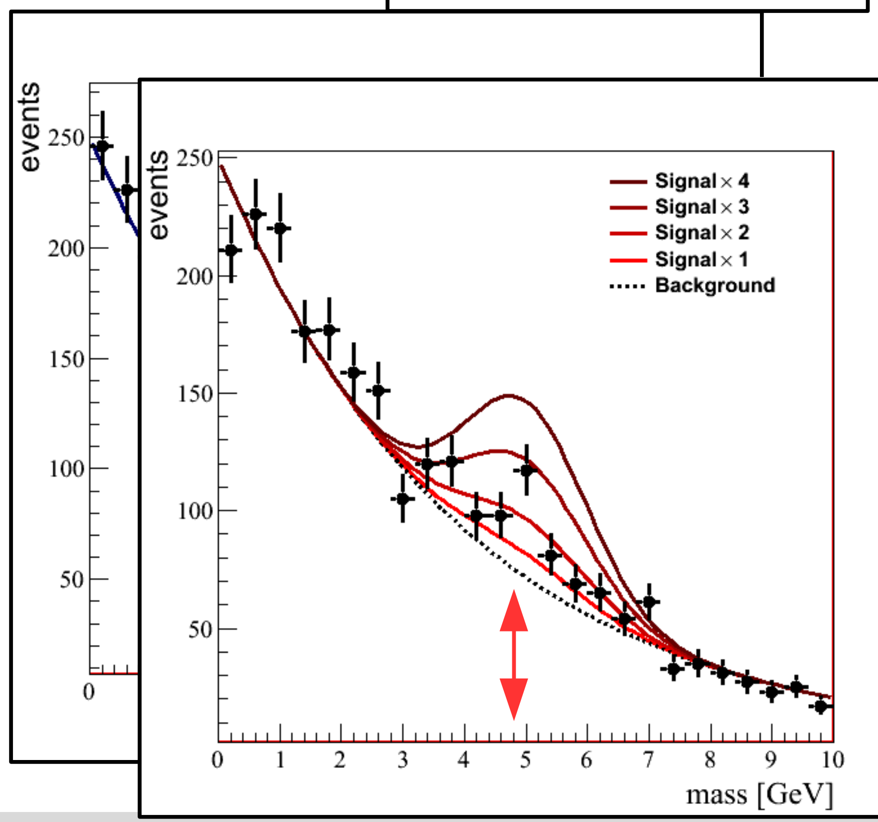
Praxisbezogene Beispiele



- Unbekanntes Signal auf einem bekannten Untergrund.
- Gebinnte Likelihood Funktion.
- Physikalisches Modell mit vier freien Parametern.
- *Parameter of interest*: Höhe des peaks bei bekannter Lage (θ_2).
- Scan der Negative Log-Likelihood.

$$\mathcal{L}(\{x_i\},\{\theta_j\}) = \prod_i \underbrace{\mathcal{P}(x_i, \mu_i(\theta_j))}_{\text{Product for each bin (Poisson).}}$$

$$\mu_i(\theta_j) = \underbrace{\theta_0 \cdot e^{-\theta_1 x_i}}_{\text{background}} + \underbrace{\theta_2 \cdot e^{-(\theta_3 - x_i)^2}}_{\text{signal}}$$



Zusammenfassung

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.