

Inhalt

Roger Wolf

- 1 Einführung und Grundlagen
Wahrscheinlichkeit, Statistik, Werkzeuge der statistischen Datenanalyse, ...
 - 2 Monte Carlo Methode als numerisches Hilfsmittel
Numerische Integration, Simulation komplexer Zusammenhänge, ...
 - 3 Parameterschätzung mit Hilfe der Maximum Likelihood Methode
Likelihood vs. Wahrscheinlichkeit, Maximum Likelihood als Optimierungsproblem, ...
 - 4 Parameterschätzung mit Hilfe der χ^2 -Methode
Ableitung aus Maximum Likelihood Methode, Optimierungsverfahren im allg., ...
 - 5 Hypothesentests in der modernen Physik
Begriffe des Hypothesentests, Beispiele, Anwendungen in der Physik, ...
-

Ralf Ulrich

- 6 Kollaboratives Arbeiten und moderne Softwarewerkzeuge
- 7 High-Performance Computing: optimales Zusammenspiel von Hard- und Software

Literaturempfehlungen

- **Einführende Literatur zu Statistik und Numerik:**
 - G. Cowan, *Statistical data analysis*, Oxford (1997) ([KIT-Bibliothek](#)).
 - G. Bohm, G. Zech, *Einführung in Statistik und Messwertanalyse für Physiker*, DESY (2006) ([eBook](#) deutsch, [eBook](#) english).
 - V. Blobel, E. Lormann, *Statistische und numerische Methoden der Datenanalyse*, DESY (2012) ([Webseite](#)).
 - R. J. Barlow, *Statistics: A Guide to the use of statistical methods in the physical sciences*, Wiley (1989) ([KIT-Bibliothek](#)).
 - W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical recipes*, Cambridge Univ. Press (2007) ([Webseite](#)).
- Skriptensammlung von Prof. G. Quast ([Link](#)):

3 Parameterschätzung mit Hilfe der ML Methode

3.1 Begriffsklärung Parameterschätzung

In diesem Kapitel geht es um die Abschätzung parametrisierter Eigenschaften einer Zufallsgröße auf einer Grundgesamtheit mit Hilfe von Stichproben. Dabei klären wir zunächst einige für die mathematische Formulierung notwendige Begriffe.



Beispiele für Parameterschätzungen

- **Aufgabe:** Schätzen Sie die erwartete Körpergröße aller Studierenden am KIT.
- Mögliche Vorgehensweisen nach dem Ziehen einer repräsentativen Stichprobe der Länge N :

Beispiele für Parameterschätzungen

- **Aufgabe:** Schätzen Sie die erwartete Körpergröße aller Studierenden am KIT.
- Mögliche Vorgehensweisen nach dem Ziehen einer repräsentativen Stichprobe der Länge N :
 - Addiere alle Körpergrößen der Stichprobe und teile durch N (→ arithmetische Mittel).

Beispiele für Parameterschätzungen

- **Aufgabe:** Schätzen Sie die erwartete Körpergröße aller Studierenden am KIT.
- Mögliche Vorgehensweisen nach dem Ziehen einer repräsentativen Stichprobe der Länge N :
 - Addiere alle Körpergrößen der Stichprobe und teile durch N (→ arithmetische Mittel).
 - Bilde arithmetisches Mittel nur über die ersten 10 Kandidaten und verwerfe den Rest.

Beispiele für Parameterschätzungen

- **Aufgabe:** Schätzen Sie die erwartete Körpergröße aller Studierenden am KIT.
- Mögliche Vorgehensweisen nach dem Ziehen einer repräsentativen Stichprobe der Länge N :
 - Addiere alle Körpergrößen der Stichprobe und teile durch N (\rightarrow arithmetische Mittel).
 - Bilde arithmetisches Mittel nur über die ersten 10 Kandidaten und verwerfe den Rest.
 - Addiere die kleinste und größte Körpergröße und teile durch 2.

Beispiele für Parameterschätzungen

- **Aufgabe:** Schätzen Sie die erwartete Körpergröße aller Studierenden am KIT.
- Mögliche Vorgehensweisen nach dem Ziehen einer repräsentativen Stichprobe der Länge N :
 - Addiere alle Körpergrößen der Stichprobe und teile durch N (\rightarrow arithmetische Mittel).
 - Bilde arithmetisches Mittel nur über die ersten 10 Kandidaten und verwerfe den Rest.
 - Addiere die kleinste und größte Körpergröße und teile durch 2.
 - Addiere alle Körpergrößen und teile durch $N-1$.

Beispiele für Parameterschätzungen

- **Aufgabe:** Schätzen Sie die erwartete Körpergröße aller Studierenden am KIT.
- Mögliche Vorgehensweisen nach dem Ziehen einer repräsentativen Stichprobe der Länge N :
 - Addiere alle Körpergrößen der Stichprobe und teile durch N (\rightarrow arithmetische Mittel).
 - Bilde arithmetisches Mittel nur über die ersten 10 Kandidaten und verwerfe den Rest.
 - Addiere die kleinste und größte Körpergröße und teile durch 2.
 - Addiere alle Körpergrößen und teile durch $N-1$.
 - Ignoriere alle Daten und nehme 1.8 m an.

Beispiele für Parameterschätzungen

- **Aufgabe:** Schätzen Sie die erwartete Körpergröße aller Studierenden am KIT.
- Mögliche Vorgehensweisen nach dem Ziehen einer repräsentativen Stichprobe der Länge N :
 - Addiere alle Körpergrößen der Stichprobe und teile durch N (\rightarrow arithmetische Mittel).
 - Bilde arithmetisches Mittel nur über die ersten 10 Kandidaten und verwerfe den Rest.
 - Addiere die kleinste und größte Körpergröße und teile durch 2.
 - Addiere alle Körpergrößen und teile durch $N-1$.
 - Ignoriere alle Daten und nehme 1.8 m an.
 - Multipliziere alle Körpergrößen und ziehe die N -te Wurzel.

Beispiele für Parameterschätzungen

- **Aufgabe:** Schätzen Sie die erwartete Körpergröße aller Studierenden am KIT.
- Mögliche Vorgehensweisen nach dem Ziehen einer repräsentativen Stichprobe der Länge N :
 - Addiere alle Körpergrößen der Stichprobe und teile durch N (\rightarrow arithmetische Mittel).
 - Bilde arithmetisches Mittel nur über die ersten 10 Kandidaten und verwerfe den Rest.
 - Addiere die kleinste und größte Körpergröße und teile durch 2.
 - Addiere alle Körpergrößen und teile durch $N-1$.
 - Ignoriere alle Daten und nehme 1.8 m an.
 - Multipliziere alle Körpergrößen und ziehe die N -te Wurzel.
 - Middle die Körpergrößen, die am häufigsten auftreten.

Beispiele für Parameterschätzungen

- **Aufgabe:** Schätzen Sie die erwartete Körpergröße aller Studierenden am KIT.
- Mögliche Vorgehensweisen nach dem Ziehen einer repräsentativen Stichprobe der Länge N :
 - Addiere alle Körpergrößen der Stichprobe und teile durch N (\rightarrow arithmetische Mittel).
 - Bilde arithmetisches Mittel nur über die ersten 10 Kandidaten und verwerfe den Rest.
 - Addiere die kleinste und größte Körpergröße und teile durch 2.
 - Addiere alle Körpergrößen und teile durch $N-1$.
 - Ignoriere alle Daten und nehme 1.8 m an.
 - Multipliziere alle Körpergrößen und ziehe die N -te Wurzel.
 - Middle die Körpergrößen, die am häufigsten auftreten.
 - Berücksichtige nur jeden zweiten Kandidaten bei der Mittelung.

Beispiele für Parameterschätzungen

- **Aufgabe:** Schätzen Sie die erwartete Körpergröße aller Studierenden am KIT.
- Mögliche Vorgehensweisen nach dem Ziehen einer repräsentativen Stichprobe der Länge N :
 - Addiere alle Körpergrößen der Stichprobe und teile durch N (\rightarrow arithmetische Mittel).
 - Bilde arithmetisches Mittel nur über die ersten 10 Kandidaten und verwerfe den Rest.
 - Addiere die kleinste und größte Körpergröße und teile durch 2.
 - Addiere alle Körpergrößen und teile durch $N-1$.
 - Ignoriere alle Daten und nehme 1.8 m an.
 - Multipliziere alle Körpergrößen und ziehe die N -te Wurzel.
 - Middle die Körpergrößen, die am häufigsten auftreten.
 - Berücksichtige nur jeden zweiten Kandidaten bei der Mittelung.

Alle diese Methoden führen auf unterschiedliche Schätzwerte der Zufallsgröße „Körpergröße der Studenten am KIT“. Aber welche dieser Schätzwerte sind sinnvoll?

Wahrscheinlichkeitsrechnung vs. Statistik

- In der Wahrscheinlichkeitsrechnung wird die wahre Wahrscheinlichkeitsverteilung als bekannt vorausgesetzt.
- In der Realität ist der Wahrscheinlichkeitsraum normalerweise nicht (vollständig) bekannt. Er muss mit Hilfe begrenzter **Stichproben** charakterisiert 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 $\vec{x} = (x_1, x_2, \dots, x_n)$ in einem neuen Ereignisraum aufgefasst werden. Da die Einzelereignisse $\{x_i\}$ stochastisch unabhängig sind ist die Wahrscheinlichkeitsverteilung über dem neuen Ereignisraum gegeben durch:

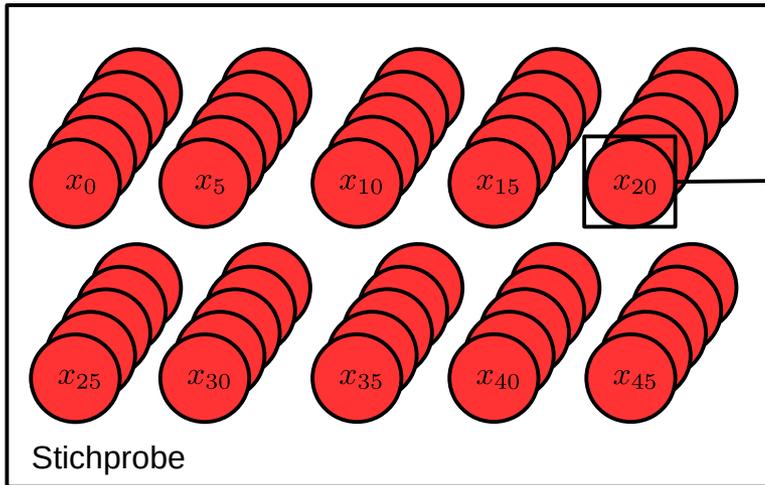
$$\mathcal{P}(\vec{x}) = \prod_{1 \leq i \leq n} p(x_i).$$

- $\vec{x} = (x_1, x_2, \dots, x_n)$ entspricht statistisch einer Stichprobe. Diese kann als Ausgang eines Experiments oder einer Messreihe betrachtet werden.

Einzelmessung vs. Stichprobe

- **Beispiel:** Messreihe aus 50 Einzelmessungen

Messreihe aus 50 Einzelmessungen:



Ausgang der Einzelmessung:
folgt $p(x)$

Ausgang der Stichprobenmessung:
folgt $\mathcal{P}(\{x_i\}) = \prod_{1 \leq i \leq n} p(x_i)$

● Einzelmessung

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

Teststatistik

Eine Funktion einer beliebigen Anzahl beobachteter Einzelereignisse \vec{x} heißt Teststatistik.

- Eine Teststatistik kann beliebig viele freie Parameter besitzen, sie kann selbst ein mehrdimensionaler Vektor $\vec{t} = (t_1, t_2, \dots, t_m)$ (für $n \geq m$ Einzelmessungen \vec{x}) sein.
- Im Extremfall kann sie den Einzelmessungen selbst entsprechen. Üblicherweise ist t aber eine einfache skalare Funktion. Die Reduktion der Dimension erfolgt i.a. aus Gründen der besseren Handhabbarkeit.
- Die Teststatistik ist eine Funktion der zufallsverteilten \vec{x} und folgt damit selbst einer **Wahrscheinlichkeitsverteilung** $g(t(\vec{x}))$.

Schätzfunktion

Eine Statistik zur Bestimmung der Eigenschaften einer Wahrscheinlichkeitsdichte heißt Schätzfunktion (oder Abschätzung), zur Bestimmung eines Schätzwertes.

- Die Schätzfunktion $\hat{\theta}(\cdot)$ einer Eigenschaft einer Wahrscheinlichkeitsdichte und/oder der Schätzwert $\hat{\theta}$ werden oft mit einem $\hat{\cdot}$ bezeichnet, um sie vom wahren Wert θ zu unterscheiden.
- Wenn $\lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| > \epsilon) = 0$ dann heissen $\hat{\theta}(\cdot)$ oder $\hat{\theta}$ **konsistent**, dabei ist n die Länge der Stichprobe.

Die Schätzfunktion als Zufallsvariable

- Da die Schätzfunktion $\hat{\theta}(\vec{x})$ eine Funktion der \vec{x} ist ist sie selbst eine Zufallsvariable. D.h. nach vielfacher Wiederholung des Experiments folgt der Ausgang der Experimente selbst einer Wahrscheinlichkeitsverteilung $g(\hat{\theta}, \theta)$ (vgl. Folie 5, letzter Eintrag).

Wir bezeichnen $g(\hat{\theta}, \theta)$ als Stichprobenverteilung. Der Erwartungswert von $\hat{\theta}$ ist definiert als:

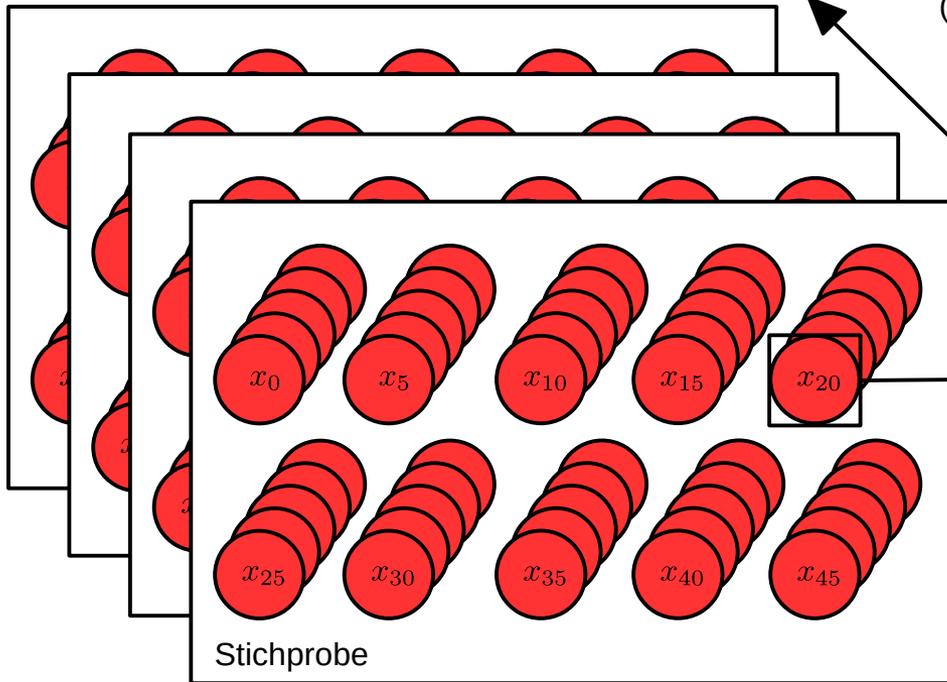
$$E[\hat{\theta}(\vec{x}, \theta)] = \int \hat{\theta} \cdot g(\hat{\theta}, \theta) d\hat{\theta} = \int \cdots \int \hat{\theta}(\vec{x}) \cdot \prod p(\vec{x}, \theta) dx_i.$$

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

- Beachten Sie, dass $g(\hat{\theta}, \theta)$ i.a. auch von den wahren Werten θ abhängt.
- Im Englischen bezeichnet man die Stichprobenverteilung als *sample distribution*.

Zwischenergebnis – Begriffsklärung

Experiment aus 50 Einzelmessungen:



Verteilung der
Stichprobenmessungen
(Stichprobenverteilung):

$$\hat{\theta}(\vec{x}) \text{ folgt } g(\hat{\theta}(\vec{x}), \theta)$$

NB: Dieser Raum ist in den folgenden Folien oftmals abstrakt, kann aber gerade mit Hilfe von MC Methoden selbst gesampled werden.

Ausgang der Einzelmessung:
folgt $p(x, \theta)$

Ausgang der Stichprobenmessung:
folgt $\mathcal{P}(\{x_i\}) = \prod_{1 \leq i \leq n} p(x_i, \theta)$

Auf diesem Raum schätzen wir Eigenschaften von $p(x)$, parametrisiert durch θ ab, selbst wenn die eigentliche Form von $p(x)$ unbekannt bleibt.

Verzerrung (engl. *bias*)

Die Größe

$$b = E[\hat{\theta}(\vec{x}, \theta)] - \theta$$

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 dem wahren Wert θ .
- Ein Schätzwert für den der *bias* unabhängig von der Stichprobenlänge der Messreihe 0 ist heißt **erwartungstreu**, ein Schätzwert für den $b = 0$ für $n \rightarrow \infty$ heißt **asymptotisch erwartungstreu**.

Verzerrung (engl. *bias*)

- **Frage:** Wie kann ich den *bias* bestimmen, wenn mir der wahre Wert des Parameters θ nicht bekannt ist?

Verzerrung (engl. *bias*)

- **Frage:** Wie kann ich den *bias* bestimmen, wenn mir der wahre Wert des Parameters θ nicht bekannt ist?
- **Antwort:**
 - In einigen wenigen Fällen läßt sich (asymptotische) Erwartungstreue analytisch beweisen. Wir haben dies bereits in [VL-01](#) für die folgenden Schätzfunktionen gemacht:
 - Mittelwert (\bar{x}) der Stichprobe.
 - Varianz (s^2) der Stichprobe.
 - Korrelationskoeffizient (r) der Stichprobe.

Verzerrung (engl. *bias*)

- **Frage:** Wie kann ich den *bias* bestimmen, wenn mir der wahre Wert des Parameters θ nicht bekannt ist?
- **Antwort:**
 - In einigen wenigen Fällen läßt sich (asymptotische) Erwartungstreue analytisch beweisen. Wir haben dies bereits in [VL-01](#) für die folgenden Schätzfunktionen gemacht:
 - Mittelwert (\bar{x}) der Stichprobe.
 - Varianz (s^2) der Stichprobe.
 - Korrelationskoeffizient (r) der Stichprobe.
 - **NB:** Bei diesen besonderen Schätzfunktionen verzichtet man oft auf den $\hat{\cdot}$.

Verzerrung (engl. *bias*)

- **Frage:** Wie kann ich den *bias* bestimmen, wenn mir der wahre Wert des Parameters θ nicht bekannt ist?
- **Antwort:**
 - In einigen wenigen Fällen läßt sich (asymptotische) Erwartungstreue analytisch beweisen. Wir haben dies bereits in [VL-01](#) für die folgenden Schätzfunktionen gemacht:
 - Mittelwert (\bar{x}) der Stichprobe.
 - Varianz (s^2) der Stichprobe.
 - Korrelationskoeffizient (r) der Stichprobe.
 - **NB:** Bei diesen besonderen Schätzfunktionen verzichtet man oft auf den $\hat{\cdot}$.
 - **Beachten Sie:** Alle drei Größen sind (asymptotisch) erwartungstreu. Sie können also aus der Stichprobe (asymptotisch) die wahren Werte für Erwartungswert, Varianz und Korrelationskoeffizient von $p(x)$ ($p(x, y)$) ermitteln ohne irgendetwas über $p(x)$ ($p(x, y)$) zu wissen!

Verzerrung (engl. *bias*)

- **Frage:** Wie kann ich den *bias* bestimmen, wenn mir der wahre Wert des Parameters θ nicht bekannt ist?
- **Antwort:**
 - Im allgemeinen testen Sie die Erwartungstreue einer Schätzfunktion z.B. mit Hilfe eines Ensemble Tests unter Annahme eines wahren Wertes θ .

Statistische und systematische Unsicherheiten

Die Größe

$$\text{MSE} = E[(\hat{\theta} - \theta)^2] = \underbrace{E[(\hat{\theta} - E[\hat{\theta}])^2]}_{\text{var}[\hat{\theta}]} + \underbrace{E^2[\hat{\theta} - \theta]}_{b^2} = \sigma_{\hat{\theta}}^2 + b_{\hat{\theta}}^2$$

heißt Genauigkeit (*mean squared error*) der Schätzfunktion. In einer physikalischen Messung beschreibt $\text{var}[\hat{\theta}]$ die statistische Unsicherheit der Messung. Die systematische Unsicherheit leitet man aus b ab.

- **Beachten Sie:** Der MSE ist a priori eine abstrakte Größe, weil Ihnen θ in der Gleichung a priori nicht bekannt ist. Sie kann z.B. dadurch konkret werden, dass Sie ihr eine hypothetische Wahrheit unterlegen (siehe [Folie 12](#)). Für erwartungstreue Schätzfunktionen ist der zweite Summand der oberen Gleichung allerdings 0. Das lässt Aussagen über den Schätzwert zu, ohne etwas über θ zu wissen.

Statistische und systematische Unsicherheiten

- Wir weisen die angegebene Beziehung der vorherigen Folie im folgenden explizit nach:

$$\text{MSE} = \underbrace{E[(\hat{\theta} - \theta)^2]}_{\text{LHS}} = \underbrace{E[(\hat{\theta} - E[\hat{\theta}])^2] + E^2[\hat{\theta} - \theta]}_{\text{RHS}} = \sigma_{\hat{\theta}}^2 + b_{\hat{\theta}}^2$$

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

LHS:

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

RHS:

$$\underbrace{E[(\hat{\theta} - E[\hat{\theta}])^2]}_{(1)} = E[\hat{\theta}^2] - E[\hat{\theta}]^2 \quad ; \quad \underbrace{E^2[\hat{\theta} - \theta]}_{(2)} = E[\hat{\theta}]^2 - 2\theta E[\hat{\theta}] + \theta^2$$

$$E[(\hat{\theta} - E[\theta])^2] + E^2[\hat{\theta} - \theta] = \underbrace{E[\hat{\theta}^2] - E[\hat{\theta}]^2}_{(1)} + \underbrace{E[\hat{\theta}]^2 - 2\theta E[\hat{\theta}] + \theta^2}_{(2)} = E[\hat{\theta}^2] - 2\theta E[\hat{\theta}] + \theta^2 \quad \text{vergleiche mit } (*)$$

3 Parameterschätzung mit Hilfe der ML Methode

3.2 Maximum Likelihood Methode

In diesem Kapitel führen wir die Maximum Likelihood Methode als allgemeines Prinzip zur Parameterschätzung begrifflich ein.



Hypothese

- Im weiteren werden wir uns mit der Auswahl und Bewertung von Modellen in Form von Hypothesen beschäftigen:

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, oder **zusammengesetzt** wenn sie unbekannte (d.h. noch zu bestimmende) Parameter θ_i enthält.

Likelihood Funktion

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

$$P(x_i, \theta) = p(x_i, \theta) dx_i$$

Die Wahrscheinlichkeit für den Ausgang des Experimentes ist gegeben durch

$$P'(\{x_i\}, \theta) = \prod_{i \leq n} p(x_i, \theta) dx_i$$

Die Funktion

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

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

Maximum Likelihood (ML)

- 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 der Einzelmessungen $\{x_i\}$ für den korrekten Wert von θ höher ist, als für jeden anderen Wert.
- Da die dx_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 $\hat{\theta}_{\text{ML}}$ des Parameters θ . Wenn $\mathcal{L}(\theta)$ (mindestens zwei mal) stetig differenzierbar ist ist $\hat{\theta}_{\text{ML}}$ gegeben durch:

$$\frac{\partial}{\partial \theta} \mathcal{L}(\theta) = 0 \quad \frac{\partial^2}{(\partial \theta)^2} \mathcal{L}(\theta) < 0$$

NB: Beachten Sie, dem ML Prinzip liegt immer eine Wahrheitshypothese zugrunde.

Negative Log Likelihood (NLL)

- Für feste Werte von θ ist $\mathcal{L}(\theta)$ selbst eine Wahrscheinlichkeitsdichte (als Funktion der $\{x_i\}$). Insbesondere beim Vorliegen vieler Einzelmessungen kann $\mathcal{L}(\theta)$ sehr kleine Werte annehmen.
- Es erweist sich sehr oft als praktikabel statt $\mathcal{L}(\theta)$ den Logarithmus $\ln(\mathcal{L}(\theta))$ zu verwenden.

Vorteile:

- Kleine Zahlen lassen sich besser darstellen und sind numerisch besser zu verarbeiten.
- Produkte transformieren sich in Summen.

Zuordnung:

- 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 (NLL), $-\ln(\mathcal{L}(\theta))$. (Siehe [VL-04 Folie 27ff](#))

Logarithmische Ableitung

- Beweisen Sie durch Ableitung von $\ln(\mathcal{L}(\theta))$, dass der Wert $\hat{\theta}_{\text{ML}}$ für $\mathcal{L}(\theta)$ und für $\ln(\mathcal{L}(\theta))$ identisch ist.

Logarithmische Ableitung

- Beweisen Sie durch Ableitung von $\ln(\mathcal{L}(\theta))$, dass der Wert $\hat{\theta}_{\text{ML}}$ für $\mathcal{L}(\theta)$ und für $\ln(\mathcal{L}(\theta))$ identisch ist.

$$\frac{\partial}{\partial \theta} \left(\ln(\mathcal{L}(\theta)) \right) = \frac{\frac{\partial}{\partial \theta} \mathcal{L}(\theta)}{\mathcal{L}(\theta)} = 0$$

Da Sie davon ausgehen können, dass $\mathcal{L}(\theta) \neq 0$ ist die Position des Maximums von $\mathcal{L}(\theta)$ mit der Position des Maximums von $\ln(\mathcal{L}(\theta))$ identisch.

Transformationsinvarianz

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

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

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

$$\frac{\partial}{\partial \theta} \mathcal{L}(a(\theta)) = 0 \quad \Leftrightarrow \quad \frac{\partial}{\partial a} \mathcal{L}(a(\theta)) = 0$$

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

NB: D.h. der ML Schätzwert von a ergibt sich aus dem ML Schätzwert von θ und dem vorgegebenen funktionalen Zusammenhang.

3 Parameterschätzung mit Hilfe der ML Methode

3.3 Beispiele für ML-Abschätzungen

Wir geben im folgenden einige Beispiele für die Anwendung der ML-Abschätzung.



Beispiel – 1: Stichprobe gaußverteilter Zufallsgrößen

- Wir betrachten den Fall einer Stichprobe gaußverteilter Zufallsgrößen der Länge N :

- Likelihood Funktion:

$$\mathcal{L}(\{x_i\}, \mu) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$$

$$\ln(\mathcal{L}(\{x_i\}, \mu)) = -\frac{1}{2} \sum_{i=1}^N \left(\ln(2\pi) + \ln(\sigma^2) + \frac{(x_i - \mu)^2}{\sigma^2} \right)$$

- ML Schätzfunktion $\hat{\mu}_{\text{ML}}$:

$$\frac{\partial}{\partial \mu} \ln(\mathcal{L}(\{x_i\}, \mu)) = \sum_{i=1}^N \frac{x_i - \mu}{\sigma^2} = 0;$$

$\hat{\mu}_{\text{ML}} = \frac{1}{N} \sum_{i=1}^N x_i$	ML Schätzfunktion entspricht arithmetischem Mittel.
--------------------------------------------------------	-----------------------------------------------------

Beispiel – 2: Stichprobe poissonverteilter Zufallsgrößen

- Wir betrachten den Fall einer Stichprobe poissonverteilter Zufallsgrößen der Länge N :
- Likelihood Funktion:

$$\mathcal{L}(\{k_i\}, \mu) = \prod_{i=1}^N \frac{\mu^{k_i}}{k_i!} e^{-\mu}$$

$$\ln(\mathcal{L}(\{k_i\}, \mu)) = \sum_{i=1}^N \left(k_i \ln(\mu) - \ln(k_i!) - \mu \right) = -N\mu + \sum_{i=1}^N \left(k_i \ln(\mu) - \ln(k_i!) \right)$$

- ML Schätzfunktion $\hat{\mu}_{\text{ML}}$:

$$\frac{\partial}{\partial \mu} \ln(\mathcal{L}(\{k_i\}, \mu)) = -N + \sum_{i=1}^N \frac{k_i}{\mu} = 0;$$

$\hat{\mu}_{\text{ML}} = \frac{1}{N} \sum_{i=1}^N k_i$	<p>ML Schätzfunktion entspricht arithmetischem Mittel.</p>
--------------------------------------------------------	------------------------------------------------------------

Beispiel – 3: Radioaktiver Zerfall

- Wir werden im folgenden das Erlernte auf ein konkretes und einfaches Beispiel anwenden. Es handelt sich um die Bestimmung der Lebensdauer eines radioaktiven Präparats.
- Unsere Wahrheitshypothese ist, dass die Wahrscheinlichkeitsdichte dieses Zerfalls die Exponentialverteilung $\exp(x, \theta)$ ist.

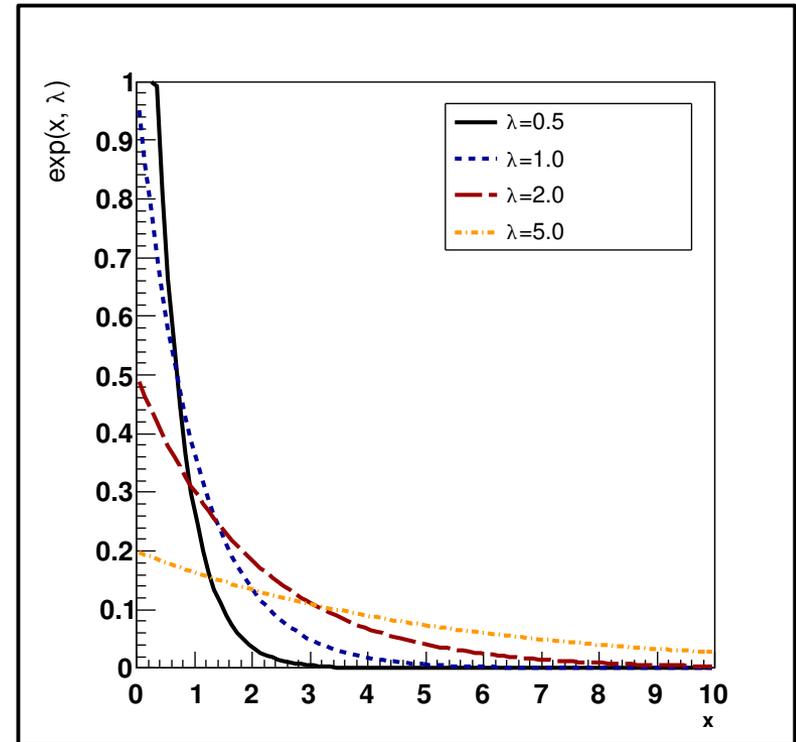


Erinnerung: Eigenschaften der Exponentialverteilung

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

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

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



- Wir betrachten eine Messreihe bestehend aus 50 Ereignissen verteilt nach $\exp(x, \theta)$ mit $\theta = 1$.

ML Schätzfunktion

- Likelihood Funktion:

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

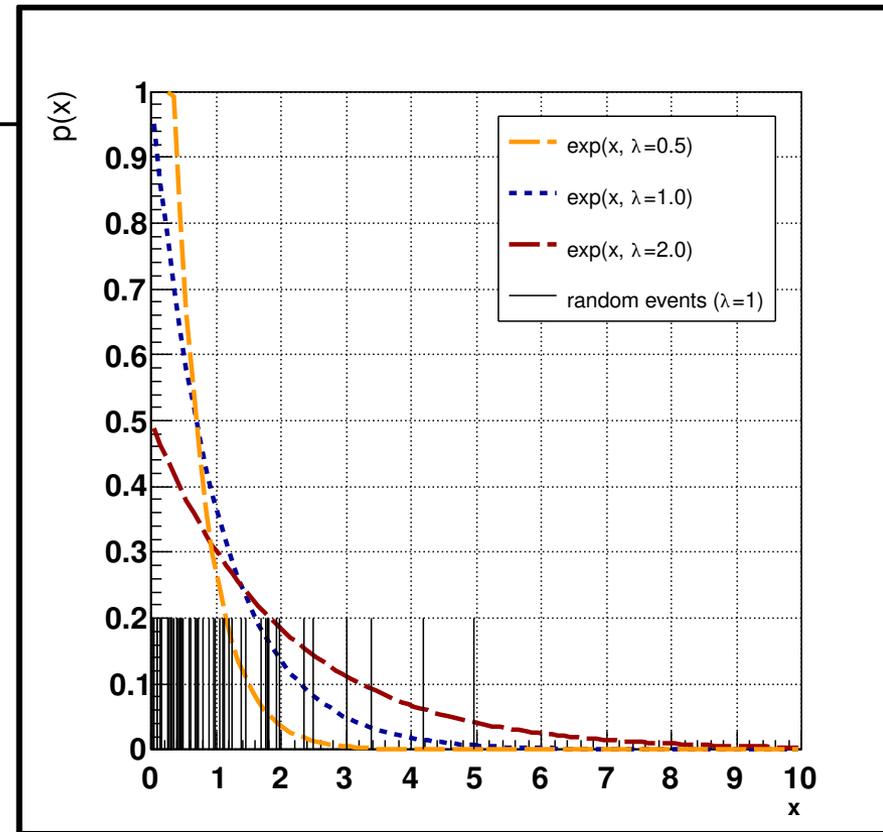
$$\ln(\mathcal{L}(\{x_i\}, \theta)) = \sum_{i=1}^{50} \left(-\ln(\theta) - \frac{x_i}{\theta} \right)$$

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

$$\begin{aligned} \frac{\partial}{\partial \theta} \ln(\mathcal{L}(\{x_i\}, \theta)) &= \sum_{i=1}^{50} \left(-\frac{1}{\theta} + \frac{x_i}{\theta^2} \right) \\ &= \sum_{i=1}^{50} \frac{-\theta + x_i}{\theta^2} = 0; \end{aligned}$$

$$\sum_{i=1}^{50} x_i = 50 \hat{\theta}_{\text{ML}};$$

$$\hat{\theta}_{\text{ML}} = \frac{1}{50} \sum_{i=1}^{50} x_i$$



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

ML Schätzfunktion

- Likelihood Funktion:

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

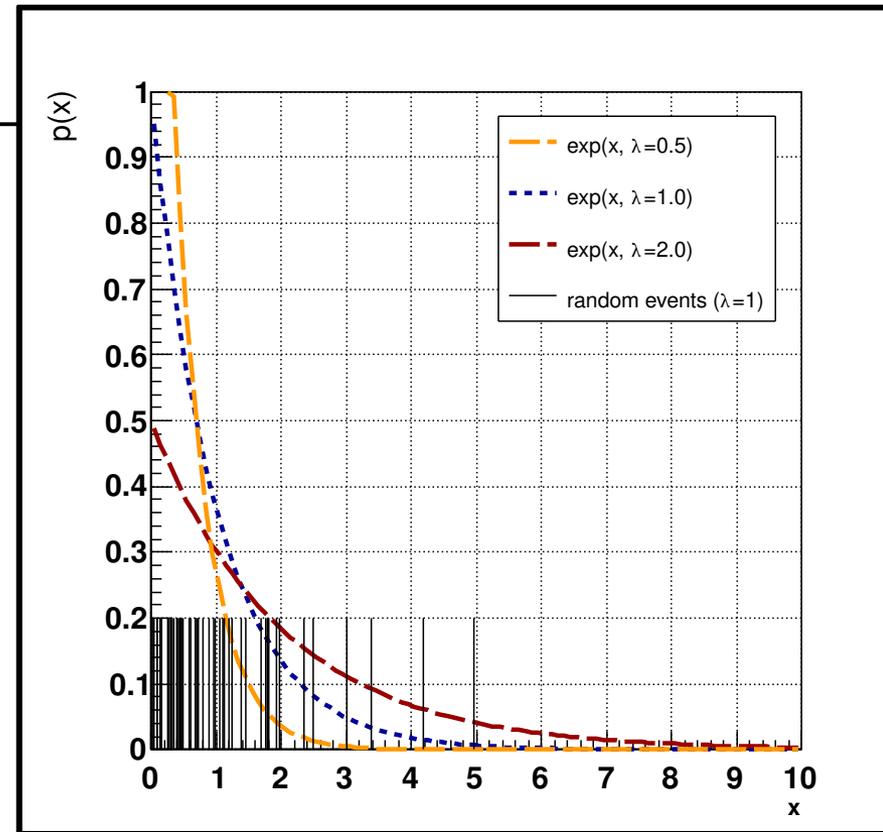
$$\ln(\mathcal{L}(\{x_i\}, \theta)) = \sum_{i=1}^{50} \left(-\ln(\theta) - \frac{x_i}{\theta} \right)$$

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

$$\begin{aligned} \frac{\partial}{\partial \theta} \ln(\mathcal{L}(\{x_i\}, \theta)) &= \sum_{i=1}^{50} \left(-\frac{1}{\theta} + \frac{x_i}{\theta^2} \right) \\ &= \sum_{i=1}^{50} \frac{-\theta + x_i}{\theta^2} = 0; \end{aligned}$$

$$\sum_{i=1}^{50} x_i = 50 \hat{\theta}_{\text{ML}};$$

$$\hat{\theta}_{\text{ML}} = \frac{1}{50} \sum_{i=1}^{50} x_i \quad \text{ML Schätzfunktion entspricht arithmetischem Mittel.}$$



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

ML Schätzfunktion

- Likelihood Funktion:

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

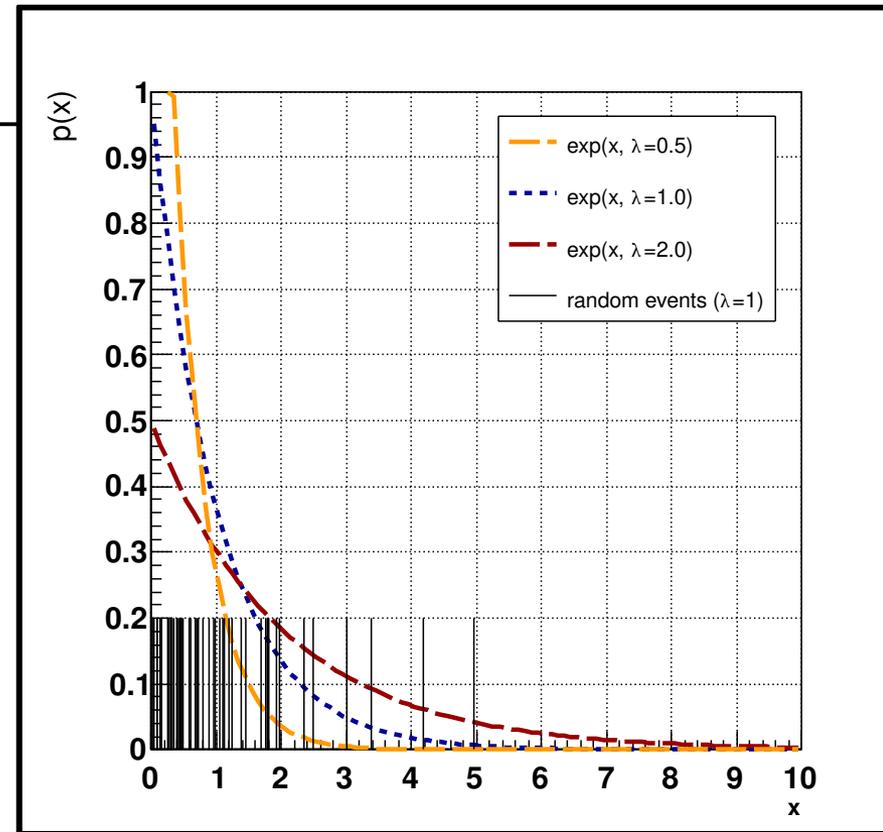
$$\ln(\mathcal{L}(\{x_i\}, \theta)) = \sum_{i=1}^{50} \left(-\ln(\theta) - \frac{x_i}{\theta} \right)$$

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

$$\begin{aligned} \frac{\partial}{\partial \theta} \ln(\mathcal{L}(\{x_i\}, \theta)) &= \sum_{i=1}^{50} \left(-\frac{1}{\theta} + \frac{x_i}{\theta^2} \right) \\ &= \sum_{i=1}^{50} \frac{-\theta + x_i}{\theta^2} = 0; \end{aligned}$$

$$\sum_{i=1}^{50} x_i = 50 \hat{\theta}_{\text{ML}};$$

$$\hat{\theta}_{\text{ML}} = \frac{1}{50} \sum_{i=1}^{50} x_i \quad \text{ML Schätzfunktion entspricht arithmetischem Mittel.}$$



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

- Beispielwerte für $\ln(\mathcal{L}(\theta))$:

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

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

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

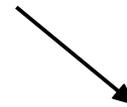
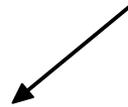
Erwartungswert von $\hat{\theta}_{\text{ML}}$

$\hat{\theta}_{\text{ML}}$ ist konsistent
& erwartungstreu.

- Erwartungswert von $\hat{\theta}_{\text{ML}}$ am Beispiel $p(x) = \exp(x, \theta)$:

[Zurück zu Folie 32](#)
[Zurück zu Folie 34](#)

$$\begin{aligned}
 E[\hat{\theta}_{\text{ML}}(\vec{x})] &= \int \cdots \int \hat{\theta}_{\text{ML}}(\vec{x}) \mathcal{L}(\vec{x}, \theta) dx_1 \cdots dx_n \\
 &= \int \cdots \int \frac{1}{n} \left(\sum_{i=1}^n x_i \right) \frac{1}{\theta} e^{-\frac{x_1}{\theta}} \cdots \frac{1}{\theta} e^{-\frac{x_n}{\theta}} dx_1 \cdots dx_n \\
 &= \frac{1}{n} \sum_{i=1}^n \left(\underbrace{\int x_i \frac{1}{\theta} e^{-\frac{x_i}{\theta}} dx_i}_{=1} \prod_{i \neq j} \underbrace{\int \frac{1}{\theta} e^{-\frac{x_j}{\theta}} dx_j}_{=1} \right) = \frac{1}{n} \sum_{i=1}^n \theta = \theta
 \end{aligned}$$



$$\int_0^{\infty} \frac{x}{\theta} e^{-\frac{x}{\theta}} dx = \theta \int_0^{\infty} x' e^{-x'} dx' = \theta$$

mit: $x' = \frac{x}{\theta}$, $dx = \theta dx'$

$$\int_0^{\infty} x' e^{-x'} dx' = \underbrace{\left[-x' e^{-x'} \right]_0^{\infty}}_{=0} + \underbrace{\int_0^{\infty} e^{-x'} dx'}_{=1} = 1$$

$$= 1$$

Erwartungswert von $\hat{\theta}_{\text{ML}}$

$\hat{\theta}_{\text{ML}}$ ist konsistent
& erwartungstreu.

- Erwartungswert von $\hat{\theta}_{\text{ML}}$ am Beispiel $p(x) = \exp(x, \theta)$:

$$\begin{aligned} E[\hat{\theta}_{\text{ML}}(\vec{x})] &= \int \cdots \int \hat{\theta}_{\text{ML}}(\vec{x}) \mathcal{L}(\vec{x}, \theta) dx_1 \cdots dx_n \\ &= \int \cdots \int \frac{1}{n} \left(\sum_{i=1}^n x_i \right) \frac{1}{\theta} e^{-\frac{x_1}{\theta}} \cdots \frac{1}{\theta} e^{-\frac{x_n}{\theta}} dx_1 \cdots dx_n \\ &= \frac{1}{n} \sum_{i=1}^n \left(\int x_i \frac{1}{\theta} e^{-\frac{x_i}{\theta}} dx_i \prod_{i \neq j} \int \frac{1}{\theta} e^{-\frac{x_j}{\theta}} dx_j \right) = \frac{1}{n} \sum_{i=1}^n \theta = \theta \end{aligned}$$

- Dieses Ergebnis war zu erwarten – warum?

Erwartungswert von $\hat{\theta}_{\text{ML}}$

$\hat{\theta}_{\text{ML}}$ ist konsistent
& erwartungstreu.

- Erwartungswert von $\hat{\theta}_{\text{ML}}$ am Beispiel $p(x) = \exp(x, \theta)$:

$$\begin{aligned}
 E[\hat{\theta}_{\text{ML}}(\vec{x})] &= \int \cdots \int \hat{\theta}_{\text{ML}}(\vec{x}) \mathcal{L}(\vec{x}, \theta) dx_1 \cdots dx_n \\
 &= \int \cdots \int \frac{1}{n} \left(\sum_{i=1}^n x_i \right) \frac{1}{\theta} e^{-\frac{x_1}{\theta}} \cdots \frac{1}{\theta} e^{-\frac{x_n}{\theta}} dx_1 \cdots dx_n \\
 &= \frac{1}{n} \sum_{i=1}^n \left(\int x_i \frac{1}{\theta} e^{-\frac{x_i}{\theta}} dx_i \prod_{i \neq j} \int \frac{1}{\theta} e^{-\frac{x_j}{\theta}} dx_j \right) = \frac{1}{n} \sum_{i=1}^n \theta = \theta
 \end{aligned}$$

- Dieses Ergebnis war zu erwarten – warum?
 - $\hat{\theta}_{\text{ML}}$ erweist sich als das arithmetische Mittel der Einzelmessungen $\{x_i\}$, die nach $p(x, \theta) = \exp(x, \theta)$ verteilt sind. (Siehe [Folie 22](#))

Erwartungswert von $\hat{\theta}_{\text{ML}}$

$\hat{\theta}_{\text{ML}}$ ist konsistent
& erwartungstreu.

- Erwartungswert von $\hat{\theta}_{\text{ML}}$ am Beispiel $p(x) = \exp(x, \theta)$:

$$\begin{aligned}
 E[\hat{\theta}_{\text{ML}}(\vec{x})] &= \int \cdots \int \hat{\theta}_{\text{ML}}(\vec{x}) \mathcal{L}(\vec{x}, \theta) dx_1 \cdots dx_n \\
 &= \int \cdots \int \frac{1}{n} \left(\sum_{i=1}^n x_i \right) \frac{1}{\theta} e^{-\frac{x_1}{\theta}} \cdots \frac{1}{\theta} e^{-\frac{x_n}{\theta}} dx_1 \cdots dx_n \\
 &= \frac{1}{n} \sum_{i=1}^n \left(\int x_i \frac{1}{\theta} e^{-\frac{x_i}{\theta}} dx_i \prod_{i \neq j} \int \frac{1}{\theta} e^{-\frac{x_j}{\theta}} dx_j \right) = \frac{1}{n} \sum_{i=1}^n \theta = \theta
 \end{aligned}$$

- Dieses Ergebnis war zu erwarten – warum?
 - $\hat{\theta}_{\text{ML}}$ erweist sich als das arithmetische Mittel der Einzelmessungen $\{x_i\}$, die nach $p(x, \theta) = \exp(x, \theta)$ verteilt sind. (Siehe [Folie 22](#))
 - Wir haben gelernt, dass das arithmetische Mittel der Einzelmessungen allg. eine erwartungstreue Schätzfunktion für den Erwartungswert von $p(x, \theta)$ ist. (Siehe [VL-01 Folie 47ff](#))

Erwartungswert von $\hat{\theta}_{\text{ML}}$

$\hat{\theta}_{\text{ML}}$ ist konsistent
& erwartungstreu.

- Erwartungswert von $\hat{\theta}_{\text{ML}}$ am Beispiel $p(x) = \exp(x, \theta)$:

$$\begin{aligned}
 E[\hat{\theta}_{\text{ML}}(\vec{x})] &= \int \cdots \int \hat{\theta}_{\text{ML}}(\vec{x}) \mathcal{L}(\vec{x}, \theta) dx_1 \cdots dx_n \\
 &= \int \cdots \int \frac{1}{n} \left(\sum_{i=1}^n x_i \right) \frac{1}{\theta} e^{-\frac{x_1}{\theta}} \cdots \frac{1}{\theta} e^{-\frac{x_n}{\theta}} dx_1 \cdots dx_n \\
 &= \frac{1}{n} \sum_{i=1}^n \left(\int x_i \frac{1}{\theta} e^{-\frac{x_i}{\theta}} dx_i \prod_{i \neq j} \int \frac{1}{\theta} e^{-\frac{x_j}{\theta}} dx_j \right) = \frac{1}{n} \sum_{i=1}^n \theta = \theta
 \end{aligned}$$

- Dieses Ergebnis war zu erwarten – warum?
 - $\hat{\theta}_{\text{ML}}$ erweist sich als das arithmetische Mittel der Einzelmessungen $\{x_i\}$, die nach $p(x, \theta) = \exp(x, \theta)$ verteilt sind. (Siehe [Folie 22](#))
 - Wir haben gelernt, dass das arithmetische Mittel der Einzelmessungen allg. eine erwartungstreue Schätzfunktion für den Erwartungswert von $p(x, \theta)$ ist. (Siehe [VL-01 Folie 47ff](#))
 - θ ist der Erwartungswert von $\exp(x, \theta)$. (Siehe [Folie 21](#))

Varianz von $\hat{\theta}_{\text{ML}}$

- **Statistische Unsicherheit** auf $\hat{\theta}_{\text{ML}}$ am Beispiel $p(x) = \exp(x, \theta)$:

$$\begin{aligned}\text{var}[\hat{\theta}_{\text{ML}}(\vec{x})] &= E[\hat{\theta}_{\text{ML}}^2(\vec{x})] - E[\hat{\theta}_{\text{ML}}(\vec{x})]^2 \\ &= \int \cdots \int \left(\hat{\theta}_{\text{ML}}(\vec{x}) \right)^2 \mathcal{L}(\vec{x}, \theta) \prod_{i \leq n} dx_i \\ &\quad - \left(\int \cdots \int \hat{\theta}_{\text{ML}}(\vec{x}) \mathcal{L}(\vec{x}, \theta) \prod_{i \leq n} dx_i \right)^2 = \frac{\theta^2}{n}\end{aligned}$$

Varianz von $\hat{\theta}_{\text{ML}}$

- **Statistische Unsicherheit** auf $\hat{\theta}_{\text{ML}}$ am Beispiel $p(x) = \exp(x, \theta)$:

$$\begin{aligned} \text{var}[\hat{\theta}_{\text{ML}}(\vec{x})] &= E[\hat{\theta}_{\text{ML}}^2(\vec{x})] - E[\hat{\theta}_{\text{ML}}(\vec{x})]^2 \\ &= \int \cdots \int \left(\hat{\theta}_{\text{ML}}(\vec{x}) \right)^2 \mathcal{L}(\vec{x}, \theta) \prod_{i \leq n} dx_i \\ &\quad - \left(\int \cdots \int \hat{\theta}_{\text{ML}}(\vec{x}) \mathcal{L}(\vec{x}, \theta) \prod_{i \leq n} dx_i \right)^2 = \boxed{\frac{\theta^2}{n}} \end{aligned}$$



- Auch dieses Ergebnis können Sie 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 $\exp(x, \theta)$.

3 Parameterschätzung mit Hilfe der ML Methode

3.4 Methoden zur Varianzbestimmung

Im Folgenden werden wir die Varianz der ML Schätzfunktion noch etwas genauer diskutieren.



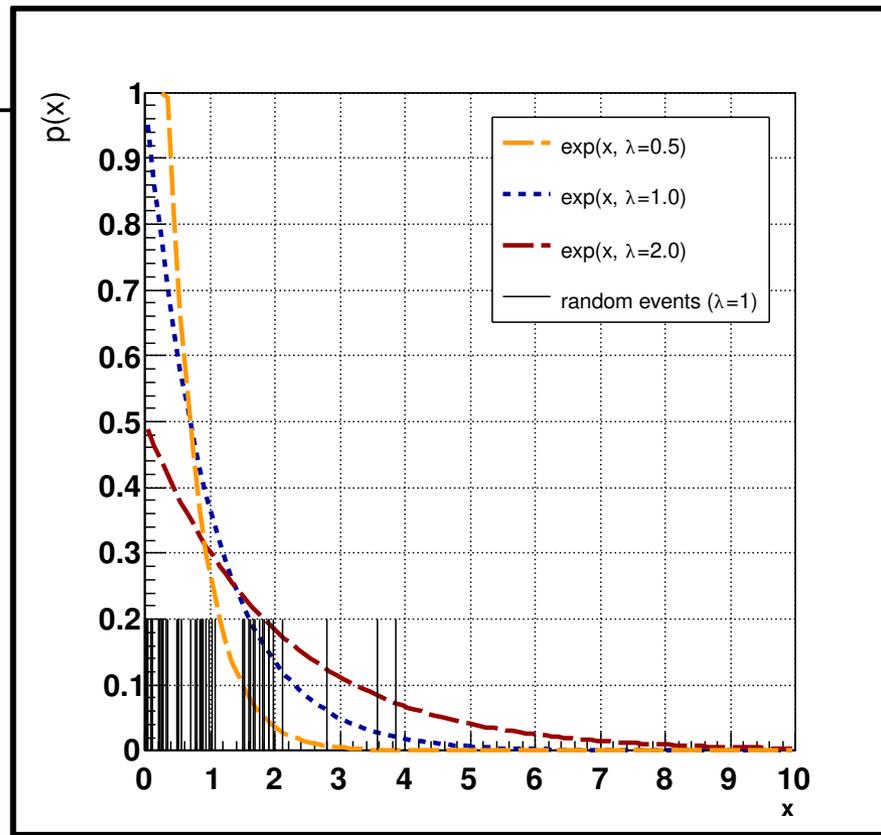
Konkreter Ausgang ...

- Wir gehen im weiteren vom folgenden Ausgang der Stichprobe aus 50 Einzelmessungen aus:

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

Maximum Likelihood
Schätzwert.

Standardabweichung des
Schätzwertes.



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

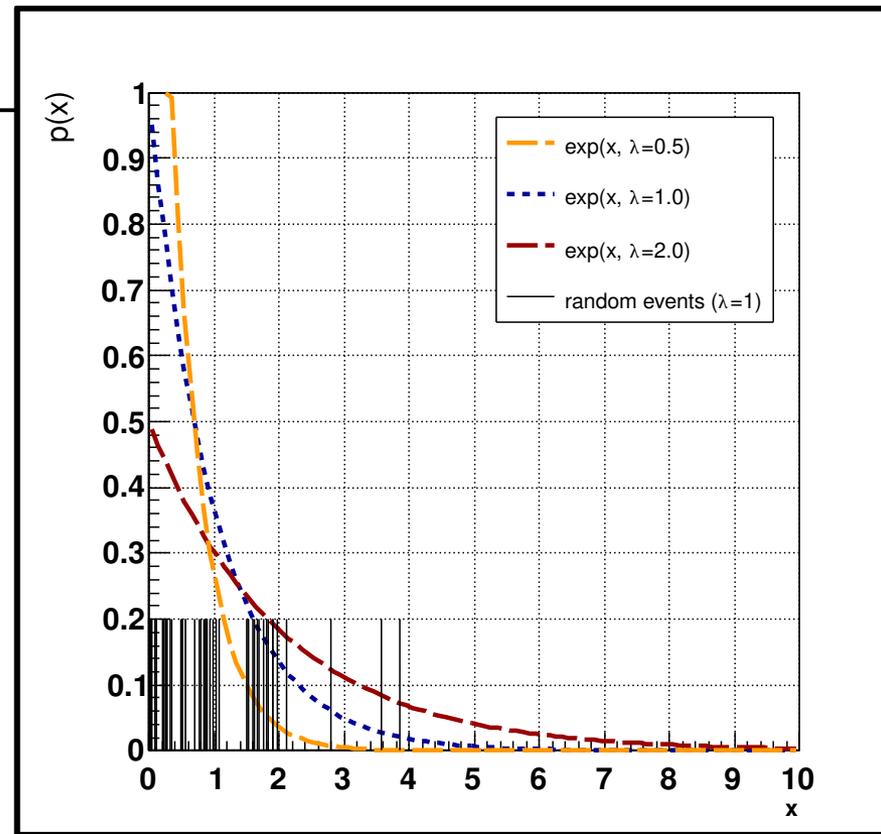
Konkreter Ausgang ...

- Wir gehen im weiteren vom folgenden Ausgang der Stichprobe aus 50 Einzelmessungen aus:

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

Maximum Likelihood
Schätzwert.

Standardabweichung des
Schätzwertes.



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

- Interpretation:**

Aus der Messreihe die Sie vorgenommen haben haben Sie den Wert 1.022 s für die Lebensdauer des Präparats erhalten. Wenn Sie die Messreihe (aus 50 Einzelmessungen) sehr oft wiederholen würde dieser Wert innerhalb von ± 0.145 s streuen.

Konkreter Ausgang ...

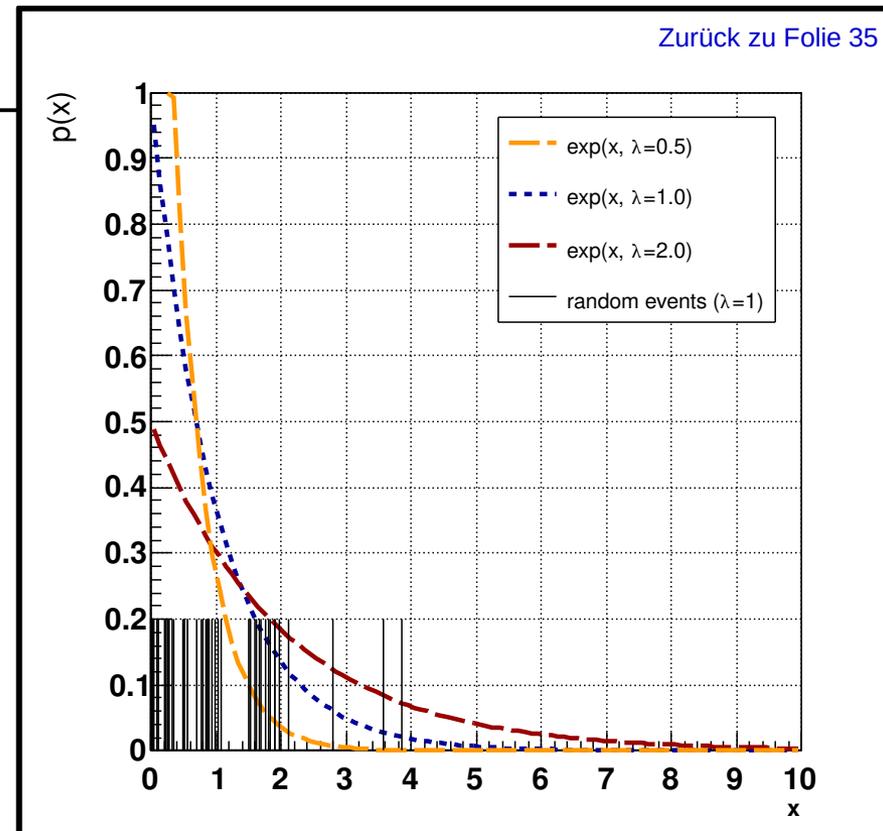
- Wir gehen im weiteren vom folgenden Ausgang der Stichprobe aus 50 Einzelmessungen aus:

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

Maximum Likelihood
Schätzwert.

Standardabweichung des
Schätzwertes.

Die Formel für die Varianz enthält den wahren Wert θ . Wir haben hier θ in der Formel durch $\hat{\theta}$ ersetzt.



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

• Interpretation:

Aus der Messreihe die Sie vorgenommen haben haben Sie den Wert 1.022 s für die Lebensdauer des Präparats erhalten. Wenn Sie die Messreihe (aus 50 Einzelmessungen) sehr oft wiederholen würde dieser Wert innerhalb von ± 0.145 s streuen.

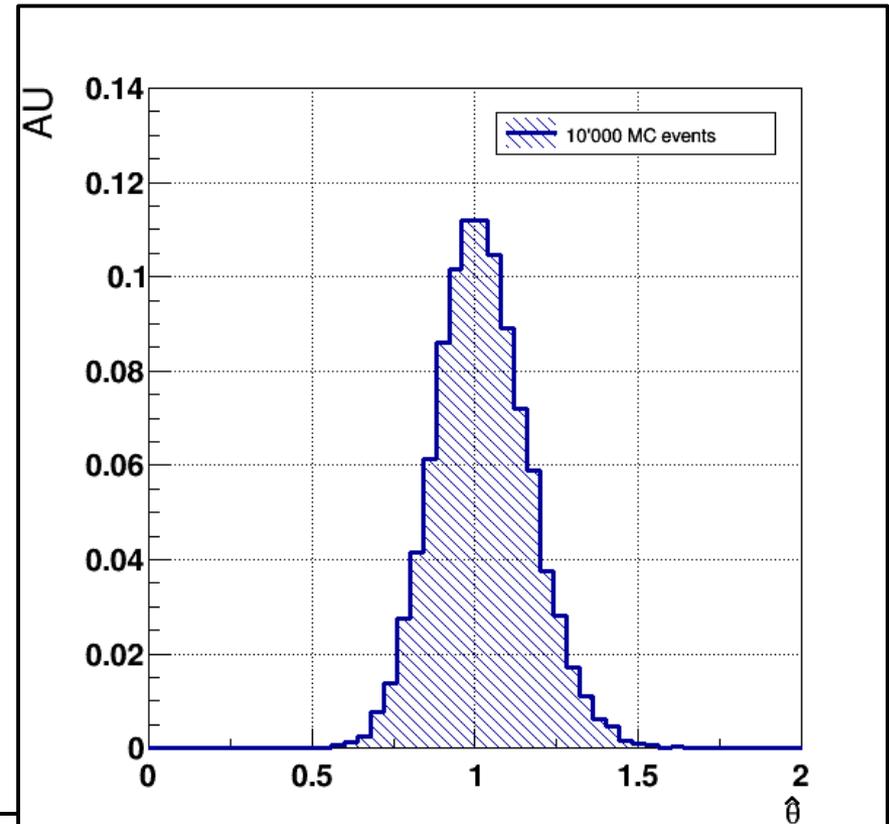
Anwendung der Monte Carlo Methode

- 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 in einem Ensemble Test so viele Pseudo-Experimente (aus jeweils $n = 50$ Einzelmessungen) wie möglich.
- Für die wahre Lebensdauer setzen Sie den Schätzwert aus der Messung $\hat{\theta}_{ML} = 1.022 \text{ s}$.
- Ergebnis für diesen Satz von 10000 Pseudo-Experimenten:

$$\langle \hat{\theta}_{MC} \rangle = 1.022 \quad \sqrt{\text{var}(\hat{\theta}_{MC})} = 0.144$$

d.h. deckungsgleich mit dem analytischen Ergebnis.

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



RCF-Ungleichung

- In Fällen in denen sowohl die analytische Berechnung, als auch eine Abschätzung durch Monte Carlo Methoden zu aufwändig ist, ist es möglich die Varianz durch die **Rao-Cramér-Frechet (RCF) Ungleichung** (a.k.a. Informationsungleichung) abzuschätzen:

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

RCF-Ungleichung

- In Fällen in denen sowohl die analytische Berechnung, als auch eine Abschätzung durch Monte Carlo Methoden zu aufwändig ist, ist es möglich die Varianz durch die **Rao-Cramér-Frechet (RCF) Ungleichung** (a.k.a. Informationsungleichung) abzuschätzen:

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

Verzerrung/
bias.

RCF-Ungleichung

- In Fällen in denen sowohl die analytische Berechnung, als auch eine Abschätzung durch Monte Carlo Methoden zu aufwändig ist, ist es möglich die Varianz durch die **Rao-Cramér-Frechet (RCF) Ungleichung** (a.k.a. Informationsungleichung) abzuschätzen:

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

Verzerrung/
bias.

Likelihood
Funktion.

RCF-Ungleichung

- In Fällen in denen sowohl die analytische Berechnung, als auch eine Abschätzung durch Monte Carlo Methoden zu aufwändig ist, ist es möglich die Varianz durch die **Rao-Cramér-Frechet (RCF) Ungleichung** (a.k.a. Informationsungleichung) abzuschätzen:

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

Verzerrung/
bias.

Likelihood
Funktion.

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

Beweis: RCF-Ungleichung – 1 –

- Ausgangspunkt ist die Ableitung der Definitionsgleichung der Verzerrung (vgl. Folie 9) nach θ :

$$b = E[\hat{\theta}] - \theta = E[\hat{\theta} - \theta] = \int_{\Omega} (\hat{\theta} - \theta) \mathcal{L}(x, \theta) dx$$

$$\begin{aligned} \frac{\partial b}{\partial \theta} &= \int_{\Omega} (\hat{\theta} - \theta) \frac{\partial \mathcal{L}}{\partial \theta} dx - \int_{\Omega} \mathcal{L}(x, \theta) dx = \int_{\Omega} (\hat{\theta} - \theta) \mathcal{L}(x, \theta) \frac{\partial (\ln \mathcal{L})}{\partial \theta} dx - \int_{\Omega} \mathcal{L}(x, \theta) dx \\ &= \int_{\Omega} (\hat{\theta} - \theta) \mathcal{L}(x, \theta) \frac{\partial (\ln \mathcal{L})}{\partial \theta} dx - 1 \end{aligned}$$

$$1 + \frac{\partial b}{\partial \theta} = \int_{\Omega} \underbrace{\left((\hat{\theta} - \theta) \sqrt{\mathcal{L}} \right)}_{\equiv x} \underbrace{\left(\sqrt{\mathcal{L}} \frac{\partial (\ln \mathcal{L})}{\partial \theta} \right)}_{\equiv y} dx = \langle x, y \rangle$$

- $\langle x, y \rangle$ ist ein Skalarprodukt aus der Funktionalanalysis, wie Sie es aus der QM kennen. Hierfür gilt die **Cauchy-Schwartzsche Ungleichung**:

$$|\langle x, y \rangle|^2 \leq \langle x, x \rangle \langle y, y \rangle$$

Beweis: RCF-Ungleichung – 2 –

- Anwendung der Cauchy-Schwartzschen Ungleichung:

$$|\langle x, y \rangle|^2 \leq \langle x, x \rangle \langle y, y \rangle$$

$$\left(1 + \frac{\partial b}{\partial \theta}\right)^2 \leq \underbrace{\left(\int_{\Omega} (\hat{\theta} - \theta)^2 \mathcal{L} \, dx\right)}_{\equiv \text{var}[\hat{\theta}]} \underbrace{\left(\int_{\Omega} \mathcal{L} \left(\frac{\partial(\ln \mathcal{L})}{\partial \theta}\right)^2 dx\right)}_{\equiv E \left[\left(\frac{\partial(\ln \mathcal{L})}{\partial \theta}\right)^2 \right]}$$

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

Beweis: RCF-Ungleichung – 3 –

- Umformung des Erwartungswerts:

$$\text{var}[\hat{\theta}] \geq \frac{\left(1 + \frac{\partial b}{\partial \theta}\right)^2}{E \left[\left(\frac{\partial (\ln \mathcal{L})}{\partial \theta} \right)^2 \right]} \quad ; \quad \int_{\Omega} \mathcal{L} \, dx = 1$$

$$\frac{\partial}{\partial \theta} \int_{\Omega} \mathcal{L} \, dx = \int_{\Omega} \frac{\partial \mathcal{L}}{\partial \theta} \, dx = \int_{\Omega} \mathcal{L} \frac{\partial (\ln \mathcal{L})}{\partial \theta} \, dx = E \left[\frac{\partial (\ln \mathcal{L})}{\partial \theta} \right] = 0$$

$$\begin{aligned} \frac{\partial^2}{\partial \theta^2} \int_{\Omega} \mathcal{L} \, dx &= \int_{\Omega} \frac{\partial^2 \mathcal{L}}{\partial \theta^2} \, dx = \int_{\Omega} \underbrace{\frac{\partial \mathcal{L}}{\partial \theta} \frac{\partial (\ln \mathcal{L})}{\partial \theta}} \, dx + \int_{\Omega} \mathcal{L} \frac{\partial^2 (\ln \mathcal{L})}{\partial \theta^2} \, dx = 0 \\ &\equiv \mathcal{L} \left(\frac{\partial (\ln \mathcal{L})}{\partial \theta} \right)^2 \quad \text{mit: } \frac{\partial (\ln \mathcal{L})}{\partial \theta} = \frac{1}{\mathcal{L}} \frac{\partial \mathcal{L}}{\partial \theta} \end{aligned}$$

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

Beispiel: Auswertung der RCF-Ungleichung

- In unserem Beispiel des radioaktiven Zerfalls mit $p(x) = \exp(x, \theta)$ lässt sich die untere Schwelle der RCF-Ungleichung explizit ausrechnen:

$$\begin{aligned} \text{var}[\hat{\theta}_{\text{ML}}] &\geq \frac{\left(1 + \frac{\partial b}{\partial \theta} \Big|_{\theta=\hat{\theta}_{\text{ML}}}\right)^2}{E\left[-\frac{\partial^2}{\partial \theta^2} \ln \mathcal{L}\right]} \\ &\geq \frac{1}{E\left[-\frac{n}{\theta^2} \left(1 - \frac{2\hat{\theta}_{\text{ML}}}{\theta}\right)\right]} = \frac{1}{-\frac{n}{\theta^2} \left(1 - \frac{2E[\hat{\theta}_{\text{ML}}]}{\theta}\right)} = \frac{\theta^2}{n} \end{aligned}$$

mit:

$$b = 0 \text{ (const)} \quad E[\hat{\theta}_{\text{ML}}] = \theta \quad (\text{Siehe Folie 23})$$

$$\frac{\partial^2}{\partial \theta^2} \ln \mathcal{L} = \frac{n}{\theta^2} \left(1 - \frac{2}{\theta} \frac{1}{n} \sum_{i \leq n} x_i\right) = \frac{n}{\theta^2} \left(1 - \frac{2\hat{\theta}_{\text{ML}}}{\theta}\right) \quad (\text{Siehe Folie 22})$$

Beispiel: Auswertung der RCF-Ungleichung

- In unserem Beispiel des radioaktiven Zerfalls mit $p(x) = \exp(x, \theta)$ lässt sich die untere Schwelle der RCF-Ungleichung explizit ausrechnen:

$$\begin{aligned} \text{var}[\hat{\theta}_{\text{ML}}] &\geq \frac{\left(1 + \frac{\partial b}{\partial \theta} \Big|_{\theta=\hat{\theta}_{\text{ML}}}\right)^2}{E\left[-\frac{\partial^2}{\partial \theta^2} \ln \mathcal{L}\right]} \\ &\geq \frac{1}{E\left[-\frac{n}{\theta^2} \left(1 - \frac{2\hat{\theta}_{\text{ML}}}{\theta}\right)\right]} = \frac{1}{-\frac{n}{\theta^2} \left(1 - \frac{2E[\hat{\theta}_{\text{ML}}]}{\theta}\right)} = \frac{\theta^2}{n} \end{aligned}$$

(Siehe Folie 24)

mit:

$$b = 0 \text{ (const)} \quad E[\hat{\theta}_{\text{ML}}] = \theta \quad \text{(Siehe Folie 23)}$$

$$\frac{\partial^2}{\partial \theta^2} \ln \mathcal{L} = \frac{n}{\theta^2} \left(1 - \frac{2}{\theta} \frac{1}{n} \sum_{i \leq n} x_i\right) = \frac{n}{\theta^2} \left(1 - \frac{2\hat{\theta}_{\text{ML}}}{\theta}\right) \quad \text{(Siehe Folie 22)}$$

$\hat{\theta}_{\text{ML}} = \frac{1}{n} \sum x_i$ ist also eine effiziente Schätzfunktion.

Beispiel: Auswertung der RCF-Ungleichung

- In unserem Beispiel des radioaktiven Zerfalls mit $p(x) = \exp(x, \theta)$ lässt sich die untere Schwelle der RCF-Ungleichung explizit ausrechnen:

$$\begin{aligned} \text{var}[\hat{\theta}_{\text{ML}}] &\geq \frac{\left(1 + \frac{\partial b}{\partial \theta} \Big|_{\theta=\hat{\theta}_{\text{ML}}}\right)^2}{E\left[-\frac{\partial^2}{\partial \theta^2} \ln \mathcal{L}\right]} \\ &\geq \frac{1}{E\left[-\frac{n}{\theta^2} \left(1 - \frac{2\hat{\theta}_{\text{ML}}}{\theta}\right)\right]} = \frac{1}{-\frac{n}{\theta^2} \left(1 - \frac{2E[\hat{\theta}_{\text{ML}}]}{\theta}\right)} = \frac{\theta^2}{n} \quad (\text{Siehe Folie 24}) \end{aligned}$$

mit:

$$b = 0 \text{ (const)} \quad E[\hat{\theta}_{\text{ML}}] = \theta \quad (\text{Siehe Folie 23})$$

$$\frac{\partial^2}{\partial \theta^2} \ln \mathcal{L} = \frac{n}{\theta^2} \left(1 - \frac{2}{\theta} \frac{1}{n} \sum_{i \leq n} x_i\right) = \frac{n}{\theta^2} \left(1 - \frac{2\hat{\theta}_{\text{ML}}}{\theta}\right) \quad (\text{Siehe Folie 22})$$

- Ohne Beweis:** Wenn es für ein Problem überhaupt effiziente Schätzfunktionen gibt, dann ist die ML Schätzfunktion effizient. D.h. für eine ML Schätzfunktion wird die RCF-Ungleichung zur Gleichung.

Graphische Abschätzung der Unsicherheit

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

(Siehe Folie 23)

$$\text{var}[\hat{\theta}_{\text{ML}}] = -\frac{1}{E\left[-\frac{\partial^2}{\partial\theta^2} \ln \mathcal{L}\right]} \longrightarrow \hat{\text{var}}[\hat{\theta}_{\text{ML}}] = \left[-\frac{1}{\frac{\partial^2}{\partial\theta^2} \ln \mathcal{L}}\right]_{\theta=\hat{\theta}_{\text{ML}}} \equiv \hat{\sigma}^2$$

Graphische Abschätzung der Unsicherheit

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

(Siehe Folie 23)

$$\text{var}[\hat{\theta}_{\text{ML}}] = -\frac{1}{E\left[-\frac{\partial^2}{\partial\theta^2} \ln \mathcal{L}\right]} \longrightarrow \hat{\text{var}}[\hat{\theta}_{\text{ML}}] = \left[-\frac{1}{\frac{\partial^2}{\partial\theta^2} \ln \mathcal{L}}\right]_{\theta=\hat{\theta}_{\text{ML}}} \equiv \hat{\sigma}^2$$

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

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

$$\mathcal{L}(\theta) = \text{const.} \exp\left(\frac{1}{2} \left[\frac{\partial^2}{\partial\theta^2} \ln \mathcal{L}\right]_{\theta=\hat{\theta}_{\text{ML}}} (\theta - \hat{\theta}_{\text{ML}})^2\right) \equiv \text{const.} \exp\left(-\frac{1}{2\hat{\sigma}^2} (\theta - \hat{\theta}_{\text{ML}})^2\right)$$

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

Graphische Abschätzung der Unsicherheit

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

(Siehe Folie 23)

$$\text{var}[\hat{\theta}_{\text{ML}}] = -\frac{1}{E\left[-\frac{\partial^2}{\partial\theta^2} \ln \mathcal{L}\right]} \longrightarrow \text{vâr}[\hat{\theta}_{\text{ML}}] = \left[-\frac{1}{\frac{\partial^2}{\partial\theta^2} \ln \mathcal{L}}\right]_{\theta=\hat{\theta}_{\text{ML}}} \equiv \hat{\sigma}^2$$

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

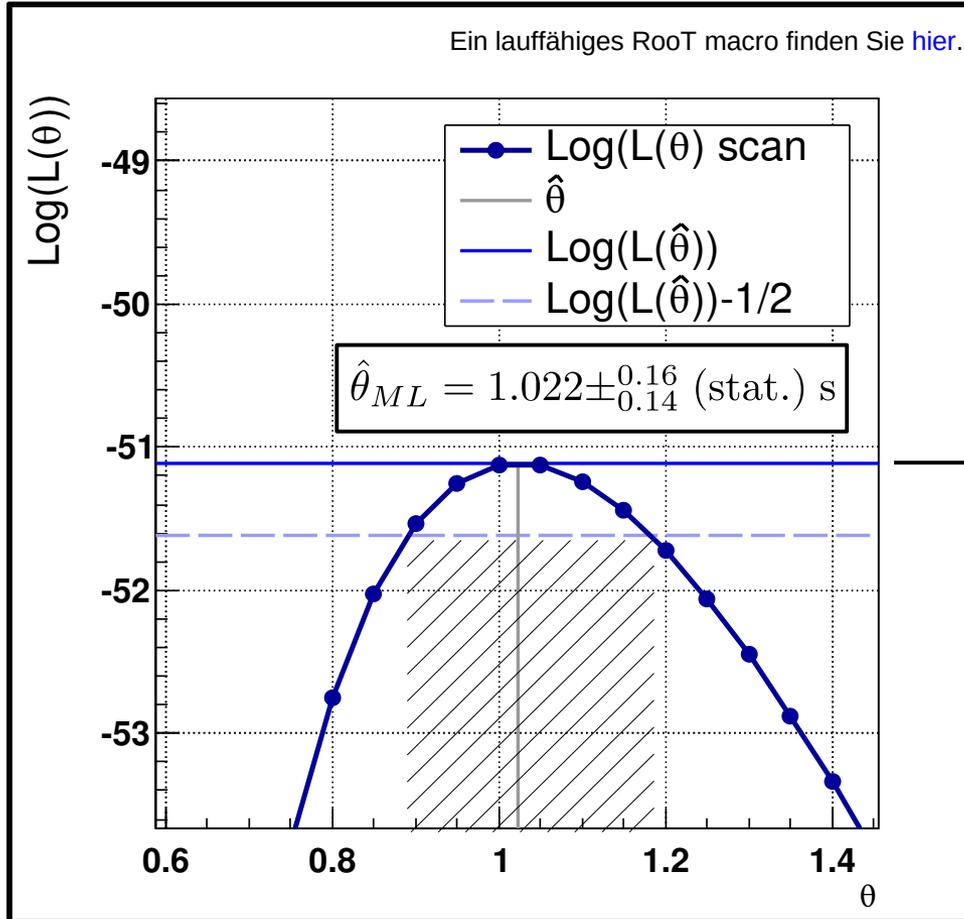
$$\ln \mathcal{L}(\theta) = \underbrace{\ln \mathcal{L}(\hat{\theta}_{\text{ML}})}_{\equiv \mathcal{L}_{\text{max}}} + \underbrace{\left[\frac{\partial}{\partial\theta} \ln \mathcal{L}\right]_{\theta=\hat{\theta}_{\text{ML}}}}_{\equiv 0} (\theta - \hat{\theta}_{\text{ML}}) + \frac{1}{2} \underbrace{\left[\frac{\partial^2}{\partial\theta^2}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

$$\mathcal{L}(\theta) = \text{const.} \exp\left(\frac{1}{2} \left[\frac{\partial^2}{\partial\theta^2} \ln \mathcal{L}\right]_{\theta=\hat{\theta}_{\text{ML}}} (\theta - \hat{\theta}_{\text{ML}})^2\right) \equiv \text{const.} \exp\left(-\frac{1}{2\hat{\sigma}^2} (\theta - \hat{\theta}_{\text{ML}})^2\right)$$

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

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

Anwendung der graphischen Abschätzung



Angewandt auf das
Beispiel aus [Folie 26](#).

(Vgl. mir [Folie 22](#))

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

3 Parameterschätzung mit Hilfe der ML Methode

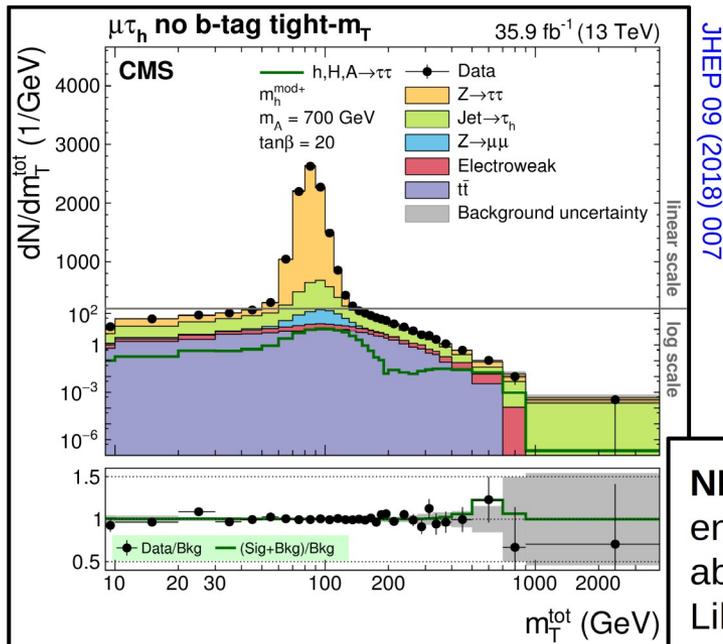
3.5 Weitere Anwendungen

Zum Schluss dieses Kapitels fassen wir noch einige weitere wichtige Anwendungen der ML-Abschätzung zusammen.



Binned Likelihood

- In Fällen, in denen die Länge der Stichprobe sehr groß wird kann es sich als aufwändig erweisen die vollständige Likelihood zu formulieren.
- 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 (→ **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.

Konstruktion der binned Likelihood

- Man unterscheidet bei der Konstruktion der Likelihood zwei Fälle:
- **(i) Normierung des Histogramms nicht Teil des Modells**, d.h. $\sum_{k=1}^N \nu_k = n_{\text{tot}}$:

$$n_{\text{tot}} = \sum_{k=1}^N n_k \quad (\text{Beobachtete Häufigkeit})$$

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

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

$$\ln \left(\mathcal{L}(\vec{n} | \vec{\nu}(\vec{\theta})) \right) = \sum_{k=1}^N n_k \ln \left(\nu_k(\vec{\theta}) \right)^{(*)}$$

Konstruktion der erweiterten binned Likelihood

- Man unterscheidet bei der Konstruktion der Likelihood zwei Fälle:

- (ii) Normierung des Histogramms **Teil des Modells** $\sum_{k=1}^N \nu_k = \nu_{\text{tot}}$:

$$n_{\text{tot}} = \sum_{k=1}^N n_k \quad (\text{Beobachtete H\u00e4ufigkeit})$$

$$\nu_k(\vec{\theta}) = \nu_{\text{tot}} \int_{x_k^{\min}}^{x_k^{\max}} p(x, \vec{\theta}) dx \quad (\text{Erwartete H\u00e4ufigkeit/Bin})$$

$$\nu_{\text{tot}}(\vec{\theta}) = \sum_{k=1}^N \nu_k(\vec{\theta}) \quad (\text{Erwartete H\u00e4ufigkeit})$$

$$\mathcal{L}(\vec{n} | \vec{\nu}(\vec{\theta})) = \underbrace{\frac{\nu_{\text{tot}}^{n_{\text{tot}}}}{n_{\text{tot}}!} e^{-\nu_{\text{tot}}}}_{\text{Erweiterung}} \underbrace{\frac{n_{\text{tot}}!}{n_1! \dots n_N!} \left(\frac{\nu_1}{\nu_{\text{tot}}}\right)^{n_1} \dots \left(\frac{\nu_N}{\nu_{\text{tot}}}\right)^{n_N}}_{\text{Binned Likelihood}} = \prod_{k=1}^N \frac{\nu_k^{n_k}}{n_k!} e^{-\nu_k}$$

$$\ln \left(\mathcal{L}(\vec{n} | \vec{\nu}(\vec{\theta})) \right) = \sum_{k=1}^N \left(-\nu_k(\vec{\theta}) + n_k \ln \left(\nu_k(\vec{\theta}) \right) \right) = -\nu_{\text{tot}}(\vec{\theta}) + \sum_{k=1}^N \left(n_k \ln \left(\nu_k(\vec{\theta}) \right) \right)^{(*)}$$

Binned vs. unbinned Likelihood

- Für sehr viele schmale Bins geht die binned Likelihood in die unbinned Likelihood über, d.h. im Gegensatz zur χ^2 -Statistik, die wir im weiteren Verlauf dieses Kurses noch diskutieren werden) sind leere oder schwach populierte Bins bei der Anpassung einer binned Likelihood Funktion **kein Problem**.

Beispiele für binned Likelihoods

- Wir geben zwei **konkrete Beispiele** für die Konstruktion und Anpassung einer (erweiterten) binned Likelihood aus der Teilchenphysik:
 - Bestimmung der Masse (\hat{m}_{ML}) eines neuentdeckten Teilchens bei vorgegebener Signalstärke.
 - Bestimmung der Signalstärke (\hat{s}_{ML}) eines neuentdeckten Teilchens bei vorgegebener Masse (m).
- Für beide Fälle stellen Sie sich zwei rekonstruierte Teilchen in einem Detektor der Teilchenphysik vor, deren Energien und Impulsvektoren vollständig vermessen werden konnten. Falls diese Teilchen aus dem Zerfall eines neuen Teilchens stammen gilt für die Summe der Vierervektoren:

$$p_1^\mu = \begin{pmatrix} E_1 \\ \vec{p}_1 \end{pmatrix} \quad p_2^\mu = \begin{pmatrix} E_2 \\ \vec{p}_2 \end{pmatrix}$$

$$(p_1^\mu + p_2^\mu)^2 = (E_1 + E_2)^2 - (\vec{p}_1 + \vec{p}_2)^2 = m^2$$

- Diese Größe finden Sie im folgenden histogrammiert aufgetragen.

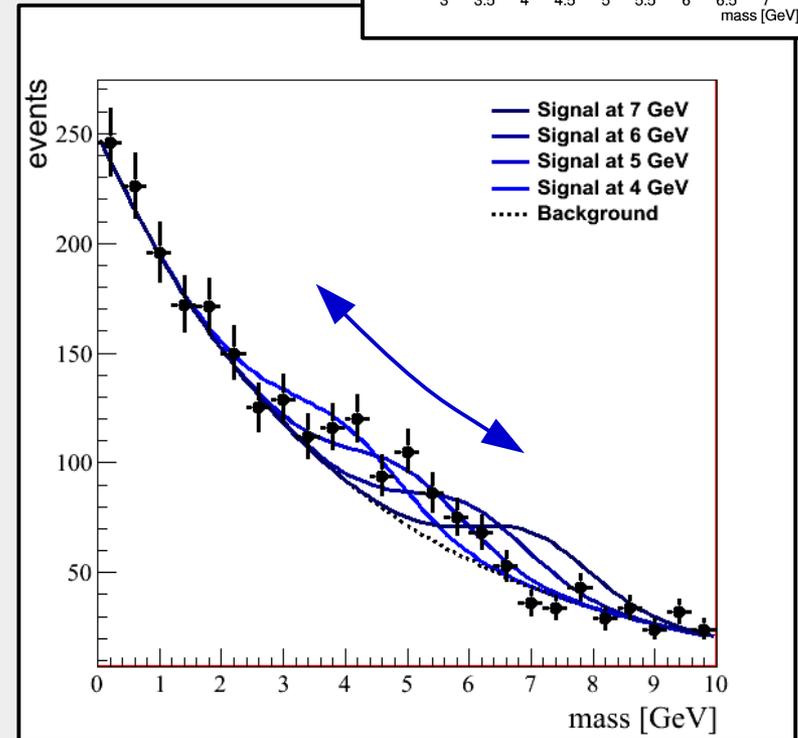
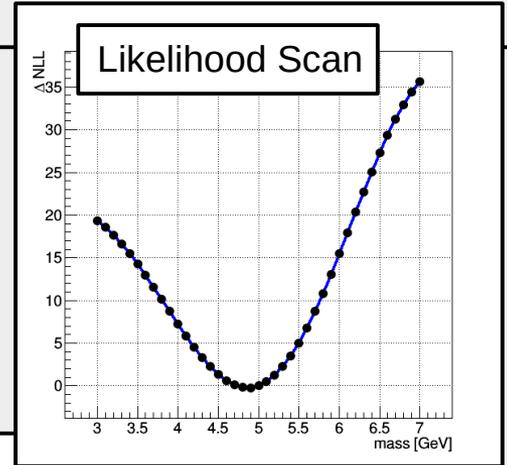
Beispiel – 1

- Signal an unbekannter Stelle über bekanntem Untergrund:
 - Stat. Modell: Erweiterte binned Likelihood.
 - Physikalisches Modell mit vier freien Parametern.
 - *Parameter of interest (POI): Lage des peaks* (θ_3).

$$\mathcal{L}(\{x_k\}, \{\theta_j\}) = \prod_{k=1}^{N=25} P(x_k, \nu_k(\{\theta_j\}))$$

Produkt der Poisson-
dichte 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

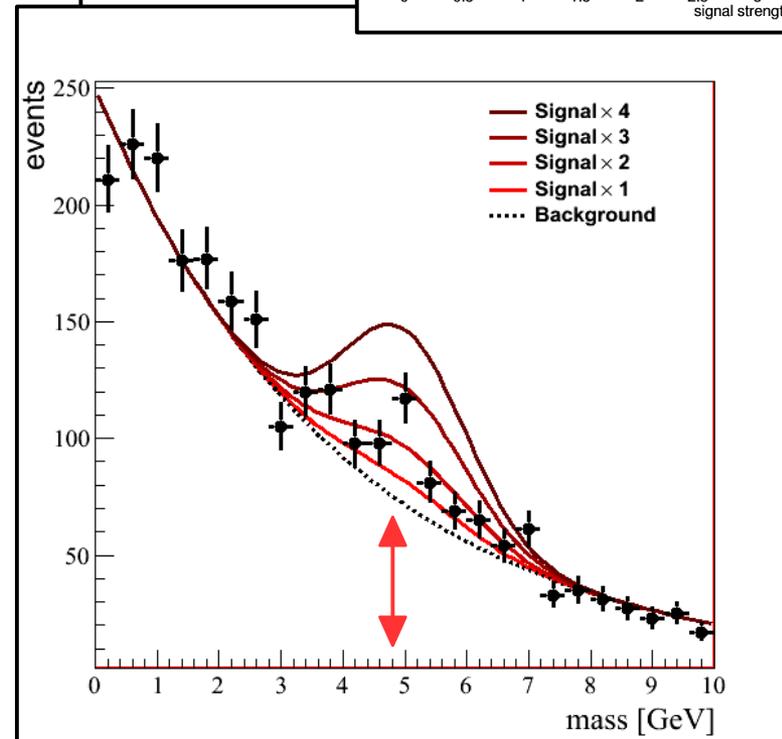
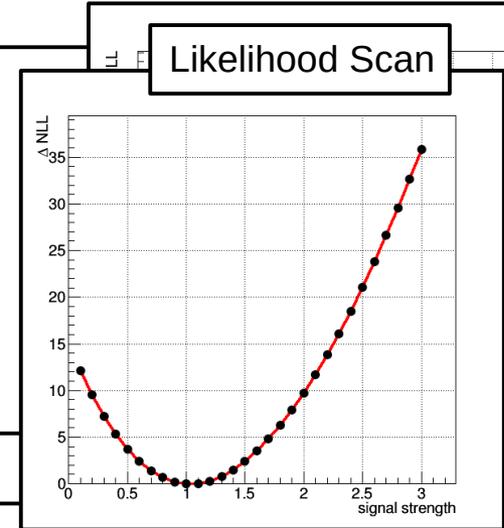
- Signal von unbekannter Stärke über bekanntem Untergrund:
- Stat. Modell: Erweiterte binned Likelihood.
- Physikalisches Modell mit vier freien Parametern.
- *Parameter of interest (POI): Höhe des peaks* bei bekannter Lage (θ_2).

$$\mathcal{L}(\{x_k\}, \{\theta_j\}) = \prod_{k=1}^{N=25} P(x_k, \nu_k(\{\theta_j\}))$$

Produkt der Poisson-
dichte 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](#).



Likelihoods mit vielen Parametern

- In realen Anwendungen kann die Likelihood aus sehr vielen Parametern bestehen, von denen nur einige wenige für die physikalische Messung von Relevanz sind.
 - Man bezeichnet die Parameter, die von physikalischer Relevanz sind als ***parameters of interest (POI)***.
 - Man bezeichnet die Parameter, die nicht von physikalischer Relevanz sind als **Störparameter (*nuisance parameters*)**.
- Man folgt zwei gängigen Ansätzen, um die Anzahl der Parameter in der Likelihood zu reduzieren:
 - Am offensichtlichsten ist es die Störparameter (bei angenommener Wahrscheinlichkeitsdichte) auszuintegrieren → siehe Marginalisierung ([VL-01 Folie 26](#)).
 - Ein zweiter Ansatz besteht darin die Störparameter auf die Werte zu setzen, die die Likelihood maximieren. Man erhält so eine neue Funktion (\mathcal{L}_p), die die Eigenschaften der Likelihood erhält deren Dimension des Parameterraums jedoch um die Anzahl der Störparameter reduziert ist. Die POIs können auf dieser Untermanigfaltigkeit des Parameter-raums gescannt werden. → Man bezeichnet \mathcal{L}_p als **Profile Likelihood**.

Beispiel – 2 mit Störparameter

- Wir diskutieren ein Beispiel für Störparameter in der Likelihood anhand von Beispiel – 2 von [Folie 43](#):

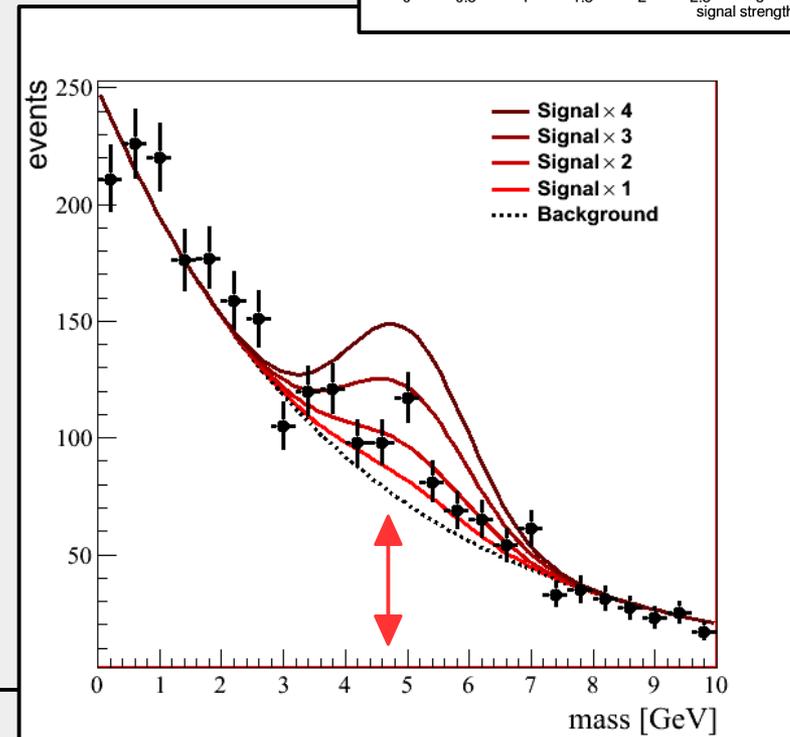
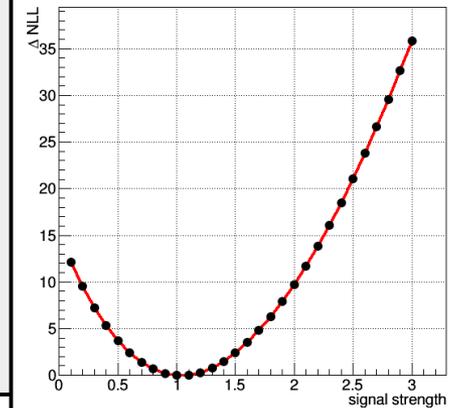
$$\mathcal{L}(\{x_k\}, \{\theta_j\}) = \prod_{k=1}^{N=25} P(x_k, \nu_k(\{\theta_j\}))$$

Produkt der Poisson-
dichte 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](#).

Likelihood Scan



Beispiel – 2 mit Störparameter

- Wir führen einen Störparameter für die Normierung der Untergrundverteilung ein. Hierfür nehmen wir an, dass die Untergrund-Normierung θ_0 in unserem Modell nicht exakt bekannt ist, sondern nur mit einer rel. Unsicherheit von $\sigma(\theta_0)/\theta_0 = \pm 5\%$.

$$\nu_k(\{\theta_j\}) = \underbrace{\tilde{\theta}_0 f(\theta_0/\tilde{\theta}_0, 1, \sigma(\tilde{\theta}_0)/\tilde{\theta}_0)}_{\equiv \rho(\theta_0|\tilde{\theta}_0)} e^{-\theta_1 x_k} + \theta_2 e^{-(\theta_3 - x_k)^2}$$

Wahrscheinlichkeitsdichte für θ_0 bei gegebener Abschätzung $\tilde{\theta}_0$.

Mögliche wahre Werte für θ_0 (auszuintegrieren).

Beste Abschätzung des wahren Wertes in Likelihood

$$\mathcal{L}(\{x_k\}, \{\theta_j\}) = \prod_{k=1}^{N=25} P(x_k, \nu_k(\{\theta_j\}))$$

Produkt der Poisson-
dichte 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](#).

Beispiel – 2 mit Störparameter

- Wir führen einen Störparameter für die Normierung der Untergrundverteilung ein. Hierfür nehmen wir an, dass die Untergrund-Normierung θ_0 in unserem Modell nicht exakt bekannt ist, sondern nur mit einer rel. Unsicherheit von $\sigma(\theta_0)/\theta_0 = \pm 5\%$.

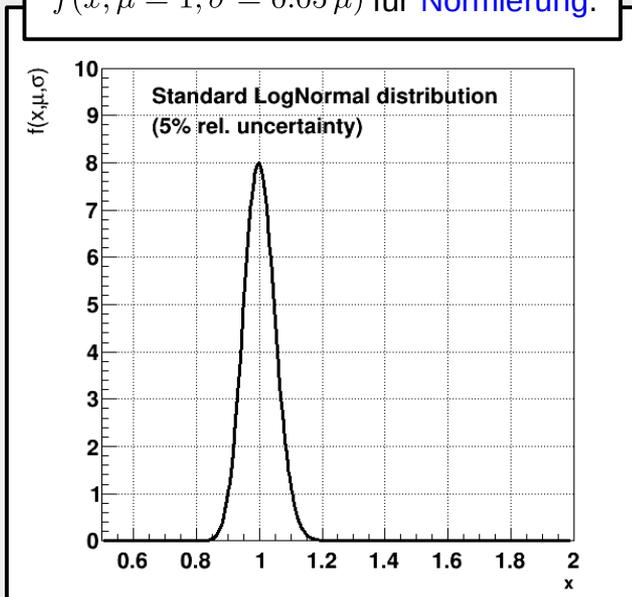
$$\nu_k(\{\theta_j\}) = \underbrace{\tilde{\theta}_0 f(\theta_0/\tilde{\theta}_0, 1, \sigma(\tilde{\theta}_0)/\tilde{\theta}_0)}_{\equiv \rho(\theta_0|\tilde{\theta}_0)} e^{-\theta_1 x_k} + \theta_2 e^{-(\theta_3 - x_k)^2}$$

Wahrscheinlichkeitsdichte für θ_0 bei gegebener Abschätzung $\tilde{\theta}_0$.

Mögliche wahre Werte für θ_0 (auszuintegrieren).

Gewählte Wahrscheinlichkeitsdichte
 $f(x, \mu = 1, \sigma = 0.05 \mu)$ für **Normierung**.

Beste Abschätzung des wahren Wertes in Likelihood



Beispiel – 2 mit Störparameter

- Wir führen einen Störparameter für die Normierung der Untergrundverteilung ein. Hierfür nehmen wir an, dass die Untergrund-Normierung θ_0 in unserem Modell nicht exakt bekannt ist, sondern nur mit einer rel. Unsicherheit von $\sigma(\theta_0)/\theta_0 = \pm 5\%$.

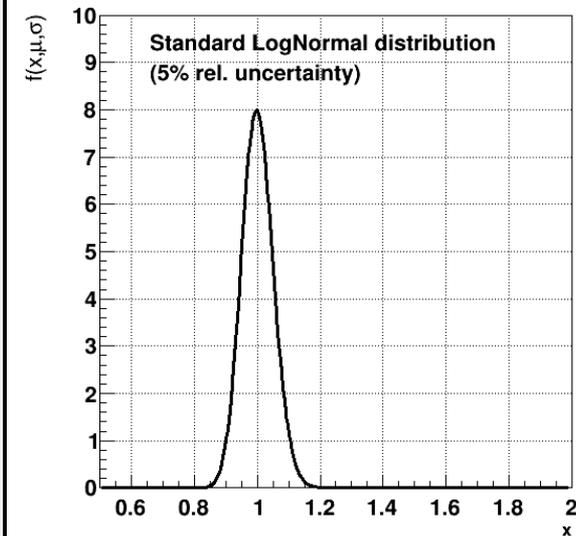
$$\nu_k(\{\theta_j\}) = \underbrace{\tilde{\theta}_0 f(\theta_0/\tilde{\theta}_0, 1, \sigma(\tilde{\theta}_0)/\tilde{\theta}_0)}_{\equiv \rho(\theta_0|\tilde{\theta}_0)} e^{-\theta_1 x_k} + \theta_2 e^{-(\theta_3 - x_k)^2}$$

Wahrscheinlichkeitsdichte für θ_0 bei gegebener Abschätzung $\tilde{\theta}_0$.

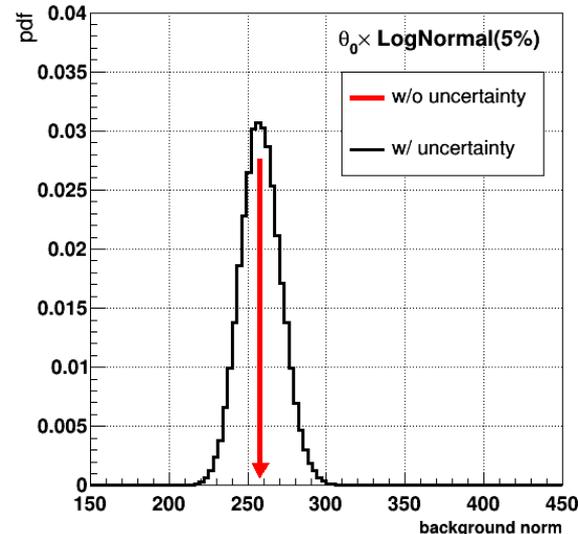
Mögliche wahre Werte für θ_0 (auszuintegrieren).

Beste Abschätzung des wahren Wertes in Likelihood

Gewählte Wahrscheinlichkeitsdichte $f(x, \mu = 1, \sigma = 0.05 \mu)$ für Normierung.



Gewählte Wahrscheinlichkeitsdichte $\rho(\theta_0|\tilde{\theta}_0)$ für θ_0 .



- Für $\rho(\theta_0|\tilde{\theta}_0)$ ist θ_0 der (wahre) Parameter in der Likelihood und $\tilde{\theta}_0$ der in der Likelihood angenommene Schätzwert.
- Bei der Bestimmung der Testverteilung wird der Parameter θ_0 ausintegriert (marginalisiert).
- Dies erfolgt i.d.R. mit Hilfe der MC Methode.

Beispiel – 2 mit Störparameter

- Wir führen einen Störparameter für die Normierung der Untergrundverteilung ein. Hierfür nehmen wir an, dass die Untergrund-Normierung θ_0 in unserem Modell nicht exakt bekannt ist, sondern nur mit einer rel. Unsicherheit von $\sigma(\theta_0)/\theta_0 = \pm 5\%$.

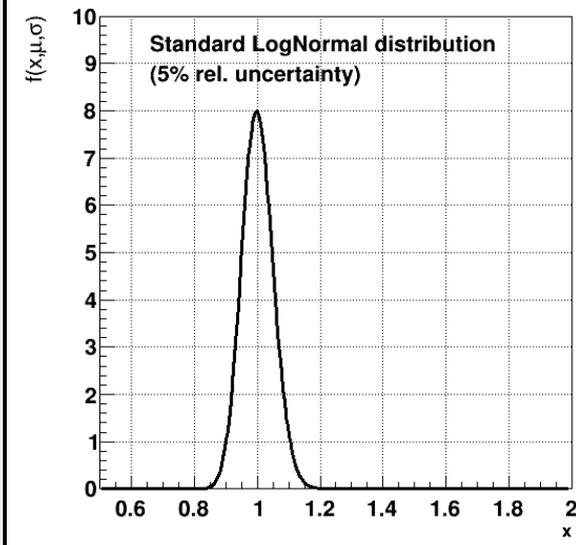
$$\nu_k(\{\theta_j\}) = \underbrace{\tilde{\theta}_0 f(\theta_0/\tilde{\theta}_0, 1, \sigma(\tilde{\theta}_0)/\tilde{\theta}_0)}_{\equiv \rho(\theta_0|\tilde{\theta}_0)} e^{-\theta_1 x_k} + \theta_2 e^{-(\theta_3 - x_k)^2}$$

Wahrscheinlichkeitsdichte für θ_0 bei gegebener Abschätzung $\tilde{\theta}_0$.

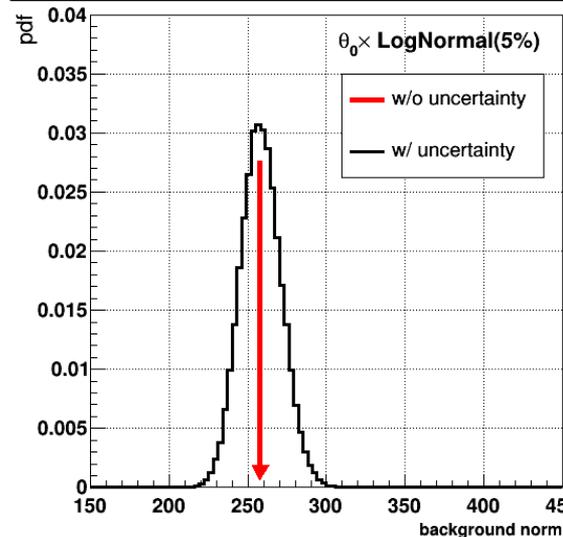
Mögliche wahre Werte für θ_0 (auszuintegrieren).

Beste Abschätzung des wahren Wertes in Likelihood

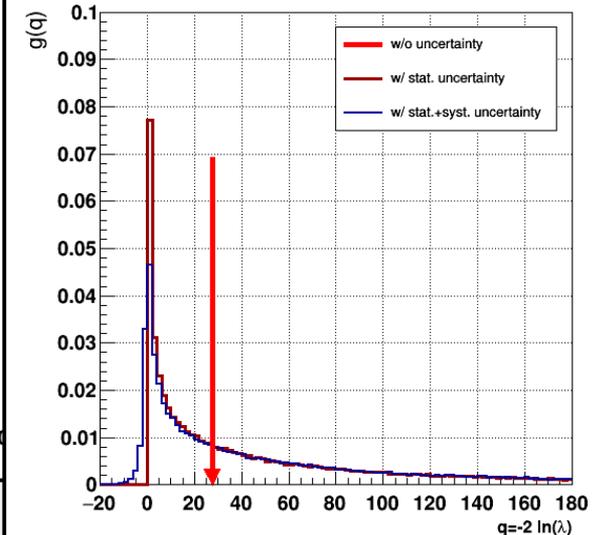
Gewählte Wahrscheinlichkeitsdichte $f(x, \mu = 1, \sigma = 0.05 \mu)$ für **Normierung**.



Gewählte Wahrscheinlichkeitsdichte $\rho(\theta_0|\tilde{\theta}_0)$ für θ_0 .



Die Testverteilung $g(q)$ weitet sich durch systematische Unsicherheiten erwartungsgemäß aus.



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

Beispiel – 3 Higgs Kopplungen

- Ereigniskategorien : 574
 - Störparameter : 4268
- $\mu = \sigma/\sigma_{SM} = 1.09 \pm 0.11$

• ATLAS+CMS LHC Run-1 Kombination:

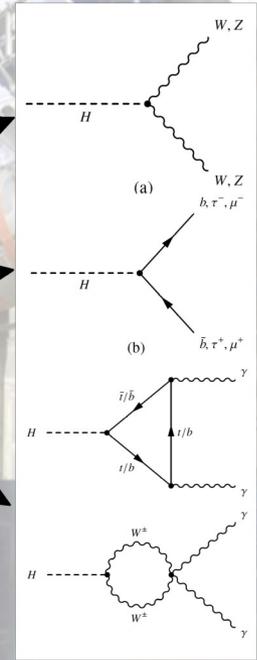
CMS-PAS-HIG-15-002

Berücksichtigte Produktionsmoden:

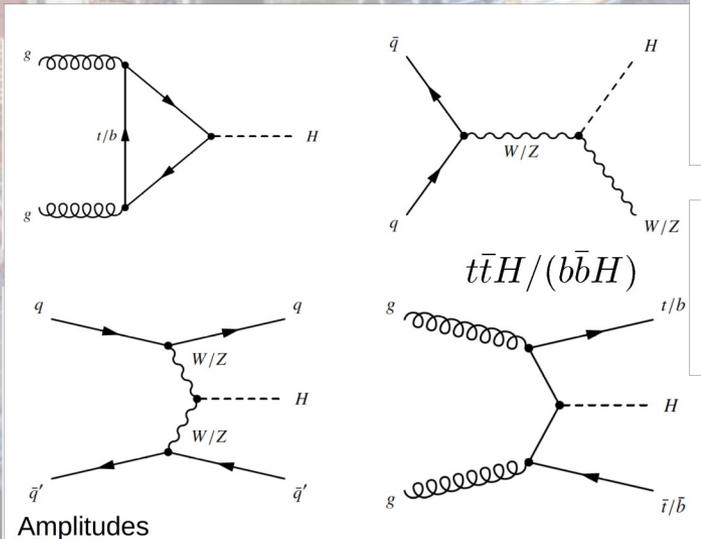
Production process	Cross section [pb]		Order of calculation
	$\sqrt{s} = 7$ TeV	$\sqrt{s} = 8$ TeV	
ggF	15.0 ± 1.6	19.2 ± 2.0	NNLO(QCD)+NLO(EW)
VBF	1.22 ± 0.03	1.58 ± 0.04	NLO(QCD+EW)+~NNLO(QCD)
WH	0.577 ± 0.016	0.703 ± 0.018	NNLO(QCD)+NLO(EW)
ZH	0.334 ± 0.013	0.414 ± 0.016	NNLO(QCD)+NLO(EW)
[ggZH]	0.023 ± 0.007	0.032 ± 0.010	NLO(QCD)
bbH	0.156 ± 0.021	0.203 ± 0.028	5FS NNLO(QCD) + 4FS NLO(QCD)
ttH	0.086 ± 0.009	0.129 ± 0.014	NLO(QCD)
tH	0.012 ± 0.001	0.018 ± 0.001	NLO(QCD)
Total	17.4 ± 1.6	22.3 ± 2.0	

Berücksichtigte Zerfallskanäle:

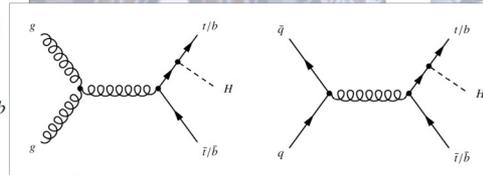
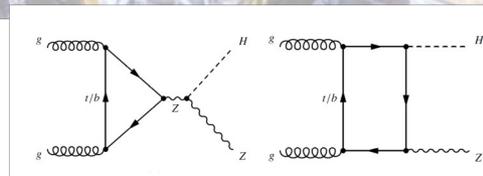
Decay channel	Branching ratio [%]
$H \rightarrow bb$	57.5 ± 1.9
$H \rightarrow WW$	21.6 ± 0.9
$H \rightarrow gg$	8.56 ± 0.86
$H \rightarrow \tau\tau$	6.30 ± 0.36
$H \rightarrow cc$	2.90 ± 0.35
$H \rightarrow ZZ$	2.67 ± 0.11
$H \rightarrow \gamma\gamma$	0.228 ± 0.011
$H \rightarrow Z\gamma$	0.155 ± 0.014
$H \rightarrow \mu\mu$	0.022 ± 0.001



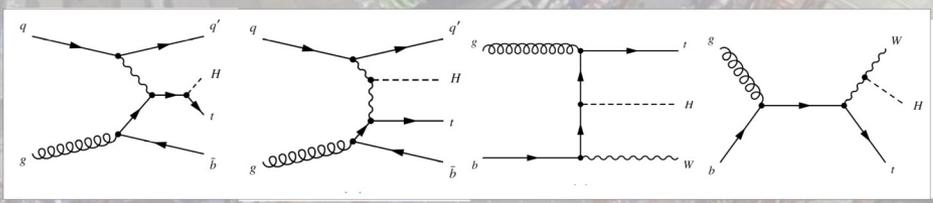
Dominante Produktionsmoden:



$gg \rightarrow ZH$ (10% to $ZHbb$)



$tqH + tHW$



Zusammenfassung

- Mathematische Formulierung der Parameterschätzung.
 - Verzerrung (bias).
- Die Maximum Likelihood Methode.
- Anwendungsbeispiele.
- Methoden zur Varianzbestimmung der ML-Abschätzung.
- Weitere Anwendungen der ML Methode.