Moderne Methoden der Datenanalyse
– Parameterschätzung –

Roger Wolf

14. Mai 2020

Inhalt der nächsten (drei) Vorlesungen

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

Stochastik vs. Statistik

- In der Stochastik wird die wahre Wahrscheinlichkeitsdichte als bekannt vorrausgesetzt.

- In der Realität ist der Wahrscheinlichkeitsraum normalerweise nicht (vollständig) bekannt. Er muss mit Hilfe begrenzter Stichproben (engl. *sample*) charakterisiert werden:

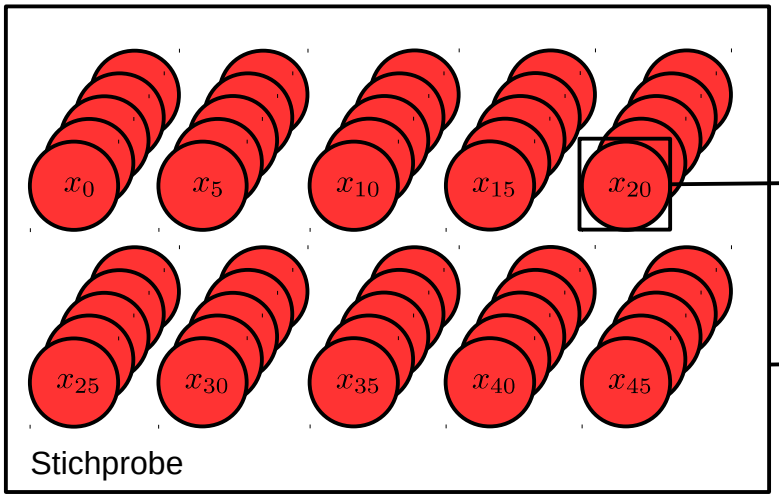
Sei x eine Zufallsvariable, die nach der Wahrscheinlichkeitsdichte $p(x)$ verteilt ist und A eine Menge von n unabhängigen Zufallsexperimenten. Die n Werte $\{x_i\}$ können als n-dimensionaler Vektor $\vec{x} = (x_1, x_2, \ldots, x_n)$ in einem neuen Ereignisraum aufgefasst werden. Da die Einzelereignisse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

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

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

Einzelmessung vs. Messreihe

Messreihe aus 50 Einzelmessungen:



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

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

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

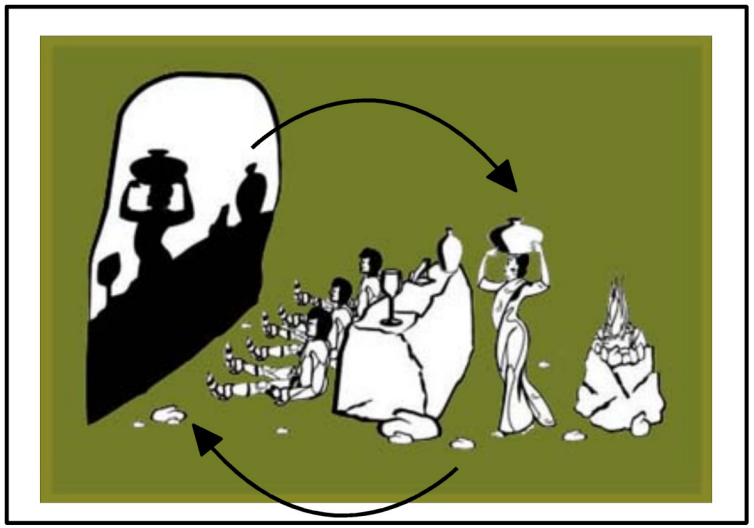
Das stochastische Modell

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

Realität:

Mit welcher Wahscheinlichkeit tritt ein reales Ereignis A_r ein?

Konstruktion des Modells.



Übertragung des mathematischen Resultats in die Realität.

Modell:

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

Die 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 $\vec{t} = (t_1, t_2, \ldots, t_m)$ (für $n \geq m$ Einzelmessungen $\vec{x}$) sein.

- Im Extremfall kann sie den Einzelmessungen selbst entprechen. Üblicherweise ist t aber eine einfache skalare Funktion. Die Reduktion der Dimension erfolgt i.a. aus Gründen der besseren Handhabarkeit.

- Die Teststatistik ist eine Funktion der zufallsverteilten $\vec{x}$ und folgt damit selbst einer Wahrscheinlichkeitsverteilung $g(t(\vec{x}))$.

Die 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 $\lim_{n\to\infty} P(|\hat{\theta} - \theta| > \epsilon) = 0$ 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 Wahrscheinlichkeitverteilung $g(\hat{\theta}, \theta)$ (vgl. Folie 6).

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

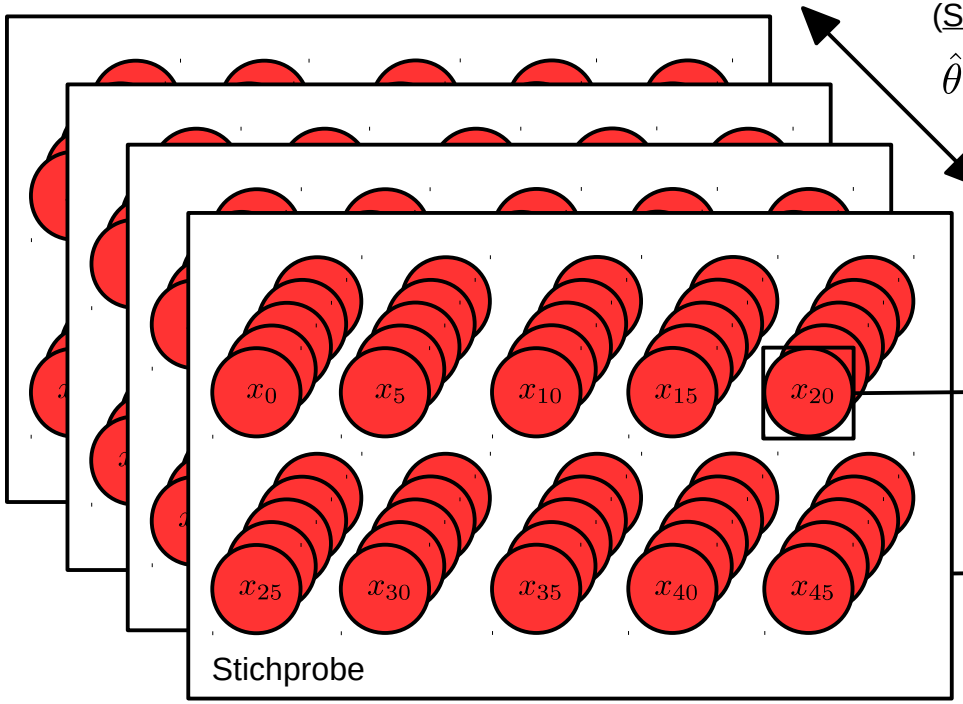
$$E[\hat{\theta}(\vec{x}, \theta)] = \int \hat{\theta} \cdot g(\hat{\theta}, \theta)\, \mathrm{d}\hat{\theta} = \int \cdots \int \hat{\theta}(\vec{x}) \cdot \prod p(\vec{x}, \theta)\, \mathrm{d}x_i \,.$$

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.

Klärung der Begriffe („Was ist was“)

Experiment aus 50 Einzelmessungen:



Verteilung der Stichprobenmessungen (Stichprobenbverteilung):

$\hat{\theta}(\vec{x})$ 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 $P(\{x_i\}) = \prod_{1 \leq i \leq n} p(x_i, \theta)$

. . .

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

Verzerrung (engl. bias)

Die Größe

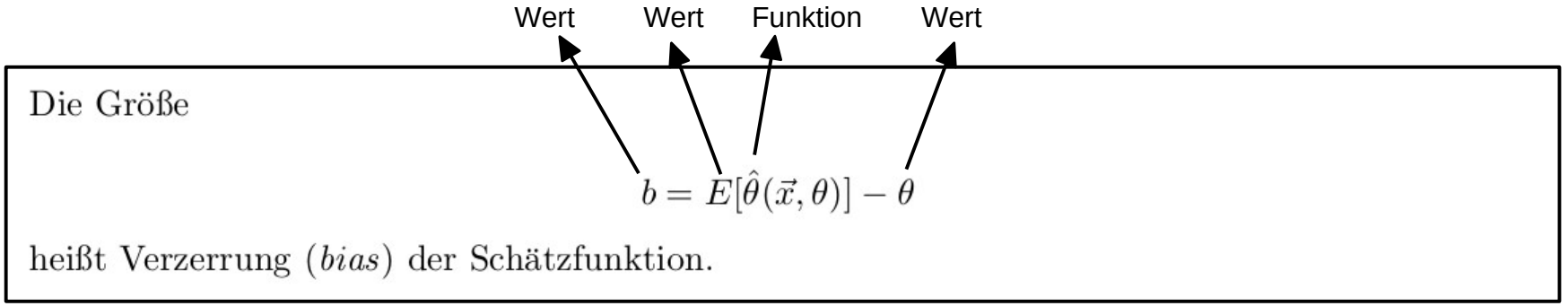
$$b = E[\hat{\theta}(\vec{x}, \theta)] - \theta$$

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 den wahren Werten $\vec{\theta}$.

- Ein Parameter für den der *bias* unabhängig von der Stichprobenlänge der Meßreihe 0 ist heißt erwartungstreu, ein Parameter für den $b = 0$ für $n \to \infty$ heißt asymptotisch erwartungstreu.

Verzerrung (engl. bias)



- **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 erwartungstreu, ein Parameter für den $b = 0$ für $n \to \infty$ heißt asymptotisch erwartungstreu.

Verzerrung (engl. bias)

- **Frage:** Kann ein geschätzter Parameter verzerrt sein, selbst wenn seine Schätzfunktion konsistent ist?

Verzerrung (engl. bias)

- **Frage:** Kann ein geschätzter Parameter verzerrt sein, selbst wenn seine Schätzfunktion konsistent ist?

- **Antwort:** Ja. Frage für zu Hause: Ist der Parameter asymptotisch erwartungstreu, wenn seine Schätzfunktion konsistent ist?

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:** Das ist in der Tat die Kunst der Statistik und von Fall zu Fall verschieden.

  In dem Beispiel, das wir im folgenden diskutieren werden werden wir die Erwartungstreue der Schätzfunktion sogar analystisch beweisen können. In der Praxis ist das i.a. nicht möglich. In solchen Fällen überprüft man diese Eigenschaft mit Hilfe von Modellen, z.B. mit Hilfe einer MC basierten Simulation.

Statistische und systematische Unsicherheiten

Die Größe

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

heißt Genauigkeit (*mean squared error*) der Schätzfunktion. In einer physikalischen Messung beschreibt $\mathrm{var}[\hat{\theta}]$ 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 21). Für erwartungstreue Schätzfunktionen ist der zweite Summand der oberen Gleichung allerginds 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 hier nochmal explizit nach:

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

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

LHS:

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

RHS:

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

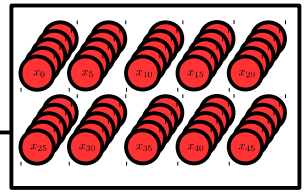
$$E[(\hat{\theta}-E[\theta])^2] + E^2[\hat{\theta}-\theta]] = \underbrace{E[\hat{\theta^2}] - E[\hat{\theta}]^2}_{(1)} + \underbrace{E[\hat{\theta}]^2 - 2\theta E[\hat{\theta}] + \theta^2}_{(2)} = E[\hat{\theta^2}] - 2\theta E[\hat{\theta}] + \theta^2$$

vergleiche mit (*)

Besondere Schätzfunktionen für Stichproben

- Im folgenden werden wir drei besondere Schätzfunktionen für Stichproben diskutieren:
  - Mittelwert ($\overline{x}$) der Stichprobe.
  - Varianz (s^2) der Stichprobe.
  - Korrelationskoeffizient (r) der Stichprobe.
- **NB**: Bei diesen besonderen Schätzfunktionen verzichtet man oft auf den $\hat{.}$ und setzt sie ohne weitere Anmerkung mit den entsprechenden Größen der Einzelmessung gleich.
- **Beachten Sie**: Alle drei Größen sind (asymptotisch) erwartungstreu. Sie können also aus der Stichprobe (asymptotisch) die wahren Werte für Mittelwert, Varianz und Korrelationskoeffizient von $p(x)$ $(p(x,y))$ ermitteln ohne irgendetwas über $p(x)$ $(p(x,y))$ zu wissen.

Mittelwert der Stichprobe


Stichprobe

Die Größe $\overline{x} = \frac{1}{n}\sum_{i\leq n} x_i$ heißt Mittelwert der Stichprode. Ihr Erwartungswert ($E[\overline{x}]$) und ihre Varianz ($\mathrm{var}[\overline{x}]$) sind:

$$E[\overline{x}] = E\left[\frac{1}{n}\sum_{i\leq n} x_i\right] = \frac{1}{n}\sum_{i\leq n}\underbrace{E[x_i]}_{\equiv\mu} = \frac{1}{n}\sum_{i\leq n}\mu = \mu$$

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

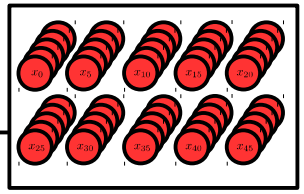
$$= \frac{1}{n^2}\left[\underbrace{(n^2-n)\mu^2}_{n(n-1)} + \underbrace{n(\mu^2+\sigma^2)}_{n}\right] - \mu^2 = \sigma^2/n\,.$$

Dabei sind μ und σ^2 Länge Erwartungswert und Varianz der Einzelmessung $p(x)$, und n die Länge der Stichprobe.

$n(n-1)$ 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:
$E[x^2] = \sigma^2 + \mu^2$

Mittelwert der Stichprobe



Stichprobe

Die Größe $\overline{x} = \frac{1}{n}\sum_{i\leq n} x_i$ heißt Mittelwert der Stichprode. Ihr Erwartungswert ($E[\overline{x}]$) und ihre Varianz ($\mathrm{var}[\overline{x}]$) sind:

$$E[\overline{x}] = E\left[\frac{1}{n}\sum_{i\leq n} x_i\right] = \frac{1}{n}\sum_{i\leq n}\underbrace{E[x_i]}_{\equiv\mu} = \frac{1}{n}\sum_{i\leq n}\mu = \mu$$

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

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

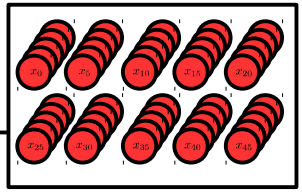
Dabei sind μ und σ^2 Länge Erwartungswert und Varianz der Einzelmessung $p(x)$, und n die Länge der Stichprobe.

$n(n-1)$ 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 und unabhängig von $p(x)$.

Varianz der Stichprobe



Stichprobe

Die Größe

$$s^2 = \frac{1}{n-1}\sum_{i\leq n}(x_i - \overline{x})^2 = \frac{n}{n-1}\left(\overline{x^2} - \overline{x}^2\right)$$

heißt Varianz der Stichprobe. Ihr Erwartungswert ($E[s^2]$) und ihre Varianz sind:

$$E[s^2] = \sigma^2$$

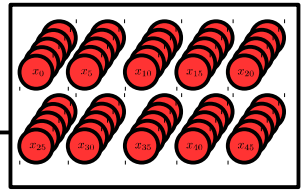
$$\mathrm{var}[s^2] = \frac{1}{n}\left(\mu_4 - \frac{n-3}{n-4}\mu_2^2\right),$$

wobei μ_k das k-te zentrale Moment um μ, und μ und σ^2 der Erwartungswert und die Varianz der Einzelmessung sind.

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

NB: Beachten Sie im Nenner steht in diesem Fall $n-1$.

Varianz der Stichprobe



Stichprobe

Die Größe

$$s^2 = \frac{1}{n-1}\sum_{i\leq n}(x_i - \overline{x})^2 = \frac{n}{n-1}\left(\overline{x^2} - \overline{x}^2\right)$$

heißt Varianz der Stichprobe. Ihr Erwartungswert ($E[s^2]$) und ihre Varianz sind:

$$E[s^2] = \sigma^2$$

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

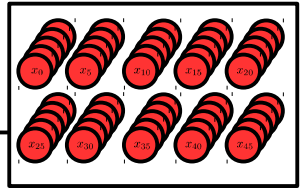
wobei μ_k das k-te zentrale Moment um μ, und μ und σ^2 der Erwartungswert und die Varianz der Einzelmessung sind.

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

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

NB: Beachten Sie im Nenner steht in diesem Fall $n-1$.

Korrelationskoeffizient der Stichprobe


Stichprobe

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 einzelner Zufallsvariablen x und y. Sie hat den Erwartungswert und die Varianz:

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

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

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

Parameterschätzung mit Hilfe der Maximum Likelihood Methode

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.

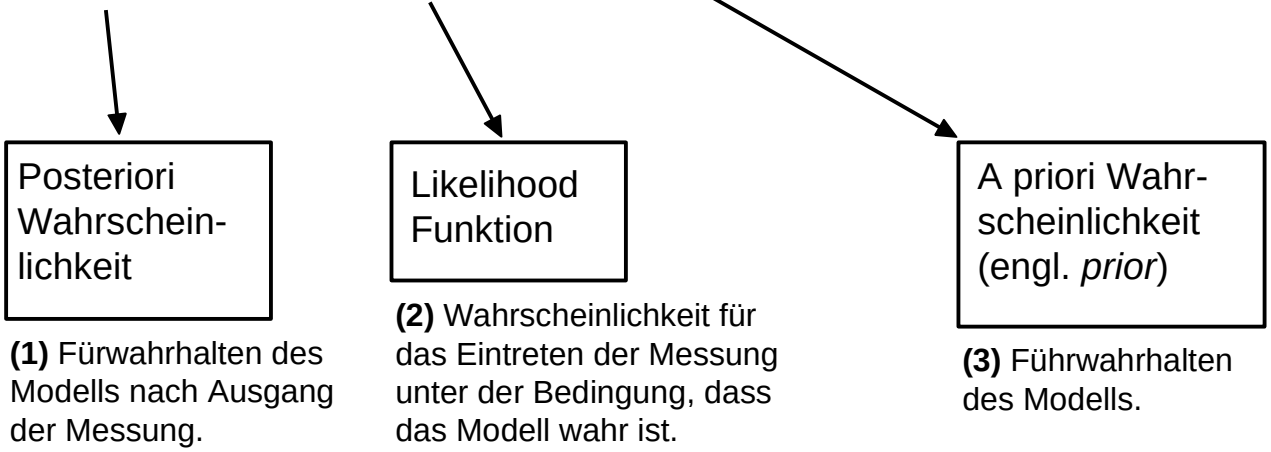
- Eine Hypothese heißt einfach, wenn sie direkt Wahrscheinlichkeitsaussagen zulässt, oder zusammengesetzt wenn sie unbekannte (d.h. noch zu bestimmende) Parameter θ_i enthält.

Hypothese, Bayesianische Statistik, Likelihood

- In der Bayesianischen Statistik wird Wahrscheinlichkeit interpretiert als das *Fürwahrhalten* einer Hypothese.

- Die Likelihood steht in direktem Zusammenhang zum Satz von Bayes:

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



(1) Fürwahrhalten des Modells nach Ausgang der Messung.

(2) Wahrscheinlichkeit für das Eintreten der Messung unter der Bedingung, dass das Modell wahr ist.

(3) Führwahrhalten des Modells.

NB: Sie können sehen, dass die Likelihood Idee zutiefst Bayesianische Züge trägt. Für **(1)** und **(3)** gibt es keine frequentistischen Entsprechungen. Sie können ebenfalls ein gewisses iteratives Vorgehen aus diesen Definitionen ableiten. Es handelt sich dabei um ein fundamentales Vorgehen in der heutigen (modellbasierten) modernen Wissenschaft (oft mit dem Begriff der Falsifizierbarkeit verknüpft).

Die Likelihood Funktion

Sei x eine Zufallsvariable, die nach einer Wahrscheinlichkeitsdichte $p(x, \theta)$ verteilt ist. Dabei sei θ (mindestens) ein unbestimmter Parameter. Seien weiterhin $\{x_i\}, i = 1 \ldots, n$ 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

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

Die Wahrscheinlichkeit für den Ausgang des Experimentes ist gegeben durch

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

Die Funktion

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

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

Das Maximum Likelihood (ML) 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, $\mathcal{P}(\{x_i\}, \theta)$ für das Vorliegen der Einzelmessungen $\{x_i\}$ für den korrekten Wert von θ höher ist, als für jeden anderen Wert.

- Da die $\mathrm{d}x_i$ 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 $\hat{\theta}_{\mathrm{ML}}$ des Parameters θ. Wenn $\mathcal{L}(\theta)$ (mindestens zwei mal) stetig differenzierbar ist ist $\hat{\theta}_{\mathrm{ML}}$ gegeben durch:

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

NB: Beachten Sie, dem ML Prinzip liegt immer eine Wahrheitshypothese zugrunde (vgl. mit Folie 13).

Das Maximum Likelihood (ML) Prinzip

- 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 $\ln\left(\mathcal{L}(\theta)\right)$ 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) $-\ln(\mathcal{L}(\theta))$.

Das Maximum Likelihood (ML) Prinzip

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

Das Maximum Likelihood (ML) Prinzip

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

$$\frac{\partial}{\partial\theta}\Big(\ln\big(\mathcal{L}(\theta)\big)\Big) = \frac{\frac{\partial}{\partial\theta}\mathcal{L}(\theta)}{\mathcal{L}(\theta)} = 0$$

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 $\ln\left(\mathcal{L}(\theta)\right)$ 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:

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

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

NB: D.h. der ML Schätzwert von a ergibt sich aus dem ML Schätzwert von θ.

Anwendungsbeispiel: 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.

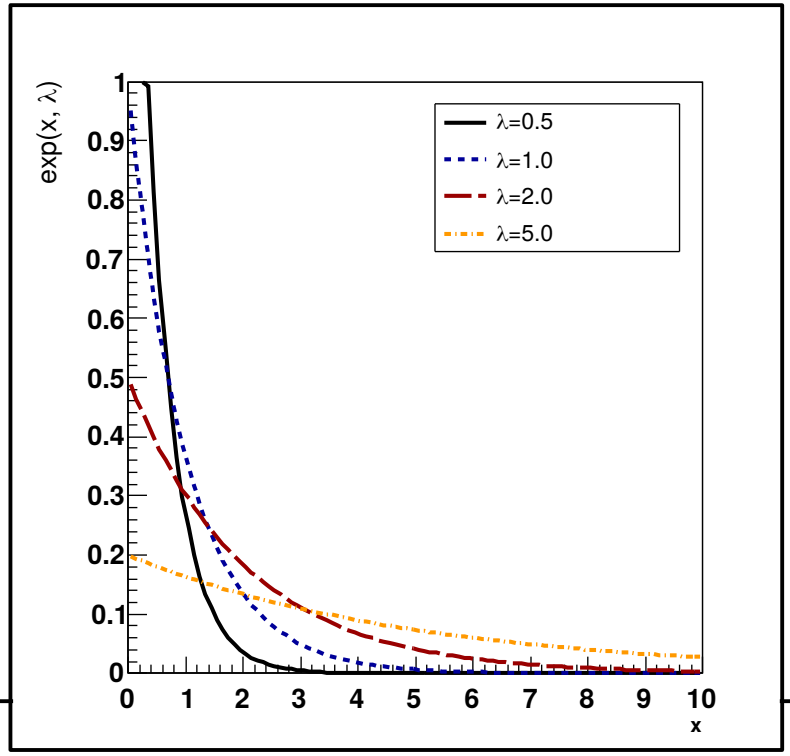
- Unsere Wahrheitshypothese ist, dass die Wahrscheinlichkeitsdichte dieses Zerfalls die Exponentialverteilung $\exp(x,\theta)$ ist (d.h. Differenzen von Zufallszahlen sind gleichverteilt).

- Eine kurze Erinnerung an die Eigenschaften der Exponentialverteilung:

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

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

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



ML Schätzfunktion $\hat{\theta}_{\mathrm{ML}}$

- 50 Ereignisse verteilt nach $\exp(x,\theta)$ mit $\theta = 1$ (Z.B. radioaktiver Zerfall mit Halb-wertszeit $\theta = 1$).

- **Likelihood Funktion**:

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

$$\ln(\mathcal{L}(\{x_i\},\theta)) = \sum_{i\leq 50} (-\ln(\theta) - \frac{x_i}{\theta})$$

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

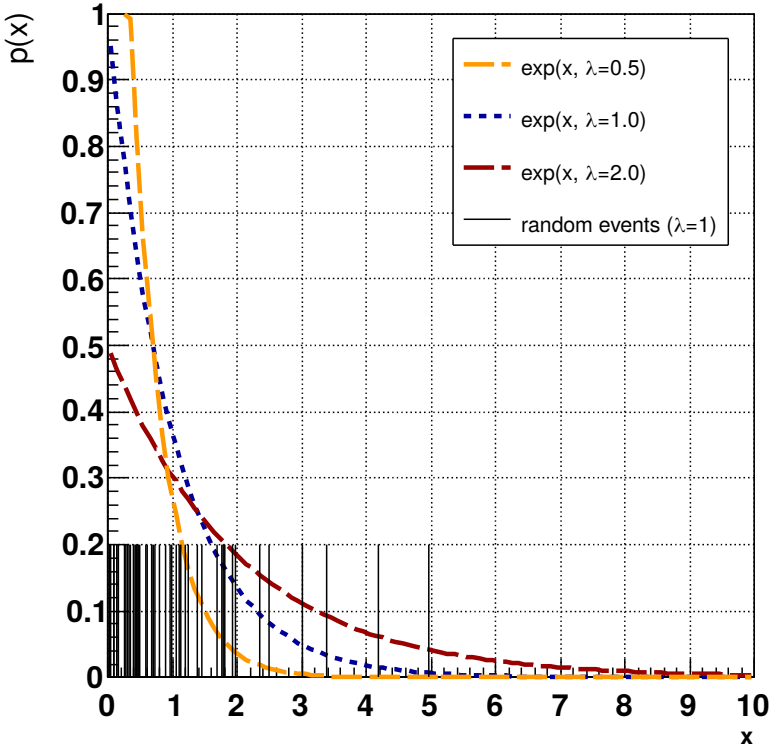
$$\frac{\partial}{\partial\theta} \ln(\mathcal{L}(\{x_i\},\theta)) = \sum_{i\leq 50} \left(-\frac{1}{\theta} + \frac{x_i}{\theta^2}\right)$$

$$= \sum_{i\leq 50} \frac{-\theta + x_i}{\theta^2} = 0\,;$$

$$\sum_{i\leq 50} x_i = n\,\hat{\theta}_{\mathrm{ML}}\,;$$

$$\hat{\theta}_{\mathrm{ML}} = \frac{1}{n} \sum_{i\leq 50} x_i$$

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



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 $\hat{\theta}_{\mathrm{ML}}$

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

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

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

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

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

$= 1$

$$\int_0^\infty \frac{x}{\theta} e^{-\frac{x}{\theta}}\, \mathrm{d}x = \theta \int_0^\infty x' e^{-x'}\, \mathrm{d}x' = \theta$$

mit: $x' = \frac{x}{\theta}\,, \quad \mathrm{d}x = \theta \mathrm{d}x'$

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

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

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

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

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

- Dieses Ergebnis war zu erwarten – warum?

NB: Versuchen Sie diese Frage zunächst aufrichtig für sich zu beantworten. Sie enthält eine Menge Transfer von dem, was wir in dieser Vorlesung diskutiert haben.

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

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

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

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

- Dieses Ergebnis war zu erwarten – warum?
  - $\hat{\theta}_{\mathrm{ML}}$ erweist sich als das arithmetische Mittel der Einzelmessungen $\{x_i\}$, die nach $p(x, \theta) = \exp(x, \theta)$ verteilt sind (siehe Folie 28).
  - Wir haben gelernt, dass das arithmetische Mittel der Einzelmessungen allg. eine erwartungstreue Schätzfunktion für den Erwartungswert von $p(x, \theta)$ ist (siehe Folie 16).
  - θ ist der Erwartungswert von $\exp(x, \theta)$ (siehe Folie 27).

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

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

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

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

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

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

- 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 (siehe Folie 17).
  - θ^2 ist die Varianz von $\exp(x, \theta)$ (siehe Folie 27).

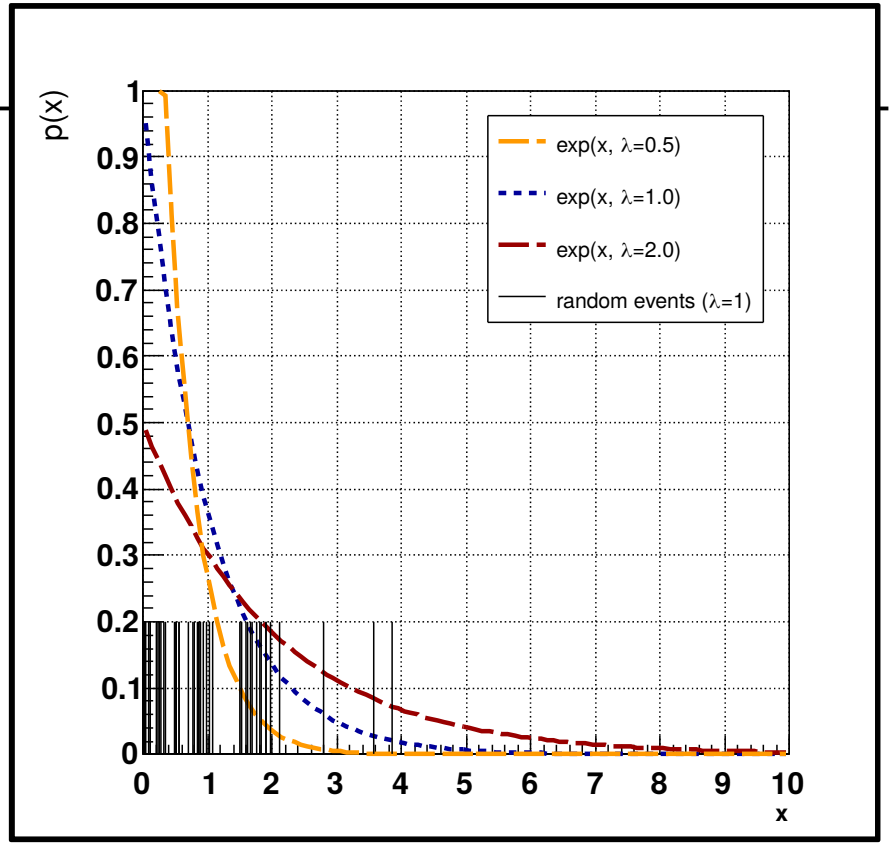
Ein konkreter Ausgang

- 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 sei:

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

Schätzwert aus Maximum der Likelihood.

Standardabweichung des Schätzwertes.



Ein lauffähiges RooT macro finden Sie hier.

- **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 ($\rightarrow$ unendlich oft) wiederholen würden würde dieser Wert innerhalb von $\pm 0.143\ \mathrm{s}$ streuen.

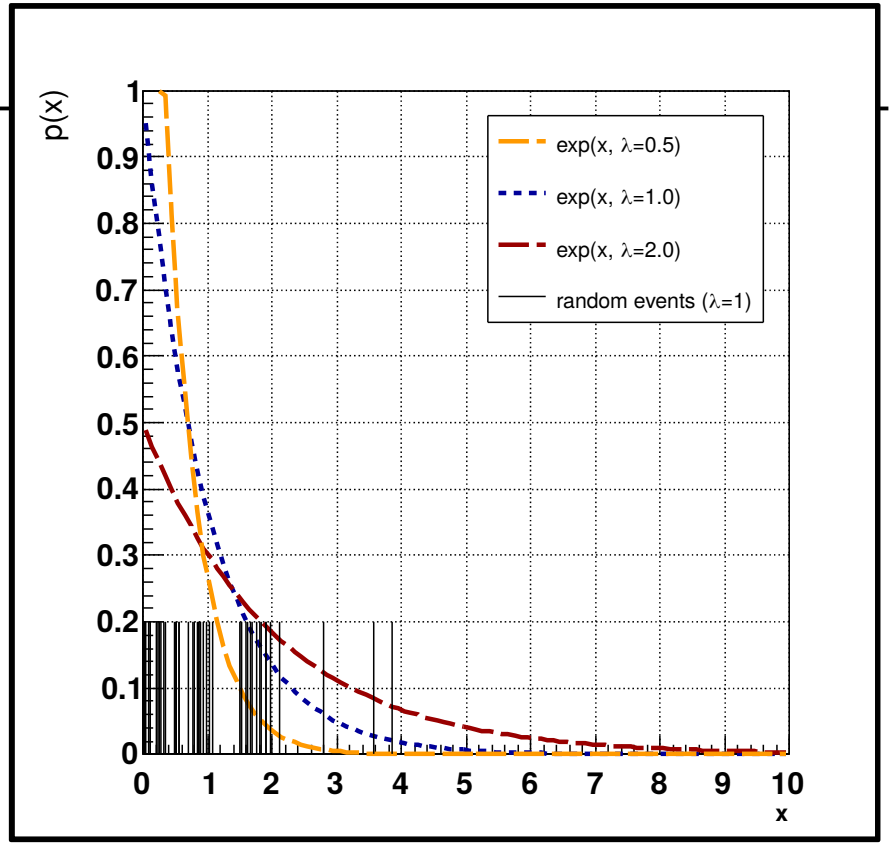
Ein konkreter Ausgang

- 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 sei:

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

Schätzwert aus Maximum der Likelihood.

Standardabweichung des Schätzwertes. (*)



Ein lauffähiges RooT macro finden Sie hier.

- **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 ($\rightarrow$ unendlich oft) wiederholen würden würde dieser Wert innerhalb von $\pm 0.143\ \mathrm{s}$ streuen.

(*)
Die Varianz enthält den wahren Wert von θ (siehe Folie 31). Was machen, wenn der wahre Wert eben nicht bekannt ist?!? In diesem Fall ersetzen Sie den wahren Wert durch den Schätzwert.

Anwendung der MC Methode

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

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

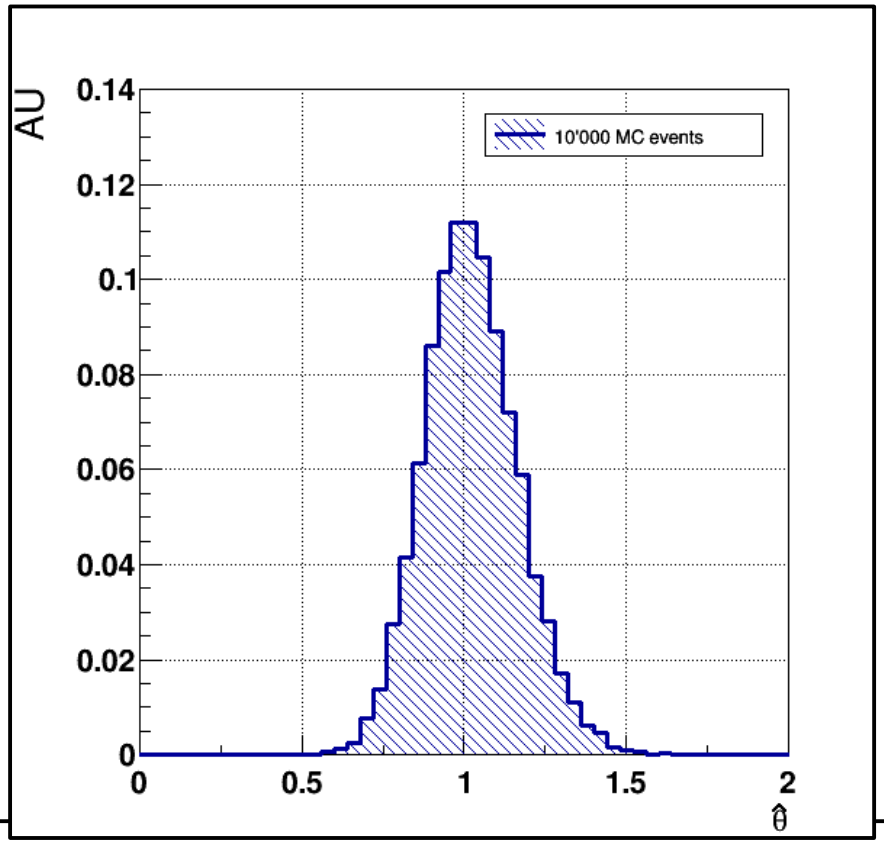
- Für die wahre Lebensdauer setzen Sie den Schätzwert des Experiments $\hat{\theta}_{\mathrm{ML}} = 1.022\,\mathrm{s}$.

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

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

  d.h. deckungsgleich mit dem analytischen Ergebnis.

Ein lauffähiges RooT macro finden Sie hier.



Die RCF-Ungleichung

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

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

Verzerrung.

Likelihood Funktion.

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

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:

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

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

mit:

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

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

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:

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

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

vgl. mit Folie 31.

mit:

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

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

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

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:

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

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

vgl. mit Folie 31.

mit:

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

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

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

- **NB (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 immer 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:

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

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

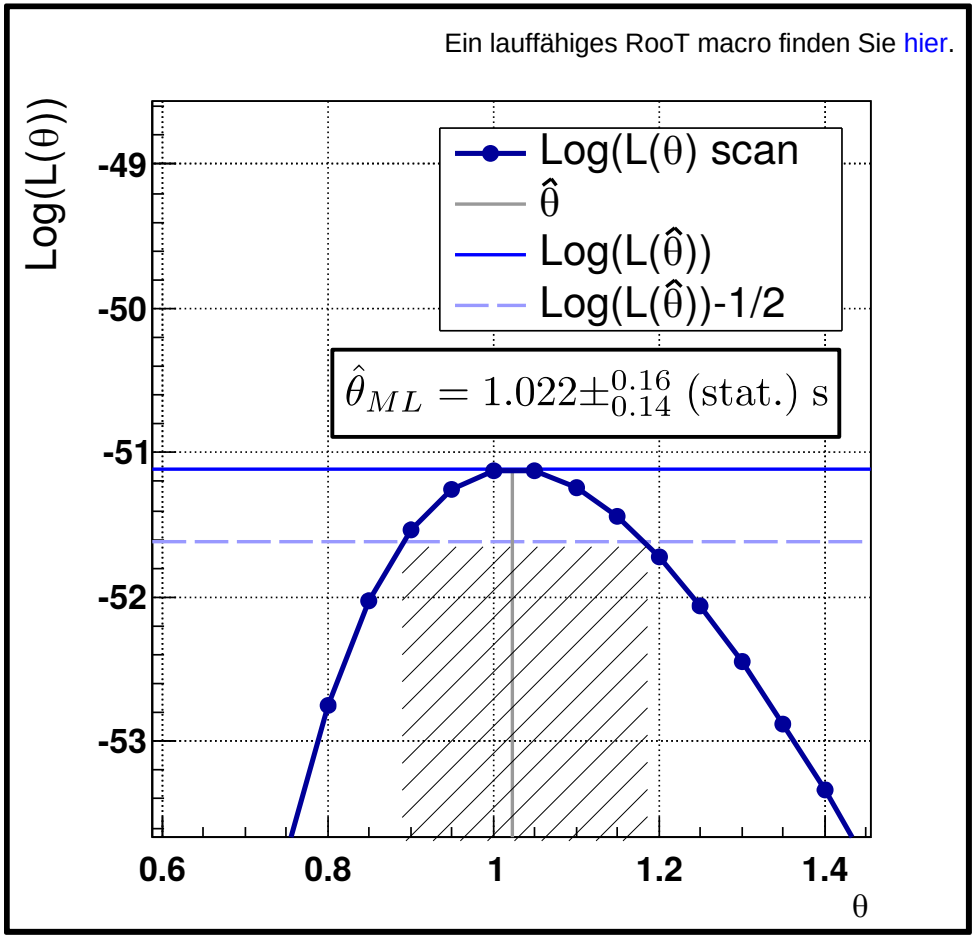
$$\ln\mathcal{L}(\theta) = \underbrace{\ln\mathcal{L}(\hat{\theta}_{\mathrm{ML}})}_{\equiv\,\mathcal{L}_{\max}} + \underbrace{\left[\frac{\partial}{\partial\theta}\ln\mathcal{L}\right]_{\theta=\hat{\theta}_{\mathrm{ML}}}}_{\equiv\,0}\left(\theta-\hat{\theta}_{ML}\right) + \frac{1}{2}\underbrace{\left[\frac{\partial^2}{\partial\theta^2}\ln\mathcal{L}\right]_{\the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

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

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

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

Anwendung der graphischen Abschätzung



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

Zusammenfassung

- Grundlagen der Parameterschätzung.
  - Definitionen: Stichprobe, Teststatistik, Schätzfunktion, Verzerrung.
  - Erwartungswert, Varianz und Korrelationskoeffizient der Stichprobe (mit Unsicherheiten).
- Parameterschätzung mit Hilfe der Maximum Likelihood (ML) Methode:
  - Definitionen: Hypothese und Likelihood.
  - ML Prinzip.
  - RCF-Ungleichung.
  - Graphische Abschätzung der Varianz.

} Am Beispiel des radioaktiven Zerfalls mit $p(x) = \exp(x, \theta)$.