Universität Zürich[UZH]

Datenanalyse
(PHY231)
Herbstsemester 2017

Olaf Steinkamp

36-J-05 olafs@physik.uzh.ch 044 63 55763

Vorlesungsprogramm

- **Einführung, Messunsicherheiten, Darstellung von Messdaten**
- **Grundbegriffe der Wahrscheinlichkeitsrechnung und Statistik**
  - Mittelwert, Standardabweichung, Kovarianz und Korrelation
- **Fehlerfortpflanzungsgesetz**
- **Wahrscheinlichkeitsverteilungen**
  - diskrete Verteilungen, kontinuierliche Verteilungen
  - zentraler Grenzwertsatz
- **Monte-Carlo Methode**
- **Wahrscheinlichkeitsverteilungen II**
  - Faltung zweier Verteilungen
  - Verteilungen zweier Variablen
- **Stichproben und Schätzfunktionen**
  - Maximum-Likelihood Methode
  - Methode der kleinsten Quadrate
- **Interpretation von Messergebnissen**
  - Konfidenzintervalle, Testen von Hypothesen

Beispielprogramme im Verzeichnis
`/disk/puma/da/vorl/ml`

Stichproben und Schätzungen

<u>Zufallsvariable *x* folgt einer Wahrscheinlichkeitsverteilung *f(x|a)*</u>

- ***f(x|a)* hängt von einem oder mehreren Parametern *a* ab**
  - kenne die funktionale Form von *f(x|a)*, aber <u>nicht</u> die Werte der Parameter *a*
- **habe eine endliche Anzahl Werte der Zufallsvariable *x* ("Stichprobe")**
  - will aus dieser Stichprobe die Werte der Parameter *a* <u>schätzen</u>

***a*: wahrer Wert des Parameters**

***â*: aus der Stichprobe berechneter Schätzwert**

<u>Beispiel: Zerfall eines radioaktiven Isotops</u>

- **Zerfallszeiten folgen einer Exponentialverteilung,** $f(t\,|\,\lambda) = \lambda \cdot e^{-\lambda t}$
  - der Wert der Zerfallskonstante λ ist nicht bekannt und soll gemessen werden
- **Experiment ergibt eine endliche Anzahl gemessener Zerfallszeiten**
  - bestimme aus diesen gemessenen Zerfallszeiten einen <u>**Schätzwert** $\hat{\lambda}$</u> für λ

Schätzwerte sind Zufallsvariablen

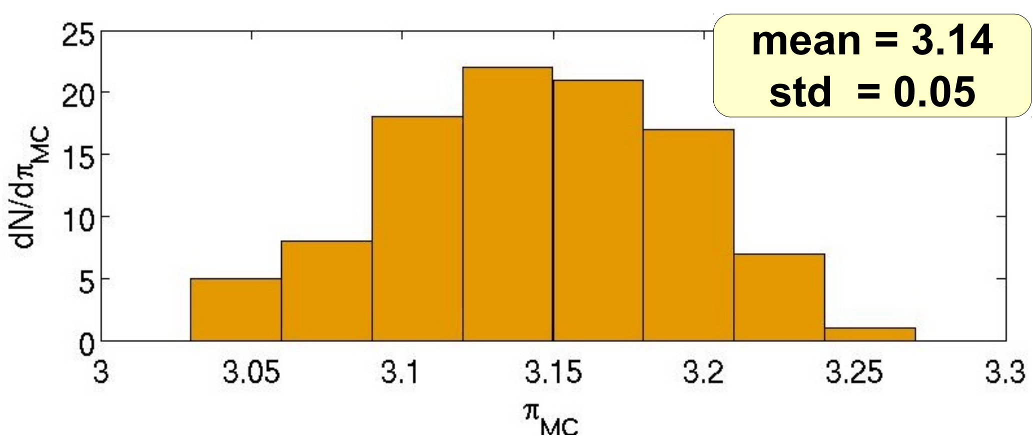
Stichprobe = zufällige Werte aus der Wahrscheinlichkeitsverteilung

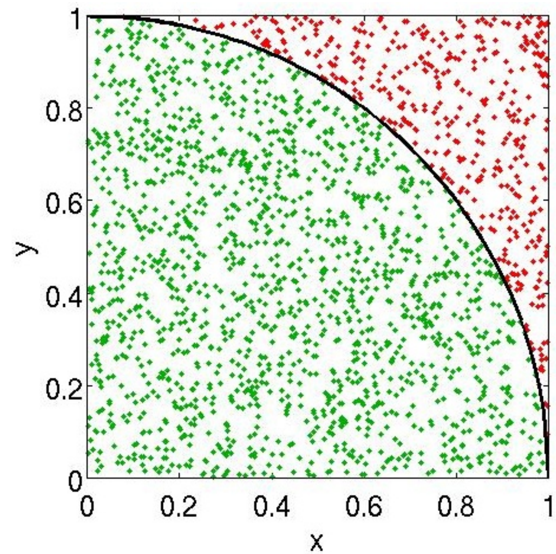
- **neue Stichprobe → andere zufällige Werte → anderer Schätzwert *â***
- **viele Stichproben → Zufallsverteilung für *â***

Schätzwert hat einen Erwartungswert ‹*â*› und eine Varianz *V(â)*

Beispiel Monte-Carlo Bestimmung von π (s. Vorlesung vor zwei Wochen)

- **eine Stichprobe aus 1000 Zufallszahlenpaaren → ein Schätzwert für π**
- **100 Stichproben → 100 Schätzwerte für π**





Schätzfunktionen

Schätzfunktion = Algorithmus zur Berechnung von Schätzwerten

- **Ansprüche an "gute" Schätzfunktionen:**
  - **Konsistenz**: Schätzwert soll gegen den wahren Wert des Parameters streben, wenn der Umfang der Stichprobe gegen unendlich strebt
  - **Erwartungstreue**: Schätzwert soll "im Mittel" über viele Stichproben den wahren Wert ergeben, egal wie groß der Umfang der Stichproben ist
  - **Effizienz**: Streuung der Schätzwerte soll für gegebene Größe der Stichproben möglichst klein sein
  - **Robustheit**: Schätzwert soll möglichst unempfindlich gegen "Ausreisser" in der Stichprobe sein
- **verschiedene Ansätze zur Herleitung von Schätzfunktionen**
  - bekannteste: "Maximum Likelihood", "Methode kleinster Quadrate" (χ^2-Methode)
- **können verschiedene Schätzfunktionen ergeben**
  - darum: zusammen mit dem Ergebnis immer auch die gewählte Methode angeben

Beispiel: Erwartungswert einer Verteilung

Schätzfunktion für den Erwartungswert μ einer Verteilung

- **naheliegender Ansatz: Mittelwert der Stichprobe**

$$\hat{\mu} \equiv \overline{x} = \frac{1}{N} \cdot \sum_{i=1}^{N} x_i$$

- **$\hat{\mu}$ ist "konsistent", d.h. $\hat{\mu} \to \mu$ für $N \to \infty$**

Gesetz großer Zahlen: $\hat{\mu} = \overline{x} \to \langle x \rangle \equiv \mu$ **für** $N \to \infty$

- **$\hat{\mu}$ ist "erwartungstreu", d.h. $\langle \hat{\mu} \rangle = \mu$ egal wie groß N ist**

$$\langle \hat{\mu} \rangle = \langle \overline{x} \rangle = \langle x \rangle = \mu$$

Beispiel: Varianz einer Verteilung

Schätzfunktion für die Varianz $V(x)$ einer Verteilung

- **naheliegender Ansatz: Varianz der Stichprobe**
- **ist konsistent (**Gesetz großer Zahlen**), aber nicht "erwartungstreu":**

$$
\begin{aligned}
\langle \hat{V} \rangle &= \left\langle \frac{1}{N} \sum (x_i - \bar{x})^2 \right\rangle \\
&= \underbrace{\left\langle \frac{1}{N} \sum x_i^2 \right\rangle}_{=\langle x^2 \rangle} - 2 \cdot \underbrace{\left\langle \frac{1}{N} \sum x_i \bar{x} \right\rangle}_{=\langle \bar{x}^2 \rangle} + \underbrace{\left\langle \frac{1}{N} \sum \bar{x}^2 \right\rangle}_{=\langle \bar{x}^2 \rangle} \\
&= \langle x^2 \rangle - \langle \bar{x}^2 \rangle \qquad \Big| \ \textbf{addiere} \ \langle \bar{x} \rangle^2 - \langle x \rangle^2 = 0 \\
&= \underbrace{\left(\langle x^2 \rangle - \langle x \rangle^2 \right)}_{= V(x)} - \underbrace{\left(\langle \bar{x}^2 \rangle - \langle \bar{x} \rangle^2 \right)}_{= V(\bar{x}) = V(x)/N} = \boxed{\frac{N-1}{N} \cdot V(x)}
\end{aligned}
$$

- **Erwartungstreue Schätzfunktion für die Varianz der wahren Verteilung:**

$$
\hat{V} \equiv \frac{N}{N-1} \cdot \left\{ \frac{1}{N} \cdot \sum_{i=1}^{N} (x_i - \bar{x})^2 \right\} = \frac{1}{N-1} \cdot \sum_{i=1}^{N} (x_i - \bar{x})^2
$$

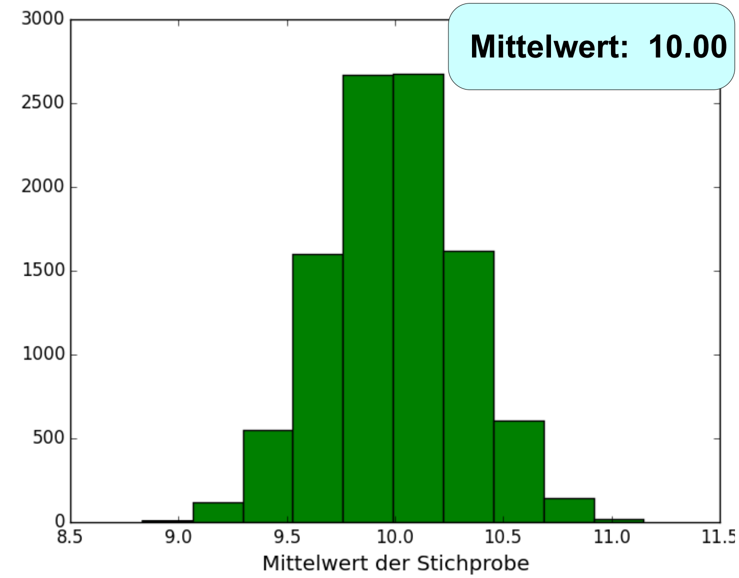
Beispiel: Stichproben einer Gaußverteilung

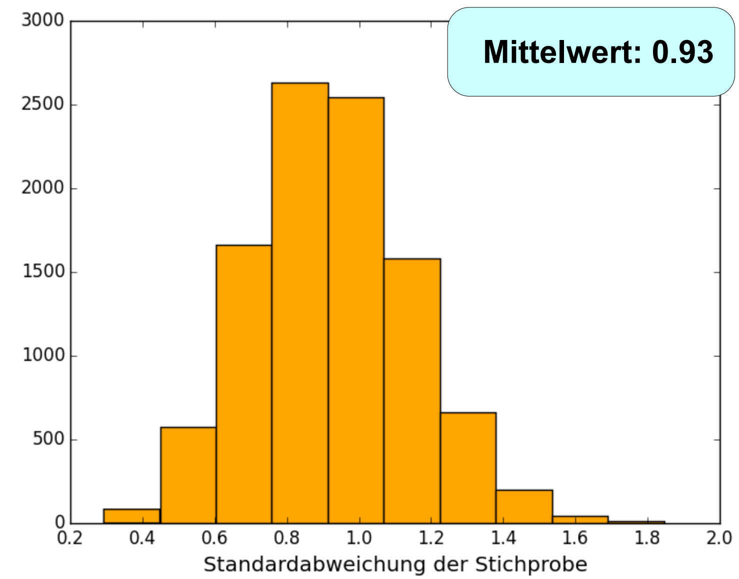
Erzeuge Stichproben von je 10 Zufallszahlen

- gaußverteilt mit $\mu = 10$, $\sigma = 1$
- berechne Mittelwert, Standardabweichung für jede der Stichproben
- Mittelwert der Mittelwerte → 10
- Mittelwert der Standardabweichungen → 0.93

erwartungstreu.py

```
#!/usr/bin/env python
from pylab import *
import scipy.stats as stats

x = stats.norm(10.0,1.0).rvs([10,10000])

meansample = mean(x,0)
sigmasample = std(x,0)

hist(meansample,facecolor='green')
xlabel("Mittelwert der Stichprobe",fontsize='large')
figure()
hist(sigmasample,facecolor='orange')
xlabel("Standardabweichung der Stichprobe",fontsize='large')

print("Mean of means:  %5.2f"%mean(meansample))
print("Mean of sigmas: %5.2f"%mean(sigmasample))
```





Zwei Standardabweichungen

(1) Standardabweichung der x_i selbst

$$\sigma \equiv \sqrt{\boxed{\frac{1}{N}} \cdot \sum_{i=1}^{N} (x_i - \bar{x})^2}$$

- z.B. messe den Wert eines Widerstandes und bestimme aus der Streuung der Messergebnisse die Genauigkeit meiner Messung

(2) Schätzfunktion für die Standardabweichung der zugrundeliegenden "theoretischen" Verteilung

$$\sigma \equiv \sqrt{\boxed{\frac{1}{N-1}} \cdot \sum_{i=1}^{N} (x_i - \bar{x})^2}$$

- z.B. überwache eine Produktionslinie für elektrische Widerstände: messe für eine Stichprobe der produzierten Bauteile deren Widerstand und will aus der Streuung der Messergebnisse die Präzision der Produktion abschätzen

je kleiner der Umfang der Stichprobe, desto größer der Unterschied zwischen den zwei Definitionen

Vorlesungsprogramm

- **Einführung, Messunsicherheiten, Darstellung von Messdaten**
- **Grundbegriffe der Wahrscheinlichkeitsrechnung und Statistik**
  - Mittelwert, Standardabweichung, Kovarianz und Korrelation
- **Fehlerfortpflanzungsgesetz**
- **Wahrscheinlichkeitsverteilungen**
  - diskrete Verteilungen, kontinuierliche Verteilungen
  - zentraler Grenzwertsatz
- **Monte-Carlo Methode**
- **Wahrscheinlichkeitsverteilungen II**
  - Faltung zweier Verteilungen
  - Verteilungen zweier Variablen
- **Stichproben und Schätzfunktionen**
  - Maximum-Likelihood Methode
  - Methode der kleinsten Quadrate
- **Interpretation von Messergebnissen**
  - Konfidenzintervalle, Testen von Hypothesen

Beispielprogramme im Verzeichnis
`/disk/puma/da/vorl/ml`

Maximum-Likelihood Methode

Sehr häufig benutzte Methode zur Herleitung von Schätzfunktionen

- habe Wahrscheinlichkeitsverteilung $f(x|a)$
- habe Stichprobe aus N Werten x_i ($i = 1,\ldots,N$)

suche Schätzfunktion $\hat{a}$ für die Parameter a

- Ansatz: bestimme die Likelihood Funktion ("likelihood" = Wahrscheinlichkeit)

$$L(a) \equiv f(x_1|a) \cdot f(x_2|a) \cdot \ldots \cdot f(x_N|a) = \prod_{i=1}^{N} f(x_i|a)$$

- $L(a)$ gibt die Wahrscheinlichkeit an, die gegebene Stichprobe zu erhalten
- $L(a)$ ist eine Funktion der Parameter a
- wähle als Schätzwerte $\hat{a}$ diejenigen Werte der Parameter, die $L(a)$ maximieren
  - d.h. wähle als Schätzwerte diejenigen Werte der Parameter, die die größte Wahrscheinlichkeit ergeben, die gegebene Stichprobe zu erhalten
- rechentechnisch vorteilhaft: suche Maximum der "Log-likelihood" Funktion

$$\boxed{\ln L(a) = \ln f(x_1|a) + \ln f(x_2|a) + \ldots + \ln f(x_N|a) = \sum_{i=1}^{N} \ln f(x_i|a)}$$

Einfaches Beispiel

Aus dem Buch "Statistics" von Roger Barlow

- **Zufallsvariable *x* folgt**

$$f(x|a) = 1 + a \cdot (x - 0.5)$$

- **Stichprobe aus 5 Werten:**

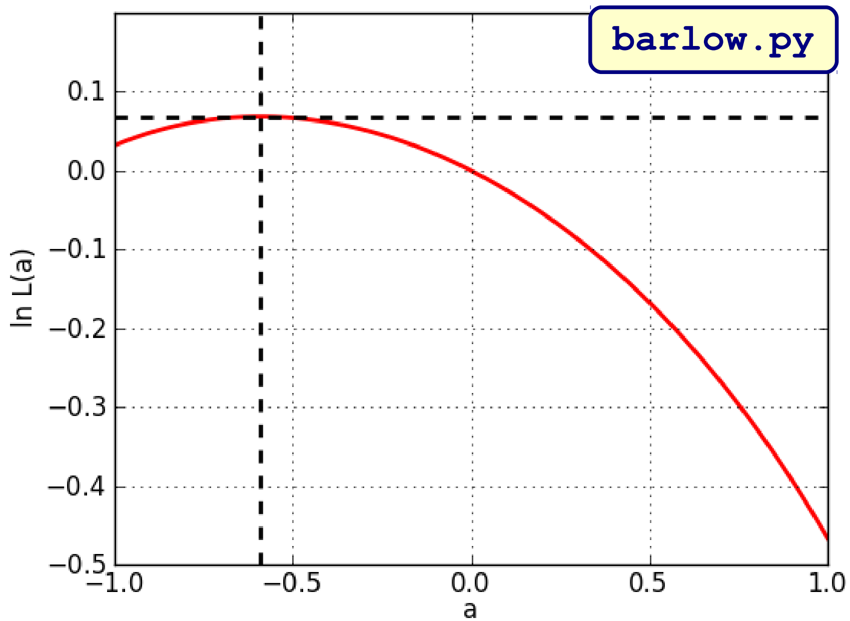
$$x_i = 0.89,\ 0.03,\ 0.50,\ 0.36,\ 0.49$$

- **Log-likelihood Funktion:**

$$\ln L(a) = \sum \ln\left(1 + a \cdot (x_i - 0.5)\right)$$

- **Maximum von ln *L*(*a*) bei *a* = – 0.6**

  ⇛ **Schätzwert $\hat{a}$ = – 0.6**



Finden des Maximums:

- **analytisch durch Lösen von** $\dfrac{d(\ln L)}{da} = 0$
- **numerisch**
- **durch Abtasten des Parameterraums**

numerische Algorithmen suchen meist Minima einer Funktion ⇛ suche Minimum von – ln L

Beispiel mittlere Lebensdauer

Radioaktive Quelle: exponentialverteilte Zerfallszeiten

- Wahrscheinlichkeitsdichte $f(t\,|\,\tau) = 1/\tau \cdot e^{-t/\tau}$
- Stichprobe aus N gemessenen Zerfallszeiten t_i

suche Schätzfunktion für mittlere Lebensdauer τ

- Log-likelihood Funktion

$$\ln L(\tau) = \sum_{i=1}^{N} \ln f(t_i\,|\,\tau) = \sum_{i=1}^{N} \ln\left(\frac{1}{\tau}\cdot e^{-t_i/\tau}\right) = \sum_{i=1}^{N}\left(-\ln\tau - \frac{t_i}{\tau}\right)$$

- Maximum von ln $L(\tau)$ durch Lösen von $d(\ln L)/d\tau = 0$

$$\frac{d\,(\ln L)}{d\,\tau} = \sum_{i=1}^{N}\left\{-\frac{1}{\tau}+\frac{t_i}{\tau^2}\right\} = -\frac{N}{\tau}+\frac{1}{\tau^2}\cdot\sum_{i=1}^{N} t_i = 0 \quad\Rightarrow\quad \boxed{\hat{\tau} = \frac{1}{N}\cdot\sum_{i=1}^{N} t_i}$$

Maximum Likelihood Schätzfunktion für mittlere Lebensdauer = Mittelwert der gemessenen Zerfallszeiten

Beispiel mittlere Lebensdauer (II)

Aber: messe im Experiment keine beliebig langen Zerfallszeiten

- **muss in der Definition der Wahrscheinlichkeitsdichte berücksichtigt werden**
  - Erinnerung: Wahrscheinlichkeitsdichten müssen auf 1 normiert sein !
- **Zerfallszeiten zwischen $t = 0$ und $t = t_{max}$**

$$\int_0^{t_{max}} \frac{1}{\tau} \cdot e^{-t/\tau}\, dt = 1 - e^{-t_{max}/\tau} \quad \Rightarrow \quad g(t\,|\,\tau) = \frac{1}{\left(1 - e^{-t_{max}/\tau}\right)} \cdot \frac{1}{\tau} \cdot e^{-t/\tau}$$

- **Log-likelihood Funktion**

$$\ln L(\tau) = \sum_{i=1}^{N} \ln g(t_i\,|\,\tau) = -N \cdot \ln\left(1 - e^{-t_{max}/\tau}\right) - N \cdot \ln \tau - \frac{1}{\tau} \cdot \sum_{i=1}^{N} t_i$$

- **suche Maximum:** $d\left(\ln L(\tau)\right)/d\tau = 0$

$$\boxed{\hat{\tau} = t_{max} \cdot \frac{e^{-t_{max}/\hat{\tau}}}{1 - e^{-t_{max}/\hat{\tau}}} + \frac{1}{N} \cdot \sum_{i=1}^{N} t_i}$$

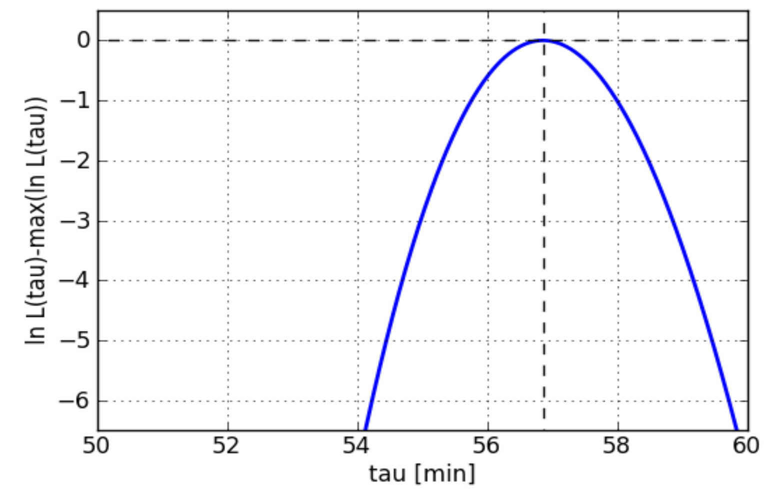
Gleichung für $\hat{\tau}$, kann aber nicht analytisch gelöst werden

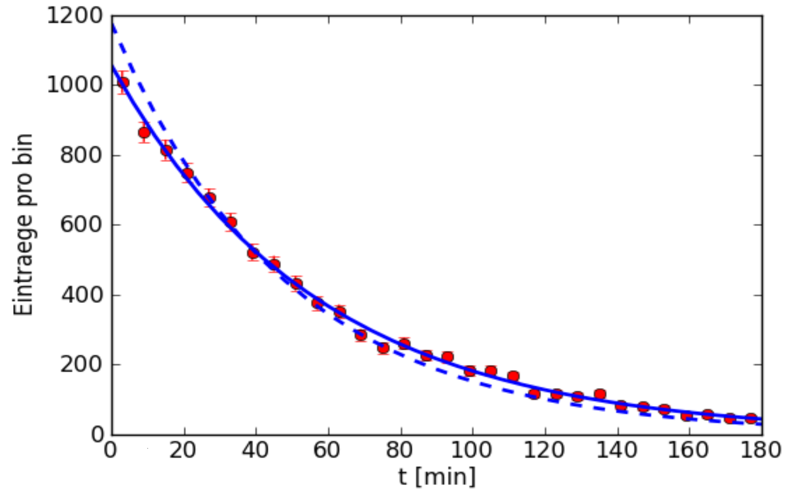
Beispiel mittlere Lebensdauer (III)

Monte-Carlo Beispiel: finde Maximum durch Abtasten von τ

mlexpo.py

```
#!/usr/bin/env python
from pylab import *
import scipy.stats as stats
#
# erzeuge exponentialverteilte Zufallszahlen
#
tautrue = 56.0 ; tmax = 180.0
t = stats.expon(0.0,tautrue).rvs(10000)
#
# verwerfe alle Zufallszahlen > tmax
#
t = t[t<tmax]
#
# einfacher Mittelwert der verbleibenden Zufallszahlen
#
tmean = mean(t)
#
# berechne ln L als Funktion der angenommenen Lebensdauer
# und finde das Maximum
#
nt = len(t)
tau = frange(50.0,60.0,0.02)
logl = nt*(-log(tau)-log(1-exp(-tmax/tau)))-sum(t)/tau
tauml = tau[argmax(logl)]

hold(True)
plot(tau,logl-max(logl))
plot([tauml,tauml],[-7.0,1.0],'k--')

print("ML estimator for tau:    %5.2f min"%tauml)
print("Mean of measured times:  %5.2f min"%tmean)
```






```
True mean decay time    = 56.00 min
ML estimator for tau    = 56.70 min
Mean of measured times = 48.27 min
```

Beispiel gewichteter Mittelwert

Habe N Messungen x_i einer Grösse X mit Messunsicherheiten σ_i

- **Annahme 1: die N Messungen sind voneinander unabhängig**
- **Annahme 2: Messabweichungen sind gaußverteilt**

Messergebnis x_i entstammt einer Gaußverteilung

mit Erwartungswert X und Standardabweichung σ_i

- **Wahrscheinlichkeit für das Messergebnis x_i**

$$f(x_i \mid X, \sigma_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \cdot e^{-\frac{(x_i - X)^2}{2\sigma_i^2}}$$

- **Log-likelihood als Funktion von X:**

$$\ln L(X) = \sum_{i=1}^{N} \ln\left(f(x_i \mid X, \sigma_i)\right) = -\sum_{i=1}^{N} \left(\ln\sqrt{2\pi\sigma_i^2}\right) - \sum_{i=1}^{N} \frac{(x_i - X)^2}{2\sigma_i^2}$$

- **Maximum-Likelihood Schätzfunktion für X:**

$$\frac{d(\ln L)}{dX} = \sum_{i=1}^{N} \frac{x_i - X}{\sigma_i^2} = \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} - X \cdot \sum_{i=1}^{N} \frac{1}{\sigma_i^2} = 0 \quad \Rightarrow \quad \boxed{\hat{X} = \frac{\sum_{i=1}^{N} x_i/\sigma_i^2}{\sum_{i=1}^{N} 1/\sigma_i^2}}$$

Unsicherheit auf dem Schätzwert

Taylor-Entwicklung der log-likelihood Funktion um $a = \hat{a}$

$$\ln L(a) \approx \ln L(\hat{a}) + \underbrace{\left[\frac{d(\ln L)}{da}\right]_{a=\hat{a}}}_{=0 \text{ weil } \hat{a} \text{ Maximum}} \cdot (a-\hat{a}) + \frac{1}{2} \cdot \left[\frac{d^2(\ln L)}{da^2}\right]_{a=\hat{a}} \cdot (a-\hat{a})^2 + \ldots$$

- nehme an, Abweichungen des Schätzwerts vom wahren Wert sind nicht zu gross → breche Entwicklung nach dem quadratischen Term ab

$$\Rightarrow \quad L(a) \approx L(\hat{a}) \cdot e^{\boxed{\frac{1}{2}\left[\frac{d^2(\ln L)}{da^2}\right]_{a=\hat{a}}(a-\hat{a})^2}}$$

- für großen Stichprobenumfang (N → ∞) gilt zentraler Grenzwertsatz: Wahrscheinlichkeitsverteilung für Schätzwert nähert sich Gaußverteilung an

$$p(\hat{a}) \approx \frac{1}{\sqrt{2\pi\sigma_{\hat{a}}^2}} \cdot e^{\boxed{-\frac{(\hat{a}-a)^2}{2\sigma_{\hat{a}}^2}}}$$

- definiere $\sigma_{\hat{a}}$ durch Gleichsetzen

$$-\frac{1}{\sigma_{\hat{a}}^2} = \left[\frac{d^2(\ln L)}{da^2}\right]_{a=\hat{a}} \quad \Leftrightarrow \quad \sigma_{\hat{a}}^2 = \left(-\left[\frac{d^2(\ln L)}{da^2}\right]_{a=\hat{a}}\right)^{-1}$$

Unsicherheit auf dem Schätzwert

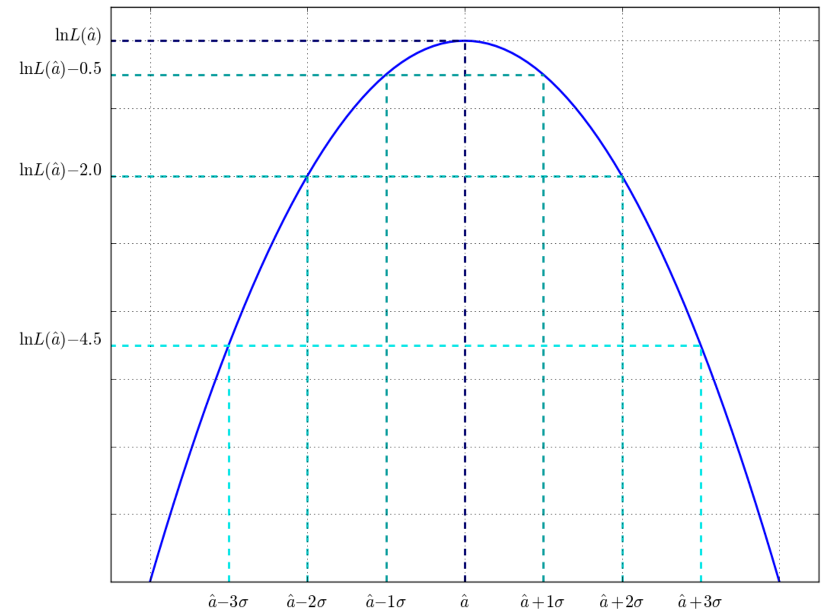
Großer Stichprobenumfang ($N \to \infty$)

- **log-likelihood Funktion parabelförmig**

$$\ln L(a) = \ln L(\hat{a}) - \frac{(a-\hat{a})^2}{2\sigma_{\hat{a}}^2}$$

- **Unsicherheit $\sigma_{\hat{a}}$ aus der Breite der Parabel**

$$\ln L(\hat{a} \pm \sigma_{\hat{a}}) = \ln L(\hat{a}) - 0.5$$
$$\ln L(\hat{a} \pm 2\sigma_{\hat{a}}) = \ln L(\hat{a}) - 2.0$$
$$\ln L(\hat{a} \pm 3\sigma_{\hat{a}}) = \ln L(\hat{a}) - 4.5$$

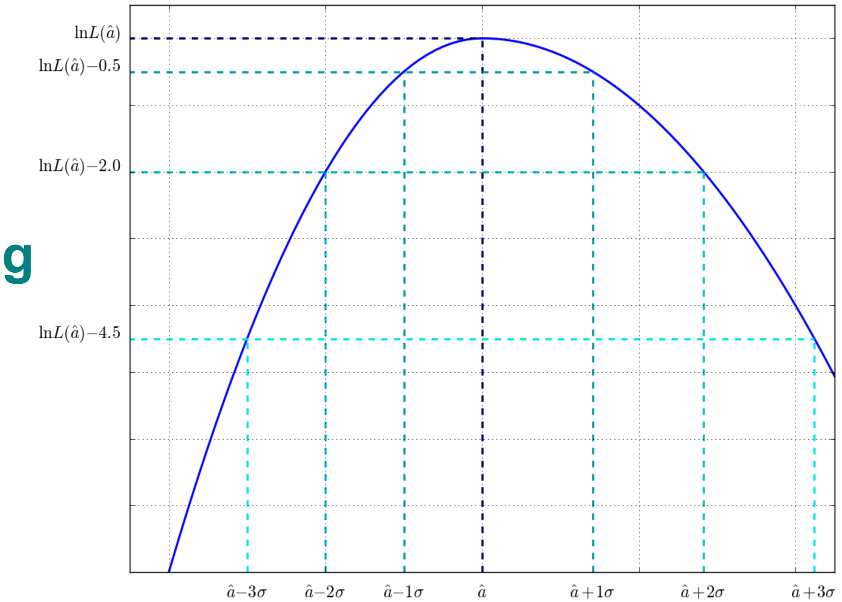


Stichprobenumfang *N* nicht groß genug:

- **log-likelihood Funktion i.a. nicht parabelförmig**
- **definiere Unsicherheit auf $\hat{a}$ dennoch als**

$$\ln L(\hat{a} \pm n \cdot \sigma_{\hat{a}}) \equiv \ln L(\hat{a}) - n^2/2$$

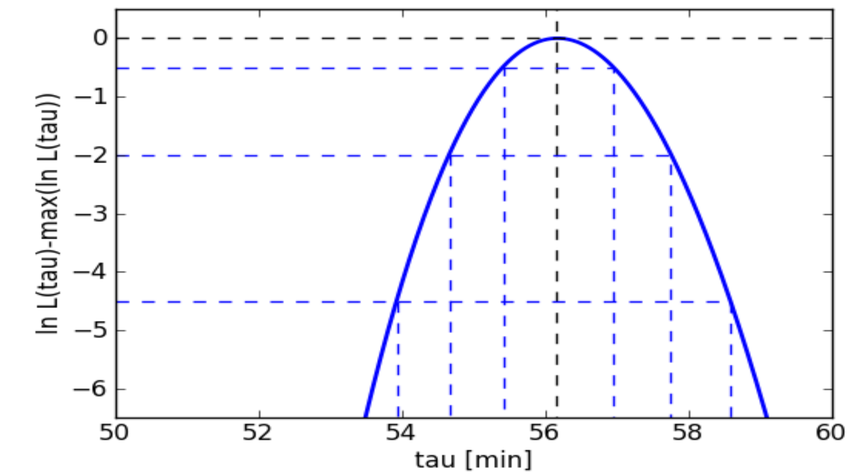
- **ergibt u.U. asymmetrische Unsicherheiten**

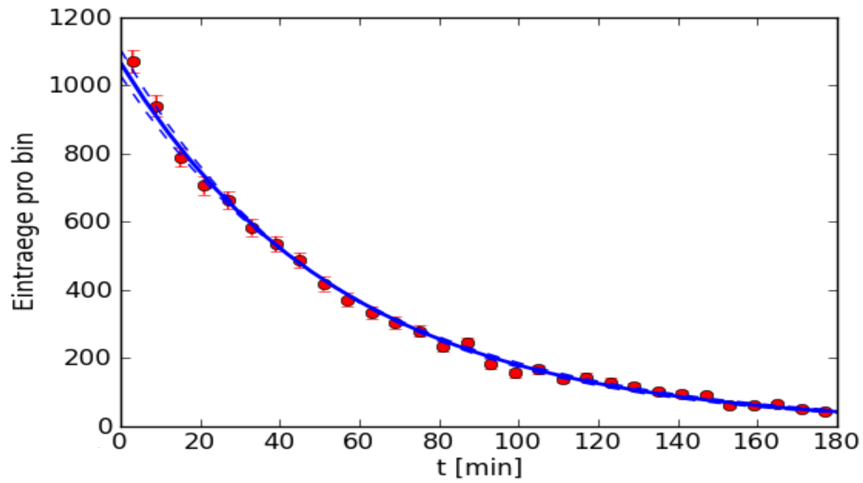


Monte-Carlo Beispiel: Unsicherheit auf $\hat{\tau}$ aus der likelihood-Parabel

mlexperr.py

```
#!/usr/bin/env python
from pylab import *
import scipy.stats as stats
#
# erzeuge exponentialverteilte Zufallszahlen
#
tautrue = 56.0 ; tmax = 180.0
t       = stats.expon(0.0,tautrue).rvs(10000)
#
# verwerfe alle Zufallszahlen > tmax
#
t       = t[t<tmax]
#
# berechne ln L als Funktion der angenommenen Lebensdauer
# und finde das Maximum
#
nt      = len(t)
tau     = frange(50.0,60.0,0.02)
logl    = nt*(-log(tau)-log(1-exp(-tmax/tau)))-sum(t)/tau
tauml   = tau[argmax(logl)]
#
# finde Werte von tau, bei denen der Wert von ln L um
# 0.5, 2.0, 4.5 etc vom Wert Maximum abgefallen ist
#
for i in range(1,4):
    c = logl-max(logl)>-i**2/2.0
    taulo = min(tau[c])
    tauhi = max(tau[c])
#
# weitere Befehle hier nicht gezeigt ...
```






```
ML estimator for lifetime tau  = 55.60 min
-1 sigma = -0.70 min, +1 sigma =  0.70 min
-2 sigma = -1.40 min, +2 sigma =  1.50 min
-3 sigma = -2.20 min, +3 sigma =  2.30 min
```

Verallgemeinerung auf n Parameter

Wahrscheinlichkeitsverteilung hängt von n Parametern a_i ab ($i = 1,\ldots,n$)

- **finde Schätzfunktionen $\hat{a}_1 , \ldots , \hat{a}_n$ durch Lösen des Gleichungssystems**

$$\frac{\partial \ln L(a_1, a_2, \ldots a_n)}{\partial a_i} = 0 \quad \text{für } i = 1, \ldots , n$$

- **Fehlermatrix für die $\hat{a}_i$ durch Invertieren der Matrix**

$$V_{ij}^{-1} = cov^{-1}(a_i, a_j) = \left[-\frac{\partial^2 \ln L}{\partial a_i \partial a_j} \right]_{a_i = \hat{a}_i, a_j = \hat{a}_j}$$

- **definiere $n \cdot \sigma$ Konfidenzregionen auf den $\hat{a}_i$ durch**

$$\ln L(a_i) = \ln L(\hat{a}_i) - n^2 / 2$$

- **großer Stichprobenumfang ($N \to \infty$): $L(a_1,\ldots,a_n) \to n$-dim Gaußverteilung**
- $n = 2$: $L(a_1,a_2) \to$ 2-dim Gaußverteilung; Konfidenzregionen $\to$ konzentrische Ellipsen
- **N nicht groß genug: kompliziertere Konturen, eventuell lokale Maxima**

Beispiel Anpassung einer Geraden

Passe Geradengleichung $y = a_0 + a_1 \cdot x$ an N Messpaare (x_i, y_i) an

- Annahme 1: Messunsicherheiten auf den x_i sind vernachlässigbar klein
- Annahme 2: Abweichungen der gemessen y_i von den erwarteten Werten

$$y = a_0 + a_1 \cdot x_i$$

sind gaußverteilt mit Standardabweichungen σ_i

- Wahrscheinlichkeit für eines der Wertepaare

$$p(y_i \mid a_0, a_1) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(y_i - a_0 - a_1 \cdot x_i)^2}{2\sigma_i^2}}$$

- Annahme 3: keine Korrelationen zwischen den Abweichungen auf den y_i
- log-likelihood Funktion für alle Wertepaare

$$\ln L(a_0, a_1) = \underbrace{\sum_{i=1}^{N} \ln\left(\frac{1}{\sqrt{2\pi\sigma_i^2}}\right)}_{=\,\mathrm{konst}} - \sum_{i=1}^{N} \frac{(y_i - a_0 - a_1 \cdot x_i)^2}{2\sigma_i^2}$$

Beispiel Anpassung einer Geraden

Vereinfachung: Messunsicherheiten auf allen y_i gleich gross ($\sigma_i = \sigma$)

$$\ln L(a_0, a_1) = -\sum_{i=1}^{N} \frac{(y_i - a_0 - a_1 \cdot x_i)^2}{2\sigma^2} + \text{konst}$$

- **zwei gekoppelte Gleichungen für zwei Schätzfunktionen $\hat{a}_0$ und $\hat{a}_1$**

$$\left[\frac{\partial(\ln L)}{\partial a_1}\right]_{\substack{a_0=\hat{a}_0 \\ a_1=\hat{a}_1}} = \sum_{i=1}^{N} \left\{ \frac{(y_i - \hat{a}_0 - \hat{a}_1 \cdot x_i) \cdot x_i}{\sigma^2} \right\} = 0 \quad \Rightarrow \quad \sum_{i=1}^{N} x_i y_i - \hat{a}_0 \cdot \sum_{i=1}^{N} x_i - \hat{a}_1 \cdot \sum_{i=1}^{N} x_i^2 = 0$$

$$\left[\frac{\partial(\ln L)}{\partial a_0}\right]_{\substack{a_0=\hat{a}_0 \\ a_1=\hat{a}_1}} = \sum_{i=1}^{N} \left\{ \frac{y_i - \hat{a}_0 - \hat{a}_1 \cdot x_i}{\sigma^2} \right\} = 0 \quad \Rightarrow \quad \sum_{i=1}^{N} y_i - \hat{a}_0 \cdot \sum_{i=1}^{N} 1 - \hat{a}_1 \cdot \sum_{i=1}^{N} x_i = 0$$

- **in Matrixschreibweise:**

$$\begin{pmatrix} \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} 1 \\ \sum_{i=1}^{N} x_i^2 & \sum_{i=1}^{N} x_i \end{pmatrix} \cdot \begin{pmatrix} \hat{a}_1 \\ \hat{a}_0 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_i y_i \end{pmatrix} \quad \underbrace{\Leftrightarrow}_{\times 1/N} \quad \begin{pmatrix} \overline{x} & 1 \\ \overline{x^2} & \overline{x} \end{pmatrix} \cdot \begin{pmatrix} \hat{a}_1 \\ \hat{a}_0 \end{pmatrix} = \begin{pmatrix} \overline{y} \\ \overline{x\,y} \end{pmatrix}$$

Beispiel Anpassung einer Geraden

- **Auflösen nach $\hat{a}_0$ und $\hat{a}_1$**

$$\hat{a}_0 = \frac{\overline{x^2}\,\overline{y} - \overline{x}\,\overline{xy}}{\overline{x^2} - \overline{x}^2} \qquad \hat{a}_1 = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2}$$

- **Kovarianzmatrix**

$$cov^{-1}(\hat{a}_0, \hat{a}_1) = -\begin{pmatrix} \frac{\partial^2 \ln L}{\partial a_0^2} & \frac{\partial^2 \ln L}{\partial a_0 \partial a_1} \\ \frac{\partial^2 \ln L}{\partial a_0 \partial a_1} & \frac{\partial^2 \ln L}{\partial a_1^2} \end{pmatrix}_{\substack{a_0 = \hat{a}_0 \\ a_1 = \hat{a}_1}} = \frac{1}{\sigma^2} \cdot \begin{pmatrix} \sum_{i=1}^{N} 1 & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i^2 \end{pmatrix} = \frac{N}{\sigma^2} \cdot \begin{pmatrix} 1 & \overline{x} \\ \overline{x} & \overline{x^2} \end{pmatrix}$$

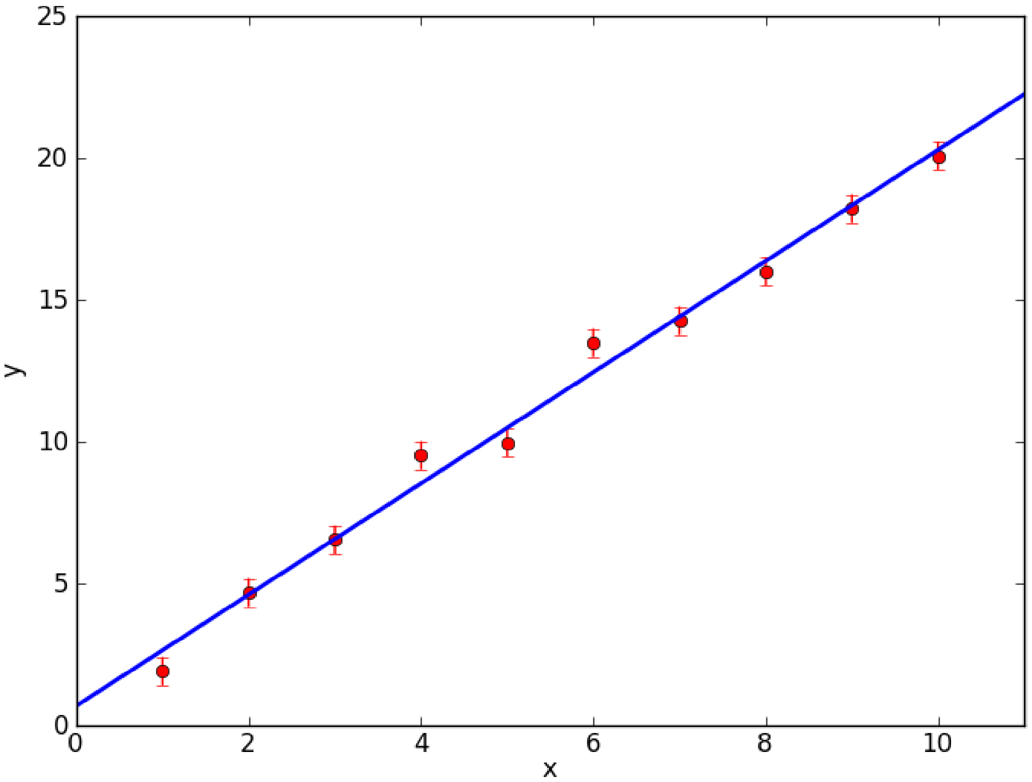
$$\Rightarrow \quad cov(\hat{a}_0, \hat{a}_1) = \frac{\sigma^2}{N \cdot \left(\overline{x^2} - \overline{x}^2\right)} \cdot \begin{pmatrix} \overline{x^2} & -\overline{x} \\ -\overline{x} & 1 \end{pmatrix}$$

Beispiel in pylab

Analytische Bestimmung der Schätzwerte und ihrer Unsicherheiten

mlline.py

```
#!/usr/bin/env python
from pylab import *
import scipy.stats as stats
#
# erzeuge Wertepaare gemaess Geradengleichgung
# mit gaussverteilten Abweichungen auf den y
#
a = 2.0 ; b = 0.5 ; sig = 0.5
x     = frange(1.0,10.0,1.0)
nx    = len(x)
y     = a*x+b+stats.norm(0.0,sig).rvs(nx)
#
# berechne ML Schaetzwerte und Kovarianzmatrix
#
xmean  = mean(x)
ymean  = mean(y)
xymean = mean(x*y)
xxmean = mean(x*x)
varx   = xxmean-xmean**2
ahat   = (xymean-xmean*ymean)/varx
bhat   = ymean - ahat*xmean
vara   = sig**2/(nx*varx)
varb   = vara*xxmean
covab  = -xmean*vara
#
# zeichne Wertepaare und angepasste Gerade
#
errorbar(x,y,yerr=sig,fmt='ro')
hold(True)
plot(x,x*ahat+bhat,'b-',linewidth=2)
#
# weitere Befehle hier nicht gezeigt ...
```



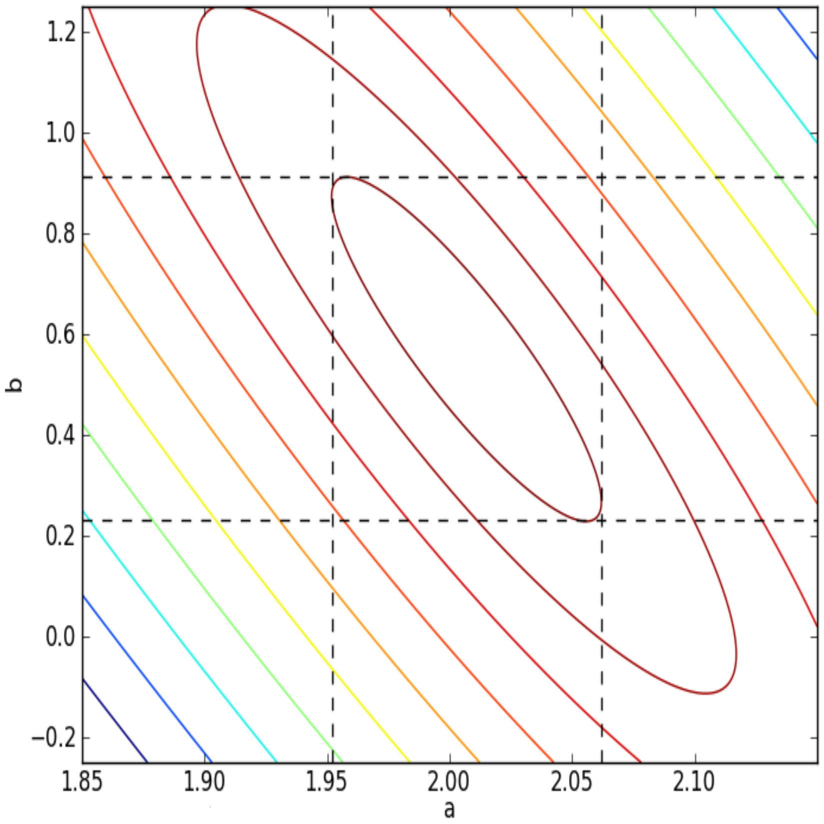

```
Estimate for a = 2.012 +/- 0.055
Estimate for b = 0.608 +/- 0.342
Corr coeff     = -0.8864
Cov matrix     = 0.003   -0.017
                 -0.017  0.117
```

Beispiel in pylab

Schätzwerte und Unsicherheiten durch Abtasten in der (*a*,*b*)-Ebene

mlline2.py

```
#!/usr/bin/env python
from pylab import *
import scipy.stats as stats
#
# erzeuge Wertepaare gemaess Geradengleichgung
# mit gaussverteilten Abweichungen auf den y
#
a = 2.0 ; b = 0.5 ; sig = 0.5
x    = frange(1.0,10.0,1.0)
nx   = len(x)
y    = a*x+b+stats.norm(0.0,sig).rvs(nx)
#
# berechne extended log likelihood auf Raster angenommener
# Werte der zwei Parameter, suche Maximum
#
amin  =  1.85 ; amax = 2.15 ; abin = 0.0003
bmin  = -0.25 ; bmax = 1.25 ; bbin = 0.0015
am,bm = meshgrid(frange(amin,amax,abin),frange(bmin,bmax,bbin))
ll      = zeros(shape(am))
for i in range(0,nx):
    ll -= (y[i]-am*x[i]-bm)**2/(2.0*sig**2)
ahat    = am.flatten()[argmax(ll)]
bhat    = bm.flatten()[argmax(ll)]
#
# bestimme Unsicherheit auf den Schaetzwerten aus Breite
# der log-likelihood als Funktion jedes der zwei Parameter
#
ll -= ll.max()
ahi = am[ll>-0.5].max()-ahat ; alo = ahat-am[ll>-0.5].min()
bhi = bm[ll>-0.5].max()-bhat ; blo = bhat-bm[ll>-0.5].min()
#
# weitere Befehle hier nicht gezeigt ...
```




```
estimate for a = 2.012 +/- 0.055
estimate for b = 0.608 +/- 0.342
```

“Binned” Maximum-Likelihood

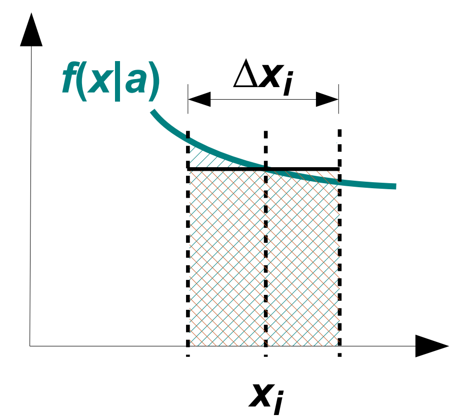
Stichprobe besteht aus Histogramm mit *N* Intervallen

- **Intervall-Zentren:** x_i $\quad (i = 1, \ldots, N)$
- **Intervall-Breiten:** Δx_i $\quad (i = 1, \ldots, N)$
- **Anzahl Einträge pro Intervall:** n_i $\quad$ (gesamte Anzahl Einträge: $n = \sum_{i=1}^{N} n_i$)

Zufallsvariable *x* folgt Wahrscheinlichkeitsverteilung *f*(*x*|*a*)

- **erwartete Anzahl Einträge im Intervall *i* des Histogramms**

$$\nu_i(a) = n \cdot \int_{x_i - \Delta x_i/2}^{x_i + \Delta x_i/2} f(x \mid a)\, dx \approx n \cdot \Delta x_i \cdot f(x_i \mid a)$$



- **Wahrscheinlichkeit, dann n_i Einträge zu beobachten**

$$P_i(a) = \frac{\left[\nu_i(a)\right]^{n_i}}{n_i!} \cdot e^{-\nu_i(a)}$$

- **log-likelihood als Funktion der Parameter *a***

$$L(a) = \prod_{i=1}^{N} \left\{ \frac{\left[\nu_i(a)\right]^{n_i}}{n_i!} \cdot e^{-\nu_i(a)} \right\} \Rightarrow \ln L(a) = \underbrace{\sum_{i=1}^{N} \frac{1}{n_i!}}_{=\text{konst}} + \sum_{i=1}^{N} \left\{ n_i \cdot \ln \nu_i(a) - \nu_i(a) \right\}$$

Erweiterte Maximum-Likelihood

Betrachte Zählexperiment, das über eine feste Zeit läuft

- **Gesamtzahl *n* der beobachteten Ereignisse ist dann auch eine Zufallsvariable**
- **sollte in der likelihood-Funktion berücksichtigt werden**

- **nehme Poissonverteilung an und nehme deren Erwartungswert ν als zusätzlichen Parameter in die likelihood-Funktion**

$$L(a, \nu(a)) = \frac{\nu^n(a)}{n!} \cdot e^{-\nu(a)} \cdot \prod_{i=1}^{n} f(x_i \,|\, a)$$

aufgepasst: ν kann von den Parametern *a* abhängen !

- **erweiterte log-likelihood Funktion:**

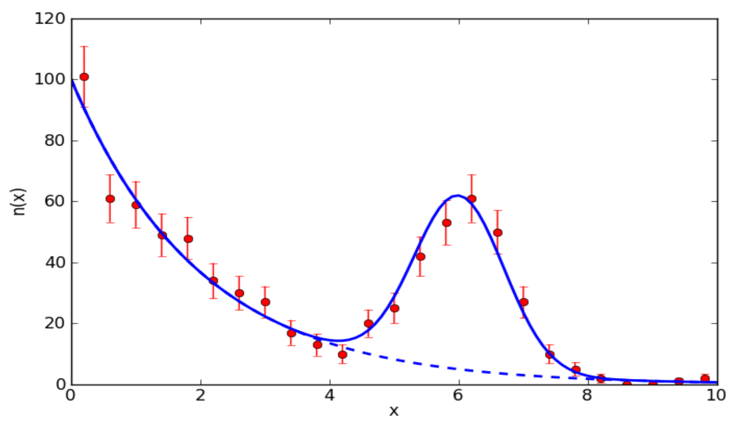
$$\ln L(a, \nu(a)) = n \cdot \ln \nu(a) - \nu(a) + \sum_{i=1}^{n} \ln \{ f(x_i \,|\, a) \}$$

$$= -\nu(a) + \sum_{i=1}^{n} \ln \{ \nu(a) \cdot f(x_i \,|\, a) \}$$

Beispiel Signal und Untergrund

Verteilung aus n_s Signalereignissen und n_b Untergrundereignissen

- **Form der Signalverteilung $s(x)$ sei bekannt**
  - z.B. Gaußverteilung mit bekanntem μ und σ
- **Form der Untergrundverteilung $b(x)$ sei bekannt**
  - z.B. Exponentialverteilung mit bekanntem λ
- **suche Schätzwerte für n_s und n_b**



- **normale maximum likelihood: Gesamtzahl der Ereignisse $n_{tot} = n_s + n_b$ fest**

$$\ln L(n_s) = \sum_{i=1}^{n_{tot}} \ln \left\{ \frac{n_s}{n_{tot}} \cdot s(x_i) + \frac{n_{tot} - n_s}{n_{tot}} \cdot b(x_i) \right\}$$

ein freier Parameter

- **erweiterte maximum likelihood: $n_s + n_b$ folgt Poissonverteilung**

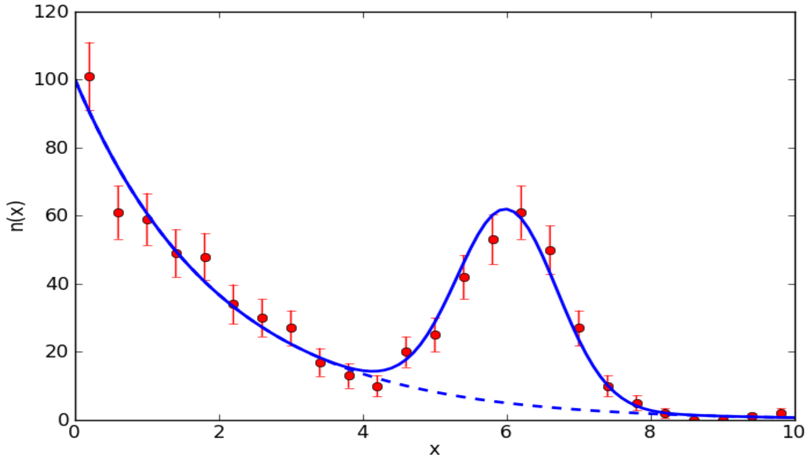
$$\ln L(n_s, n_b) = -(n_s + n_b) + \sum_{i=1}^{n_{tot}} \ln \{ n_s \cdot s(x_i) + n_b \cdot b(x_i) \}$$

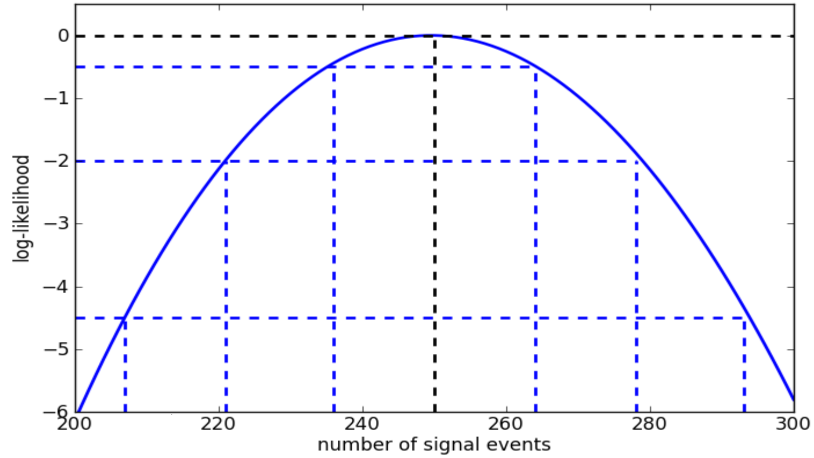
zwei freie Parameter

Normale maximum likelihood: $n_s + n_b$ fest, nur ein freier Parameter (n_s)

mlgaussexp.py

```
#!/usr/bin/env python
from pylab import *
import scipy.stats as stats
#
# erzeuge Zufallszahlen
#
nstrue = 250 ; mu = 6.0 ; sig = 0.7
nbtrue = 500 ; tau = 2.0
x      = concatenate([stats.norm(mu,sig).rvs(nstrue),\
                      stats.expon(0.0,tau).rvs(nbtrue)])
#
# log likelihood als Funktion der angenommenen Anzahl
# Signalereignisse; totale Anzahl Ereignisse fest
#
ntot  = nstrue + nbtrue
nsmin = 200 ; nsmax = 300 ; ns = frange(nsmin,nsmax,1.0)
ll     = zeros(len(ns))
for i in range(0,ntot):
    ll  += log((ns*stats.norm(mu,sig).pdf(x[i])+\
               (ntot-ns)*stats.expon(0.0,tau).pdf(x[i]))/ntot)
ll      -= ll.max()
nshat    = ns[argmax(ll)]
#
# Unsicherheit auf Schaetzwert aus Breite der log likelihood
#
for nsigma in range(1,4):
    c = ll >= -nsigma**2/2.0
    nslo  = nshat - min(ns[c])
    nsup  = max(ns[c]) - nshat
#
# Ausgabe-Befehle hier nicht gezeigt ...
```



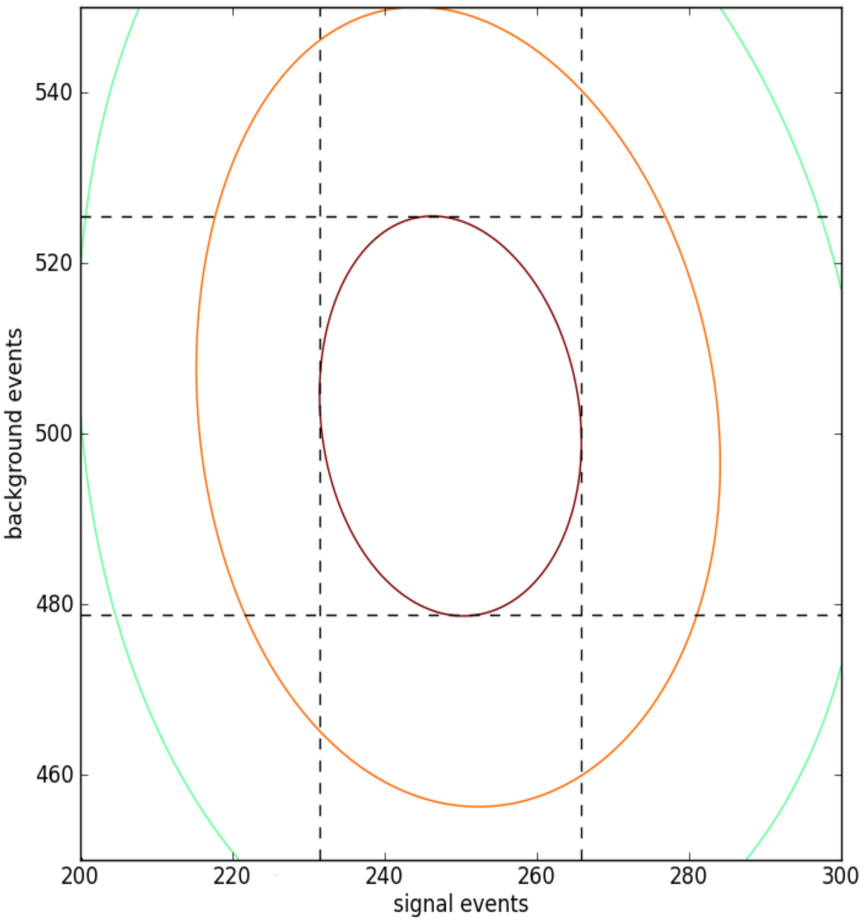



```
total number of events          =  750
ml estimate of signal events =  250
+/- 1 sigma = + 14/- 14 events
+/- 2 sigma = + 29/- 28 events
+/- 3 sigma = + 44/- 42 events
```

Erweiterte maximum likelihood: $n_s + n_b$ folgt Poissonverteilung

emlgaussexp.py

```
#!/usr/bin/env python
from pylab import *
import scipy.stats as stats
#
# erzeuge Zufallszahlen
#
nstrue = 250 ; mu = 6.0 ; sig = 0.7
nbtrue = 500 ; tau = 2.0
x      = concatenate([stats.norm(mu,sig).rvs(nstrue),\
                      stats.expon(0.0,tau).rvs(nbtrue)])
#
# extended log likelihood auf Raster angenommener Anzahlen
# Signal- und Untergrundereignisse
#
nsmin = 200 ; nsmax = 300 ; nstest = frange(nsmin,nsmax,1.0)
nbmin = 450 ; nbmax = 550 ; nbtest = frange(nbmin,nbmax,1.0)
ns,nb = meshgrid(nstest,nbtest)
ll    = zeros(shape(ns))
for i in range(0,ntot):
    ll += log(ns*stats.norm(mu,sig).pdf(x[i])+\
           nb*stats.expon(0.0,tau).pdf(x[i])) ;
ll     -= (ns+nb)
nshat  = ns.flatten()[argmax(ll)]
nbhat  = nb.flatten()[argmax(ll)]
#
# Unsicherheit auf den Schaetzwerten aus Breite der
# log-likelihood als Funktion jedes der zwei Parameter
#
ll     -= ll.max()
nslo   = ns[ll>-0.5].min() ; nsup = ns[ll>-0.5].max()
nblo   = nb[ll>-0.5].min() ; nbup = nb[ll>-0.5].max()
```



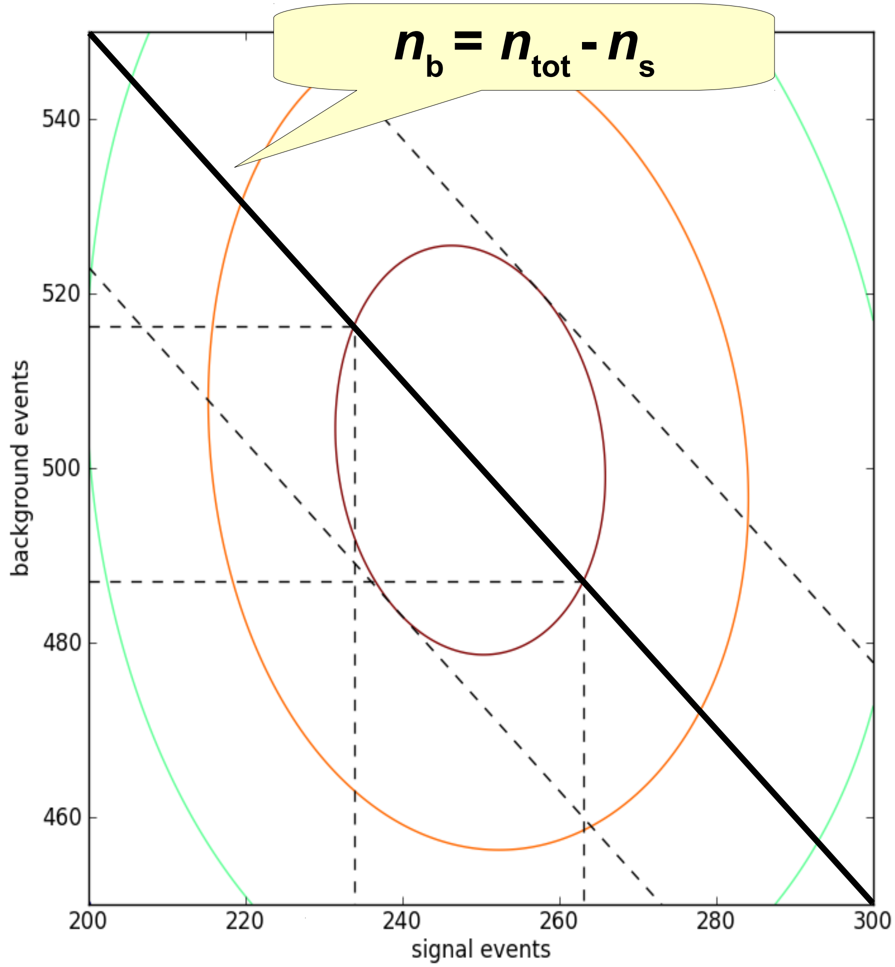

```
ns = 255 + 17 /- 18
nb = 495 + 23 /- 24
```

Beispiel Signal und Untergrund

Zwinge $n_s + n_b$ auf den wahren Wert

emlgaussexp.py

```
#
# oberer Teil wie auf vorheriger Folie ...
#








nshat  = ns.flatten()[argmax(ll)]
nbhat  = nb.flatten()[argmax(ll)]
#
# bestimme Unsicherheit auf den Parametern aus Breite der
# log-likelihood als Funktion jedes der zwei Parameter
# fuer Werte auf der Achse ns+nb=nstrue+nbtrue
#
ll     -= ll.max()
nse = ns[ll>-0.5]
nbe = nb[ll>-0.5]
ntotfixed = nse+nbe==ntot
nslo = nse[ntotfixed].min() ; nsup = nse[ntotfixed].max()
nblo = nbe[ntotfixed].min() ; nbup = nbe[ntotfixed].max()
```




```
ns = 245 + 14 /- 14
nb = 505 + 14 /- 14
```

Zusammenfassung

Maximum Likelihood Methode zur Herleitung von Schätzfunktionen

- wähle als Schätzwerte der Parameter diejenigen Werte, die die Wahrscheinlichkeit maximieren, die gegebene Stichprobe zu erhalten

Vorteile:

- Schätzfunktionen sind konsistent und für großen Stichprobenumfang ($N \to \infty$) asymptotisch erwartungstreu
- Schätzwerte sind für $N \to \infty$ gaußverteilt um den wahren Wert des Parameters
- Daten brauchen nicht histogrammiert zu werden: kein Informationsverlust
- Methode kann leicht auf mehrere Parameter erweitert werden

Nachteile:

- Schätzfunktionen sind nicht immer erwartungstreu
- Resultate lassen keinen Rückschluss auf die Qualität der Anpassung zu