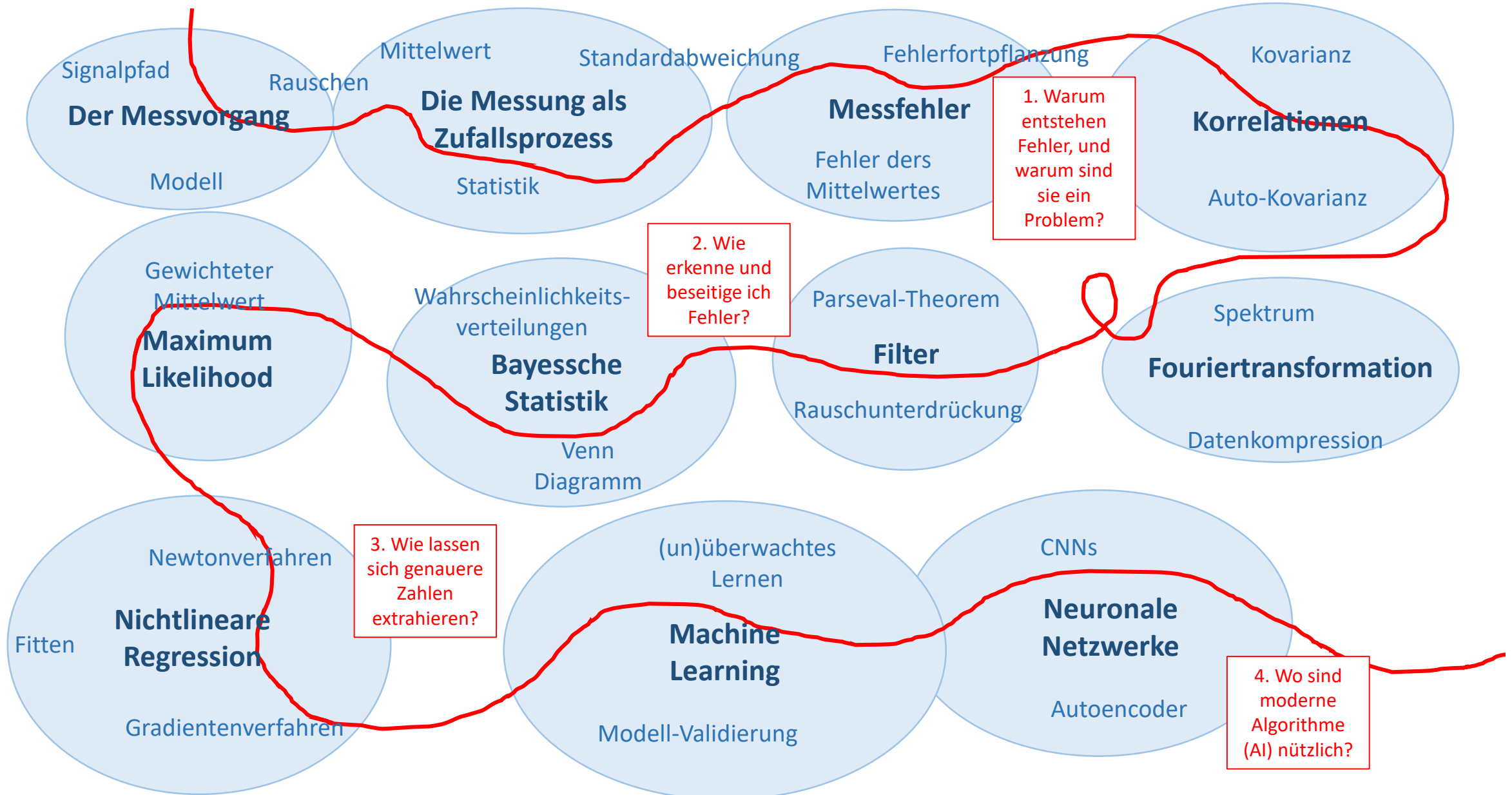


Maximum Likelihood

Der rote Faden: wie erhalte ich aussagekräftige Zahlen aus komplexen Daten?



Themen dieser Vorlesung:

- Die Likelihood Funktion
- Maximum Likelihood Methode
- Der gewichtete Mittelwert
- Die Methode der kleinsten Quadrate
- Lineares Fitten von Polynomen
- Die Kovarianzmatrix
- Die Güte des Fits

Repetition zu Wahrscheinlichkeitsverteilungen

Welche Zufallsprozesse lassen sich durch eine Binomialverteilung beschreiben?

Wann können wir eine Poissonverteilung benutzen?

Und wann eine Normalverteilung?

Die Likelihood Funktion

Wir haben gesehen, dass eine einzelne Messung äquivalent ist zu einer Ziehung einer Zufallszahl aus einer Wahrscheinlichkeitsverteilung:

Die Wahrscheinlichkeitsverteilung einer Zufallsvariable ζ ist: $f(\zeta, a_0, a_1, \dots, a_M) = f(\zeta, \mathbf{a})$

Hier ist \mathbf{a} ein Vektor der die Wahrscheinlichkeitsverteilung parametrisiert. (Im Falle einer Gaussverteilung wäre: $\mathbf{a} = (\mu, \sigma)$)

Für mehrere Variablen schreiben wir: $f(\zeta_0, \zeta_1, \dots, \zeta_N, a_0, a_1, \dots, a_M) = f(\boldsymbol{\zeta}, \mathbf{a})$

Die Wahrscheinlichkeitsverteilung $f(\boldsymbol{\zeta}, \mathbf{a})$ ist eine Funktion der Daten $\boldsymbol{\zeta}$. Warum ist das bei einer Anwendung zur Datenanalyse eigentlich nicht hilfreich?

Die Likelihood Funktion

Wir haben gesehen, dass eine einzelne Messung äquivalent ist zu einer Ziehung einer Zufallszahl aus einer Wahrscheinlichkeitsverteilung:

Die Wahrscheinlichkeitsverteilung einer Zufallsvariable ζ ist: $f(\zeta, a_0, a_1, \dots, a_M) = f(\zeta, \mathbf{a})$

Hier ist \mathbf{a} ein Vektor der die Wahrscheinlichkeitsverteilung parametrisiert. (Im Falle einer Gaussverteilung wäre: $\mathbf{a} = (\mu, \sigma)$)

Für mehrere Variablen schreiben wir: $f(\zeta_0, \zeta_1, \dots, \zeta_N, a_0, a_1, \dots, a_M) = f(\boldsymbol{\zeta}, \mathbf{a})$

Die Wahrscheinlichkeitsverteilung $f(\boldsymbol{\zeta}, \mathbf{a})$ ist eine Funktion der Daten $\boldsymbol{\zeta}$. Warum ist das bei einer Anwendung zur Datenanalyse eigentlich nicht hilfreich?

In einem Experiment sind die Werte der a_i unbekannt, aber wir kennen (aus der Messung) die ζ_i
Z.B. Messen wir die Werte $\zeta_n = x_n$

Wir brauchen also eine Funktion der a_m nicht der ζ_n , bzw der x_n

Wir definieren also: $f(x_0, x_1, \dots, x_N, \mathbf{a}) = f(\mathbf{x}, \mathbf{a}) \stackrel{\text{def}}{=} L(\mathbf{x}, \mathbf{a})$

Die Likelihood Funktion L

Die Likelihood Funktion

Wir haben gesehen, dass eine einzelne Messung äquivalent ist zu einer Ziehung einer Zufallszahl aus einer Wahrscheinlichkeitsverteilung:

Die Wahrscheinlichkeitsverteilung einer Zufallsvariable ζ ist: $f(\zeta, a_0, a_1, \dots, a_M) = f(\zeta, \mathbf{a})$

Hier ist \mathbf{a} ein Vektor der die Wahrscheinlichkeitsverteilung parametrisiert. (Im Falle einer Gaussverteilung wäre: $\mathbf{a} = (\mu, \sigma)$)

Für mehrere Variablen schreiben wir: $f(\zeta_0, \zeta_1, \dots, \zeta_N, a_0, a_1, \dots, a_M) = f(\boldsymbol{\zeta}, \mathbf{a})$

Die Wahrscheinlichkeitsverteilung $f(\boldsymbol{\zeta}, \mathbf{a})$ ist eine Funktion der Daten $\boldsymbol{\zeta}$. Warum ist das bei einer Anwendung zur Datenanalyse eigentlich nicht hilfreich?

In einem Experiment sind die Werte der a_i unbekannt, aber wir kennen (aus der Messung) die ζ_i
Z.B. Messen wir die Werte $\zeta_n = x_n$

Wir brauchen also eine Funktion der a_m nicht der ζ_n , bzw der x_n

Wir definieren also: $f(x_0, x_1, \dots, x_N, \mathbf{a}) = f(\mathbf{x}, \mathbf{a}) \stackrel{\text{def}}{=} L(\mathbf{x}, \mathbf{a})$ Die Likelihood Funktion L

In der Likelihood Funktion sind die x_n bekannt und die a_m sind die Variablen.

Die Likelihood Funktion

Beispiel: Wir messen zwei unabhängige Datenpunkte x_1 und x_2 aus einer Gaussverteilung:

$$f(\zeta, a_0, a_1) = \frac{1}{\sqrt{2\pi a_1}} \exp\left(-\frac{1}{2} \frac{(\zeta - a_0)^2}{a_1}\right)$$

Die Parameter $a_0 = \mu$ (der Erwartungswert) und $a_1 = \sigma^2$ (die Varianz) sind unbekannt.

Da die Messungen unabhängig sind, ist die gemeinsame Wahrscheinlichkeitsverteilung:

$$L(x_1, x_2, a_0, a_1) = f(x_1, x_2, a_0, a_1) = f(x_1, a_0, a_1) f(x_2, a_0, a_1) = \frac{1}{2\pi a_1} \exp\left(-\frac{1}{2} \frac{\sum_{n=1}^2 (x_i - a_0)^2}{a_1}\right)$$

In diesem Beispiel sieht die Likelihood Funktion der Wahrscheinlichkeitsverteilung sehr ähnlich. Aber Vorsicht:

Die Likelihood Funktion ist **keine Wahrscheinlichkeitsverteilung**

Insbesondere ist sie nicht normiert.

Die Likelihood Funktion

Beispiel: Für n Datenpunkte: $\mathbf{x} = (x_1, x_2, \dots, x_N)$ gilt dann:

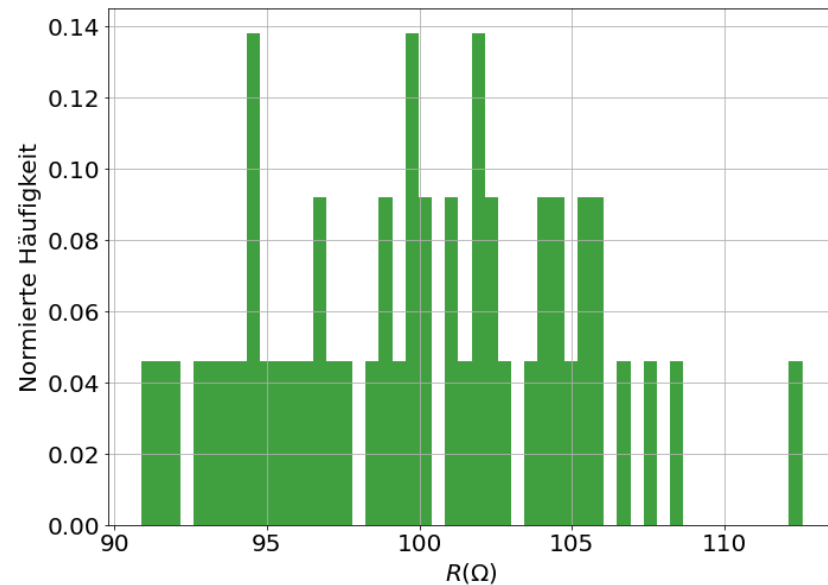
$$L(\mathbf{x}, a_0, a_1) = \prod_{n=1}^N f(x_n, a_0, a_1) = \left(\frac{1}{2\pi a_1} \right)^{N/2} \exp \left(-\frac{1}{2} \frac{\sum_{n=1}^N (x_n - a_0)^2}{a_1} \right)$$

Die Likelihood Funktion

Beispiel: Wir messen einen Satz von 100 Ω Widerständen mit einer unbekannten Toleranz. Wir nehmen an, dass die Widerstandswerte Normalverteilt sind.

$$L(\mathbf{x}, a_0, a_1) = \prod_{n=1}^N f(x_n, a_0, a_1) = \left(\frac{1}{2\pi a_1}\right)^{N/2} \exp\left(-\frac{1}{2} \frac{\sum_{n=1}^N (x_n - a_0)^2}{a_1}\right)$$

Wir wollen diesen Datensatz mit der Likelihood Funktion analysieren.

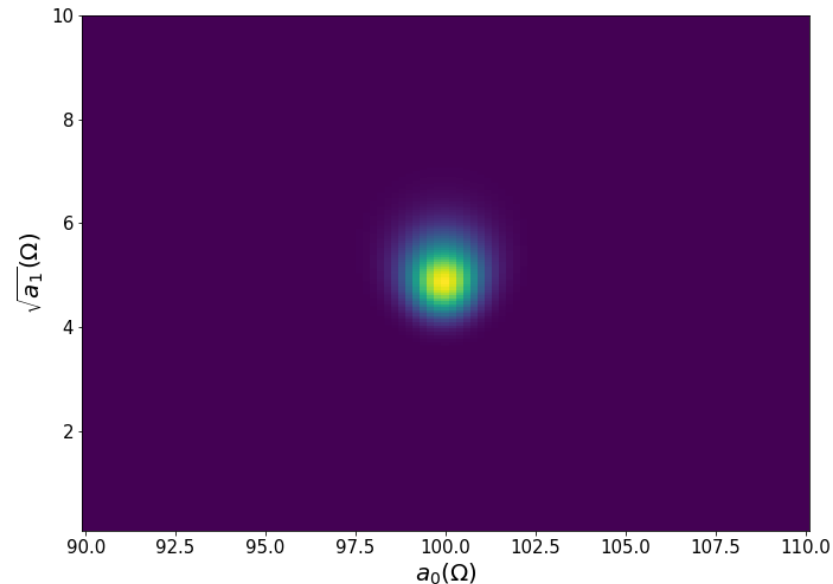
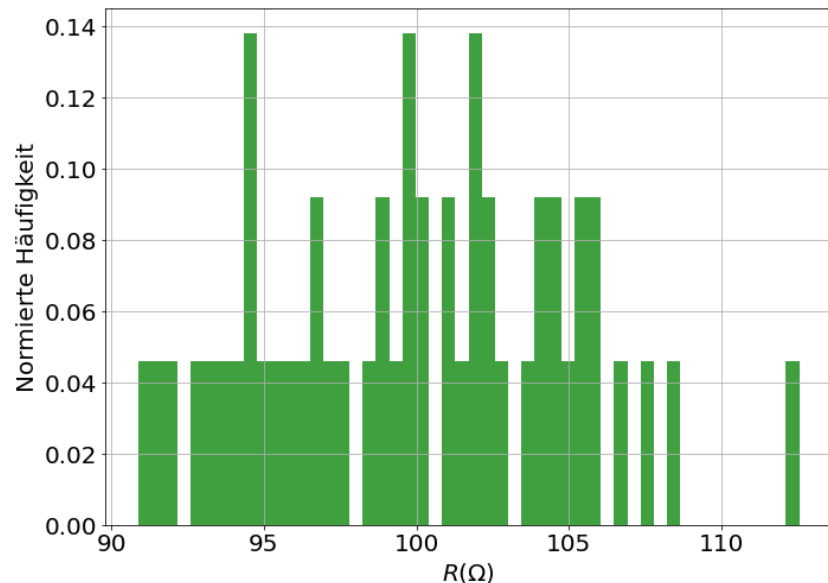


Die Likelihood Funktion

Beispiel: Wir messen einen Satz von 100Ω Widerständen mit einer unbekannten Toleranz. Wir nehmen an, dass die Widerstandswerte Normalverteilt sind.

$$L(\mathbf{x}, a_0, a_1) = \prod_{n=1}^N f(x_n, a_0, a_1) = \left(\frac{1}{2\pi a_1}\right)^{N/2} \exp\left(-\frac{1}{2} \frac{\sum_{n=1}^N (x_n - a_0)^2}{a_1}\right)$$

Wir wollen diesen Datensatz mit der Likelihood Funktion analysieren.



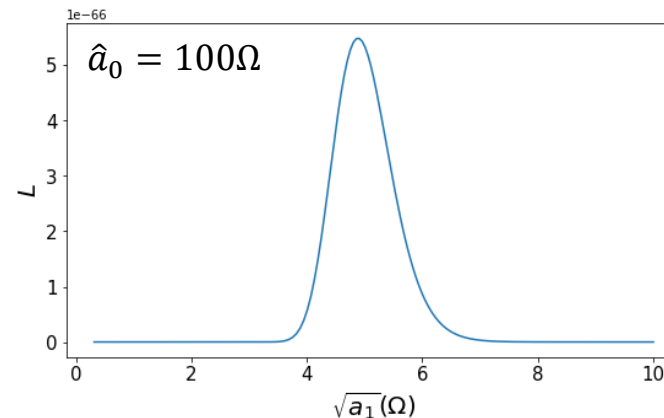
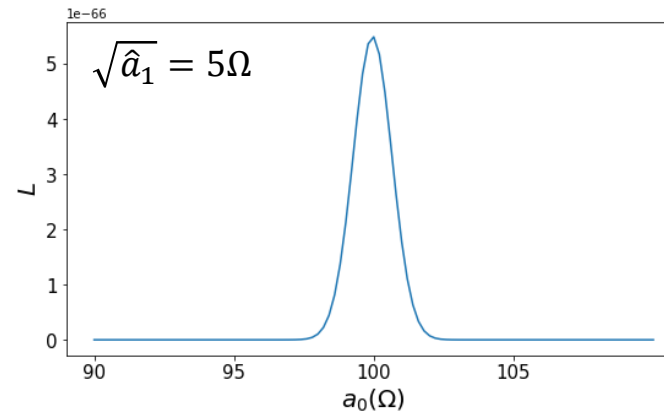
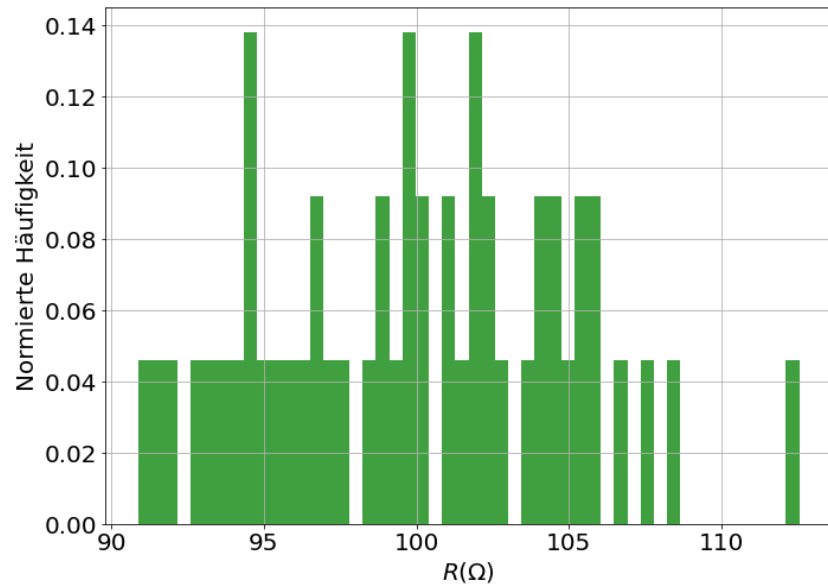
Wir berechnen die Likelihood Funktion für „alle“ Werte von a_0 und a_1 . Wir finden die maximale Likelihood für eine Verteilung bei $\hat{a}_0 = 100\Omega$ und $\sqrt{\hat{a}_1} = 5\Omega$

Die Likelihood Funktion

Beispiel: Wir messen einen Satz von 100Ω Widerständen mit einer unbekannten Toleranz. Wir nehmen an, dass die Widerstandswerte Normalverteilt sind.

$$L(\mathbf{x}, a_0, a_1) = \prod_{n=1}^N f(x_n, a_0, a_1) = \left(\frac{1}{2\pi a_1}\right)^{N/2} \exp\left(-\frac{1}{2} \frac{\sum_{n=1}^N (x_n - a_0)^2}{a_1}\right)$$

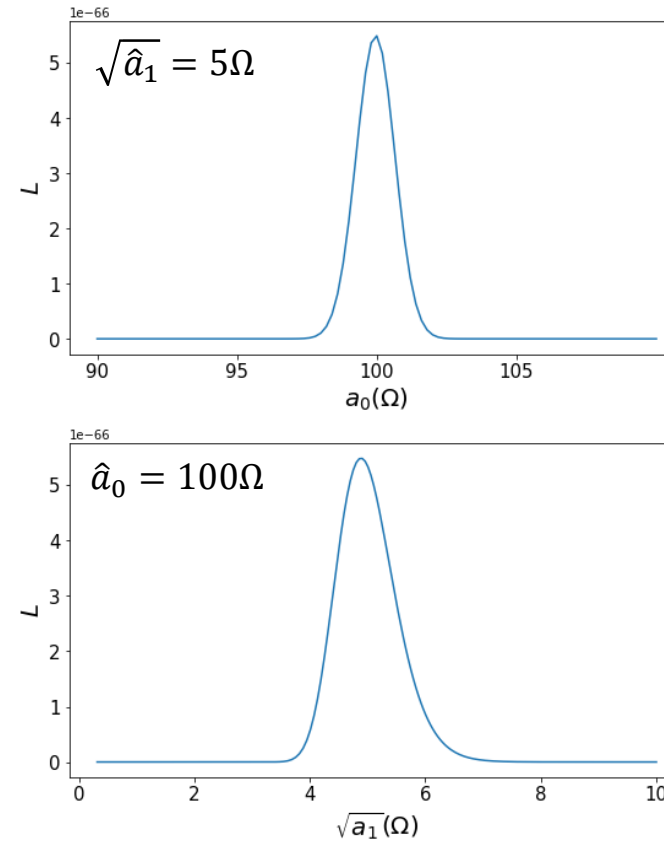
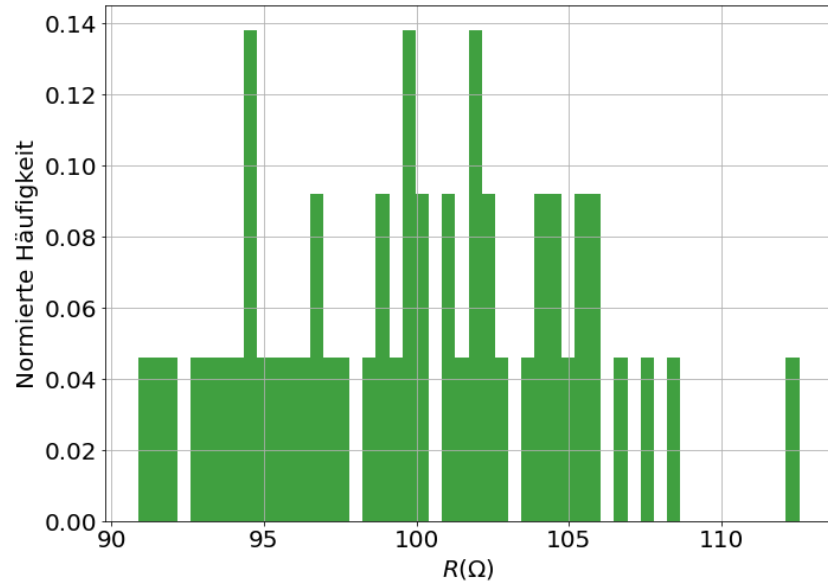
Wir wollen diesen Datensatz mit der Likelihood Funktion analysieren.



Wir berechnen die Likelihood Funktion für „alle“ Werte von a_0 und a_1 . Wir finden die maximale Likelihood für eine Verteilung bei $\hat{a}_0 = 100\Omega$ und $\sqrt{\hat{a}_1} = 5\Omega$

Die Likelihood Funktion

Beispiel: Wir wollen diesen Datensatz mit der Likelihood Funktion analysieren.



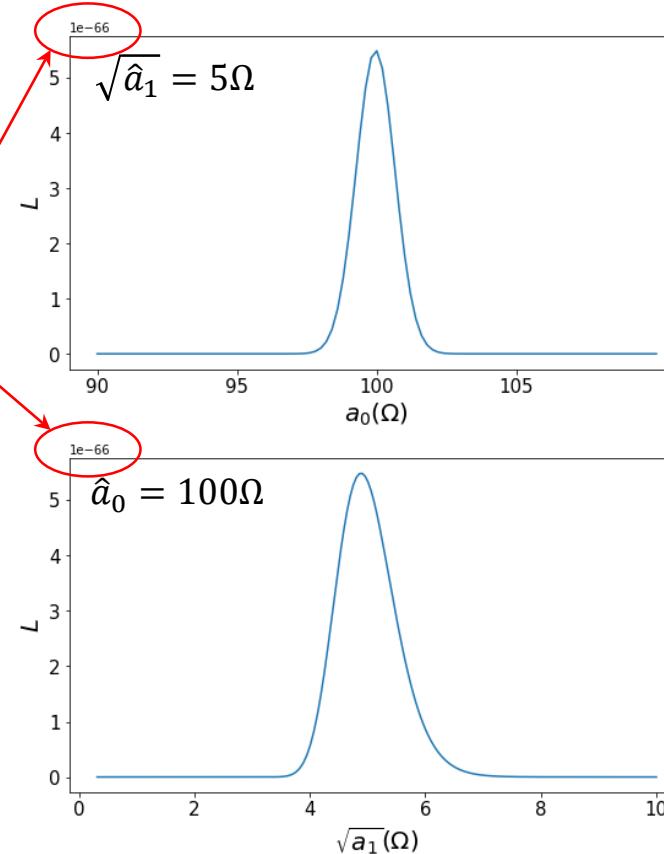
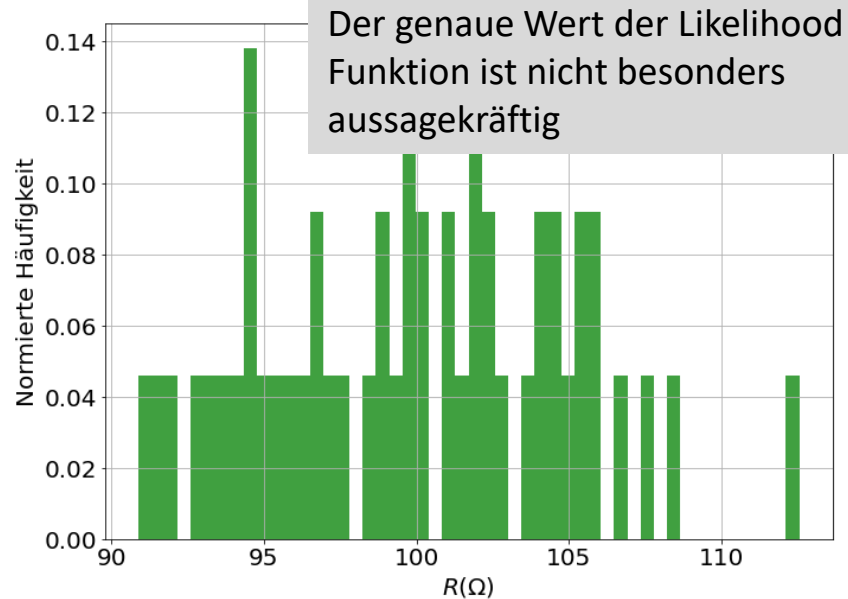
Wir berechnen die Likelihood Funktion für „alle“ Werte von a_0 und a_1 . Wir finden die maximale Likelihood für eine Verteilung bei $\hat{a}_0 = 100\Omega$ und $\sqrt{\hat{a}_1} = 5\Omega$

In diesem Fall, hätten wir auch die Ausdrücke für die Abschätzung von Erwartungswert und Standardabweichung benutzen können.

Aber die Analyse der Likelihood Funktion funktioniert auch bei beliebigen anderen Problemen.

Die Likelihood Funktion

Beispiel: Wir wollen diesen Datensatz mit der Likelihood Funktion analysieren.



Wir berechnen die Likelihood Funktion für „alle“ Werte von a_0 und a_1 . Wir finden die maximale Likelihood für eine Verteilung bei $\hat{a}_0 = 100\Omega$ und $\sqrt{\hat{a}_1} = 5\Omega$

In diesem Fall, hätten wir auch die Ausdrücke für die Abschätzung von Erwartungswert und Standardabweichung benutzen können.

Aber die Analyse der Likelihood Funktion funktioniert auch bei beliebigen anderen Problemen.

Maximum Likelihood

Der Wert der Likelihood Funktion an sich ist also nicht sehr aussagekräftig.

Aber: Die Likelihood Funktion beinhaltet Information über die Parameter der Wahrscheinlichkeitsverteilung:
→ Werte von ***a*** die eine grosse Likelihood haben sind wahrscheinlicher, also zu bevorzugen.

Wir können Likelihoods über ihr Verhältnis vergleichen

$$LR = \frac{L(\boldsymbol{x}, \boldsymbol{a})}{L(\boldsymbol{x}, \boldsymbol{b})}$$

Das Verhältnis der „Wahrscheinlichkeiten“, dass die Parameter ***a*** und ***b*** die gemessenen Daten erklären.

Insbesondere wollen wir die Werte für ***a*** finden, die die Likelihood Funktion maximieren.

Das sogenannte **Maximum Likelihood** Prinzip

Maximum Likelihood

Beispiel: Wir beginnen wieder mit gaussverteilten Daten $\mathbf{x} = (x_1, x_2, \dots, x_N)$:

Die zugrundeliegende Likelihood Funktion ist:

$$f(\zeta, a_0, a_1) = \frac{1}{\sqrt{2\pi a_1}} \exp\left(-\frac{1}{2} \frac{(\zeta - a_0)^2}{a_1}\right)$$

Die Parameter a_0 (der Erwartungswert) und a_1 (die Varianz) sind unbekannt und wir wollen diese Werte abschätzen.

Die kombinierte Likelihood Funktion ist dann das Produkt der einzelnen Likelihood Funktionen:

$$L(\mathbf{x}, a_0, a_1) = \left(\frac{1}{2\pi a_1}\right)^{N/2} \exp\left(-\frac{1}{2} \frac{\sum_{n=1}^N (x_n - a_0)^2}{a_1}\right)$$

Wir sind ausschliesslich an den Werten von a_0, a_1 interessiert die diese Funktion maximieren, aber nicht am Wert des Maximums. Wir können also den Logarithmus nehmen:

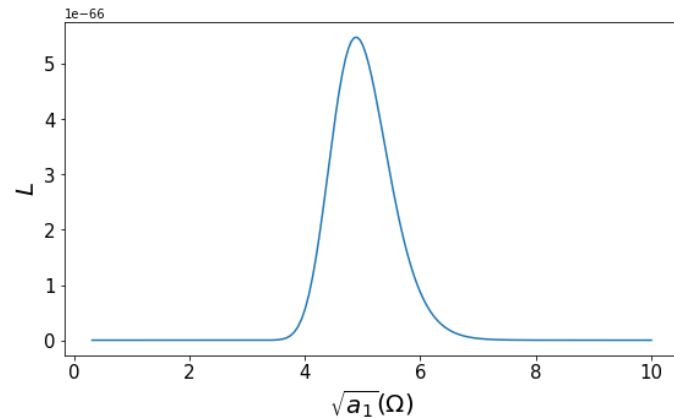
$$l(\mathbf{x}, a_0, a_1) = -\frac{N}{2} \ln(2\pi a_1) - \frac{1}{2} \frac{\sum_{n=1}^N (x_n - a_0)^2}{a_1}$$

Die sogenannte **Log-Likelihood Funktion** (damit lässt sich leichter rechnen...)

Maximum Likelihood

Beispiel: Wir beginnen wieder mit gaussverteilten Daten $\mathbf{x} = (x_1, x_2, \dots, x_n)$:

Likelihood Funktion

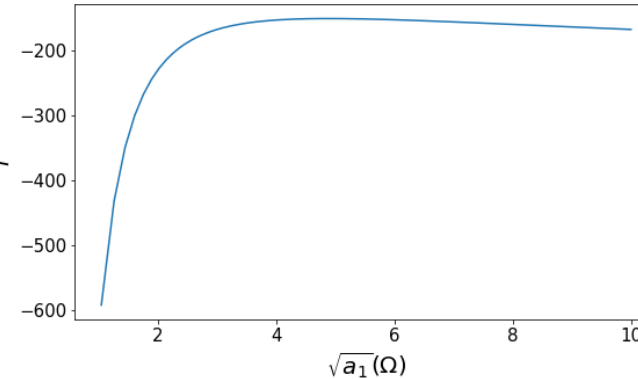


Log-Likelihood Funktion

tion ist:

$l(\mathbf{x}, a_0, a_1) = -\frac{1}{2} \ln(2\pi a_1) - \frac{1}{2} \frac{\sum_{n=1}^N (x_n - a_0)^2}{a_1}$

st dann das Prod



! Werte

$$L(\mathbf{x}, a_0, a_1) = \left(\frac{1}{2\pi a_1} \right)^{N/2} \exp \left(-\frac{1}{2} \frac{\sum_{n=1}^N (x_n - a_0)^2}{a_1} \right)$$

Wir sind ausschliesslich an den Werten von a_0, a_1 interessiert die diese Funktion maximieren, aber nicht am Wert des Maximums. Wir können also den Logarithmus nehmen:

$$l(\mathbf{x}, a_0, a_1) = -\frac{N}{2} \ln(2\pi a_1) - \frac{1}{2} \frac{\sum_{n=1}^N (x_n - a_0)^2}{a_1}$$

Die sogenannte **Log-Likelihood Funktion** (damit lässt sich leichter rechnen...)

Maximum Likelihood

Beispiel: Wir wollen also die Log-Likelihood Funktion maximieren:

$$l(x, a_0, a_1) = -\frac{N}{2} \ln(2\pi a_1) - \frac{1}{2} \frac{\sum_{n=1}^N (x_n - a_0)^2}{a_1}$$

Dafür berechnen wir die Ableitung von l nach a_0 und a_1 an den Stellen \hat{a}_0 und \hat{a}_1 und setzen diese gleich Null, so dass \hat{a}_0 und \hat{a}_1 dann die Werte sind die l maximieren:

$$\frac{\partial l}{\partial a_0} = \frac{\partial}{\partial a_0} \bigg|_{\hat{a}_0, \hat{a}_1} -\frac{N}{2} \ln(2\pi a_1) - \frac{1}{2} \frac{\sum_{n=1}^N (x_n - a_0)^2}{a_1} = 0$$

$$\rightarrow 0 = \frac{1}{\hat{a}_1} \sum_{n=1}^N (x_n - \hat{a}_0) = \frac{1}{\hat{a}_1} \left(\sum_{n=1}^N x_n - N \hat{a}_0 \right)$$

$$\rightarrow \hat{a}_0 = \frac{1}{N} \sum_{n=1}^N x_n$$

Maximum Likelihood

Beispiel: Wir wollen also die Log-Likelihood Funktion maximieren:

$$l(x, a_0, a_1) = -\frac{n}{2} \ln(2\pi a_1) - \frac{1}{2} \frac{\sum_{i=1}^n (x_i - a_0)^2}{a_1}$$

Also berechnen wir die Ableitung von l nach a_0 und a_1 an den Stellen \hat{a}_0 und \hat{a}_1 und setzen diese gleich Null, so dass \hat{a}_0 und \hat{a}_1 dann die Werte sind die l maximieren:

$$\frac{\partial l}{\partial a_0} = \frac{\partial}{\partial a_0} \bigg|_{\hat{a}_0, \hat{a}_1} - \frac{N}{2} \ln(2\pi a_1) - \frac{1}{2} \frac{\sum_{n=1}^N (x_n - a_0)^2}{a_1} = 0$$

$$\rightarrow 0 = \frac{1}{\hat{a}_1} \sum_{n=1}^N (x_n - \hat{a}_0) = \frac{1}{\hat{a}_1} \left(\sum_{n=1}^N x_n - N \hat{a}_0 \right)$$

$$\rightarrow \hat{a}_0 = \frac{1}{N} \sum_{n=1}^N x_n$$

$$\frac{\partial l}{\partial a_1} = \frac{\partial}{\partial a_1} \bigg|_{\hat{a}_0, \hat{a}_1} - \frac{N}{2} \ln(2\pi a_1) - \frac{1}{2} \frac{\sum_{n=1}^N (x_n - a_0)^2}{a_1} = 0$$

$$\rightarrow 0 = -\frac{N}{2} \frac{1}{\hat{a}_1} + \frac{1}{2\hat{a}_1^2} \sum_{n=1}^N (x_n - \hat{a}_0)^2$$

$$\rightarrow \hat{a}_1 = \frac{1}{N} \sum_{n=1}^N (x_n - \hat{a}_0)^2$$

Maximum Likelihood

Beispiel: Wir wollen also die Log-Likelihood Funktion maximieren:

$$l(x, a_0, a_1) = -\frac{n}{2} \ln(2\pi a_1) - \frac{1}{2} \frac{\sum_{i=1}^n (x_i - a_0)^2}{a_1}$$

Also berechnen wir die Ableitung von l nach a_0 und a_1 an den Stellen \hat{a}_0 und \hat{a}_1 und setzen diese gleich Null, so dass \hat{a}_0 und \hat{a}_1 dann die Werte sind die l maximieren:

$$\begin{aligned} \frac{\partial l}{\partial a_0} &= \frac{\partial}{\partial a_0} \bigg|_{\hat{a}_0, \hat{a}_1} - \frac{N}{2} \ln(2\pi a_1) - \frac{1}{2} \frac{\sum_{n=1}^N (x_n - a_0)^2}{a_1} = 0 \\ \rightarrow 0 &= \frac{1}{\hat{a}_1} \sum_{n=1}^N (x_n - \hat{a}_0) = \frac{1}{\hat{a}_1} \left(\sum_{n=1}^N x_n - N \hat{a}_0 \right) \end{aligned}$$

Der Mittelwert, bzw. der Erwartungswert

$$\rightarrow \hat{a}_0 = \frac{1}{N} \sum_{n=1}^N x_n$$

$$\begin{aligned} \frac{\partial l}{\partial a_1} &= \frac{\partial}{\partial a_1} \bigg|_{\hat{a}_0, \hat{a}_1} - \frac{N}{2} \ln(2\pi a_1) - \frac{1}{2} \frac{\sum_{n=1}^N (x_n - a_0)^2}{a_1} = 0 \\ \rightarrow 0 &= -\frac{N}{2} \frac{1}{\hat{a}_1} + \frac{1}{2\hat{a}_1^2} \sum_{n=1}^N (x_n - \hat{a}_0)^2 \end{aligned}$$

Und die Varianz

$$\rightarrow \hat{a}_1 = \frac{1}{N} \sum_{n=1}^N (x_n - \hat{a}_0)^2$$

Eine weitere Methode den Mittelwert und die Varianz zu bestimmen

Maximum Likelihood

Beispiel: Wir wollen also die Log-Likelihood Funktion maximieren:

$$l(x, a_0, a_1) = -\frac{n}{2} \ln(2\pi a_1) - \frac{1}{2} \frac{\sum_{i=1}^n (x_i - a_0)^2}{a_1}$$

Also berechnen wir die Ableitung von l nach a_0 und a_1 an den Stellen \hat{a}_0 und \hat{a}_1 und setzen diese gleich Null, so dass \hat{a}_0 und \hat{a}_1 dann die Werte sind die l maximieren:

$$\begin{aligned} \frac{\partial l}{\partial a_0} &= \frac{\partial}{\partial a_0} \bigg|_{\hat{a}_0, \hat{a}_1} - \frac{N}{2} \ln(2\pi a_1) - \frac{1}{2} \frac{\sum_{n=1}^N (x_n - a_0)^2}{a_1} = 0 \\ \rightarrow 0 &= \frac{1}{\hat{a}_1} \sum_{n=1}^N (x_n - \hat{a}_0) = \frac{1}{\hat{a}_1} \left(\sum_{n=1}^N x_n - N \hat{a}_0 \right) \end{aligned}$$

Der Mittelwert, bzw. der Erwartungswert

$$\rightarrow \hat{a}_0 = \frac{1}{N} \sum_{n=1}^N x_n$$

$$\begin{aligned} \frac{\partial l}{\partial a_1} &= \frac{\partial}{\partial a_1} \bigg|_{\hat{a}_0, \hat{a}_1} - \frac{N}{2} \ln(2\pi a_1) - \frac{1}{2} \frac{\sum_{n=1}^N (x_n - a_0)^2}{a_1} = 0 \\ \rightarrow 0 &= -\frac{N}{2} \frac{1}{\hat{a}_1} + \frac{1}{2\hat{a}_1^2} \sum_{n=1}^N (x_n - \hat{a}_0)^2 \end{aligned}$$

Und die Varianz

$$\rightarrow \hat{a}_1 = \frac{1}{N} \sum_{n=1}^N (x_n - \hat{a}_0)^2$$

Bemerkung:

Der Wert der Varianz hat einen Bias im Vergleich zu dem Wert für die empirische Varianz

Der gewichtete Mittelwert

Wir betrachten nun ein häufiges Szenario

Wir messen mit einer uns bekannten und charakterisierten Messapparatur. Das heisst die Messunsicherheit (die Standardabweichung) σ_i jeder Messung ist bekannt. Wir messen den Datensatz $\mathbf{x} = (x_1, x_2, \dots, x_N)$, $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \dots, \sigma_N)$ und wollen den Mittelwert bestimmen:

Die zugrundeliegende Likelihood Funktion jeder Messung ist:

$$f_n(\zeta_n, \sigma_n, a) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{1}{2} \frac{(\zeta_n - a)^2}{\sigma_n^2}\right)$$

Damit ist die Likelihood Funktion:

$$L(\mathbf{x}, \boldsymbol{\sigma}, a) = \prod_{n=1}^N \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{1}{2} \frac{(x_n - a)^2}{\sigma_n^2}\right)$$

Und die Log-Likelihood Funktion:

$$l(\mathbf{x}, \boldsymbol{\sigma}, a) = \sum_{n=1}^N \ln\left(\frac{1}{\sqrt{2\pi}\sigma_n}\right) - \frac{1}{2} \sum_{n=1}^N \frac{(x_n - a)^2}{\sigma_n^2}$$

Der gewichtete Mittelwert

Wir betrachten nun ein häufiges Szenario

Wir messen mit einer uns bekannten und charakterisierten Messapparatur. Das heisst die Messunsicherheit (die Standardabweichung) σ_i jeder Messung ist bekannt. Wir messen den Datensatz $\mathbf{x} = (x_1, x_2, \dots, x_n)$ und wollen den Mittelwert bestimmen:

Die Log-Likelihood Funktion:

$$l(\mathbf{x}, \boldsymbol{\sigma}, a) = \sum_{n=1}^N \ln \left(\frac{1}{\sqrt{2\pi}\sigma_n} \right) - \frac{1}{2} \sum_{n=1}^N \frac{(x_n - a)^2}{\sigma_n^2}$$

Wir bestimmen das Maximum:

$$\begin{aligned} \frac{\partial l}{\partial a} \Big|_{\hat{a}} &= \frac{\partial}{\partial a} \Big|_{\hat{a}} \left(-\frac{1}{2} \sum_{n=1}^N \frac{(x_n - a)^2}{\sigma_n^2} \right) = 0 \\ \rightarrow \sum_{n=1}^N \frac{x_n - \hat{a}}{\sigma_n^2} &= \sum_{n=1}^N \frac{x_n}{\sigma_n^2} - \hat{a} \sum_{n=1}^N \frac{1}{\sigma_n^2} = 0 \\ \rightarrow \hat{a} &= \frac{\sum_{n=1}^N x_n / \sigma_n^2}{\sum_{n=1}^N 1 / \sigma_n^2} = \frac{\sum_{n=1}^N w_n x_n}{\sum_{n=1}^N w_n} \end{aligned}$$

Der gewichtete Mittelwert

Der gewichtete Mittelwert

Wir betrachten nun ein häufiges Szenario

Wir messen mit einer uns bekannten und charakterisierten Messapparatur. Das heisst die Messunsicherheit (die Standardabweichung) σ_i jeder Messung ist bekannt. Wir messen den Datensatz $\mathbf{x} = (x_1, x_2, \dots, x_n)$ und wollen den Mittelwert bestimmen:

Die Log-Likelihood Funktion:

$$l(\mathbf{x}, \boldsymbol{\sigma}, a) = \sum_{n=1}^N \ln \left(\frac{1}{\sqrt{2\pi}\sigma_n} \right) - \frac{1}{2} \sum_{n=1}^N \frac{(x_n - a)^2}{\sigma_n^2}$$

Wir bestimmen das Maximum:

$$\left. \frac{\partial l}{\partial a} \right|_{\hat{a}} = \left. \frac{\partial}{\partial a} \right|_{\hat{a}} \left(-\frac{1}{2} \sum_{n=1}^N \frac{(x_n - a)^2}{\sigma_n^2} \right) = 0$$

$$\rightarrow \sum_{n=1}^N \frac{x_n - \hat{a}}{\sigma_n^2} = \sum_{n=1}^N \frac{x_n}{\sigma_n^2} - \hat{a} \sum_{n=1}^N \frac{1}{\sigma_n^2} = 0$$

Hier definieren wir das Gewicht der einzelnen Messwerte als $w_n = 1/\sigma_n^2$

$$\rightarrow \hat{a} = \frac{\sum_{n=1}^N x_n / \sigma_n^2}{\sum_{n=1}^N 1 / \sigma_n^2} = \frac{\sum_{n=1}^N w_n x_n}{\sum_{n=1}^N w_n}$$

Der gewichtete Mittelwert

Die Likelihood Funktion - Zusammenfassung

Die Likelihood Funktion:

1. Ist eine nicht-normierte Wahrscheinlichkeitsverteilung
2. Ist nur ein anderer Name für eine Wahrscheinlichkeitsfunktion
3. Ist eine Funktion der Parameter der zugrunde liegenden Wahrscheinlichkeitsfunktion
4. Gibt die Wahrscheinlichkeit für die Parameter der zugrunde liegenden Wahrscheinlichkeitsfunktion

Welche der Aussagen sind richtig?

Die Methode der kleinsten Quadrate

Bisher haben wir die Likelihood Funktion maximiert um Parameter zu bestimmen die die Wahrscheinlichkeitsverteilung (PDF) eines Datensatzes zu beschreiben.

Insbesondere die Log-Likelihood Funktion ist einfach zu maximieren wenn die PDF eine Gaussverteilung ist.

In diesem Fall ist die Maximierung der Log-Likelihood Funktion äquivalent zur Minimierung der Abstandsquadrats der einzelnen Messpunkte zum Erwartungswert (also dem Modell). Dies ist die **Methode der kleinsten Quadrate**:

$$\left. \frac{\partial l}{\partial a} \right|_{\hat{a}} = \left. \frac{\partial}{\partial a} \right|_{\hat{a}} \left(-\frac{1}{2} \sum_{n=1}^N \frac{(x_n - a)^2}{\sigma_n^2} \right) = 0$$

Wir definieren: $S = \sum_{n=1}^N \frac{(x_n - a)^2}{\sigma_n^2}$ und damit schreiben wir:

$$\left. \frac{\partial S}{\partial a} \right|_{\hat{a}} = \left. \frac{\partial}{\partial a} \right|_{\hat{a}} \sum_{n=1}^N \frac{(x_n - a)^2}{\sigma_n^2} = 0$$

Bemerkung: Wenn die σ_n bekannt sind, dann ist dies äquivalent zur Minimierung von χ^2 . Dann ist $S = \chi^2$. Das ist jedoch oft nicht der Fall. Wir bleiben also zunächst bei S .

Lineares Fitten von Polynomen

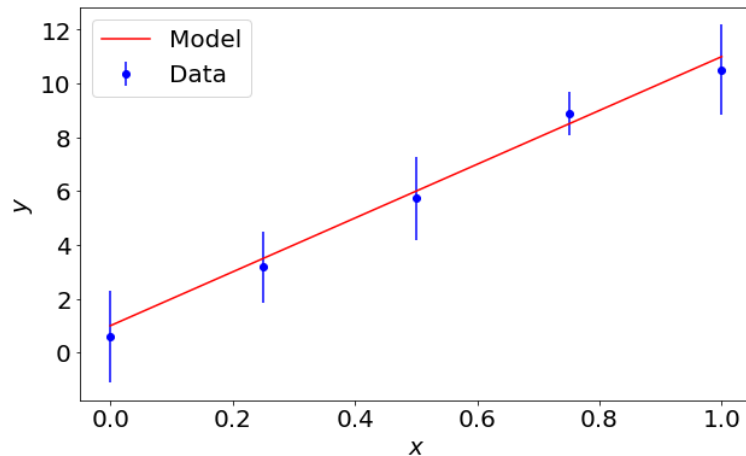
1) Fitten einer Ausgleichsgeraden

Wir haben die Werte y_1, y_2, \dots, y_N als Funktion der Kontrollparameters x_1, x_2, \dots, x_N gemessen. Die x_n haben keine Unsicherheit, während die Unsicherheiten der y_n durch σ_n gegeben sind.

Das bedeutet, dass die Wahrscheinlichkeit, dass der Wert y gemessen wird, gegeben ist durch:

$$f(y) \sim \exp\left(-\frac{(y - \mu)^2}{2\sigma^2}\right)$$

Wobei beide Werte y und σ von x abhängen. Ausserdem nehmen wir an dass $\mu = a_0 + a_1x$, also dass die Erwartungswerte von y linear von x abhängen.



Wir wollen die Werte von a_0 und a_1 finden die die Daten am besten wiedergeben.

Lineares Fitten von Polynomen

1) Fitten einer Ausgleichsgeraden

Wir haben die Werte y_1, y_2, \dots, y_N als Funktion der Kontrollparameters x_1, x_2, \dots, x_N gemessen. Die x_n haben keine Unsicherheit, während die Unsicherheiten der y_n durch σ_n gegeben sind.

Das bedeutet, dass die Wahrscheinlichkeit, dass der Wert y gemessen wird, gegeben ist durch:

$$f(y) \sim \exp\left(-\frac{(y - \mu)^2}{2\sigma^2}\right)$$

Wobei beide Werte y und σ von x abhängen. Ausserdem nehmen wir an dass $\mu = a_0 + a_1x$, also dass die Erwartungswerte von y linear von x abhängen.

Also:

$$f(y) \sim \exp\left(-\frac{(y - a_0 - a_1x_n)^2}{2\sigma_n^2}\right)$$

Ist die Wahrscheinlichkeitsverteilung der y_n .

Dann ist die Likelihood Funktion für jeden Datenpunkt:

$$L(x_n, y_n, \sigma_n, a_0, a_1) \sim \exp\left(-\frac{(y_n - a_0 - a_1x_n)^2}{2\sigma_n^2}\right)$$

Und die kombinierte Log-Likelihood Funktion ist dann:

$$l(\mathbf{x}, \mathbf{y}, \boldsymbol{\sigma}, a_0, a_1) = -\frac{1}{2} \sum_{n=1}^N \frac{(y_n - a_0 - a_1x_n)^2}{\sigma_n^2} + C$$

Mit einer Konstanten C .

Lineares Fitten von Polynomen

1) Fitten einer Ausgleichsgeraden

Wir haben die Werte y_1, y_2, \dots, y_N als Funktion der Kontrollparameters x_1, x_2, \dots, x_N gemessen. Die x_n haben keine Unsicherheit, während die Unsicherheiten der y_n durch σ_n gegeben sind.

Die kombinierte Log-Likelihood Funktion ist dann:

$$l(\mathbf{x}, \mathbf{y}, \boldsymbol{\sigma}, a_0, a_1) = -\frac{1}{2} \sum_{n=1}^N \frac{(y_n - a_0 - a_1 x_n)^2}{\sigma_n^2} + C$$

Mit einer Konstanten C . Wir benutzen die Gewichte $w_n = 1/\sigma_n^2$ und damit ist:

$$l(\mathbf{x}, \mathbf{y}, \mathbf{w}, a_0, a_1) = -\frac{1}{2} \sum_{n=1}^N w_n (y_n - a_0 - a_1 x_n)^2 + C$$

Wir können die besten Abschätzungen für a_0 und a_1 finden indem wir die partiellen Ableitungen von l nach a_0 und a_1 gleich Null setzen.

Da die Unsicherheiten einer Normalverteilung zu Grunde liegen, können wir auch die gewichtete Summe der Abstandsquadrate (der Residuen) minimieren:

$$S = \sum_{n=1}^N w_n (y_n - a_0 - a_1 x_n)^2$$

Lineares Fitten von Polynomen

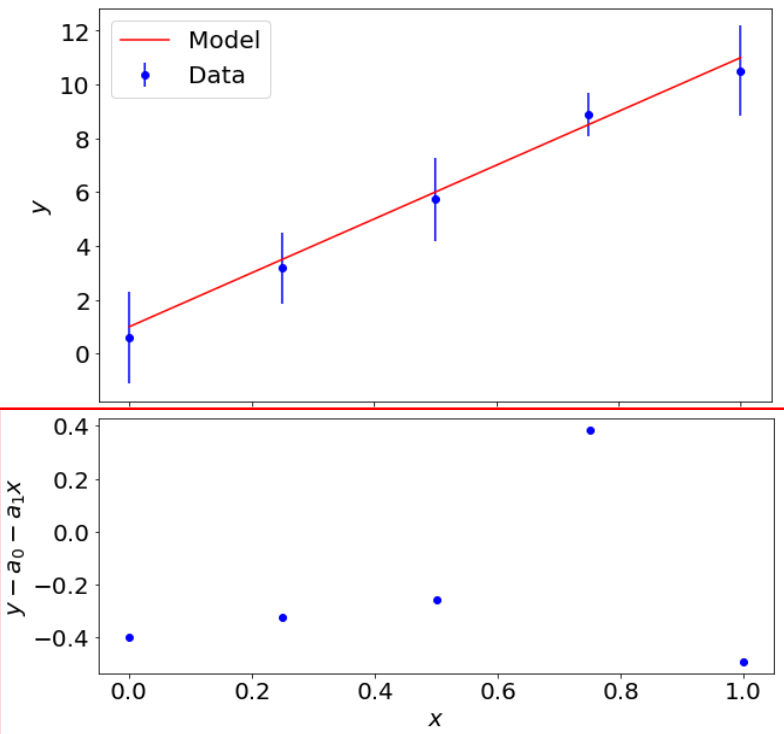
1) Fitten einer Ausgleichsgeraden

Wir haben die Werte y_1, y_2, \dots, y_N als Funktion der Kontrollparameters x_1, x_2, \dots, x_N gemessen. Die x_n haben keine Unsicherheit, während die Unsicherheiten der y_n durch σ_n gegeben sind.

Da die Unsicherheiten einer Normalverteilung zu Grunde liegen, können wir auch die gewichtete Summe der Abstandsquadrate (der Residuen) minimieren:

$$S = \sum_{n=1}^N w_n (y_n - a_0 - a_1 x_n)^2$$

Die Residuen sind die Abstände der gemessenen Datenpunkte vom Modell.



Lineares Fitten von Polynomen

1) Fitten einer Ausgleichsgeraden

Wir haben die Werte y_1, y_2, \dots, y_N als Funktion der Kontrollparameters x_1, x_2, \dots, x_N gemessen. Die x_n haben keine Unsicherheit, während die Unsicherheiten der y_n durch σ_n gegeben sind.

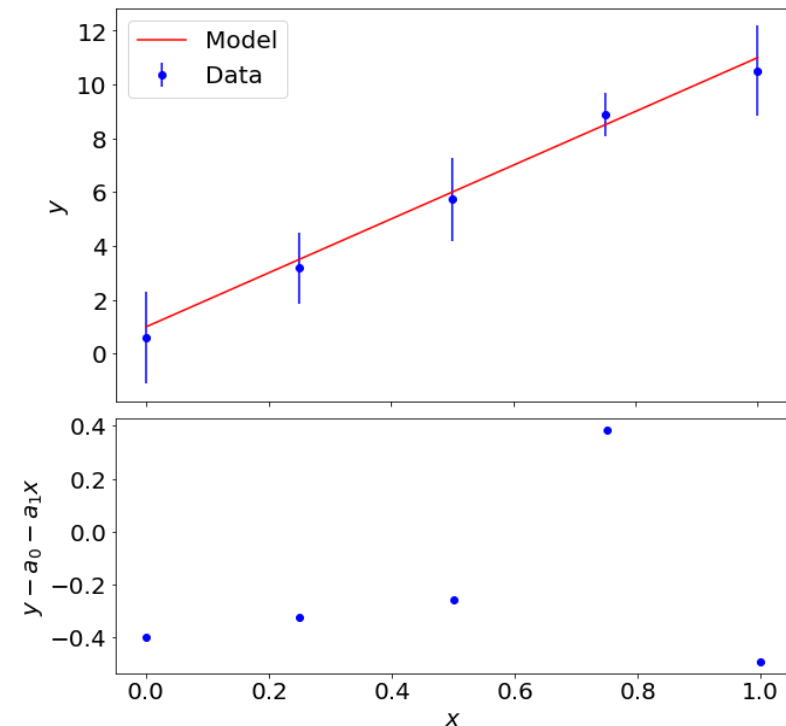
Da die Unsicherheiten einer Normalverteilung zu Grunde liegen, können wir auch die gewichtete Summe der Abstandsquadrate (der Residuen) minimieren:

$$S = \sum_{n=1}^N w_n (y_n - a_0 - a_1 x_n)^2$$

Die Residuen sind die Abstände der gemessenen Datenpunkte vom Modell.

Wir haben also zwei Gleichungen:

$$\frac{\partial S}{\partial a_0} = -2 \sum_{n=1}^N w_n (y_n - \hat{a}_0 - \hat{a}_1 x_n) = 0$$
$$\frac{\partial S}{\partial a_1} = -2 \sum_{n=1}^N w_n x_n (y_n - \hat{a}_0 - \hat{a}_1 x_n) = 0$$



Lineares Fitten von Polynomen

1) Fitten einer Ausgleichsgeraden

Wir haben die Werte y_1, y_2, \dots, y_N als Funktion der Kontrollparameters x_1, x_2, \dots, x_N gemessen. Die x_n haben keine Unsicherheit, während die Unsicherheiten der y_n durch σ_n gegeben sind.

Wir haben also zwei Gleichungen:

$$\frac{\partial S}{\partial a_0} = -2 \sum_{n=1}^N w_n (y_n - \hat{a}_0 - \hat{a}_1 x_n) = 0$$
$$\frac{\partial S}{\partial a_1} = -2 \sum_{n=1}^N w_n x_n (y_n - \hat{a}_0 - \hat{a}_1 x_n) = 0$$

Mit den Lösungen:

$$\hat{a}_0 = \frac{\sum w_n x_n^2 \sum w_n y_n - \sum w_n x_n \sum w_n x_n y_n}{\sum w_n \sum w_n x_n^2 - (\sum w_n x_n)^2}$$
$$\hat{a}_1 = \frac{\sum w_n \sum w_n x_n y_n - \sum w_n x_n \sum w_n y_n}{\sum w_n \sum w_n x_n^2 - (\sum w_n x_n)^2}$$

Lineares Fitten von Polynomen

1) Fitten einer Ausgleichsgeraden

Wir haben die Werte y_1, y_2, \dots, y_N als Funktion der Kontrollparameters x_1, x_2, \dots, x_N gemessen. Die x_n haben keine Unsicherheit, während die Unsicherheiten der y_n durch σ_n gegeben sind.

Wir haben also zwei Gleichungen:

$$\begin{aligned}\frac{\partial S}{\partial a_0} &= -2 \sum_{n=1}^N w_n (y_n - \hat{a}_0 - \hat{a}_1 x_n) = 0 \\ \frac{\partial S}{\partial a_1} &= -2 \sum_{n=1}^N w_n x_n (y_n - \hat{a}_0 - \hat{a}_1 x_n) = 0\end{aligned}$$

Mit den Lösungen:

$$\begin{aligned}\hat{a}_0 &= \frac{\sum w_n x_n^2 \sum w_n y_n - \sum w_n x_n \sum w_n x_n y_n}{\sum w_n \sum w_n x_n^2 - (\sum w_n x_n)^2} \\ \hat{a}_1 &= \frac{\sum w_n \sum w_n x_n y_n - \sum w_n x_n \sum w_n y_n}{\sum w_n \sum w_n x_n^2 - (\sum w_n x_n)^2}\end{aligned}$$

Wir können das auch in Matrixform schreiben:

$$\begin{pmatrix} \sum w_n & \sum w_n x_n \\ \sum w_n x_n & \sum w_n x_n^2 \end{pmatrix} \begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \end{pmatrix} = \begin{pmatrix} \sum w_n y_n \\ \sum w_n x_n y_n \end{pmatrix}$$

$$N \hat{\mathbf{a}} = \mathbf{Y}$$

Lineares Fitten von Polynomen

1) Fitten einer Ausgleichsgeraden

Wir haben die Werte y_1, y_2, \dots, y_N als Funktion der Kontrollparameters x_1, x_2, \dots, x_N gemessen. Die x_n haben keine Unsicherheit, während die Unsicherheiten der y_n durch σ_n gegeben sind.

In Matrixform (die sogenannte Normalform):

$$\begin{pmatrix} \sum w_n & \sum w_n x_n \\ \sum w_n x_n & \sum w_n x_n^2 \end{pmatrix} \begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \end{pmatrix} = \begin{pmatrix} \sum w_n y_n \\ \sum w_n x_n y_n \end{pmatrix}$$

$$\mathbf{N} \hat{\mathbf{a}} = \mathbf{Y}$$

Mit der Lösung:

$$\hat{\mathbf{a}} = \mathbf{N}^{-1} \mathbf{Y}$$

Lineares Fitten von Polynomen

2) Fitten eines Polynoms M -ten Grades

Dies kann nun sehr einfach auf Polynome beliebigen Grades erweitert werden:

$$y = a_0 + a_1x + a_2x^2 + \dots + a_Mx^M$$

$$\mathbf{N} \hat{\mathbf{a}} = \mathbf{Y}$$

Ist dann explizit:

$$\begin{pmatrix} \sum w_n & \sum w_n x_n & \dots & \sum w_n x_n^M \\ \sum w_n x_n & \sum w_n x_n^2 & \dots & \sum w_n x_n^{M+1} \\ \vdots & \vdots & \ddots & \vdots \\ \sum w_n x_n^M & \sum w_n x_n^{M+1} & \dots & \sum w_n x_n^{2M} \end{pmatrix} \begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \vdots \\ \hat{a}_M \end{pmatrix} = \begin{pmatrix} \sum w_n y_n \\ \sum w_n x_n y_n \\ \vdots \\ \sum w_n x_n^M y_n \end{pmatrix}$$

Und auch hier ist die Lösung:

$$\hat{\mathbf{a}} = \mathbf{N}^{-1} \mathbf{Y}$$

Siehe Beispiel Python Notebook:
„Linear Fitting - Vorlesung“

Lineares Fitten von Polynomen

2) Fitten eines Polynoms M -ter Ordnung

Dies kann nun sehr einfach auf Polynome beliebiger Ordnung erweitert werden:

$$y = a_0 + a_1x + a_2x^2 + \dots + a_Mx^M$$

$$N \hat{a} = Y$$

Typisches Beispiel: Kalibrationskurve

APPENDIX 10: Calibrations Resistance Thermometers

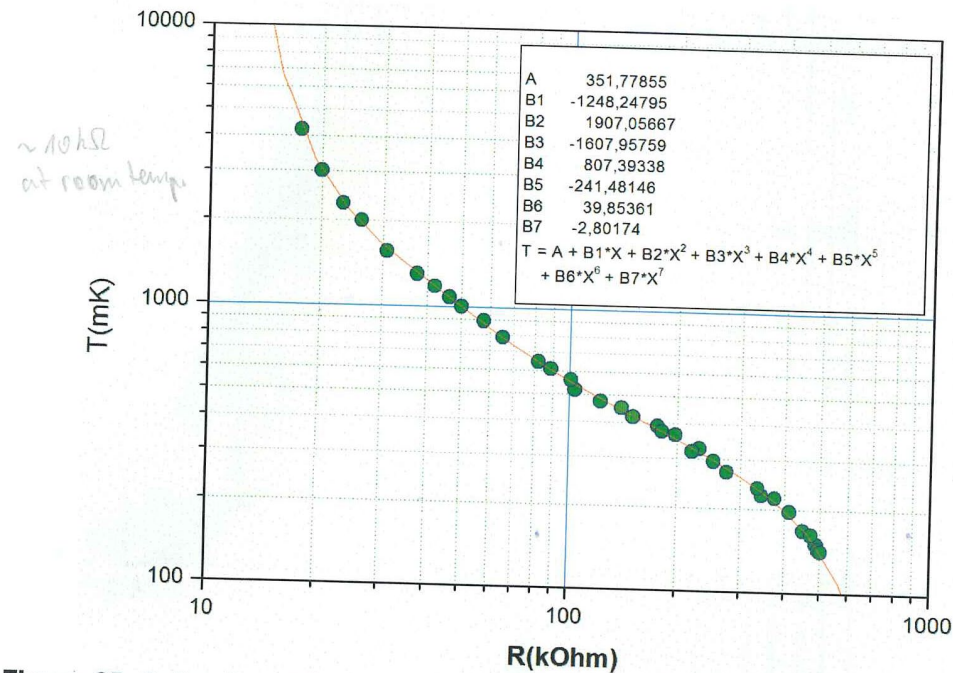


Figure 25 Calibration of the 10 kOhm RuO₂ thermometer used in the still and sorb pump. (~ 10 k - pot)

Lineares Fitten von Polynomen

Nun haben wir Abschätzung der Modellparameter \mathbf{a} bestimmt.

Z.B. Für einen linearen Zusammenhang zwischen x und y :
$$\hat{a}_0 = \frac{\sum w_n x_n^2 \sum w_n y_n - \sum w_n x_n \sum w_n x_n y_n}{\sum w_n \sum w_n x_n^2 - (\sum w_n x_n)^2}$$

Frage: Was wollen wir noch wissen?

$$\hat{a}_1 = \frac{\sum w_n \sum w_n x_n y_n - \sum w_n x_n \sum w_n y_n}{\sum w_n \sum w_n x_n^2 - (\sum w_n x_n)^2}$$

Lineares Fitten von Polynomen

3) Unsicherheit der Fitparameter

Auf diese Weise können wir die beste Abschätzung der Modellparameter \mathbf{a} bestimmen. Aber wir wollen auch wissen wie zuverlässig diese Werte sind; wir möchten deren **Varianz** und **Kovarianz** wissen.

Das Polynom berechnet mit den $\hat{\mathbf{a}}$ wird nicht durch alle Datenpunkte gehen. Also können wir schreiben:

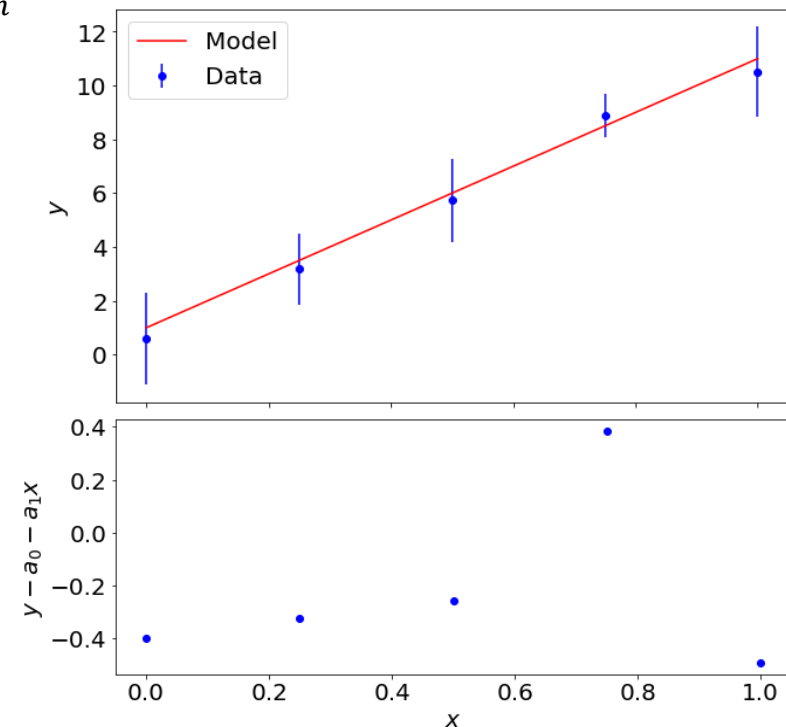
$$y_n = \hat{a}_0 + \hat{a}_1 x_n + \dots + \hat{a}_M x_n^M + \epsilon_n$$

Wobei ϵ_n der Abstand ist den der n -te Punkt von der gefitteten Funktion hat: Die **Residuen**

Wir nehmen an, dass die Unsicherheiten der y_n durch zufällige Fehler zustande kommen und normalverteilt sind. Dann sind die ϵ_n unkorreliert und ihr Erwartungswert ist Null:

$$\langle \epsilon_n \rangle = 0$$

$$\langle \epsilon_n, \epsilon_m \rangle = \sigma_n^2 \delta_{nm}$$



Lineares Fitten von Polynomen

3) Unsicherheit der Fitparameter

Auf diese Weise können wir die beste Abschätzung der Modellparameter \mathbf{a} bestimmen. Aber wir wollen auch wissen wie zuverlässig diese Werte sind; wir möchten deren **Varianz** und **Kovarianz** wissen.

Das Polynom berechnet mit den $\hat{\mathbf{a}}$ wird nicht durch alle Datenpunkte gehen. Also können wir schreiben:

$$y_n = \hat{a}_0 + \hat{a}_1 x_n + \dots + \hat{a}_M x_n^M + \epsilon_n$$

Wir definieren die Vektoren:

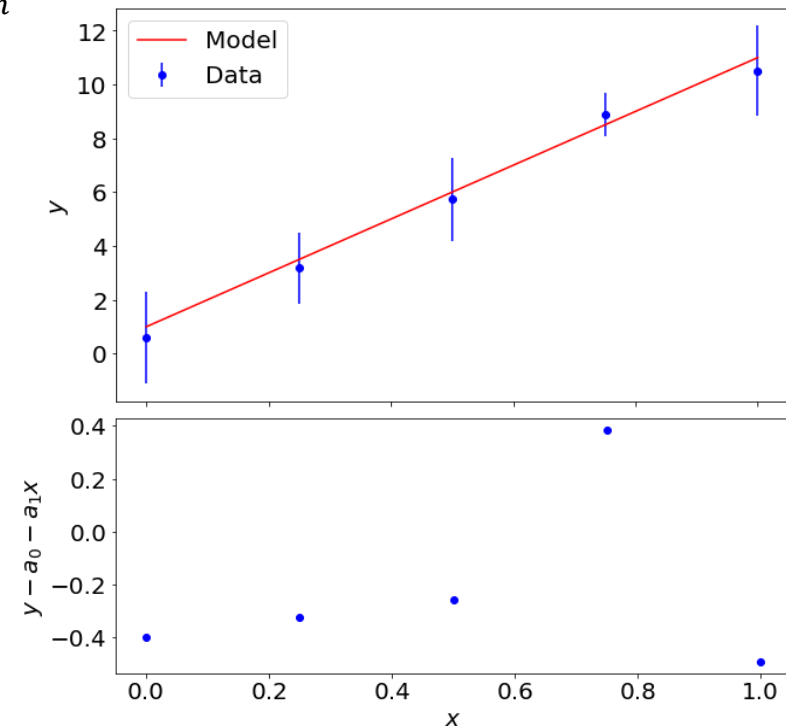
$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}, \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{pmatrix}, \mathbf{a} = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_N \end{pmatrix}$$

Und die Matrizen:

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 & \dots & x_1^M \\ 1 & x_2 & \dots & x_2^M \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_N & \dots & x_N^M \end{pmatrix} \text{ und } \mathbf{W} = \begin{pmatrix} w_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & w_N \end{pmatrix}$$

Damit wird obige Gleichung:

$$\mathbf{y} = \mathbf{X}\mathbf{a} + \boldsymbol{\epsilon}$$



Lineares Fitten von Polynomen

3) Unsicherheit der Fitparameter

Auf diese Weise können wir die beste Abschätzung der Modellparameter \mathbf{a} bestimmen. Aber wir wollen auch wissen wie zuverlässig diese Werte sind; wir möchten deren **Varianz** und **Kovarianz** wissen.

Das Polynom berechnet mit den $\hat{\mathbf{a}}$ wird nicht durch alle Datenpunkte gehen. Also können wir schreiben:

$$y_i = \hat{a}_0 + \hat{a}_1 x_n + \dots + \hat{a}_M x_n^M + \epsilon_n$$

In Matrixnotation:

$$\mathbf{y} = \mathbf{X}\mathbf{a} + \boldsymbol{\epsilon}$$

↖ Einzig verbleibende Zufallsvariable

Wir können wieder die Normalmatrix und den gewichteten Ergebnisvektor finden:

$$\mathbf{N} = \mathbf{X}^T \mathbf{W} \mathbf{X} \text{ und } \mathbf{Y} = \mathbf{X}^T \mathbf{W} \mathbf{y}$$

$\mathbf{N} = \mathbf{N}^T, \mathbf{W} = \mathbf{W}^T$, da beide Matrizen symmetrisch sind.

Der Ausdruck für S , die gewichtete Summe der Quadrate der Residuen ist dann:

$$S = (\mathbf{y} - \mathbf{X}\mathbf{a})^T \mathbf{W} (\mathbf{y} - \mathbf{X}\mathbf{a})$$

Und die Normalform:

$$(\mathbf{X}^T \mathbf{W} \mathbf{X}) \hat{\mathbf{a}} = (\mathbf{X}^T \mathbf{W} \mathbf{y})$$

Mit der Lösung:

$$\hat{\mathbf{a}} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{y}$$

Lineares Fitten von Polynomen

3) Unsicherheit der Fitparameter

Auf diese Weise können wir die beste Abschätzung der Modellparameter \mathbf{a} bestimmen. Aber wir wollen auch wissen wie zuverlässig diese Werte sind; wir möchten deren **Varianz** und **Kovarianz** wissen.

Die Unsicherheiten der Parameter $\hat{\mathbf{a}}$ sind gegeben durch deren Varianz:

$$\hat{\sigma}_n^2 = \langle (\hat{a}_n - a_n)^2 \rangle$$

Und Kovarianz:

$$\hat{\sigma}_{nm} = \langle (\hat{a}_n - a_m)(\hat{a}_m - a_n) \rangle$$

Relativ zu den (unbekannten) tatsächlichen werten \mathbf{a} .

Das können wir auch als Matrix schreiben (Die Kovarianzmatrix):

$$\mathbf{C} = \begin{pmatrix} \hat{\sigma}_0^2 & \cdots & \hat{\sigma}_{0M} \\ \vdots & \ddots & \vdots \\ \hat{\sigma}_{M0} & \cdots & \hat{\sigma}_M^2 \end{pmatrix} = \langle (\hat{\mathbf{a}} - \mathbf{a})(\hat{\mathbf{a}} - \mathbf{a})^T \rangle$$

Lineares Fitten von Polynomen

3) Unsicherheit der Fitparameter

Auf diese Weise können wir die beste Abschätzung der Modellparameter \mathbf{a} bestimmen. Aber wir wollen auch wissen wie zuverlässig diese Werte sind; wir möchten deren **Varianz** und **Kovarianz** wissen.

Die Unsicherheiten der Parameter $\hat{\mathbf{a}}$ sind gegeben durch die Kovarianz Matrix:

$$\mathbf{C} = \begin{pmatrix} \hat{\sigma}_0^2 & \cdots & \hat{\sigma}_{0M} \\ \vdots & \ddots & \vdots \\ \hat{\sigma}_{M0} & \cdots & \hat{\sigma}_M^2 \end{pmatrix} = \langle (\hat{\mathbf{a}} - \mathbf{a})(\hat{\mathbf{a}} - \mathbf{a})^T \rangle$$

Mit den Ausdrücken und Definitionen die wir eingeführt haben finden wir:

$$\mathbf{C} = \left\langle \left((\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \boldsymbol{\epsilon} \right) \left((\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \boldsymbol{\epsilon} \right)^T \right\rangle$$

Und mit $\langle \boldsymbol{\epsilon} \boldsymbol{\epsilon}^T \rangle = \mathbf{W}^{-1}$ wird das zu:

$$\mathbf{C} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} = \mathbf{N}^{-1}$$

Ein bemerkenswert einfaches und schönes Ergebnis!

Siehe Beispiel Python Notebook:
„Linear Fitting - Vorlesung“

Lineares Fitten von Polynomen

3) Unsicherheit der Fitparameter

Auf diese Weise können wir die beste Abschätzung der Modellparameter \mathbf{a} bestimmen. Aber wir wollen auch wissen wie zuverlässig diese Werte sind; wir möchten deren **Varianz** und **Kovarianz** wissen.

Die Unsicherheiten der Parameter $\hat{\mathbf{a}}$ sind gegeben durch die Kovarianz Matrix:

$$\mathbf{C} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} = \mathbf{N}^{-1}$$

Bemerkung:

Tatsächlich sind die σ_i oft nicht gut genug bekannt. Das heisst $\sigma_i \neq \langle \epsilon_i^2 \rangle$.

Das kann korrigiert werden, und man findet nach langer Rechnung:

$$\mathbf{C} = \frac{\hat{S}_{\min}}{N - M - 1} \mathbf{N}^{-1}$$

Wobei $\hat{S}_{\min} = \sum_{n=1}^N w_n (y_n - \hat{a}_0 - \hat{a}_1 x_n - \dots - \hat{a}_M x_n^M)^2$, der Minimale Wert von S ausgewertet für die Werte von $\hat{\mathbf{a}}$.

Dies ist etwas das z.B. SciPy Curve Fit per Default macht.

In allen Aufgaben, inklusive der Klausur, verlangen wird immer die unskalierte Kovarianzmatrix!
Ausser dies wird explizit verlangt.

Güte des Fits – χ^2 Test

Bisher haben wir uns nur dafür interessiert für welche Werte $\hat{\mathbf{a}} = (a_0, \dots, a_M)$, S sein Minimum erreicht:

$$S_{\min} = \sum_{n=1}^N w_n (y_n - f(x_n, \hat{\mathbf{a}}))^2$$

Aber, was bedeutet der Wert von S_{\min} ?

S ist die mit ihren Unsicherheiten normierte Summer des Abstandsquadrate, insofern folgt S einer χ^2 Verteilung mit den Freiheitsgraden:

$$k = N - M - 1$$

Wir erwarten, dass der Wert von S_{\min} z.B. mit 67% Wahrscheinlichkeit in einem Intervall $\langle \chi_k^2 \rangle \pm \sqrt{\text{var}(\chi_k^2)} = k \pm \sqrt{2k}$ liegt.

Typischerweise wird allerdings ein Konfidenzintervall von 95% angegeben:

$$0.025 = \int_0^{\chi_{0.025}^2} f_k(\chi^2) d\chi^2, \quad 0.975 = \int_{\chi_{0.975}^2}^{\infty} f_k(\chi^2) d\chi^2$$

So dass: - $S_{\min} < \chi_{0.025}^2$ \rightarrow
 - $S_{\min} > \chi_{0.975}^2$ \rightarrow

?

Zusammenfassung

- Wahrscheinlichkeitsverteilungen beschreiben die Wahrscheinlichkeiten von möglichen Ergebnissen von Zufallsexperimenten.
- Wir maximieren die (Log)Likelihood Funktion um die wahrscheinlichsten Werte für die Parameter der Wahrscheinlichkeitsverteilung zu finden.
- Die Likelihood Funktion ist nicht normiert und ist somit selbst keine Wahrscheinlichkeitsverteilung.
- Wir können Polynome m -ter Ordnung fitten indem wir die Normalmatrix N berechnen.
- Die Unsicherheit (Varianz und Kovarianz) der Fitparameter ist gegeben durch die Kovarianzmatrix $C = N^{-1}$
- Die Güte des Fits können wir mit Hilfe der χ^2 Funktion abschätzen.