

Messdatenauswertung und Messunsicherheit (MDA)

Dr. habil. Dorothee Hüser und Dr.-Ing. Gerd Ehret
Physikalisch-Technische Bundesanstalt
Modulverantwortlicher: Prof. Dr.-Ing. R. Tutsch
iprom, TU Braunschweig

6. November 2019

Inhaltsverzeichnis

1	Einführung: Messtechnik und Statistik	1
2	Lineare Regression (Ausgleichsrechnung)	3
2.1	Die Idee der linearen Regression	4
2.2	Lineare Regression als Beispiel zur Methode der kleinsten Abweichungsquadrate . . .	6
2.3	Beispiel zur linearen Regression	9
2.4	Varianz der Regressionsparameter	11
2.5	Matrixformalismus zur allgemeinen Lösung der linearen Regression	12
2.5.1	Matrixformalismus anhand der Regressionsgeraden	12
2.5.2	Matrixansatz als allgemeiner Ansatz für linearen Regression	17
2.6	Das Bestimmtheitsmaß, empirischer Korrelationskoeffizient	20
2.7	Beispiele	22
2.7.1	Fit einer Geraden	22
2.7.2	Fit einer kubischen Funktion (Polynom 3. Ordnung)	27
2.8	Aufgaben zum Selbststudium	30
2.8.1	1. Aufgabe zur Linearen Regression	30
3	Optimierung nichtlinearer Modelle	33
3.1	Konzepte von Optimierungsverfahren für nichtlineare Modelle	34
3.2	Gradientenverfahren für nichtlineare Zielfunktionale	37
3.3	Levenberg-Marquardt-Verfahren	40
3.4	Matlabfunktionen zur Berechnung nichtlinearer Fits	50

3.5	Gradientenverfahren mit unterschiedlichen Varianzen	52
3.6	Robuste Schätzverfahren	54
4	Konzepte der Statistik für die Messdatenanalyse	59
4.1	Konzepte zum Lösen inverser Probleme	60
4.2	Konzept der bayesischen Verfahren zum Lösen inverser Probleme	66
5	Wahrscheinlichkeiten und Hypothesentests	73
5.1	Konzept der statistischen Erwartung	74
5.2	Kolmogorow-Smirnow-Test	75
5.3	Wahrscheinlichkeitsdichtefunktionen und ihre Parameter	78
5.4	Vertrauensintervall und t-Verteilung	79
5.5	t-Test - Mittelwerttest	89
5.6	Chi ² -Verteilung und Varianztest	90
5.7	Anwendung von Hypothesentests	94
5.7.1	Entscheidung bei Vergleich zweier Stichproben	95
5.7.2	Konformitätswahrscheinlichkeiten	96
5.8	Übungsaufgaben zum Selbststudium	98
5.9	Quantiltabellen	100
5.9.1	Quantile der Student-t Verteilung	100
5.9.2	Quantile der χ^2 -Verteilung	101
6	Auswertung von Mess- und Ringvergleichen	103
6.1	Auswertung von Messvergleichen mit Referenzlabor	104
6.2	Auswertung von Ringvergleichen ohne Referenzlabor	106
6.2.1	Vorgehensweise	106
6.2.2	Beispiel	109
6.2.3	Identifikation von Ausreißern und Konsistenzcheck	110
6.2.4	Paule-Mandel Metjode zur Anpassung von Gewichtungsfaktoren	112

7	Messunsicherheitsfortpflanzung bei linearen Modelle	121
7.1	Konzepte der Messunsicherheitsfortpflanzung (M.U.F.)	122
7.2	M.U.F. für univariate, explizite indirekte Größe, linear von direkten Größen abhängig	127
7.3	M.U.F. für univariate, explizite indirekte Größe, Abhängigkeit von direkten Größen linearisierbar	130
7.4	Überdeckungsintervall für indirekte Messgrößen	132
7.5	Vergleich der klassischen mit der bayesischen Methode	136
7.6	Aufgabe zum Selbststudium	141
7.7	Anhang: Erwartungswerte zur χ^2 -Verteilung	143
7.8	Anhang: Beispiel indirekte Größe aus Produkt zweier direkter Größen	144
8	Numerische Werkzeuge für Messunsicherheitsberechnungen	147
8.1	Numerische Integrationsverfahren	148
8.2	Zufallszahlen und Wahrscheinlichkeitsverteilungen	151
8.3	Monte-Carlo-Verfahren gemäß GUM-supplement 1 JCGM 101	152
9	Lösungen zu den Aufgaben	155
9.1	Lösungen zu den Aufgaben aus Vorl 2	156
9.1.1	Lösung zur 1. Aufgabe: Lineare Regression	156
9.2	Hilfe zu den Übungsaufgaben 5.8 zum Selbststudium	162
9.3	Lösung zur Aufgabe 7.6	165

Kapitel 1

Einführung: Messtechnik und Statistik

Kapitel 2

Lineare Regression (Ausgleichsrechnung)

2.1 Die Idee der linearen Regression

Ziel der Regression ist es, den funktionalen Zusammenhang $Y = f(X_1, X_2, \dots, X_N)$ zwischen den Eingangs-Messgrößen X_i (Regressor) und der Ausgangsgröße Y (Regressanden) möglichst gut zu charakterisieren.

Der allgemeinste Fall der Regressionsrechnung ist der mit sowohl mehreren Regressoren als auch mehreren Regressanden. Regressionsrechnung mit einem Regressanden, wird **univariate Regression**, oder oft einfach nur lineare Regression genannt. Die Darstellung der Modellgleichung kann dabei wie folgt aussehen

$$Y_j = f(X_{1,j}, \dots, X_{i,j}, \dots, X_{N,j}) + \varepsilon_j \quad \text{mit} \quad j = 1, 2, \dots, J \quad (2.1)$$

wobei Y_j eine Beobachtung zur Zufallsgröße, dem Regressanden Y , darstellt bei Vorgabe (Einstellung an einem Gerät wie im Beispiel von Kapitel 1 der Präzisionsstromquelle) $X_{i,j}$ des Regressors X_i . Wenn es mehrere abhängige Größen, also mehrere Regressanden gibt, spricht man von **multivariater Regression**. Die Regressoren müssen voneinander **linear unabhängig** sein. Sie können aber auch eine Potenzbeziehung zueinander haben, also quadratisch, kubisch etc. Wir werden für die Methode der Regressionsrechnung zur Schätzung von Polynomkoeffizienten eine Eingangsgröße X und eine Ausgangsgröße Y betrachten, die über ein Potenzgesetz

$$Y = f(X) \propto X^k \quad \text{mit} \quad k \in \mathbb{N}_0 \quad (2.2)$$

also mit k als natürliche Zahl einschließlich der Null, verknüpft sind.

Bei der Regressionsrechnung wird vorausgesetzt, dass der Regressor X eine Größe ist, die voreingestellt ist und keiner Streuung unterliegt, und nur der Regressand Y streut und deshalb als Zufallsgröße betrachtet wird. Die Streuung kann verschiedene Ursachen haben:

- Die indirekten Messgrößen, hier die Größen, die durch die Modellparameter beschrieben werden, streuen aufgrund von nicht im Modell enthaltenden Einflüssen. Das bedeutet, dass das Modell eine Approximation des physikalischen Sachverhalts ist.
- Der Messvorgang des Regressanden führt zu Streuungen.

Die Modellparameter sind also wie die Regressanden als Zufallsvariablen zu betrachten. Die Regression soll nun Schätzwerte für die Modellparameter liefern. Das Modell wird über einen funktionalen Zusammenhang zwischen Regressor und Regressand

$$Y = f(X) \quad (2.3)$$

beschrieben.

Um eine Schätzung für die Parameter des funktionalen Zusammenhangs $Y = f(X)$ zu erhalten, wird die Eingangsgröße in einem bestimmten Bereich variiert. Man misst J Messwertpaare (X_j, Y_j) mit $j = 1, \dots, J$, die sich als Punktwolke in einem Diagramm darstellen lassen. Auf Grund der Streuungen werden die Wertepaare (X_j, Y_j) die Gleichung $Y_j = f(X_j)$ nicht exakt erfüllen, sondern etwas

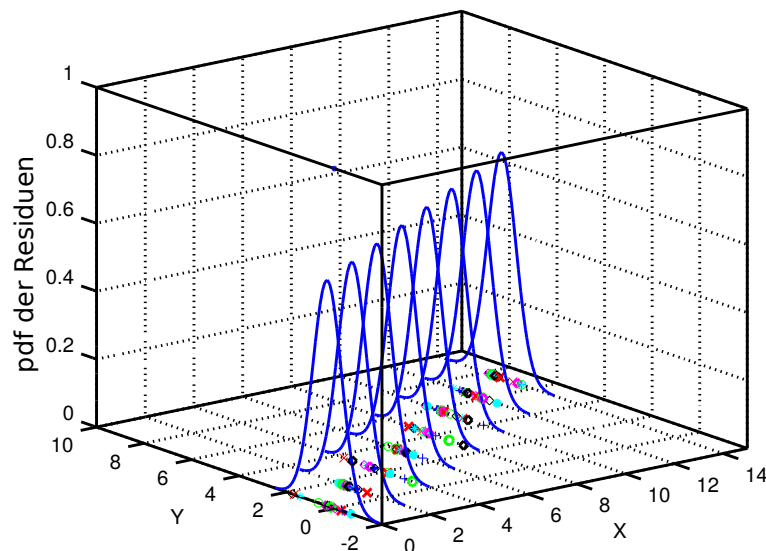


Abbildung 2.1: Lineare Regression für unabhängig und identisch verteilte Residuen.

abweichen ε_j , d. h.

$$Y_j = f(X_j) + \varepsilon_j \quad \text{mit} \quad j = 1, 2, \dots, J \quad (2.4)$$

Wir stellen hier zusammen welche Typen von Größen bei der Regressionsrechnung behandelt werden:

- Die unabhängigen Größen X werden vorgegeben. Sie haben genau bekannte Werte (engl. *exactly known values*) und sind somit keine Zufallsgrößen. Sie sind unabhängig und heißen **Regressoren**.
- Die abhängigen Größen Y , d.h. die **Regressanden** streuen und ihre Residuen sind normalverteilt mit Erwartungswert $E(\varepsilon) = 0$ und Varianz $\text{Var}(\varepsilon) = \sigma^2$.

Die Residuen sind **unabhängig und identisch verteilt**, kurz u.i.v.

$$\varepsilon \stackrel{\text{u.i.v.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon). \quad (2.5)$$

Die Regressanden werden auf Grund ihrer Streuung als Zufallsgrößen betrachtet. Zu jedem beobachteten Wert des Regressanden wird jeweils ein Regressorwert zugeordnet (engl. *observed data corresponding to the known values*). Sie sind abhängig vom Regressor und werden deshalb auch *dependent variable* oder *response variable* genannt.

- Modellparameter, die unbekannt sind (*unknown parameters*), sind Zufallsgrößen. Sie werden auch Regressionsparameter (*regression parameters*) genannt.
- Ferner gibt es zusätzliche unbekannte Parameter, und zwar die Varianzen der beteiligten Zufallsgrößen, *unknown additional parameters*.

Abb. 2.1 veranschaulicht, dass die Residuen normalverteilt sind, wobei die vertikale Achse die Wahrscheinlichkeitsdichte der Residuen darstellt. Der englische Begriff für Wahrscheinlichkeitsdichteverteilung ist *probability density function*, kurz pdf. Die unterschiedlichen Farben und Symbole sollen unterschiedliche Durchläufe gleicher Messvorgänge darstellen.

2.2 Lineare Regression als Beispiel zur Methode der kleinsten Abweichungsquadrate

Man bestimmt die Parameter der Regressionsfunktion $Y = f(X)$ nun so, dass die Summe der Quadrate der Abweichungen der Datenpunkte vom Modell (**Residuen**) ε_j möglichst klein wird (Methode der kleinsten Abweichungsquadrate), d.h. die **Kostenfunktion** Q muss minimal werden:

$$Q := \sum_{j=1}^J \varepsilon_j^2 = \sum_{j=1}^J (Y_j - f(X_j))^2 \rightarrow \min \quad (2.6)$$

Die Kostenfunktion ist ein Maß für die Abweichung der Messgrößen vom Modell in Abhängigkeit von der Wahl der Werte für die Modellparameter. Für gegebene Werte für die Modellparameter liefert die Kostenfunktion ein Maß für die Qualität der Schätzung und wird **Qualitätsmaß** genannt.

Wie in Kapitel 1 dargelegt, basiert dies auf dem Maximieren der Likelihood, der normalverteilten bzw. gaußverteilten Wahrscheinlichkeitsdichte der Residuen. Da diesem Verfahren die Annahme als Voraussetzung zugrunde liegt, dass die Likelihood gaußverteilt ist, spricht man in der ingenieurmäßigen Anwendung auch vom Ausgleichsverfahren nach Gauß.

In die Modellgleichung (2.3) setzen wir mehrere Potenzgesetze (2.2) ein, um folgenden Polynomansatz aufzustellen

$$Y = f(X) \equiv p_m = \theta_m X^m + \theta_{m-1} X^{m-1} + \dots + \theta_1 X + \theta_0 \quad (2.7)$$

m bezeichnet den Grad des Polynoms. Die Anzahl der Regressionsparameter bezeichnen wir mit $M = m + 1$. Unter Berücksichtigung der Streuung der Beobachtungen Y_j des Regressanden ist dies wie in Gl. (2.4)

$$Y_j = f(X_j) \equiv p_m = \theta_m X_j^m + \theta_{m-1} X_j^{m-1} + \dots + \theta_1 X_j + \theta_0 + \varepsilon_j \quad (2.8)$$

Für die Schätzung der Polynomkoeffizienten setzen wir den Polynomansatz in die Minimumbedingung Gl. (2.6) ein:

$$Q = Q(\theta_m, \theta_{m-1}, \dots, \theta_1, \theta_0) = \sum_{j=1}^J (Y_j - p_m(X_j))^2 \rightarrow \min \quad (2.9)$$

Die Schätzwerte der Koeffizienten $\hat{\theta}_k$, für die die Kostenfunktion Q ein Minimum annimmt, sind die beste Schätzung oder das Optimum der Parameter θ_k , weshalb im englischen Sprachgebrauch auch der Begriff *best fit* Verwendung findet. Die optimalen Schätzwerte zu den Polynomkoeffizienten $\hat{\theta}_k$ ergeben sich durch Lösen des linearen Gleichungssystems, das sich durch Null setzen der partiellen Ableitungen von Q nach den Modellparametern ergibt:

$$\left. \frac{\partial Q}{\partial \theta_k} \right|_{\theta_k = \hat{\theta}_k} = 0, \quad k = 0, 1, \dots, m \quad (2.10)$$

Die resultierende Modellgleichung ist dann die mit den geschätzten Regressionsparametern (hier den

Polynomkoeffizienten) $\hat{\theta}_k$:

$$Y = f(X) \equiv p_m = \hat{\theta}_m X^m + \hat{\theta}_{m-1} X^{m-1} + \dots + \hat{\theta}_1 X + \hat{\theta}_0 \quad (2.11)$$

Ein Schätzwert für die Varianz der Residuen, d.h. $\text{Var}(\varepsilon)$, ergibt sich durch die empirische Varianz:

$$\text{Var}(\varepsilon(\hat{\theta}_m, \hat{\theta}_{m-1}, \dots, \hat{\theta}_1, \hat{\theta}_0)) = s^2 = \frac{Q(\hat{\theta}_m, \hat{\theta}_{m-1}, \dots, \hat{\theta}_1, \hat{\theta}_0)}{J - 1 - m} \quad (2.12)$$

Wir betrachten zunächst den einfacheren Fall der **linearen Regression**, bei dem eine empirische Regressionsgerade gesucht wird, d. h. es gibt nur zwei Modellparameter θ_0 und θ_1 :

$$Y = \theta_0 + \theta_1 \cdot X \quad (2.13)$$

Die beste Schätzung für die Parameter θ_0 und θ_1 findet man durch Minimierung gemäß Gl. (2.6), d. h.

$$Q(\theta_0, \theta_1) = \sum_{j=1}^J (Y_j - (\theta_0 + \theta_1 \cdot X_j))^2 \rightarrow \min \quad (2.14)$$

Für jedes vorgegebene Messwertpaar-Ensemble (X_j, Y_j) , $j = 1, 2, \dots, J$ existiert eine eindeutige Lösung der Gl. (2.14).

Zum Auffinden des Minimums der Gl.(2.14), bildet man die partiellen Ableitungen und setzt diese gleich Null:

$$\begin{aligned} \left. \frac{\partial Q(\theta_0, \theta_1)}{\partial \theta_0} \right|_{\theta_0=\hat{\theta}_0, \theta_1=\hat{\theta}_1} &= 0 \\ \left. \frac{\partial Q(\theta_0, \theta_1)}{\partial \theta_1} \right|_{\theta_0=\hat{\theta}_0, \theta_1=\hat{\theta}_1} &= 0 \end{aligned} \quad (2.15)$$

Für das Vorliegen eines Minimums (nicht Maximum oder Sattelpunktes) müssen noch die Bedingungen

$$\left. \frac{\partial^2 Q(\theta_0, \theta_1)}{\partial \theta_0^2} \right|_{\theta_0=\hat{\theta}_0, \theta_1=\hat{\theta}_1} > 0 \quad \text{und} \quad \left. \frac{\partial^2 Q(\theta_0, \theta_1)}{\partial \theta_1^2} \right|_{\theta_0=\hat{\theta}_0, \theta_1=\hat{\theta}_1} > 0 \quad (2.16)$$

erfüllt sein.

Daraus folgt:

$$\begin{aligned} \sum_{j=1}^J Y_j - J \hat{\theta}_0 - \hat{\theta}_1 \sum_{j=1}^J X_j &= 0 \\ \sum_{j=1}^J Y_j X_j - \hat{\theta}_0 \sum_{j=1}^J X_j - \hat{\theta}_1 \sum_{j=1}^J X_j^2 &= 0 \end{aligned}$$

Mit $\sum_{j=1}^J X_j = J \bar{X}$ und $\sum_{j=1}^J Y_j = J \bar{Y}$ folgt:

$$J \bar{Y} - J \hat{\theta}_0 - J \cdot \hat{\theta}_1 \bar{X} = 0 \quad \text{bzw.} \quad \hat{\theta}_0 = \bar{Y} - \hat{\theta}_1 \bar{X}$$

$$\sum_{j=1}^J X_j Y_j - J \hat{\theta}_0 \bar{X} - \hat{\theta}_1 \sum_{j=1}^J X_j^2 = 0$$

Diese beiden Gleichungen ineinander eingesetzt liefert:

$$\left(\sum_{j=1}^J X_j Y_j \right) - J(\bar{Y} - \hat{\theta}_1 \bar{X}) \bar{X} - \hat{\theta}_1 \sum_{j=1}^J X_j^2 = 0$$

d.h.

$$\sum_{j=1}^J X_j Y_j - J \bar{Y} \bar{X} + \hat{\theta}_1 \left(J \bar{X}^2 - \sum_{j=1}^J X_j^2 \right) = 0$$

und mit $(\sum X_j^2) - J \bar{X}^2 = (\sum X_j^2) - 2 J \bar{X}^2 + J \bar{X}^2 = \sum (X_j^2 - 2 X_j \bar{X} + \bar{X}^2) = \sum (X_j - \bar{X})^2$

$$\hat{\theta}_0 = \bar{Y} - \hat{\theta}_1 \bar{X} \quad \text{und} \quad \hat{\theta}_1 = \frac{\sum_{j=1}^J (X_j Y_j - \bar{X} \bar{Y})}{\sum_{j=1}^J (X_j - \bar{X})^2}$$

Durch Verwendung der empirischen Varianz

$$s_X^2 = \frac{1}{J-1} \sum_{j=1}^J (X_j - \bar{X})^2$$

und der empirischen Kovarianz:

$$s_{XY} = \frac{1}{J-1} \sum_{j=1}^J (X_j - \bar{X})(Y_j - \bar{Y})$$

ergibt sich für die gesuchten Regressionsparameter $\hat{\theta}_0$ und $\hat{\theta}_1$:

$$\hat{\theta}_0 = \bar{Y} - \hat{\theta}_1 \bar{X} \quad \text{und} \quad \hat{\theta}_1 = \frac{s_{XY}}{s_X^2} \quad (2.17)$$

An dieser Gleichung sieht man, dass der Punkt (\bar{X}, \bar{Y}) , den man auch als Schwerpunkt bezeichnet, stets auf der Regressionsgeraden liegt. Hier wurde angenommen, dass X als unabhängige und Y als abhängige Zufallsvariable betrachtet wird. Oft ist es jedoch inhaltlich nicht klar, welche der beiden Zufallsvariablen die abhängige ist. In solch einem Fall kann man zusätzlich die Regression von X auf Y durchführen. Trägt man beide Regressionsgeraden in das gleiche Koordinatensystem ein, so schneiden sich diese im Schwerpunkt (\bar{X}, \bar{Y}) . Für den Schätzwert der Varianz der Residuen ergibt sich, siehe Gl.(2.12)

$$\hat{\sigma}_\epsilon^2 = s^2(\hat{\theta}_0, \hat{\theta}_1) = \frac{Q(\hat{\theta}_0, \hat{\theta}_1)}{J-1-m} = \frac{Q(\hat{\theta}_0, \hat{\theta}_1)}{J-2} \quad (2.18)$$

2.3 Beispiel zur linearen Regression

Beispiel: Vergleich der Modellierung mit Gerade und mit Polynom 6. Grades

Gegeben sei ein Ensemble von Werten (X_j, Y_j) , $j = 1, 2, \dots, J$ einer Eingangsgröße (Regressor) X und einer Zufallsgrößen, der abhängigen Ausgangsgröße (Regressand) Y . Beispielhaft ist in Abb. 2.2 eine entsprechende Punktwolke dargestellt: Bevor man mit der Regressionsrechnung anfängt, muss man

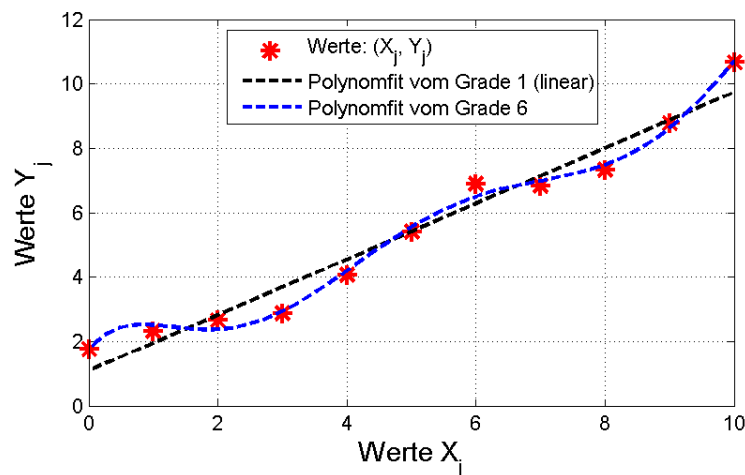


Abbildung 2.2: Beispiel für mögliche Fits, linearer Fit und Polynomfit vom Grade 6

sich im vorhinein sehr sorgfältig überlegen, in welchem physikalischen Zusammenhang die Größen stehen, also welcher Ansatz für das Messsystem als Modellgleichung sinnvoll ist. In einigen Fällen ist dies auch eine Frage, welche Näherung für welchen Messbereich ausreicht.

Beispiel: Bestimmung des Widerstandes R durch lineare Regression

Wir erinnern uns wieder an das Beispiel aus der 1. Vorlesung (siehe z. B. dort Abb. 1) und bestimmen den Ohmschen Widerstand R sowie eine Offsetspannung U_0 bei gegebenen Werten einer Präzisionsstromquelle (Regressor, genau bekannt, d.h. keine Zufallsgröße) und beobachteten Werten eines Voltmeters (Regressand, Zufallsgröße). Der Ohmsche Widerstand R und die Offsetspannung U_0 sind die zu bestimmenden Modellparameter (Zufallsgrößen)

I in mA	4.0	6.0	8.0	10.0	12.0	14.0	16.0	18.0	20.0
U in mV	62.5	51.5	96.0	140.2	138.9	195.1	225.8	207.8	223.7

Die Modellgleichung lautet mit den beiden Modellparametern $\theta_0 := U_0$ und $\theta_1 := R$

$$U = U_0 + R \cdot I \quad (2.19)$$

Die beiden Modellparameter können wir mit den beiden Gleichungen (siehe Gl. (2.17)) bestimmt werden.

$$\hat{U}_0 = \bar{U} - \hat{R} \bar{I} \quad \text{und} \quad \hat{R} = \frac{s_{IU}}{s_I^2} \quad (2.20)$$

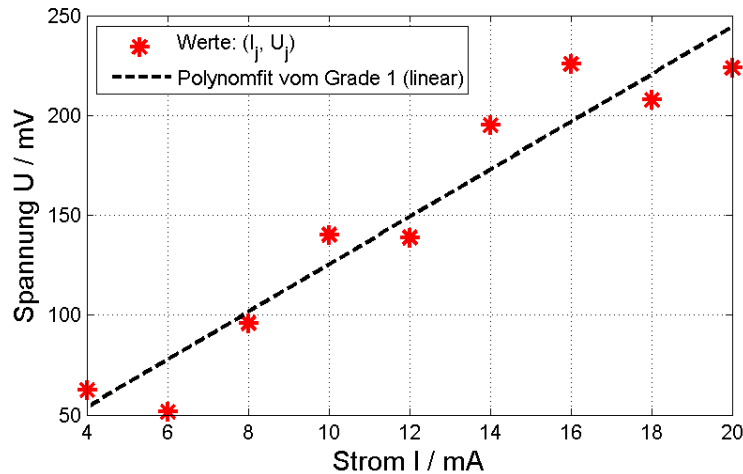


Abbildung 2.3: Beispiel: Bestimmung des Ohmschen Widerstand und einer Offset-Spannung durch lineare Regression

Zunächst bestimmen wir die beiden Mittelwerte:

$$\bar{I} := \frac{1}{J} \sum_{j=1}^J I_j = 12 \text{ mA} \quad \bar{U} := \frac{1}{J} \sum_{j=1}^J U_j = 149.0556 \text{ mV} \quad (2.21)$$

Die empirische Kovarianz mit den $J = 9$ Stichproben ergibt sich zu:

$$s_{IU} = \frac{1}{J-1} \sum_{j=1}^J (I_j - \bar{I})(U_j - \bar{U}) = 357.0500 \text{ mA} \cdot \text{mV}$$

Die Varianz ergibt sich zu:

$$s_I^2 = \frac{1}{J-1} \sum_{j=1}^J (I_j - \bar{I})^2 = 30 \text{ mA}^2$$

Somit erhalten wir als Schätzwert für R

$$\hat{R} = \frac{s_{IU}}{s_I^2} = \frac{357.050 \text{ mA} \cdot \text{mV}}{30 \text{ mA}^2} = 11.9017 \frac{\text{mV}}{\text{mA}}$$

mit $\frac{\text{mV}}{\text{mA}} = \frac{\text{V}}{\text{A}} = \Omega$ also $\hat{R} = 11.9017 \Omega$ und als Schätzwert für U_0 :

$$\hat{U}_0 = \bar{U} - \hat{R}\bar{I} = 149.0556 \text{ mV} - 11.9017 \Omega \cdot 12 \text{ mA} = 6.2356 \text{ mV}$$

Das Ergebnis für die Regressionsgerade lautet somit

$$U = 6.2356 \text{ mV} + (11.9017 \Omega) \cdot I \quad (2.22)$$

und ist in Abb. 2.3 dargestellt.

2.4 Varianz der Regressionsparameter

Es können aus den Messreihen die Varianzen der geschätzten Regressionsparameter ermittelt werden. Das Verfahren zeigen wir hier anhand der Regressionsgeraden mit den geschätzten Parametern Achsenabschnitt $\hat{\theta}_0$ und Steigung $\hat{\theta}_1$. Die Varianz der Schätzer liefert ein Maß für die statistische Sicherheit der Schätzung der Parameter θ_0 und θ_1 . Für die Varianz des Regressionsparameters Steigung $\hat{\theta}_1$ gilt:

$$\sigma_{\hat{\theta}_1}^2 = \frac{\hat{\sigma}_\varepsilon^2}{s_X^2 \cdot (J-1)} \quad (2.23)$$

Für die Varianz des Regressionsparameters Achsenabschnitt $\hat{\theta}_0$ gilt:

$$\sigma_{\hat{\theta}_0}^2 = \hat{\sigma}_\varepsilon^2 \cdot \left(\frac{1}{J} + \frac{\bar{X}^2}{(J-1)s_X^2} \right)^2 \quad (2.24)$$

Die Varianz des Achsenabschnittes wird weiter unten in der Gl.(2.72) hergeleitet. Dort ergibt sich

$$\hat{\sigma}_{\hat{\theta}_0}^2 = \hat{\sigma}_\varepsilon^2 \sum (X_j^2) / (J \sum (X_j^2) - (\sum (X_j))^2) \quad (2.25)$$

Wir bezeichnen nun das $\hat{\sigma}_\varepsilon^2$ hier mit s_ε^2 und erhalten:

$$\hat{\sigma}_{\hat{\theta}_0}^2 = s_\varepsilon^2 \sum (X_j^2) / (J \sum (X_j^2) - (\sum (X_j))^2) \quad (2.26)$$

$$\hat{\sigma}_{\hat{\theta}_0}^2 = s_\varepsilon^2 \frac{1}{J} \sum (X_j^2) / (\sum (X_j^2) - \frac{1}{J} (\sum (X_j))^2) \quad (2.27)$$

Mit der folgenden Null-Identität

$$0 = -\frac{1}{J} (\sum (X_j))^2 + \frac{1}{J} (\sum (X_j))^2 \quad (2.28)$$

gilt

$$\begin{aligned} & (1/J) \sum (X_j^2) / (\sum (X_j^2) - \frac{1}{J} (\sum (X_j))^2) = \\ & (1/J) [\sum (X_j^2) - (1/J) (\sum (X_j))^2 + \frac{1}{J} (\sum (X_j))^2] / (\sum (X_j^2) - (1/J) (\sum (X_j))^2) \end{aligned}$$

das heisst

$$\begin{aligned} & \frac{1}{J} \sum (X_j^2) / (\sum (X_j^2) - \frac{1}{J} (\sum (X_j))^2) = \\ & (1/J) [1 + \frac{1}{J} (\sum (X_j))^2 / (\sum (X_j^2) - \frac{1}{J} (\sum (X_j))^2)] \end{aligned}$$

und mit $(\sum (X_j^2) - \frac{1}{J} (\sum (X_j))^2) = (J-1)s_X^2$ gilt dann

$$\hat{\sigma}_{\hat{\theta}_0}^2 = s_\varepsilon^2 \frac{1}{J} \sum (X_j^2) / (\sum (X_j^2) - \frac{1}{J} (\sum (X_j))^2) \quad (2.29)$$

$$= s_\varepsilon^2 \frac{1}{J} \sum (X_j^2) / (J-1)s_X^2 \quad (2.30)$$

und mit der Umformung der Null-Identität gilt dann

$$\hat{\sigma}_{\theta_0}^2 = s_\varepsilon^2 \frac{1}{J} \sum (X_j^2) / \left(\sum (X_j^2) - \frac{1}{J} (\sum (X_j))^2 \right) \quad (2.31)$$

$$= s_\varepsilon^2 \frac{1}{J} \left[1 + \frac{1}{J} (\sum (X_j))^2 / \left(\sum (X_j^2) - \frac{1}{J} (\sum (X_j))^2 \right) \right] \quad (2.32)$$

$$= s_\varepsilon^2 \frac{1}{J} \left[1 + \frac{1}{J(J-1)s_X^2} \left(\sum (X_j) \right)^2 \right] \quad (2.33)$$

mit $\frac{1}{J} (\sum (X_j)) = \bar{X}$ gilt $\frac{1}{J} (\sum (X_j))^2 = J\bar{X}^2$ so dass

$$\hat{\sigma}_{\theta_0}^2 = s_\varepsilon^2 \frac{1}{J} \left[1 + \frac{1}{J(J-1)s_X^2} \left(\sum (X_j) \right)^2 \right] \quad (2.34)$$

$$= s_\varepsilon^2 \frac{1}{J(J-1)s_X^2} [1 + (J\bar{X}^2)] \quad (2.35)$$

$$= s_\varepsilon^2 \left[\frac{1}{J} + \frac{\bar{X}^2}{(J-1)s_X^2} \right] \quad (2.36)$$

Damit erhalten wir für die **Varianz bzw. die Unsicherheit des Achsenabschnittes**:

$$\hat{\sigma}_{\theta_0}^2 = s_\varepsilon^2 \left[\frac{1}{J} + \frac{\bar{X}^2}{(J-1)s_X^2} \right] \quad (2.37)$$

2.5 Matrixformalismus zur allgemeinen Lösung der linearen Regression

2.5.1 Matrixformalismus anhand der Regressionsgeraden

Als Regressormatrix verwenden wir jetzt

$$\mathbf{X} = \begin{pmatrix} 1 & X_{1,1} \\ \vdots & \vdots \\ 1 & X_{1,J} \end{pmatrix} \quad (2.38)$$

Die Residuen berechnen sich wie folgt:

$$\begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_J \end{pmatrix} = \begin{pmatrix} Y_1 - (1\theta_0 + X_{1,1}\theta_1) \\ \vdots \\ Y_J - (1\theta_0 + X_{1,J}\theta_1) \end{pmatrix} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_J \end{pmatrix} - \begin{pmatrix} 1 & X_{1,1} \\ \vdots & \vdots \\ 1 & X_{1,J} \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} \quad (2.39)$$

Die Summe der Quadrate der Residuen $\sum_{j=1}^J \varepsilon_j^2$ lässt sich ebenso mit der Rechenregel Zeile mal Spalte schreiben, als Zeilenvektor mal Spaltenvektor

$$\sum_{j=1}^J \varepsilon_j^2 = \begin{pmatrix} \varepsilon_1 & \dots & \varepsilon_J \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_J \end{pmatrix} \quad (2.40)$$

Dabei heißt der Zeilenvektor der transponierte Vektor, also

$$\boldsymbol{\varepsilon}^T = \begin{pmatrix} \varepsilon_1 & \dots & \varepsilon_J \end{pmatrix} \quad \text{und} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_J \end{pmatrix} \quad (2.41)$$

und ferner führen wir ein

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} \quad \text{und} \quad \mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_J \end{pmatrix} \quad (2.42)$$

und Gl. (2.39) sieht in transponierter Schreibweise wie folgt aus

$$\boldsymbol{\varepsilon}^T = \mathbf{Y}^T - \boldsymbol{\theta}^T \begin{pmatrix} 1 & X_{1,1} \\ \vdots & \vdots \\ 1 & X_{1,J} \end{pmatrix}^T \quad (2.43)$$

d.h.

$$\boldsymbol{\varepsilon}^T = \mathbf{Y}^T - \begin{pmatrix} \theta_0 & \theta_1 \end{pmatrix} \begin{pmatrix} 1 & \dots & 1 \\ X_{1,1} & \dots & X_{1,J} \end{pmatrix}. \quad (2.44)$$

Die Summe der Quadrate der Residuen $\sum_{j=1}^J \varepsilon_j^2$ sieht damit wie folgt aus

$$\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (\mathbf{Y}^T - \boldsymbol{\theta}^T \mathbf{X}^T)(\mathbf{Y} - \mathbf{X}\boldsymbol{\theta}) \quad (2.45)$$

d.h.

$$\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = \mathbf{Y}^T \mathbf{Y} - \boldsymbol{\theta}^T \mathbf{X}^T \mathbf{Y} - \mathbf{Y}^T \mathbf{X} \boldsymbol{\theta} + \boldsymbol{\theta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\theta} \quad (2.46)$$

Für die Ableitungen nach den θ_0, θ_1 gilt

$$\frac{\partial}{\partial \theta_l} \sum_{j=1}^J \varepsilon_j^2 = 2 \sum_{j=1}^J \varepsilon_j \frac{\partial}{\partial \theta_l} \varepsilon_j = 2 \sum_{j=1}^J \varepsilon_j (-1) \frac{\partial}{\partial \theta_l} (1 \theta_1 + X_{1,j} \theta_2) = 0 \quad (2.47)$$

wobei $l = 0, 1$ ist also

$$\sum_{j=1}^J \varepsilon_j \frac{\partial}{\partial \theta_0} (1 \theta_0 + X_{1,j} \theta_1) = 0 \quad \text{und} \quad \sum_{j=1}^J \varepsilon_j \frac{\partial}{\partial \theta_1} (1 \theta_0 + X_{1,j} \theta_1) = 0 \quad (2.48)$$

also

$$\sum_{j=1}^J \varepsilon_j (1) = 0 \quad \text{und} \quad \sum_{j=1}^J \varepsilon_j (X_{1,j}) = 0 \quad (2.49)$$

Dies sieht in Matrixschreibweise als 2×2 -Gleichungssystem wie folgt aus

$$\boldsymbol{\varepsilon}^T \mathbf{X} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (2.50)$$

d.h.

$$\left(\mathbf{Y}^\top - \boldsymbol{\theta}^\top \mathbf{X}^\top\right) \mathbf{X} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (2.51)$$

d.h.

$$\boldsymbol{\theta}^\top \mathbf{X}^\top \mathbf{X} = \mathbf{Y}^\top \mathbf{X} \quad (2.52)$$

das ist äquivalent zu

$$\mathbf{X}^\top \mathbf{X} \boldsymbol{\theta} = \mathbf{X}^\top \mathbf{Y}. \quad (2.53)$$

Die numerische Lösung des Gleichungssystems liefert dann die Schätzwerte zu den Regressionparametern $\boldsymbol{\theta}$. Ein mögliches Verfahren zum Lösen des linearen Gleichungssystems ist das Gauß-Jordan-Eliminationsverfahren, zu dem im Anhang dieses Skripts der Quellcode gemäß den *Numerical Recipes* [Fla02] abgedruckt ist. Formal notieren wir die Schätzer (als solche kenntlich durch das Dach) wie folgt

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{Y} \quad (2.54)$$

wobei hoch minus Eins soviel bedeutet wie die Inverse der Matrix.

Alternative Herleitung der Schätzparameter für den Linearen Least Square Gl.(2.54) in der Schreibweise mit Indizes

Das j-te Residuum ist gegeben durch:

$$\epsilon_j = Y_j - \sum_{k=1}^M \theta_k X_{jk} \quad \text{mit } j = 1, \dots, J \quad (2.55)$$

Das Qualitätsmaß Q oder häufig auch als Zielfunktion S bezeichnet ist gegeben durch:

$$Q = \sum_{j=1}^J \epsilon_j^2. \quad (2.56)$$

Die besten Schätzparameter $\hat{\boldsymbol{\theta}}$ erhält durch Aufsuchen des Minimums des Qualitätsmaßes Q . D. h. wir bilden die 1. Ableitung und setzen diese gleich Null:

$$\frac{\partial Q}{\partial \theta_k} = 2 \sum_{j=1}^J \epsilon_j \frac{\partial \epsilon_j}{\partial \theta_k} \quad \text{mit } (k = 1, 2, \dots, M) \quad (2.57)$$

Die partiellen Ableitungen lassen sich aus Gl.(2.55) bestimmen:

$$\frac{\partial \epsilon_j}{\partial \theta_k} = -X_{jk} \quad (2.58)$$

Somit erhalten wir:

$$\frac{\partial Q}{\partial \theta_k} = 2 \sum_{j=1}^J \left(Y_j - \sum_{l=1}^M \theta_l X_{jl} \right) (-X_{jk}) \quad \text{mit } (k = 1, 2, \dots, M). \quad (2.59)$$

Diese erste Ableitung muss gleich Null sein, d.h.

$$2 \sum_{j=1}^J \left(Y_j - \sum_{l=1}^M \hat{\theta}_l X_{jl} \right) (-X_{jk}) = 0 \quad \text{mit} \quad (k = 1, 2, \dots, M). \quad (2.60)$$

Durch Umformen erhalten wir die „Normalgleichung“:

$$\sum_{j=1}^J \sum_{l=1}^M X_{jk} X_{jl} \hat{\theta}_l = \sum_{j=1}^J X_{jk} Y_j \quad \text{mit} \quad (k = 1, 2, \dots, M). \quad (2.61)$$

Die Normalgleichung kann dann in Matrixform geschrieben werden:

$$(\mathbf{X}^T \mathbf{X}) \hat{\boldsymbol{\theta}} = \mathbf{X}^T \mathbf{Y} \quad (2.62)$$

Wir erhalten auch über diese Herleitung die Gl.(2.54), deren Lösung uns die optimalen Fitparameter $\hat{\boldsymbol{\theta}}$ liefert.

Als nächstes ermitteln wir, um das **vollständige Messergebnis** zu erhalten, auch die Schätzwerte der Kovarianzen. Die Hauptdiagonale der Kovarianzmatrix sind die Varianzen.

$$\text{Cov}(\boldsymbol{\theta}, \boldsymbol{\theta}) = \text{Cov} \left((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}, (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \right) \quad (2.63)$$

das ist

$$\text{Cov}(\boldsymbol{\theta}, \boldsymbol{\theta}) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \text{Cov}(\mathbf{Y}, \mathbf{Y}) \left((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right)^T \quad (2.64)$$

und mit Einsetzen der aus den Schätzern $\hat{\boldsymbol{\theta}}$ erhaltenen empirischen Varianz der Residuen $\text{Cov}(\mathbf{Y}, \mathbf{Y}) = \text{Var}(\varepsilon) = \hat{\sigma}_\varepsilon^2$ bekommen wir die empirische Kovarianzmatrix

$$\hat{\Sigma}_{\boldsymbol{\theta}} = \hat{\sigma}_\varepsilon^2 (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \left((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right)^T \quad (2.65)$$

mit

$$\text{Var}(\varepsilon) = \hat{\sigma}_\varepsilon^2 = \frac{1}{J - M} \hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}}$$

wobei M die Anzahl der Regressionsparameter ist, also $M = 2$ im Fall der Geraden, für die beiden Parameter Achsenabschnitt und Steigung und

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\theta}}.$$

Ferner gilt mit $\left((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right)^T = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}$

$$\hat{\Sigma}_{\boldsymbol{\theta}} = \hat{\sigma}_\varepsilon^2 (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \quad (2.66)$$

und mit $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} = \mathbb{I}$ Einheitsmatrix

$$\hat{\Sigma}_{\boldsymbol{\theta}} = \hat{\sigma}_\varepsilon^2 (\mathbf{X}^T \mathbf{X})^{-1} \quad (2.67)$$

Gleichung (2.67) liefert die Unsicherheit für die Regressionsparameter θ . Setzen wir dies nun für die Regressionsgerade ein, also setzen wir

$$\mathbf{X} = \begin{pmatrix} 1 & X_{1,1} \\ \vdots & \vdots \\ 1 & X_{1,J} \end{pmatrix}$$

ein, so erhalten wir für

$$(\mathbf{X}^T \mathbf{X})^{-1} = \left(\begin{pmatrix} 1 & \cdots & 1 \\ X_{1,1} & \cdots & X_{1,J} \end{pmatrix} \begin{pmatrix} 1 & X_{1,1} \\ \vdots & \vdots \\ 1 & X_{1,J} \end{pmatrix} \right)^{-1} \quad (2.68)$$

d.h.

$$(\mathbf{X}^T \mathbf{X})^{-1} = \begin{pmatrix} J & \sum_{j=1}^J X_{1,j} \\ \sum_{j=1}^J X_{1,j} & \sum_{j=1}^J X_{1,j}^2 \end{pmatrix}^{-1} \quad (2.69)$$

d.h.

$$(\mathbf{X}^T \mathbf{X})^{-1} = \frac{1}{J \sum_{j=1}^J X_{1,j}^2 - \left(\sum_{j=1}^J X_{1,j} \right)^2} \begin{pmatrix} \sum_{j=1}^J X_{1,j}^2 & -\sum_{j=1}^J X_{1,j} \\ -\sum_{j=1}^J X_{1,j} & J \end{pmatrix} \quad (2.70)$$

so dass wir für den Achsenabschnitt θ_0 und die Steigung θ_1 folgende empirische Varianzen und Kovarianzen erhalten

$$\begin{pmatrix} \hat{\sigma}_{\theta_0}^2 & \hat{\sigma}_{\theta_0, \theta_1} \\ \hat{\sigma}_{\theta_0, \theta_1} & \hat{\sigma}_{\theta_1}^2 \end{pmatrix} = \frac{\hat{\sigma}_\varepsilon^2}{J \sum_{j=1}^J X_{1,j}^2 - \left(\sum_{j=1}^J X_{1,j} \right)^2} \begin{pmatrix} \sum_{j=1}^J X_{1,j}^2 & -\sum_{j=1}^J X_{1,j} \\ -\sum_{j=1}^J X_{1,j} & J \end{pmatrix} \quad (2.71)$$

oder einzeln aufgeschrieben, die Varianz des Achsenabschnitts

$$\hat{\sigma}_{\theta_0}^2 = \frac{\hat{\sigma}_\varepsilon^2 \sum_{j=1}^J X_{1,j}^2}{J \sum_{j=1}^J X_{1,j}^2 - \left(\sum_{j=1}^J X_{1,j} \right)^2} \quad (2.72)$$

die Kovarianz für Achsenabschnitt und Steigung

$$\hat{\sigma}_{\theta_0, \theta_1} = \frac{-\hat{\sigma}_\varepsilon^2 \sum_{j=1}^J X_{1,j}}{J \sum_{j=1}^J X_{1,j}^2 - \left(\sum_{j=1}^J X_{1,j} \right)^2} \quad (2.73)$$

und die Varianz der Steigung

$$\hat{\sigma}_{\theta_1}^2 = \frac{\hat{\sigma}_\varepsilon^2}{\sum_{j=1}^J X_{1,j}^2 - \frac{1}{J} \left(\sum_{j=1}^J X_{1,j} \right)^2}. \quad (2.74)$$

Die Standardabweichungen zu jedem Regressionsparameter θ_l mit $l = 0, \dots, M-1$ sind die Wurzel aus den Varianzen, die auf der Hauptdiagonalen der Kovarianzmatrix $(\mathbf{X}^\top \mathbf{X})^{-1} \hat{\sigma}_\varepsilon^2$ stehen. In der üblichen englischsprachigen Literatur zur Regressionrechnung wird die Standardabweichung der Regressionsparameter *Standard Error* genannt.

2.5.2 Matrixansatz als allgemeiner Ansatz für linearen Regression

Die behandelten allgemeinen Formen von Problemstellungen zur Schätzung von Modellparametern werden wir in dieser Vorlesungsreihe nicht weiter vertiefen. Sie sollen lediglich einen Einblick und Überblick über die vielfältigen Möglichkeiten in diesem Bereich liefern.

Um die lineare Regression universeller einsetzen zu können, ob nun Polynome oder Ebenen oder Paraboloiden zu fitten sind, ist es erforderlich den allgemeinen Ansatz zu verstehen und die unterschiedlichen Anwendungen so formulieren zu können, dass sie darüber lösbar werden. Im folgenden wollen wir den Übergang auf M Modellparameter $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_M)$ vollziehen. Wir wollen also den Übergang von Gl. (2.39) mit

$$\begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_J \end{pmatrix} = \begin{pmatrix} Y_1 - (1\theta_0 + X_{1,1}\theta_1) \\ \vdots \\ Y_J - (1\theta_0 + X_{1,J}\theta_1) \end{pmatrix} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_J \end{pmatrix} - \begin{pmatrix} 1 & X_{1,1} \\ \vdots & \vdots \\ 1 & X_{1,J} \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}$$

mit

$$\mathbf{X} = \begin{pmatrix} 1 & X_{1,1} \\ \vdots & \vdots \\ 1 & X_{1,J} \end{pmatrix}$$

zu einer allgemeinen Formulierung finden.

Für M Regressoren wird die Regressormatrix zu einer $J \times M$ -Matrix, also einer Matrix aus M Spalten und J Zeilen aufgestellt:

$$\mathbf{X} = \begin{pmatrix} X_{1,1} & \dots & X_{M,1} \\ \vdots & & \vdots \\ X_{1,J} & \dots & X_{M,J} \end{pmatrix} \quad (2.75)$$

Die Indexschreibweise ist hier relativ zur Schreibweise der Matrizenrechnung vertauscht, weil wir den Index für die Messgrößen vorne haben gefolgt von dem Index für die Beobachtungen, wir aber die Messgrößen als Spaltenvektoren verwenden und somit den Index für die Beobachtungen als Zeilenindex haben.

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_J \end{pmatrix} = \begin{pmatrix} Y_1 - (X_{1,1}\theta_1 + \dots + X_{M,1}\theta_M) \\ \vdots \\ Y_J - (X_{1,J}\theta_1 + \dots + X_{M,J}\theta_M) \end{pmatrix} = \mathbf{Y} - \mathbf{X}\boldsymbol{\theta} \quad (2.76)$$

Im Fall eines Polynoms kann jeder der Regressoren Funktion derselben Messgröße sein. Dies kann beispielsweise einfach durch folgenden Potenzansatz $X_{m,j} = X_{1,j}^{m-1}$ mit $m = 1, \dots, M$ als Polynom vom Grad $M - 1$ dargestellt werden

$$\begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_J \end{pmatrix} = \begin{pmatrix} Y_1 - (X_{1,1}^0\theta_1 + \dots + X_{1,1}^{M-1}\theta_M) \\ \vdots \\ Y_J - (X_{1,J}^0\theta_1 + \dots + X_{1,J}^{M-1}\theta_M) \end{pmatrix} \quad (2.77)$$

wobei $X_{1,j}^0 = 1$ ist.

Wird die Regressormatrix zu Beginn korrekt aufgestellt, so lassen sich die M Regressionsparameter $\boldsymbol{\theta}$ schätzen durch numerische Lösung des folgenden Gleichungssystems:

$$\mathbf{X}^\top \mathbf{X} \boldsymbol{\theta} = \mathbf{X}^\top \mathbf{Y}. \quad (2.78)$$

Dass nach Lösen des Gleichungssystems Schätzwerte für die Parameter gewonnen wurden, wird durch das Dach auf dem Symbol $\boldsymbol{\theta}$ gekennzeichnet.

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} \quad (2.79)$$

wobei hoch minus Eins soviel bedeutet wie die Inverse der Matrix. Für Varianzen und Kovarianzen der M Regressionsparameter $\theta_1, \dots, \theta_M$ werden die Residuen gebraucht

$$\hat{\varepsilon} = \begin{pmatrix} \hat{\varepsilon}_1 \\ \vdots \\ \hat{\varepsilon}_J \end{pmatrix} = \begin{pmatrix} Y_1 - (X_{1,1}\hat{\theta}_1 + \dots + X_{M,1}\hat{\theta}_M) \\ \vdots \\ Y_J - (X_{1,J}\hat{\theta}_1 + \dots + X_{M,J}\hat{\theta}_M) \end{pmatrix} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}} \quad (2.80)$$

um deren empirische Varianz zu berechnen (M : Anzahl der Modellparameter; J : Anzahl der Messwertepaare)

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{J - M} \hat{\varepsilon}^\top \hat{\varepsilon} \quad (2.81)$$

so dass wir die empirische Kovarianzmatrix der Parameter gemäß Gl. (2.67) erhalten:

$$\hat{\boldsymbol{\Sigma}}_\theta = (\mathbf{X}^\top \mathbf{X})^{-1} \hat{\sigma}_\varepsilon^2$$

Um die Unsicherheit der einzelnen Regressionsparameter zu erhalten, führen wir Einheitsvektoren ein, die M -dimensional sind. Bei dem l -ten Einheitsvektor sind die ersten $l - 1$ Vektorkomponenten Nullen,

an der l -ten Stelle steht eine Eins und die $l + 1$ -te Komponente bis zur M -ten sind wieder Nullen.

$$\mathbf{e}_l = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (2.82)$$

Dann lässt sich durch folgende Multiplikation die l -te Varianz, d.h. das l -te Element der Hauptdiagonalen aus der Kovarianzmatrix herausholen

$$\hat{\sigma}_l^2 = \mathbf{e}_l^\top \left(\mathbf{X}^\top \mathbf{X} \right)^{-1} \mathbf{e}_l \hat{\sigma}_\varepsilon^2. \quad (2.83)$$

Die Standardabweichung (*Standard Error*) wird daraus durch Ziehen der Wurzel gewonnen

$$\hat{\sigma}_l = \sqrt{\hat{\sigma}_l^2}. \quad (2.84)$$

Im fünften Kapitel werden wir dann sehen, wie man mit Hilfe der Varianzen bzw. Standardabweichungen Vertrauensintervalle für die geschätzten Modellparameter bestimmt. Die Herleitung ist im Anhang aufgeschrieben.

2.6 Das Bestimmtheitsmaß, empirischer Korrelationskoeffizient

Die Güte des gewählten linearen Ansatzes kann mit dem **Bestimmtheitsmaß** ρ_{XY}^2 bzw. dem **empirischen Korrelationskoeffizienten** ρ_{XY} beurteilt werden, aber nur dann wenn ein solcher Zusammenhang auch existiert. Der empirische Korrelationskoeffizienten gibt den Grad der linearen Abhängigkeit zwischen den Messwerten (X_1, \dots, X_J) und (Y_1, \dots, Y_J) an.

Für den empirischen Korrelationskoeffizienten ρ_{XY} gilt folgendes:

$\rho_{XY} = \pm 1$	100% lineare Abhängigkeit
$\rho_{XY} = 0$	keine lineare Abhängigkeit
$\rho_{XY} > \text{bzw. } < 0$	gleich-/gegenläufige lineare Abhängigkeit

Für die Bestimmung des Bestimmtheitsmaßes oder des empirischen Korrelationskoeffizient führt man eine Streuungs- bzw. Varianzanalyse durch. Man kann bei der linearen Regression 3 Arten von Streuungen unterscheiden:

s_T^2 : die **totale empirische Varianz**

s_M^2 : die **empirische Varianz auf Grund des Modells**

s_R^2 : die **restliche empirische Varianz**

Die totale Streuung ist bei unabhängigen direkten Messgrößen X_j gegeben durch:

$$s_T^2 = \frac{1}{J-1} \sum_{j=1}^J (Y_j - \bar{Y})^2 \quad \text{mit } Y_j : \text{Messwert zu } X_j \quad (2.85)$$

Die empirische Varianz auf Grund des Modells beschreibt die Streuung der Y-Werte auf der Regressionsgeraden in Bezug zum Mittelwert \bar{Y} , d. h.

$$s_M^2 = \frac{1}{J-1} \sum_{j=1}^n (\hat{\theta}_0 + \hat{\theta}_1 X_j - \bar{Y})^2 \quad (2.86)$$

Die Differenz zwischen Y_j und $\hat{\theta}_0 + \hat{\theta}_1 X_j$ bleiben im Modell unerklärt, sie wird als restliche Streuung bezeichnet:

$$s_R^2 = \frac{1}{J-1} \sum_{j=1}^J (Y_j - (\hat{\theta}_0 + \hat{\theta}_1 X_j))^2 \quad (2.87)$$

Man kann nun folgendes Bemerkenswerte zeigen:

$$s_T^2 = s_M^2 + s_R^2 \quad (2.88)$$

Das Bestimmtheitsmaß ist definiert als das Verhältnis von empirischer Varianz des Modells zu der

totalen Varianz:

$$\rho_{XY}^2 = \frac{s_M^2}{s_T^2} = \frac{\sum_{j=1}^J (\hat{\theta}_0 + \hat{\theta}_1 X_j - \bar{Y})^2}{\sum_{j=1}^J (Y_j - \bar{Y})^2}$$

Mit dem Schwerpunkt $\bar{Y} = \hat{\theta}_0 + \hat{\theta}_1 \bar{X}$ ergibt sich:

$$\rho_{XY}^2 = \frac{\hat{\theta}_1^2 \cdot \sum_{j=1}^J (X_j - \bar{X})^2}{\sum_{j=1}^J (Y_j - \bar{Y})^2}$$

Mit den Regressionsparametern $\hat{\theta}_1 = \frac{\sum_{j=1}^J (X_j - \bar{X})(Y_j - \bar{Y})}{\sum_{j=1}^J (X_j - \bar{X})^2}$ (siehe Gl.(2.17)) erhält man für das Bestimmtheitsmaß:

$$\begin{aligned} \rho_{XY}^2 &= \frac{\left[\sum_{j=1}^J (X_j - \bar{X})(Y_j - \bar{Y}) \right]^2}{\sum_{j=1}^J (X_j - \bar{X})^2 \cdot \sum_{j=1}^J (Y_j - \bar{Y})^2} \\ &= \frac{s_{XY}^2}{s_X^2 \cdot s_Y^2} \quad \text{mit } 0 \leq \rho_{XY}^2 \leq 1 \end{aligned} \quad (2.89)$$

Den empirischen Korrelationskoeffizienten kann man somit wie folgt berechnen:

$$\rho_{XY} = \frac{s_{XY}}{s_X \cdot s_Y} \quad \text{mit } -1 \leq \rho_{XY} \leq 1 \quad (2.90)$$

2.7 Beispiele

2.7.1 Fit einer Geraden

Gegeben sind 5 Messpunkte

X_i	1	2	3	4	5
Y_i	0.4	0.55	0.70	0.75	0.8

- Führen Sie eine lineare Regression durch, indem Sie den y-Abschnitt $\hat{\theta}_0$ und die Steigung $\hat{\theta}_1$ bestimmen. Geben Sie das Bestimmtheitsmaß ρ_{XY}^2 an.
- Bestimmen Sie die Residuen ε_j und das Qualitätsmaß $Q(\hat{\theta}_0, \hat{\theta}_1)$ der Regression. Geben Sie die Varianz der Residuen $s^2(\hat{\theta}_0, \hat{\theta}_1)$ an?
- Geben Sie den Standardabweichung (standard error) der beiden Regressionsparameter $\hat{\theta}_0$ und $\hat{\theta}_1$ an?

Lösung zu Teil (a):

Gesucht:

$$Y = \hat{\theta}_1 \cdot X + \hat{\theta}_0$$

Mittelwert von X : $\bar{X} = 3$; Mittelwert von Y : $\bar{Y} = 0.64$;

Empirische Standardabweichung von X : $s_X = 1.5811$;

Empirische Standardabweichung von Y : $s_Y = 0.1636$;

Empirische Kovarianz: $s_{XY} = 0.25$

Schätzwerte $\hat{\theta}_1$ und $\hat{\theta}_0$:

$$\hat{\theta}_1 = \frac{s_{XY}}{s_X^2} = 0.100$$

$$\hat{\theta}_0 = \bar{Y} - \hat{\theta}_1 \bar{X} = 0.340$$

Bestimmtheitsmaß:

$$\rho_{XY}^2 = \frac{s_{XY}^2}{s_X^2 \cdot s_Y^2} = 0.9346$$

Mit der Gl.(2.54) können wir das Beispiel mit dem Matrixformalismus auch lösen. Der Matlab/Octave-Code lautet:

```
close all; clear all;
% Eingabe der Messwertepaare
input = [...
1, 0.4;...
2, 0.55;...
3, 0.7;...
4, 0.75; ...
```

```

5, 0.8];
m = length(input);           % Anzahl der Messpunkte
X = [ones(m,1), input(:,1)]; % Aufstellen der X-Matrix
Y = input(:,2);              % Aufstellen der Y-Matrix
% degree of freedom
df = 5-2;
% Berechnung der Schätzwerte \thetaHat;
% y-Abschnitt = \thetaHat(1); Steigung = \thetaHat(2)
thetaHat = (X' * X) \ (X' * Y);
% Anzeige der Schätzwerte
disp(thetaHat);
% Plotten der Regressionsgeraden
xx = linspace(0,6,100);
yy = thetaHat(1) + thetaHat(2)*xx;
figure(10)
plot(xx, yy, 'linewidth', 1.5)
% Plotten der Datenwerte
hold on
plot(input(:,1), input(:,2), 'or', 'linewidth', 2, 'MarkerSize', 6)
hold off
grid on;
xlabel('Werte X_j', 'fontsize', 14)
ylabel('Werte Y_j', 'fontsize', 14)
set(gcf, 'PaperUnits', 'centimeters');
x_width=15; y_width= 9;
set(gcf, 'PaperPosition', [0 0 x_width y_width]);
print '-dpng' LinRegression.png

```

Als Lösung erhalten wir den Plot in Abb.2.4. Der y-Achsenabschnitt liegt bei 0.34 und die Steigung ist 0.1.

Lösung zu Teil (b)

Residuen: $\varepsilon_j = Y_j - (\hat{\theta}_0 + \hat{\theta}_1 \cdot X_j)$ mit $j = 1, \dots, 5$

$$\varepsilon_1 = -0.0400; \quad \varepsilon_2 = 0.0100; \quad \varepsilon_3 = 0.0600; \quad \varepsilon_4 = 0.0100; \quad \varepsilon_5 = -0.0400$$

Qualitätsmaß:

$$Q(\hat{\theta}_0, \hat{\theta}_1) = \sum_{j=1}^J \varepsilon_j^2 = 0.0070$$

Varianz der Residuen:

$$s^2(\hat{\theta}_0, \hat{\theta}_1) = \frac{Q(\hat{\theta}_0, \hat{\theta}_1)}{J-2} = 0.0023 \quad \text{bzw.} \quad s(\hat{\theta}_0, \hat{\theta}_1) = 0.048$$

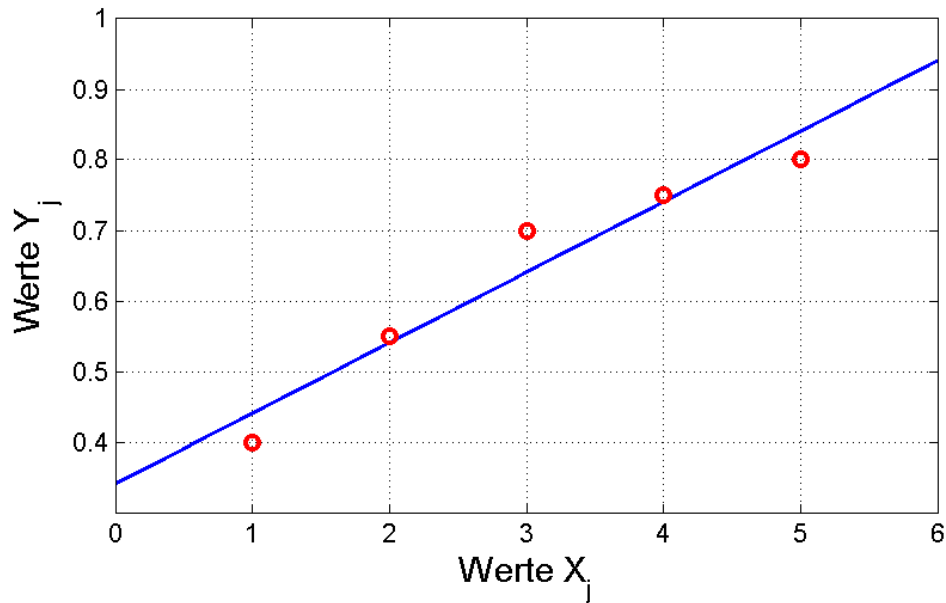


Abbildung 2.4: Lineare Regression. Als Schätzparameter erhält man $\hat{\theta} = [0.3400, 0.1000]^T$

Die Varianz der Residuen ist wie folgt gegeben:

$$\text{Var}(\epsilon) = \frac{1}{J - M} \cdot (\mathbf{Y} - \mathbf{X}\hat{\theta})^T (\mathbf{Y} - \mathbf{X}\hat{\theta}) \quad (2.91)$$

Die Kovarianzen der Schätzerparameter θ sind gegeben durch:

$$\text{Cov}(\theta) = (\mathbf{X}^T \mathbf{X})^{-1} \text{Var}(\epsilon) \quad (2.92)$$

Matlab-Code zur Bestimmung des Qualitätsmaßes und der Varianz

```
% Qualitätsmaß Q
Q = (Y-X*thetaHat)' * (Y-X*thetaHat)
% Varianz der Residuen
Var_epsilon = 1/df * (Y-X*thetaHat)' * (Y-X*thetaHat)
% Wuzel aus Var_epsilon
RMSE = sqrt(Var_epsilon)
% Kovarianzen
Cov = Var_epsilon * inv(X'*X)
```

Als Ergebnis für unser Beispiel erhalten wir:

```
Q = 0.0070
Var_epsilon = 0.0023
RMSE = 0.0483
Cov =
0.0026 -0.0007
-0.0007 0.0002
```

Lösung zu Teil (c)

Die Standardabweichung für X berechnet sich zu $s_X = 1.5811$.

Für die Varianz des y-Abschnitt $\hat{\theta}_0$ ergibt sich:

$$\begin{aligned}\hat{\sigma}_{\hat{\theta}_0}^2 &= s^2(\hat{\theta}_0, \hat{\theta}_1) \left[\frac{1}{J} + \frac{\bar{X}^2}{(J-1)s_x^2} \right] \\ &= 0.0023 \left[\frac{1}{5} + \frac{3^2}{(5-1) \cdot 1.5811^2} \right] \\ &= 0.00253\end{aligned}$$

Damit ergibt sich die Standardabweichung für $\hat{\theta}_0$:

$$s_{\hat{\theta}_0} = \sqrt{\hat{\sigma}_{\hat{\theta}_0}^2} = 0.0503$$

Für die Varianz der Steigung $\hat{\theta}_1$ ergibt sich:

$$\begin{aligned}\hat{\sigma}_{\hat{\theta}_1}^2 &= \frac{\hat{s}^2(\hat{\theta}_0, \hat{\theta}_1)}{s_X^2 \cdot (J-1)} \\ &= \frac{0.0023}{1.5811^2 \cdot (5-1)} \\ &= 0.00023\end{aligned}\tag{2.93}$$

Damit ergibt sich die Standardabweichung für $\hat{\theta}_1$

$$s_{\hat{\theta}_1} = \sqrt{\hat{\sigma}_{\hat{\theta}_1}^2} = 0.01517$$

Anmerkung: Lösung mit Matlab/Octave Bei Matlab gibt es den Befehl „polyfit“ mit dem Polynomfits durchgeführt werden können. Matlab liefert hier dasselbe Ergebnis, auch für die Vertrauensbereiche, das Qualitätsmaß Q (engl. SSE: Sum squared error) oder das Bestimmtheitsmaß ρ_{XY}^2 (R-square), siehe Abb.2.5:

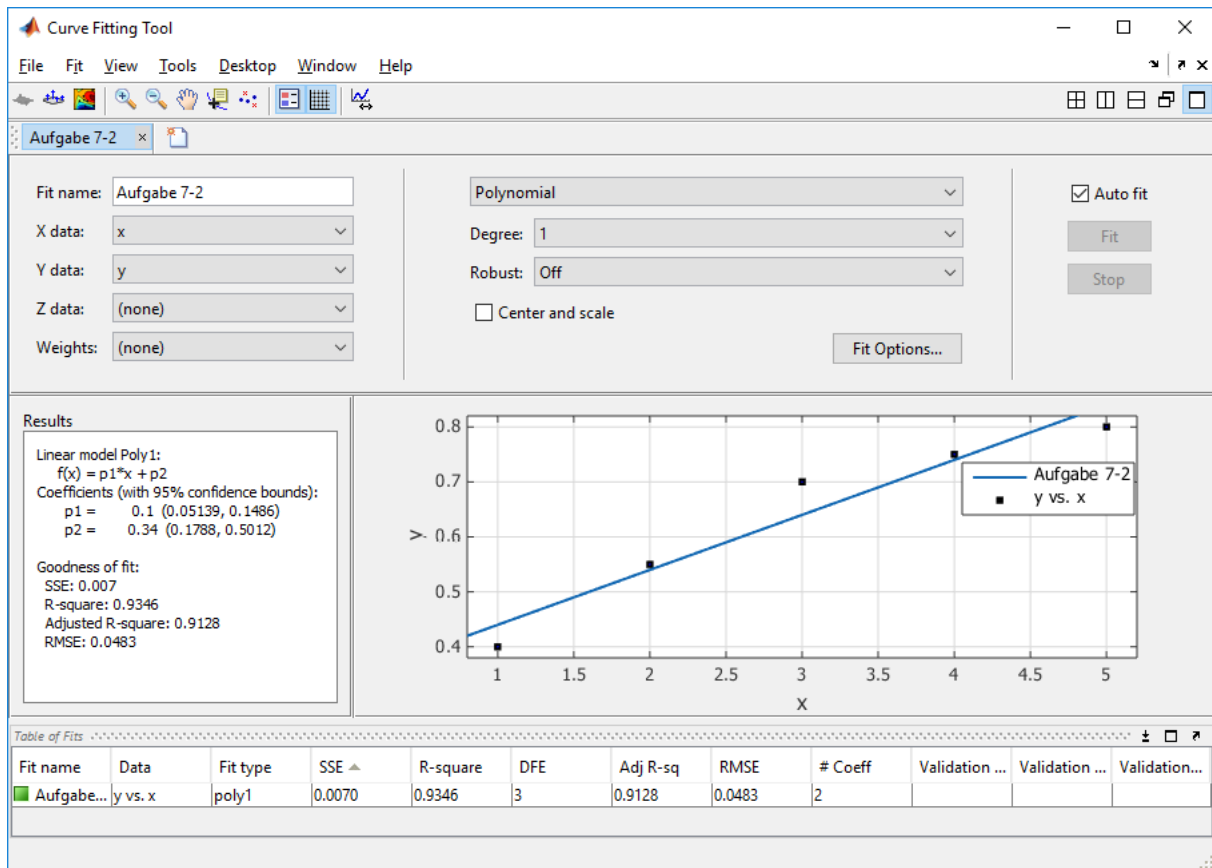


Abbildung 2.5: Lösung der Aufgabe mit dem polyfit Befehl von Matlab/Octave. Es wird hier u. a. der 95%ige Vertrauensbereich berechnet und angezeigt (siehe Mitte links in der Abbildung).

Hinweis: Vertrauensbereich

In einer der nächsten Vorlesungen werden wir sehen, wie man Vertrauensbereiche mit Hilfe der Varianzen bzw. Standardabweichungen berechnet. Dazu benötigt man noch die t-Verteilung. Da man hier 5 Messpunkte und 2 Modellparameter hat, wird die t-Verteilung mit dem Freiheitsgrad $5-2 = 3$ benötigt. 95%iger Vertrauensbereich für $\hat{\theta}_1$ mit $t_3 = 3.182$:

$$\varepsilon_{\hat{\theta}_1} = t_3 \cdot \hat{\sigma}_{\theta_1} = \frac{t_3 \cdot s(\hat{\theta}_0, \hat{\theta}_1)}{s_X \cdot \sqrt{J-1}} = 0.0486$$

95%iger Vertrauensbereich für $\hat{\theta}_0$ errechnet sich durch:

$$\varepsilon_{\hat{\theta}_0} = t_3 \cdot \hat{\sigma}_{\theta_0} = t_3 \cdot s(\hat{\theta}_0, \hat{\theta}_1) \cdot \sqrt{\frac{1}{J} + \frac{\bar{X}^2}{(J-1)s_X^2}} = 0.1612$$

Ergebnis:

Der 95% Vertrauensbereich des Schätzwertes $\hat{\theta}_1$ ist somit gegeben durch:

$$\hat{\theta}_1 = 0.1000 \pm 0.0486 \quad \text{bzw.} \quad [0.0514; 0.1486] \quad (2.94)$$

Der 95% Vertrauensbereich des Schätzwertes $\hat{\theta}_0$ ist somit gegeben durch:

$$\hat{\theta}_1 = 0.3400 \pm 0.1612 \quad \text{bzw.} \quad [0.1788; 0.5012] \quad (2.95)$$

2.7.2 Fit einer kubischen Funktion (Polynom 3. Ordnung)

Wir nehmen wieder die gleichen Messwertpaare wie eben. Jedoch fitten wir nun eine kubische Funktion durch die Messdaten, d.h.

$$Y = \hat{\theta}_3 X^3 + \hat{\theta}_2 X^2 + \hat{\theta}_1 X^1 + \hat{\theta}_0 \quad (2.96)$$

Für die Schätzparameter erhalten wir:

$$\hat{\theta} = (0.1900, \quad 0.2286, \quad -0.0214, \quad -0.0000)$$

Für das Qualitätsmaß ergibt sich:

$$Q = 5.7143e - 04$$

Für die Streuung der Residuen ergibt sich:

$$\text{Var_epsilon} = 5.7143e - 04$$

Standardabweichung der Residuen:

$$\text{RMSE} = 0.0239$$

Die Kovarianzmatrix ergibt sich zu:

$$\text{Cov} = \begin{bmatrix} 0.0138 & -0.0176 & 0.0063 & -0.0007 \\ -0.0176 & 0.0236 & -0.0087 & 0.0009 \\ 0.0063 & -0.0087 & 0.0033 & -0.0004 \\ -0.0007 & 0.0009 & -0.0004 & 0.0000 \end{bmatrix}$$

Der Matlab/Octave-Code lautet:

```
close all; clear all;
% Eingabe der Messwertpaare
input = [...
1, 0.4;...
2, 0.55;...
3, 0.7;...
4, 0.75; ...
5, 0.8];
m = length(input);      % Anzahl der Messpunkte
x_Werte = input(:,1);
```

```

X = [ones(m,1), x_Werte, x_Werte.^2,x_Werte.^3]; % Aufstellen der X-Matrix
Y = input(:,2); % Aufstellen der Y-Matrix
% degree of freedom
df = 5-4;
% Berechnung der Schätzwerte \thetaHat;
% y-Abschnitt = \thetaHat(1); Steigung = \thetaHat(2)
thetaHat = (X' * X) \ (X' * Y);
% Anzeige der Schätzwerte
disp(thetaHat);
% Plotten der Fitfunktion
xx = linspace(0,6,100);
yy = thetaHat(1) + thetaHat(2)*xx+ thetaHat(3)*xx.^2+thetaHat(4)*xx.^3;
figure(10)
plot(xx, yy,'linewidth',1.5)
% Plotten der Datenwerte
hold on
plot(input(:,1), input(:,2), 'or','linewidth',2, 'MarkerSize',6)
hold off
grid on;
xlabel('Werte X_j','fontsize',14)
ylabel('Werte Y_j','fontsize',14)
set(gcf, 'PaperUnits', 'centimeters');
x_width=15 ;y_width= 9;
set(gcf, 'PaperPosition', [0 0 x_width y_width]);
print '-dpng' kubischerFit.png
% Qualitätsmaß Q
Q = (Y-X*thetaHat)' * (Y-X*thetaHat)
% Varianzen der Residuen
Var_epsilon = 1/df *(Y-X*thetaHat)' * (Y-X*thetaHat)
% Wuzel aus Var_epsilon
RMSE = sqrt(Var_epsilon)
% Kovarianzen
Cov = Var_epsilon * inv(X'*X)

```

Wir können auch die Curve Fitting Toolbox von Matlab verwenden, die uns das gleiche Ergebnis anzeigt:

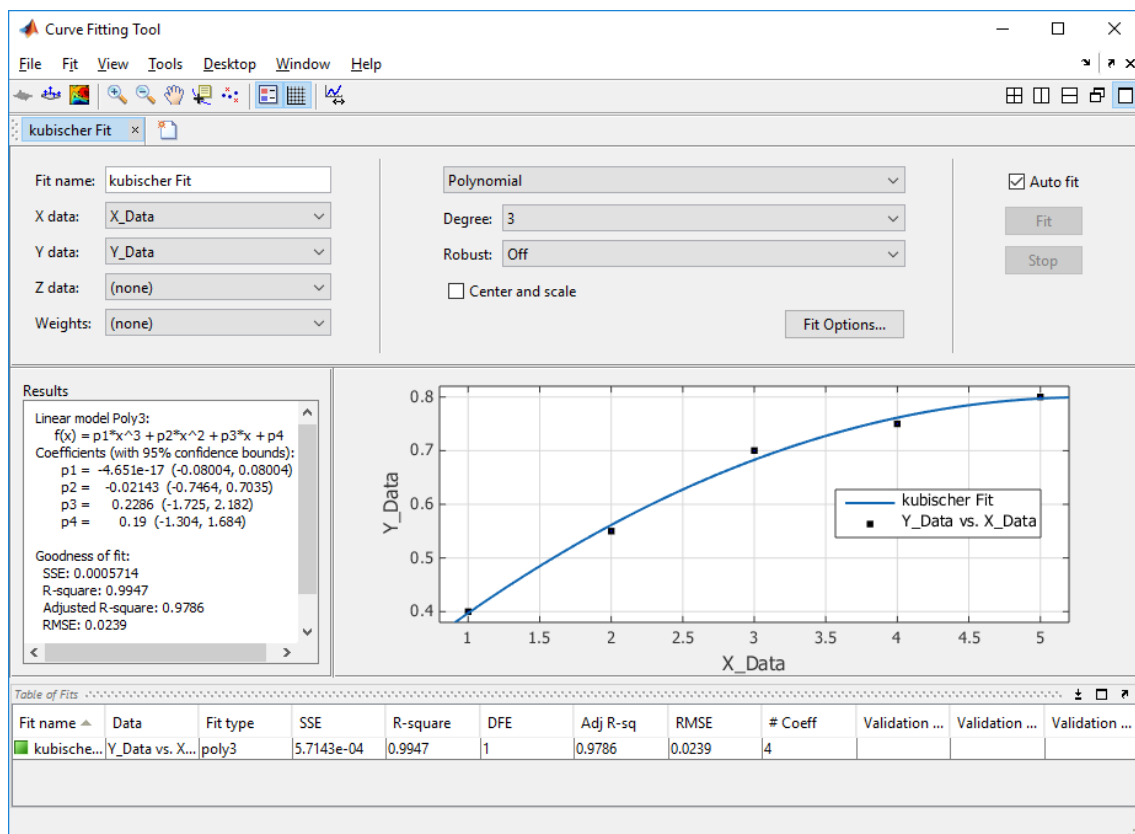
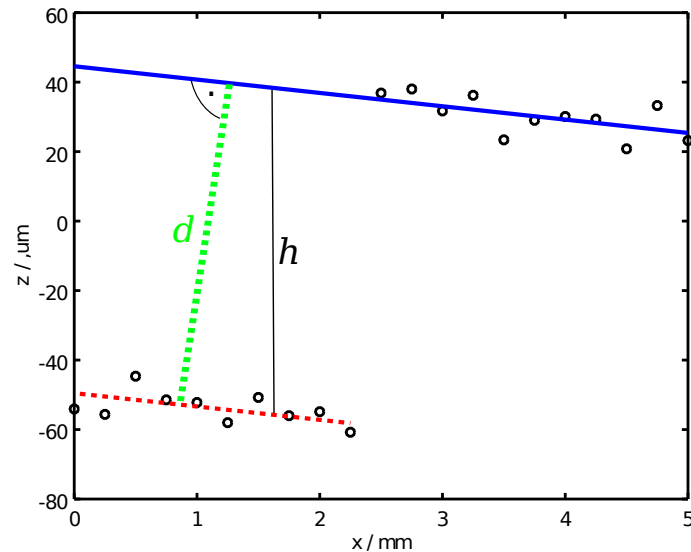


Abbildung 2.6: Kubischer Fit mit der Curve Fitting Toolbox von Matlab. Zum jedem geschätzten Parameter wird der 95%ige Vertrauensbereich angegeben. Zu sehen sind u.a. das Qualitätsmaß Q (engl. SSE: Sum squared error), das Bestimmtheitsmaß ρ_{XY}^2 (engl. R-square), Standardabweichung der Residuen (engl. RMSE: root mean square error) und die Anzahl der Freiheitsgrade (engl. DFE, degree of freedom)

2.8 Aufgaben zum Selbststudium

2.8.1 1. Aufgabe zur Linearen Regression

Es wurde eine Stufe vermessen, die nominal eine Höhe von $100\text{ }\mu\text{m}$ haben soll. Die laterale Achse x wird eingestellt (mit Präzisionstisch positioniert) und als Regressor betrachtet, also nicht als Zufallszahl. Die vertikale Achse z wird gemessen, als Regressand und damit als Zufallszahl betrachtet.



x/mm	0.00	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25
$z/\mu\text{m}$	-54.08	-55.63	-44.65	-51.44	-52.21	-58.01	-50.76	-56.01	-54.86	-60.77

x/mm	2.50	2.75	3.00	3.25	3.50	3.75	4.00	4.25	4.50	4.75	5.00
$z/\mu\text{m}$	36.85	38.02	31.71	36.21	23.39	29.01	30.11	29.35	20.81	33.27	23.19

Die Modellgleichung, mit der die Stufe beschrieben wird, ist

$$z_j = a x_j + c + h \delta_{j \in C} + \varepsilon_j$$

mit C Menge der Indizes von 1 bis J_C . Hier sei $J_C = 10$:

$$C = \{j \mid j = 1, 2, \dots, J_C\}$$

und

$$\delta_{j \in C} = \begin{cases} 1 & \text{falls } j \in C \text{ d.h. } j = 1 \text{ oder } j = 2 \text{ oder } \dots j = J_C \\ 0 & \text{sonst} \end{cases}$$

Dabei ist h noch nicht die Stufenhöhe, sondern deren Projektion auf die vertikale Achse. Die Stufenhöhe d ist der senkrechte Abstand zwischen den beiden parallelen Geraden, die jeweils durch die Punkte auf dem oberen und dem unteren Niveau gehen.

a) Stellen Sie das lineare Gleichungssystem auf, das sich durch partielles Ableiten für das Opti-

mierungsproblem

$$\min_{a,c,h} \left\{ \sum_{j=1}^{J_T} \varepsilon_j^2 \right\}$$

ergibt, mit $J_T = 21$.

- b) Schreiben Sie die Gleichung für die Stufenhöhe d als Funktion von h und a auf.
- c) Schreiben Sie die Gleichung für die Varianz der Residuen auf.
- d) Schreiben Sie die Formel für die Kovarianzmatrix der Modellparameter a, c, h auf.
- e) Verwenden Sie eine Programmierungsumgebung Ihrer Wahl, Matlab, Octave, Python, R, ... die Ihnen einen Solver für lineare Gleichungssysteme zur Verfügung stellt, sowie eine Routine zur Matrixinversion und berechnen Sie die Zahlenwerte für die Modellparameter sowie deren Kovarianzmatrix.

Anmerkung: Die inverse Matrix brauchen Sie nicht analytisch zu berechnen, sondern lediglich $(\dots)^{-1}$ zu notieren.

Kapitel 3

Optimierung nichtlinearer Modelle

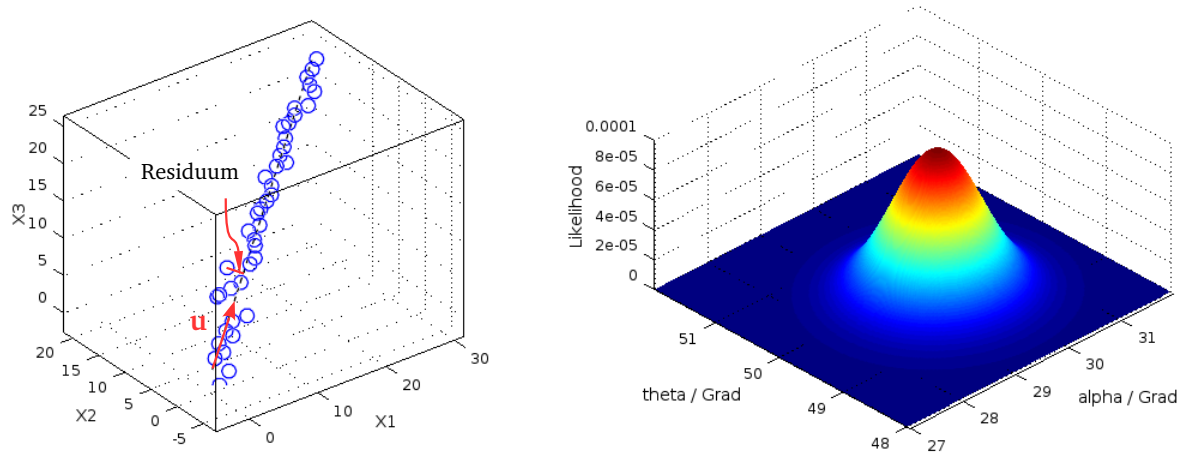


Abbildung 3.1: Modell mit 2 zu schätzenden Parametern (*links*) und der dazugehörigen Likelihood für normalverteilte Residuen (*rechts*).

3.1 Konzepte von Optimierungsverfahren für nichtlineare Modelle

Während dieser Vorlesungsreihe beleuchten wir die Thematik physikalische Größen zu ermitteln, die indirekt zugänglich sind. Sie werden über eine Modellbildung als Modellparameter approximiert. In den bisherigen Kapiteln hatten wir uns damit befasst, Modellparameter linearer Modelle, zu schätzen. Die Besonderheit dabei ist, dass die *Kostenfunktion* Q eine lineare Abhängigkeit von den Modellparametern hat, so dass die Modellparameter im Fall der Ausgleichsrechnung nach der Methode der kleinsten Quadrate durch Lösen eines linearen Gleichungssystems ermittelt werden können. Im vorangegangenen Kapitel wurde in Gl. (2.10), Gl. (2.15) und Gl. (2.78) gezeigt, dass das Gleichungssystem aufgestellt wird, indem der Gradient, der die Kostenfunktion nach den Modellparametern ableitet, gleich Null gesetzt wird, um das Maximum zu finden.

Für den Fall, dass die Modellparameter nicht linear in die Kostenfunktion Q , auch *Zielfunktional* genannt, eingehen, erfolgt die Schätzung der Modellparameter durch Variation, also durch Optimierung derart, um das Minimum der Optimierungszielfunktion Q zu finden.

Wir sind von normalverteilten Abweichungen ausgegangen und haben uns die Likelihoodverteilungen angeschaut. Als Beispiel nehmen wir ein Modell mit 2 Parametern, bei dem eine Gerade durch den Ursprung des Koordinatensystems eines dreidimensionalen Raums gehe und bei dem alle drei Achsen von der gleichen physikalischen Dimension seien. Die beiden zu schätzenden Parameter seien hier der Azimutalwinkel α und der Polarwinkel θ , die die gesuchte Richtung der Geraden beschreiben. Ein Punkt \mathbf{r} , der exakt auf der Geraden liegt, erfülle

$$\mathbf{r} = \lambda \mathbf{u} \quad \text{mit} \quad \mathbf{u}(\alpha, \theta) = \begin{pmatrix} \cos(\alpha) \sin(\theta) \\ \sin(\alpha) \sin(\theta) \\ \cos(\theta) \end{pmatrix} \quad \text{und} \quad \lambda \in \mathbb{R}. \quad (3.1)$$

Abb. 3.1 zeigt das Modell als schwarz gestrichelte Gerade und die Beobachtungen als blaue Kreise,

sowie rechts die Likelihoodverteilung. Wir hatten bereits ausführlich darüber gesprochen, dass die Suche des Maximums der Likelihood in die Suche des Minimums übergeht. Es gelte nun diejenige Geradenrichtung zu finden, für die die senkrechten Abstände, die Residuen ε_j , der beobachteten Punkte (Messpunkte) $\mathbf{r}_{\text{Mess},j} = (X_{1,j}, X_{2,j}, X_{3,j})^\top$ minimal sind. Für das Kriterium für minimale Abstände, also die Kostenfunktion Q , setzen wir wieder voraus, dass die Messpunkte gaußverteilt um eine Gerade streuen. Den Betrag für die senkrechten Abstände berechnen wir mit

$$\varepsilon(\alpha, \theta)_j = |\mathbf{r}_{\text{Mess},j} - (\mathbf{r}_{\text{Mess},j} \cdot \mathbf{u}(\alpha, \theta)) \mathbf{u}(\alpha, \theta)| \quad (3.2)$$

mit $|\cdot|$ als Betrag gemäß $|a| = \sqrt{a_1^2 + a_2^2 + a_3^2}$. Für die Likelihoodverteilung auf der rechten Seite von Abb. 3.1 verwenden wir die Wahrscheinlichkeitsdichte gemäß

$$p(\varepsilon_1, \dots, \varepsilon_J | \alpha, \theta) \propto e^{-\frac{1}{2} \sum_{j=1}^J \left(\frac{\varepsilon(\alpha, \theta)_j}{\sigma} \right)^2}. \quad (3.3)$$

Für den Schätzvorgang wählen wir wieder für das Zielfunktional Q die Summe der Quadrate der Residuen $Q = \sum \varepsilon_j^2$, also einen Optimierungsvorgang gemäß der Methode der kleinsten Residuenquadratsumme (RQS).

$$\min_{\alpha, \theta} \sum_{j=1}^J \varepsilon(\alpha, \theta)_j^2. \quad (3.4)$$

Die Optimierungsrechnung bietet sehr unterschiedliche Methoden für eine Minimumsuche.

Wir können sie grob danach einteilen

- ob das nächste Modellparametertupel gemäß einer gewissen Strategie aus mehreren Vorgängertupeln mit den dazugehörigen Werten der Kostenfunktion Q gewonnen wird
- oder ob nicht nur die Kostenfunktion Q direkt, sondern auch deren Gradienten bestimmt werden.

Ein möglicher, vielfach genutzter Ansatz für das Verwenden mehrerer Vorgängertupel mit den dazugehörigen Werten der RQS ist die Simplexmethode nach Nelder und Mead aus dem Jahr 1965. Er wird als Bibliotheksfunktion `fmins` in Matlab und Gnu-Octave zur Verfügung gestellt. Mit Simplex ist ein Vieleck gemeint, das $M + 1$ Ecken hat für M Parameter, in unserem Beispiel also drei Ecken für zwei Parameter, was in Abb. 3.2 durch schwarz umrandete weiße Punkte, die mit schwarzen, durchgezogenen Linien verbunden sind, skizziert ist. Zum Aufbau eines Startsimplex wird ein Startwerttupel $\mathbf{p}_{0,0}$ des Modellparametervektors gewählt und zu jedem Parameter ein weiterer Vektor, der eine Variation ΔP_m repräsentiert:

$$\mathbf{p}_{0,1} = \mathbf{p}_{0,0} + \begin{pmatrix} \Delta P_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \mathbf{p}_{0,m} = \mathbf{p}_{0,0} + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \Delta P_m \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \mathbf{p}_{0,M} = \mathbf{p}_{0,0} + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \Delta P_M \end{pmatrix} \quad (3.5)$$

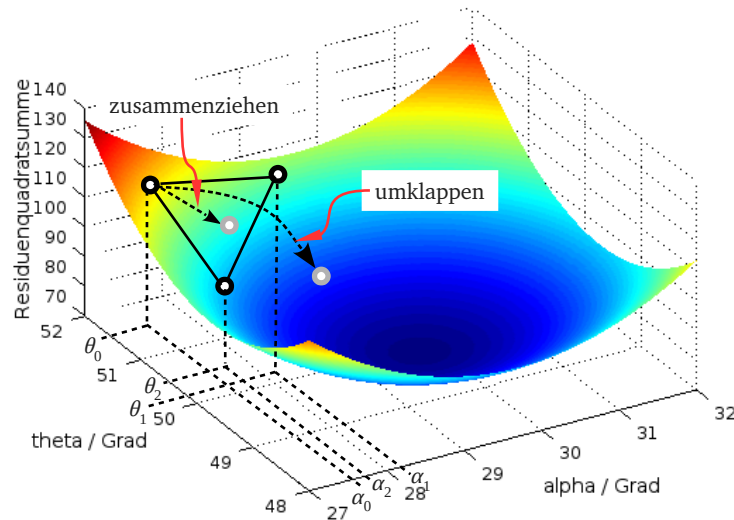


Abbildung 3.2: Residuenquadratsumme des Beispielmmodells mit 2 zu schätzenden Parametern und Simplex für Optimierungsverfahren nach Nelder und Mead.

Die Strategie ist es, diesen Simplex durch Zusammenziehen oder durch Umklappen dazu zu bewegen, in die Minimumsmulde herunter zu klettern, was in Abb. 3.2 durch hellgrau umrandete weiße Punkte und durch Pfeile skizziert ist. Die Entscheidung, wann er sich zusammen zieht und dadurch auch verkleinert und wann er umklappt und wann er notfalls auch wieder gestreckt (vergrößert) werden muss wird anhand dessen getroffen, ob die Werte des Zielfunktional Q kleiner werden oder nicht. Wenn er klein und flach im Tal angekommen ist, terminiert der Algorithmus. Hierzu gibt es dann noch verschiedene Ansätze, wie überprüft wird, ob er zu weit über dem Tal waagrecht festhakt und eigentlich noch eine zu große Mulde darunter liegt und eines der Simplexbeine angezogen werden muss.

Dieses Verfahren eignet sich auch für zu minimierende Zielfunktionale, die nicht so schön stetig differenzierbar, d.h. glatt, in der Talmulde sind. Für Messungen mit Beobachtungen, die nicht gaußverteilt streuen, sondern stark abweichende Punkte aufweisen, greift man auf andere Verteilungen zurück. Eine mögliche Verteilung kann die Laplaceverteilung sein

$$p(\varepsilon_1, \dots, \varepsilon_J | \alpha, \theta) \propto e^{-\frac{1}{2} \sum_{j=1}^J \frac{|\varepsilon(\alpha, \theta)_j|}{\sigma}} \quad (3.6)$$

was bedeutet, dass sie eine Exponentialverteilung der Absolutbeträge der Residuen ist. Bei dem zu minimierenden Exponenten

$$\min_{\alpha, \theta} \sum_{j=1}^J |\varepsilon(\alpha, \theta)_j| \quad (3.7)$$

bildet sich die Minimumsmulde durch die Betragsbildung zu einer ausgeprägt scharfen Spitze aus. Im Minimum gibt es also keine einheitliche Steigung. Je nach Richtung, von der man sich ihr nähert, gibt es sehr unterschiedliche Steigungen, was soviel bedeutet wie die Nichtstetigkeit der Steigungsfunktion. Steigungen im mehrdimensionalen Raum heißen Gradienten. Die Gradientenfunktion ist nicht stetig, die Funktion selber also nicht stetig differenzierbar. Für das Simplexverfahren spielt dies aber keine Rolle. Das Simplexverfahren hat aber den Nachteil, dass es sehr empfindlich darauf reagiert, welchen

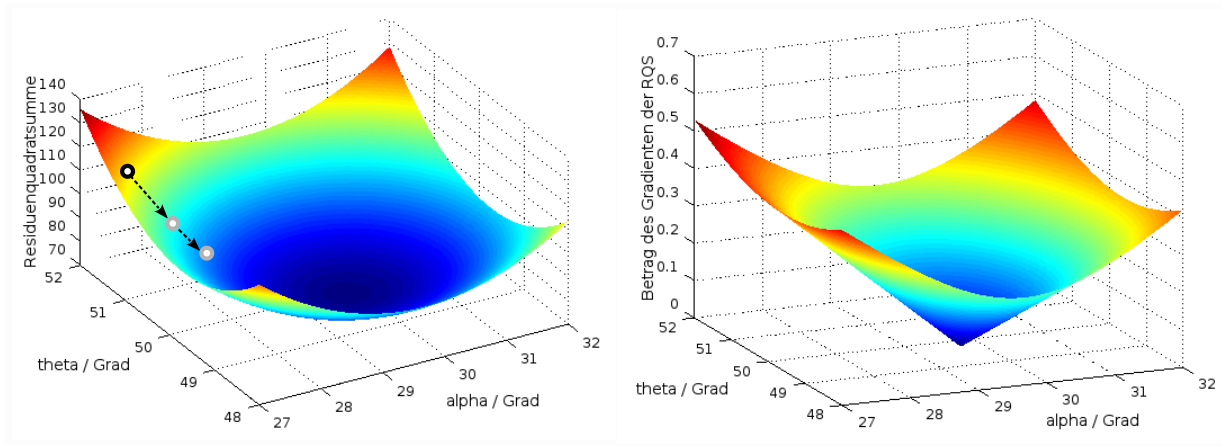


Abbildung 3.3: Residuenquadratsumme des Beispielsmodells mit 2 zu schätzenden Parametern und Gradienten eingezeichnet als Pfeile (*links*) und Beträge der Gradienten als Funktion der 2 zu schätzenden Parameter (*rechts*).

Startsimplex man zu Beginn auswählt. Es kann dadurch im harmlosen Fall relativ langsam werden, aber im ungünstigsten Fall in die Irre laufen und gar nicht konvergieren. Auch kann das Problem des Festhakens bei Optimierungsaufgaben mit vielen Parametern zu einem schwerwiegenden werden.

Für zu minimierende Zielfunktionale wie wir sie bei normalverteilten Residuen, also bei den rundlichen, stetig differenzierbaren Kostenfunktionen als Residuenquadratsummen haben, können wir Verfahren einsetzen, die nicht nur die Kostenfunktion Q direkt, sondern auch deren Gradienten ∇Q verwenden. Hierzu wird ein Startwertetupel (nicht wie bei der Simplexmethode $M + 1$) für die Modellparameter gebracht. Es wird der Gradient, also die Richtung des steilsten Abhangs der Kostenfunktion zusammen mit dem Wert der Steigung an der Position des Startwertetupels bestimmt. Dieses Verfahren beleuchten wir im folgenden Abschnitt anhand unseres Beispiels etwas genauer.

3.2 Gradientenverfahren für nichtlineare Zielfunktionale

Das Gradientenverfahren braucht wie alle anderen Methoden zur Schätzung von Parametern, die nicht linear in die Residuen eingehen, ein Tupel von Startwerten der Parameter, was wir in Abb. 3.3 als schwarz umrandeten weißen Punkt dargestellt haben. In Abb. 3.3 ist die Idee illustriert, wie sich die Position des Parametervektors verändert. Die Änderungsvektoren $\Delta \mathbf{p}$ werden aus der Richtung des steilsten Abhangs ermittelt, wie durch die Pfeile mit gestrichelter Linie skizziert. Der Gradient des Betrags eines Residuums an der Stelle des Startwertetupels (α_0, θ_0) ist der Vektor der partiellen Ableitungen

$$\nabla_{\alpha, \theta} \varepsilon_j|_{\alpha_0, \theta_0} = \begin{pmatrix} \frac{\partial}{\partial \alpha} \\ \frac{\partial}{\partial \theta} \end{pmatrix} \varepsilon_j|_{\alpha_0, \theta_0}. \quad (3.8)$$

Das auf die Spitze gestellte Dreieck heißt Nablaoperator. Die Schreibweise $|_{\alpha_0, \theta_0}$ lesen wir „an der Stelle“ (α_0, θ_0) . Für die Minimierung bestimmen wir die Stelle $(\hat{\alpha}, \hat{\theta})$, an der sich die Talsohle der Mulde der RQS-Funktion befinden. Anders gesagt suchen wir die Position im Modellparameterraum, die unten in der Mulde RQS-Funktion liegt, in der keine Steigung mehr vorhanden ist, d.h. der Gradient

ein Nullvektor ist, d.h.

$$\lim_{\alpha \rightarrow \hat{\alpha}, \theta \rightarrow \hat{\theta}} \left\{ \nabla_{\alpha, \theta} \sum_{j=1}^J \varepsilon(\alpha, \theta)_j^2 \right\} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (3.9)$$

Hierbei müssen wir uns aber darüber im Klaren sein, dass Gradient gleich Nullvektor nicht nur bei einem Minimum vorliegt, sondern auch bei einem Maximum und bei einem Sattelpunkt, weshalb wir im letzten Abschnitt ein Verfahren kurz vorstellen, das sich mit dieser Problematik besser auseinander setzt.

Wir ziehen den Nablaoperator in die Summe rein und verwenden die Kettenregel, äußere Ableitung mal innere Ableitung:

$$\nabla_{\alpha, \theta} \sum_{j=1}^J \varepsilon(\alpha, \theta)_j^2 = \sum_{j=1}^J \nabla_{\alpha, \theta} \varepsilon(\alpha, \theta)_j^2 = 2 \sum_{j=1}^J \varepsilon(\alpha, \theta)_j \nabla_{\alpha, \theta} \varepsilon(\alpha, \theta)_j \stackrel{!}{=} \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (3.10)$$

Die beiden Parameter α, θ fassen wir zusammen in einen Parametervektor \mathbf{p} und betrachten im folgenden den Allgemeinfall für M Parameter $\mathbf{p} = (P_1, \dots, P_M)^\top$. Die Residuen sind dann jeweils Funktion des Parametervektors $\boldsymbol{\varepsilon}(\mathbf{p}) = (\varepsilon_1(\mathbf{p}), \dots, \varepsilon_J(\mathbf{p}))^\top$, so dass die Kostenfunktion Q entsprechend Funktion des Parametervektors ist

$$Q(\mathbf{p}) = \boldsymbol{\varepsilon}^\top(\mathbf{p}) \boldsymbol{\varepsilon}(\mathbf{p}), \quad (3.11)$$

für die wir das Minimum suchen mit

$$\min_{\mathbf{p}} \{Q(\mathbf{p})\} \quad (3.12)$$

indem wir nach dem Nullstellenvektor der Gradientenfunktion

$$\nabla_{\mathbf{p}} Q(\mathbf{p}) = 2 \boldsymbol{\varepsilon}^\top \left(\nabla_{\mathbf{p}}^\top \boldsymbol{\varepsilon} \right) \quad (3.13)$$

suchen, für den Q ein Minimum und nicht ein Maximum oder Sattelpunkt annimmt. Dabei ist $\nabla_{\mathbf{p}}$ Spaltenvektor und $\nabla_{\mathbf{p}}^\top$ Zeilenvektor mit den partiellen Ableitungsoperatoren $\frac{\partial}{\partial P_m}$. Wir suchen dasjenige \mathbf{p} für das

$$\lim_{\mathbf{p} \rightarrow \hat{\mathbf{p}}} \nabla_{\mathbf{p}} Q(\mathbf{p}) = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \quad (3.14)$$

Bei der Suche nach dem Minimum wird mit einem Tupel von Parametern angefangen, also mit einem Startwertvektor \mathbf{p}_0 . Dann wird ein Schrittvektor $\Delta \mathbf{p}_0$ ermittelt, der gegangen wird, um zu einem Nachfolgepunkt im Raum der Modellparameter zu gelangen $\mathbf{p}_1 = \mathbf{p}_0 + \Delta \mathbf{p}_0$. Dies erfolgt solange, bis der Parametervektor beliebig dicht an das Minimum von Q gelangt sind. Es gibt damit eine gewisse Anzahl K von Iterationsschritten, derart dass der Vektor $\mathbf{p}_K = \hat{\mathbf{p}}$ für minimales Q als Tupel von Schätzwerten für die Modellparameter verwendet wird. Die einzelnen Iterationsschritte κ bedeuten eine Revision des Modellparametervektors

$$\mathbf{p}_\kappa = \mathbf{p}_{\kappa-1} + \Delta \mathbf{p}_{\kappa-1} \quad (3.15)$$

mittels *Inkrementvektoren* $\Delta \mathbf{p}_1, \dots, \Delta \mathbf{p}_\kappa, \dots, \Delta \mathbf{p}_{K-1}$.

Wir nehmen eine Taylorreihenentwicklung vor, über die wir die Gradientenfunktion des Zielfunktionals

als lineare Funktion eines *Inkrementvektors* $\Delta \mathbf{p}$ darstellen. Dazu entwickeln wir die Residuen ε_j in Taylorreihe bis zum linearen Term

$$\begin{pmatrix} \varepsilon_1(\mathbf{p}) \\ \vdots \\ \varepsilon_j(\mathbf{p}) \\ \vdots \\ \varepsilon_J(\mathbf{p}) \end{pmatrix} = \begin{pmatrix} \varepsilon_1(\mathbf{p}_\kappa) + \nabla_{\mathbf{p}} \varepsilon_1|_{\mathbf{p}_\kappa} \cdot \Delta \mathbf{p} \\ \vdots \\ \varepsilon_j(\mathbf{p}_\kappa) + \nabla_{\mathbf{p}} \varepsilon_j|_{\mathbf{p}_\kappa} \cdot \Delta \mathbf{p} \\ \vdots \\ \varepsilon_J(\mathbf{p}_\kappa) + \nabla_{\mathbf{p}} \varepsilon_J|_{\mathbf{p}_\kappa} \cdot \Delta \mathbf{p} \end{pmatrix}. \quad (3.16)$$

Hier bilden die partiellen Ableitungen der einzelnen Residuen nach allen Parametern eine $J \times M$ -Matrix, die *Jacobimatrix* genannt wird:

$$\mathbf{J}^{(\kappa)} = \begin{pmatrix} \frac{\partial \varepsilon_1}{\partial p_1} \Big|_{\mathbf{p}_\kappa} & \cdots & \frac{\partial \varepsilon_1}{\partial p_M} \Big|_{\mathbf{p}_\kappa} \\ & \ddots & \\ \frac{\partial \varepsilon_J}{\partial p_1} \Big|_{\mathbf{p}_\kappa} & \cdots & \frac{\partial \varepsilon_J}{\partial p_M} \Big|_{\mathbf{p}_\kappa} \end{pmatrix}, \text{ bzw. } J_{jm}^{(\kappa)} = \frac{\partial \varepsilon_j}{\partial p_m} \Big|_{\mathbf{p}_\kappa} \text{ mit } j = 1 \dots J; \quad m = 1 \dots M. \quad (3.17)$$

Der hochgestellte und in Klammern gesetzte Index κ an dem Symbol für die *Jacobimatrix*, soll bedeuten, dass dies die *Jacobimatrix* für den Parametervektor an der Stelle \mathbf{p}_κ ist, also für die Position der Modellparameter beim κ -ten Iterationsschritt. Der Begriff Jacobimatrix bedeutet nur, dass es sich um eine Matrix partieller Ableitungen handelt.

Die *Jacobimatrix* (benannt nach Carl Gustav Jacob Jacobi; auch *Funktionalmatrix*, *Ableitungsmatrix* oder *Jacobische* genannt) einer differenzierbaren Funktion $f: \mathbb{R}^n \mapsto \mathbb{R}^m$ ist die $m \times n$ -Matrix sämtlicher erster partieller Ableitungen.

Einsetzen dieser Jacobimatrix in Gl. (3.16) liefert dann

$$\begin{pmatrix} \varepsilon_1(\mathbf{p}) \\ \vdots \\ \varepsilon_J(\mathbf{p}) \end{pmatrix} = \begin{pmatrix} \varepsilon_1(\mathbf{p}_\kappa) \\ \vdots \\ \varepsilon_J(\mathbf{p}_\kappa) \end{pmatrix} + \mathbf{J}^{(\kappa)} \Delta \mathbf{p}. \quad (3.18)$$

Wir schreiben Gl. (3.18) in Vektorschreibweise, mit ε hier als Spaltenvektor der einzelnen Residuen.

$$\varepsilon(\mathbf{p}) = \varepsilon(\mathbf{p}_\kappa) + \mathbf{J}^{(\kappa)} \Delta \mathbf{p}. \quad (3.19)$$

Ferner setzen wir die Jacobimatrix in Gl. (3.13) ein

$$\frac{1}{2} \nabla_{\mathbf{p}} Q(\mathbf{p}) = \varepsilon^T(\mathbf{p}) \mathbf{J} \quad (3.20)$$

Wir setzen die Reihenentwicklung der Residuen Gl. (3.19) ein und den Gradienten Gl. (3.20) in Gl. (3.14) ein. So erhalten wir ein in $\Delta \mathbf{p}$ lineares Gleichungssystem:

$$\left(\varepsilon(\mathbf{p}_\kappa) + \mathbf{J}^{(\kappa)} \Delta \mathbf{p}_\kappa \right)^T \mathbf{J}^{(\kappa)} \stackrel{!}{=} (0 \dots 0) \quad (3.21)$$

wobei hier der Nullvektor wie der transponierte Modellparametervektor ein Zeilenvektor im \mathbb{R}^M ist.

Wir formen diese Gleichung, die in Zeilenvektorschreibweise ist, um in

$$\left(\mathbf{J}^{(\kappa)} \Delta \mathbf{p}_\kappa\right)^\top \mathbf{J}^{(\kappa)} \stackrel{!}{=} -\varepsilon(\mathbf{p}_\kappa)^\top \mathbf{J}^{(\kappa)} \quad (3.22)$$

und transponieren diese, um sie als lineares Gleichungssystem mit Spaltenvektoren vorliegen zu haben d.h.

$$\mathbf{J}^{(\kappa)\top} \mathbf{J}^{(\kappa)} \Delta \mathbf{p}_\kappa \stackrel{!}{=} -\mathbf{J}^{(\kappa)\top} \varepsilon(\mathbf{p}_\kappa). \quad (3.23)$$

Die Lösung dieses linearen Gleichungssystems liefert den Inkrementvektor $\Delta \mathbf{p}_\kappa$, mit dem wir den Schritt zur nächsten Position $\mathbf{p}_{\kappa+1} = \mathbf{p}_\kappa + \Delta \mathbf{p}_\kappa$ gehen. Die Iterationen enden, wenn der Betrag der Gradientenfunktion oder das Quadrat nahe Null, also kleiner als eine sehr kleine Zahl ϵ ist:

$$\varepsilon(\mathbf{p}_\kappa)^\top \mathbf{J}^{(\kappa)} \mathbf{J}^{(\kappa)\top} \varepsilon(\mathbf{p}_\kappa) < \epsilon \quad (3.24)$$

Für Kostenfunktionen, die wie in Abb. 3.3, ein sauber ausgeprägtes Minimum haben, und für die ein Startwerttupel zu den Modellparametern, das nah genug an der Mulde dran ist und sehr weit weg von irgendwelchen Maxima und Sattelpunkten, lässt sich auf diese Weise das Minimum finden.

Um eine sinnvolle Konvergenz zu erzielen, ist eine gute Vorkenntnis, also ein geeignetes Modellparametertupel \mathbf{p}_0 erforderlich, das in einigen Problemstellungen nur schwerlich, oder gar nicht verfügbar ist. Das bedeutet, dass nicht klar ist, ob der Gradient dazu führt, dass der Parametervektor in eine Mulde hinab klettert, oder auf einen Berggipfel oder entlang eines Sattelpunktes spaziert.

Die Numeriker haben Verfahren entwickelt, die auf dieser Grundidee den Gradienten der Zielfunktion (Summe der Quadrate der Residuen) zu berechnen basieren, aber ausgereifter sind hinsichtlich Stabilität und Konvergenz. Der sog. der *Levenberg-Marquardt-Algorithmus* ist eine solche gradientenbasierte Methode, die in den Numerikbibliotheken zu Matlab, Gnu-Octave, Python bis hin zu Labview, für nichtlineare Regression zur Verfügung gestellt wird. Die Grundidee dieser Methode wollen wir im Folgenden skizzieren.

3.3 Levenberg-Marquardt-Verfahren

Die Problematik, dass sich der Vektor $\Delta \mathbf{p}_\kappa$, den wir im Folgenden *Inkrementvektor* nennen, verirrt und auf einen Berggipfel oder entlang eines Sattelpunktes spaziert, soll anhand eines Beispielsignals illustriert werden. Das Beispiel ist ein zeitabhängiges Spannungssignal mit sinusförmigem Verlauf, dem als Störung zum einen ein Signal mit höherer Frequenz und kleinerer Amplitude aufmoduliert ist und zum anderen ein normalverteiltes weißes Rauschen. Die Amplitude von 5 V sei fest vorgegeben, aber die zu schätzenden Modellparameter seien die Wellenlänge und die Phasenlage des Signals. Abb. 3.4 zeigt das Signal als schwarze Rauten und in Rot das gesuchte „wahre“ Signal und als blaue, gestrichelte Kurve das Signal superponiert mit der höherfrequenten Modulation.

Abb. 3.5 zeigt die Summe der Quadrate der Residuen RQS als Kostenfunktion in drei verschiedenen Ansichten, die zeigen, dass es eine ausgeprägte Minimummulde gibt, aber ebenso ein ausgeprägtes Maximum und noch kleine Rippeln als Nebenminima und Nebenmaxima. Das Diagramm unten rechts zeigt die Betragsfunktion des Gradienten mit den vielen Nullstellen. Hier sehen wir, dass sowohl die

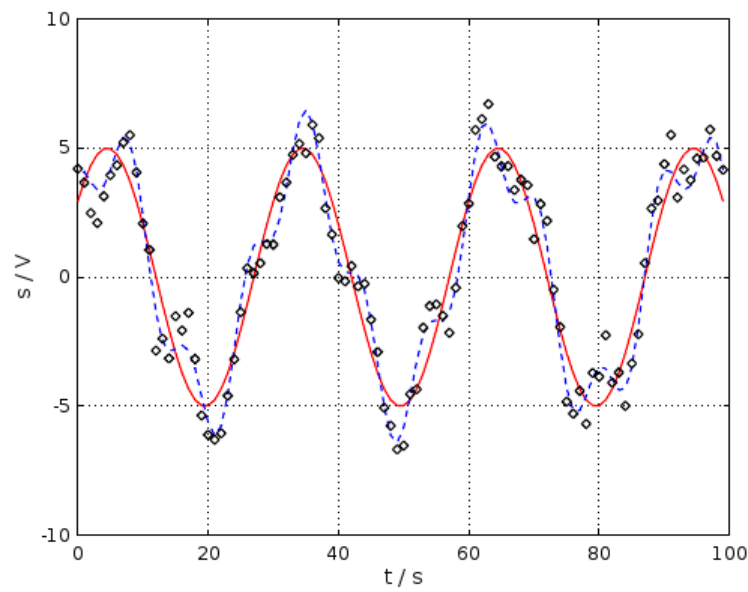


Abbildung 3.4: Beispiel eines sinusförmigen Signals.

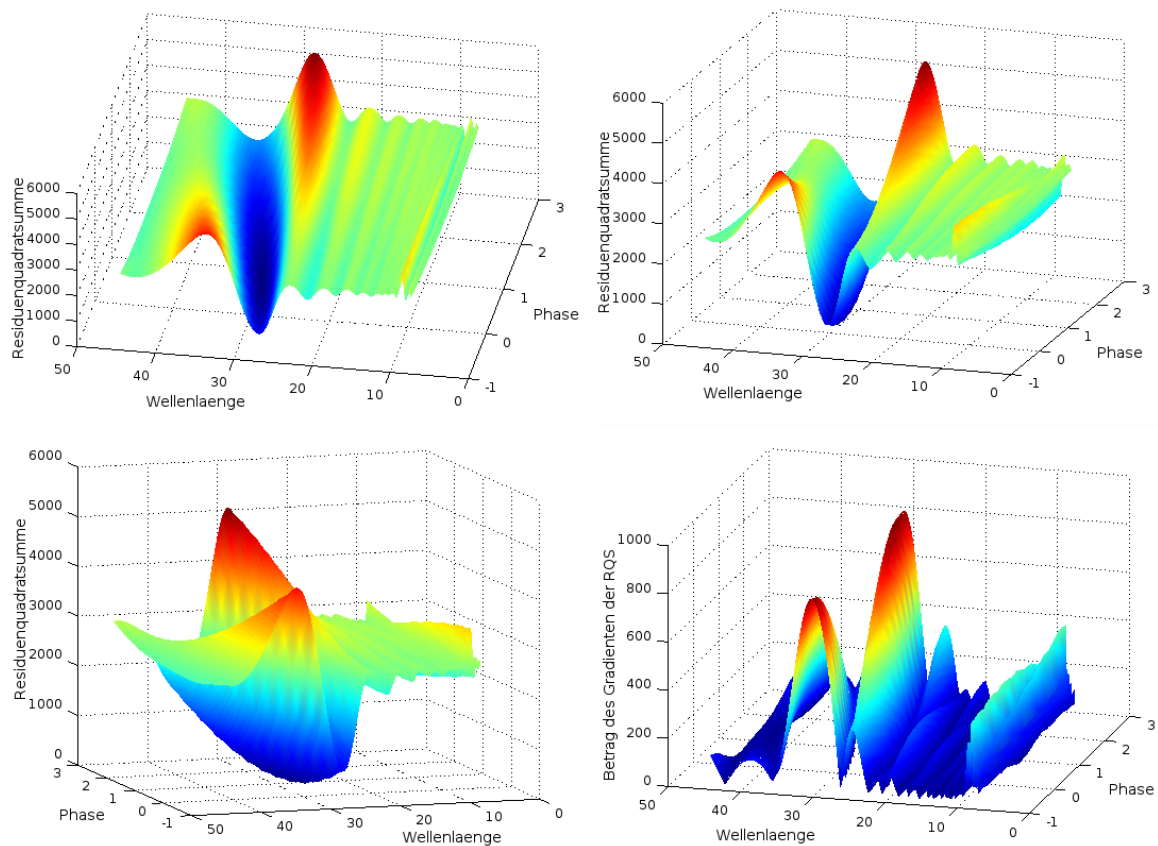


Abbildung 3.5: Residuenquadratsumme des Beispielsmodells eines sinusförmigen Signals in drei verschiedenen Perspektiven dargestellt (*oben und unten links*) und die Beträge der Gradienten als Funktion der 2 zu schätzenden Parameter (*unten rechts*).

Maxima als auch die Minima die Nullstellen bilden.

Der *Levenberg-Marquardt-Algorithmus* ist ein Gradientenverfahren, bei dem das lineare Gleichungssystem (3.23) durch Einführen eines Dämpfungsterms erweitert wird.

Mit der Gauss-Newton-Methode kann also passieren, dass die Summe der Residuen Q , siehe nicht bei jedem Iterationsschritt kleiner wird. Wenn jedoch der Richtungsvektor $\Delta \mathbf{p}_\kappa$ in die abfallende Richtung zeigt, so kann auch nur ein Bruchteil α des Richtungsvektors gegangen werden, solange gilt:

$$Q(\mathbf{p}_\kappa + \alpha \cdot \Delta \mathbf{p}_\kappa) < Q(\mathbf{p}_\kappa) \quad (3.25)$$

D. h., wenn bei der Gauß-Newton Methode Divergenz auftritt, so ist eine mögliche Verbesserung des Verfahrens, einen Bruchteil α des Richtungsvektors $\Delta \mathbf{p}_\kappa$ zu nehmen:

$$\mathbf{p}_{\kappa+1} = \mathbf{p}_\kappa + \alpha \cdot \Delta \mathbf{p}_\kappa \quad (3.26)$$

In anderen Worten, wenn der Inkrement-Vektor zu lang ist, dann muss man diesen eben kürzen. Das geht natürlich nur unter der Voraussetzung, dass die Richtung des Vektors auch die Summe der Residuen weiter verkleinert. Der Bruchteil α liegt zwischen

$$0 < \alpha < 1$$

Wenn der optimale Bruchanteil α annähernd Null ist, wird eine alternative Methode zur Behandlung der Divergenz verwendet, der sog. **Levenberg-Marquardt** Algorithmus. Die Normalengleichung wird modifiziert, indem der Inkrementvektor \mathbf{p}_κ in Richtung des steilsten Abfalls gedreht wird. Der Dämpfungsterm stellt einen Schrittweitenfaktor dar, der den Inkrementvektor so modifiziert, dass das Optimierungsverfahren besser konvergiert. Dieser Algorithmus wurde unabhängig voneinander von Levenberg (1944), Girard (1958), Wynne (1959), Morrison (1960) und Marquardt (1963) vorgeschlagen. In der Literatur wird heutzutage meist vom Levenberg-Marquardt-Algorithmus gesprochen.

Wir schreiben hier nochmal zusammen das zu minimierende Zielfunktional Q aus Gl. (3.13) links und die in Taylorreihe entwickelten Residuen $\varepsilon(\mathbf{p})$ (Gl. 3.19) auf

$$Q = \varepsilon(\mathbf{p})^\top \varepsilon(\mathbf{p}) \quad \varepsilon(\mathbf{p}) = \varepsilon(\mathbf{p}_\kappa) + \mathbf{J}^{(\kappa)} \Delta \mathbf{p}$$

und setzen diese ineinander ein, so dass die Kostenfunktion wie folgt aussieht

$$Q = \left(\varepsilon(\mathbf{p}_\kappa) + \mathbf{J}^{(\kappa)} \Delta \mathbf{p} \right)^\top \left(\varepsilon(\mathbf{p}_\kappa) + \mathbf{J}^{(\kappa)} \Delta \mathbf{p} \right). \quad (3.27)$$

Damit hatten wir im vorigen Abschnitt einen Ansatz erhalten, der ein bezüglich des Inkrementvektors $\Delta \mathbf{p}_\kappa$ lineares Gleichungssystem geliefert hatte. Wir notieren Gl. (3.23) hier noch einmal

$$\mathbf{J}^{(\kappa)\top} \mathbf{J}^{(\kappa)} \Delta \mathbf{p}_\kappa \stackrel{!}{=} -\mathbf{J}^{(\kappa)\top} \varepsilon(\mathbf{p}_\kappa).$$

Nun modifizieren wir das Gleichungssystem durch Hinzufügen eines Dämpfungsterms wie folgt

$$\left(\mathbf{J}^{(\kappa)\top} \mathbf{J}^{(\kappa)} + \mu \operatorname{diag} \left(\mathbf{J}^{(\kappa)\top} \mathbf{J}^{(\kappa)} \right) \right) \Delta \mathbf{p}_\kappa \stackrel{!}{=} -\mathbf{J}^{(\kappa)\top} \varepsilon(\mathbf{p}_\kappa) \quad (3.28)$$

wobei $\text{diag}(\mathbf{J}^{(\kappa)\top} \mathbf{J}^{(\kappa)})$ eine Diagonalmatrix ist, die aus den Matrixelementen auf der Hauptdiagonalen von $\mathbf{J}^{(\kappa)\top} \mathbf{J}^{(\kappa)}$ besteht und die Nebendiagonalelemente alle mit Null besetzt hat.

Der Parameter μ wird Dämpfungsparameter oder *Marquardtparameter* genannt. Die Dämpfungsstrategie besteht darin, dass für große Werte von μ die Länge $\sqrt{\Delta \mathbf{p}_\kappa^\top \Delta \mathbf{p}_\kappa}$ der Inkrementsschritte $\Delta \mathbf{p}_\kappa$ entsprechend verringert wird.

Diese Methode, die Länge des Inkrementvektors $\Delta \mathbf{p}_\kappa$ zu verbessern, also die Werte für den Dämpfungsparameter μ zu wählen, ist heuristisch. Es gibt eine Vielzahl von Methoden und Möglichkeiten und wir gucken uns eine davon an, um einen Einblick zu erhalten, wie soetwas aussehen kann.

Es wird eine Prüfgröße ρ_μ definiert, bei der die Differenz der Kostenfunktion Q für unterschiedliche Modellparametertupel verglichen wird mit der Differenz der die Kostenfunktion an der Nachfolgerstelle zu dem entsprechenden Punkt auf der approximierenden Tangentialebenen

$$\rho_\mu := \frac{Q(\mathbf{p}_\kappa) - Q(\mathbf{p}_\kappa + \Delta \mathbf{p}_\kappa)}{Q(\mathbf{p}_\kappa) - (\boldsymbol{\varepsilon}(\mathbf{p}_\kappa) + \mathbf{J}^{(\kappa)} \Delta \mathbf{p}_\kappa)^\top (\boldsymbol{\varepsilon}(\mathbf{p}_\kappa) + \mathbf{J}^{(\kappa)} \Delta \mathbf{p}_\kappa)} =: \frac{\Delta R(\mathbf{p}_\kappa, \Delta \mathbf{p}_\kappa)}{\Delta \tilde{R}(\mathbf{p}_\kappa, \Delta \mathbf{p}_\kappa)}. \quad (3.29)$$

Die Prüfgröße ρ_μ ist das Verhältnis

- der Differenz ΔR des Zielfunktionals Q zwischen der Vorgängerposition \mathbf{p}_κ und Q an der Nachfolgerposition $\mathbf{p}_\kappa + \Delta \mathbf{p}_\kappa$ des Parametervektors
- relativ zu der Differenz $\Delta \tilde{R}$ zwischen dem Q an der Stelle der Vorgängerposition \mathbf{p}_κ und dem Wert der Residuenquadratsumme an der Position auf der Tangentialfläche die an der Stelle \mathbf{p}_κ an die Kostenfunktion Q angelegt wird,

um dort das Zielfunktional zu approximieren.

Jetzt wird noch eine Vergleichsgröße, ein Schwellwert gebraucht, mit der die Prüfgröße verglichen wird, oder auch mehrere Schwellwerte. Hier stellen wir eine Methode vor, bei der zwei Schwellwerte β_0, β_1 als Entscheidungsgrundlage für die Wahl des Wertes μ verwendet werden, um den Inkrementvektor geeignet zu manipulieren. Für diese beiden heuristischen Schwellwertparameter β_0, β_1 können wir beispielsweise die Werte $\beta_0 = 0.2, \beta_1 = 0.8$ wählen.

Tabelle 1: Entscheidungen zur Behandlung des Marquadtparameters

$\rho_\mu \leq \beta_0$:	$\Delta \mathbf{p}_\kappa$ wird nicht akzeptiert <i>Gewährleistung der Konvergenz</i> μ wird vergrößert, z.B. durch Verdoppeln $\mu \rightarrow 2\mu$, und die neue zugehörige Korrektur $\Delta \mathbf{p}_\kappa$ wird berechnet
$\beta_0 < \rho_\mu < \beta_1$:	$\Delta \mathbf{p}_\kappa$ wird akzeptiert bei der Berechnung von $\Delta \mathbf{p}_{\kappa+1}$ wird als Anfangswert dasselbe μ gewählt
$\rho_\mu \geq \beta_1$:	$\Delta \mathbf{p}_\kappa$ wird akzeptiert <i>Effektivität</i> bei der Berechnung von $\Delta \mathbf{p}_\kappa$ wird als Anfangswert ein kleinerer Wert für μ gewählt, z.B. durch Halbieren $\mu/2$

Ein möglicher Algorithmus des Levenberg-Marquardt-Verfahrens

Input: Startwert \mathbf{p}_0 , Zielfunktion $\epsilon(\mathbf{p}_\kappa)$, Partielle Ableitungen $J = \epsilon'(\mathbf{p}_\kappa)$, Marquardt-Parameter μ , weitere Parameter $0 < \beta_0 < \beta_1 < 1$

for $k = 0, 1, \dots$

i) Berechne $\epsilon(\mathbf{p}_\kappa)$, $F'(\mathbf{p}_\kappa)$

ii) Bestimme den Korrekturvektor $\Delta\mathbf{p}_\kappa$

iii) Test, ob die Korrektur $\Delta\mathbf{p}_\kappa$ akzeptabel ist:

- $\rho_\mu \leq \beta_0$: Setze $\mu = 2\mu$ und berechne $\Delta\mathbf{p}_\kappa$ gemäß ii) neu
- $\rho_\mu \geq \beta_1$: Setze $\mu = \mu/2$ und behalte $\Delta\mathbf{p}_\kappa$

iv) Setze $\mathbf{p}_{\kappa+1} = \mathbf{p}_\kappa + \Delta\mathbf{p}_\kappa$

end

Diese Methode wurde hier auf das zuvor beschriebene Beispiel des in Abb. 3.4 dargestellten sinusförmigen Signals mit folgendem Gnu-Octave/Matlab-Skript ausprobiert:

```
function fit_sin_LMAC(gen_new)
% Stand: 191105
% Läuft unter Matlab und Octave
t = [0:99]'; % Sekunden
N = length(t);
a = 5; % Volt
lam = 30; % Sekunden
sigma = 0.13*a;
phi = 0.2*pi;
%
am = 0.3*a;
phim = 0.7*pi;
lamm = 0.3*lam;
%
st = a*sin(2*pi*t/lam + phi);
if gen_new
smod = st + am*sin(2*pi*t/lamm + phim);
s = smod + sigma*0.5*randn(1,N)';
save('pltSS_nonlin_leastsquare_sin_zB2.mat','t','s','sigma','a');
else
load('pltSS_nonlin_leastsquare_sin_zB2.mat');
end
%
% Startwerte geraten
%
```

```

p0 = [33; 0.23*pi];
x = [t(:) s(:)];
%
% Residuen, muessen als Spaltenvektor vorliegen
%
epsi = @(x, p) (x(:,2) - a*sin(2*pi*x(:,1)/p(1) + p(2)));
%
% partielle Ableitungen der Residuen epsi
% muss M Spalten und J Zeilen haben, wobei M die Anzahl der Fitparameter
% und J die Anzahl der Messpunkte ist
%
Jac = @(x, p) [(4*a*pi/p(1)^2)*x(:,1).*cos(2*pi*x(:,1)/p(1)+p(2)),...
-a*cos(2*pi*x(:,1)/p(1)+p(2))];
%
maxit = 5000; % Anzahl der Iterationen
tol = 1E-15; % Abbruchbedingung
beta0 = 0.2;
beta1 = 0.8;
%
% Startwert des Marquardtparameters
%
mu0 = 10;
%
% Optimierungsrechnung
%
p1=levenberg_marquardt(epsi,Jac,x,p0,mu0,beta0,beta1,maxit,tol);
%
% Ergebnisausgabe
%
disp(['gefittete Werte: lambda = ', ...
num2str(p1(1)),'; phi/pi = ',num2str(p1(2)/pi)]);

sf = @(x, p) a*sin(2*pi*x/p(1) + p(2));
figure(1);
plot( t, s,'MarkerFaceColor',[0 0 1],'MarkerSize',8, ...
'Marker','diamond','LineStyle','none')
hold on;
plot(t, st, 'g-', 'linewidth', 3)
plot(t, sf(t,p0), 'b-', 'linewidth', 2);
plot(t, sf(t,p1), 'r--', 'linewidth', 2);
legend('Messwerte','Sollfunktion','Startwerte','Fit' )

xlabel('t / s','fontsize', 14);
ylabel('s / V','fontsize', 14);

```

```

set(gca,'fontsize',12);
grid on;
print('-dpng','pltSS_nonlin_leastsquare_sin_zB2.png');
end
%
function p=levenberg_marquardt(F,DF,x,p0,mu0,beta0,beta1,maxit,tol)
k = 0;
mu= mu0;
p = p0;
gradQabs = abbruchkrit(F,DF,x,p);
sflag = 1;

while (gradQabs>tol) && (k<maxit) && (mu>tol) && (sflag<2)
[s,mu,sflag] = korrektur(F,DF,x,p,mu,beta0,beta1,tol);
if sflag==1
p=p+s;
end
gradQabs = abbruchkrit(F,DF,x,p);
k=k+1;
end
disp(['Anzahl Iterationen: ', num2str(k)]);
end
%
function [s,mu,sflag] = korrektur(F,DF,x,p,mu,beta0,beta1,tol)
s = Delta_p(F,DF,x,p,mu);
Q = F(x,p)'*F(x,p);
DR = Q - F(x,p+s)'*F(x,p+s);
DRtilde = Q - (F(x,p)+DF(x,p)*s)'*(F(x,p)+DF(x,p)*s);
sflag = 1;
if DRtilde > tol
if DR <= beta0*DRtilde
sflag = 0;
mu = 2*mu;
elseif DR >= beta1*DRtilde
mu = mu/2;
end
else
sflag = 2;
end
end
%
function s = Delta_p(F,DF,x,p,mu)
JTJ = DF(x,p)'*DF(x,p);
s = -(JTJ + mu*diag(diag(JTJ)))\ (DF(x,p)'*F(x,p));

```

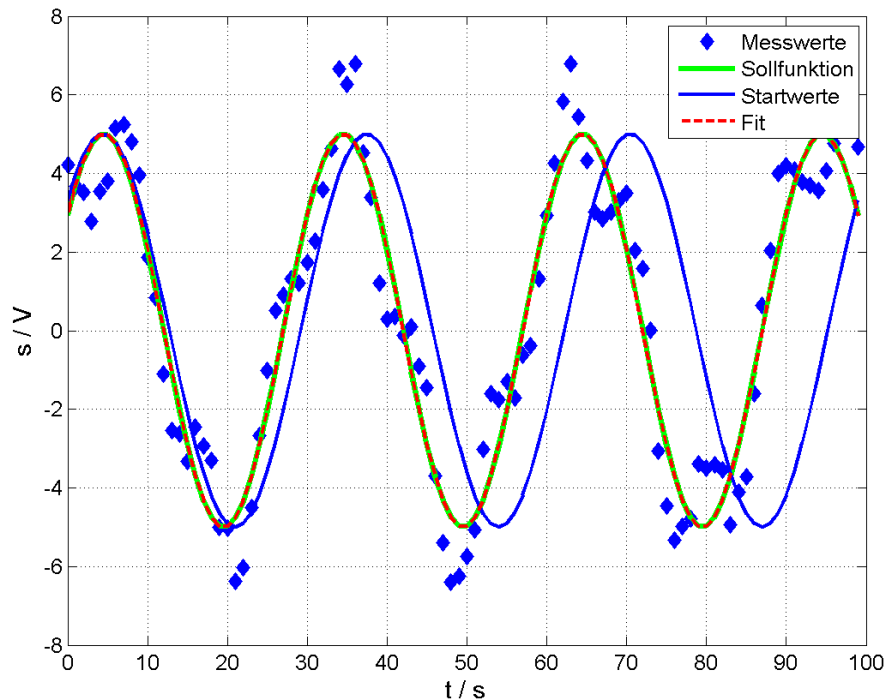


Abbildung 3.6: Beispiel eines sinusförmigen Signals mit Levenberg-Marquadt-Fit gemäß Gln. (3.28) und (3.29) zusammen mit den Entscheidungen aus Tabelle 1.

```
end
%
function gradQabs = abbruchkrit(F,DF,x,p)
gradQabs = ( F(x,p)'*DF(x,p) * DF(x,p)'*F(x,p) );
end
```

Dieses Gnu-Octave-/Matlab-Skript hat dazu das in Abb. 3.6 dargestellte Ergebnis mit den Werten

```
Anzahl Iterationen: 1878
gefittete Werte: lambda = 30.0505; phi/pi = 0.2051
```

geliefert.

Dieses Beispiel hat einen Modellansatz

$$s(t) = a \sin(2\pi \frac{t}{\lambda} + \varphi) + \varepsilon \quad \text{mit} \quad \varepsilon \sim \mathcal{N}(0, \sigma) \quad (3.30)$$

bei dem die Residuen in Richtung von s streuen und bei dem die Abtastpunkte entlang der Zeitachse t als Regressor vorgegeben sind, so dass wir einen Regressionsansatz der Gestalt

$$\epsilon_j = Y_j - f(X_j, \boldsymbol{\theta}) \quad \text{mit} \quad j = 1, 2, \dots, J \quad (3.31)$$

vorliegen haben. Die Kostenfunktion sieht damit wie folgt aus

$$Q = \sum_{j=1}^J (Y_j - f(X_j, \boldsymbol{\theta}))^2. \quad (3.32)$$

Für die nichtlineare Regression gibt es in Gnu-Octave beispielsweise folgende Funktion:

```
[f, p, cvg, iter] =leasqr (x, y, init, F);
```

Angewendet auf unser Beispiel vom sinusförmigen Signal kann sie in folgender Weise verwendet werden

```
function fit_sin(gen_new)
% Stand: 191105
% Läuft nur unter Octave, da leasqr nur unter Octave
% frei zur Verfügung steht. Bei Matlab gibt es
% einen ähnlichen Befehl, der heisst nlinfit
load('pltSS_nonlin_leastsquare_sin_zB2.mat');
p0 = [34; 0.23*pi];
sf = @(x, p) a*sin(2*pi*x/p(1) + p(2));
[f1, p1, kvg1, iter1, corp1, covp1, covr1, stdresid1, Z1, r21] = ...
leasqr (t(:), s(:), p0, sf);
fprintf(stdout,'gefittete Werte: lambda = %1.2f; phi/pi = %1.2f\n', ...
p1(1), p1(2)/pi );
t = [0:99]'; % Sekunden
N = length(t);
%
a = 5; % Volt
lam = 30; % Sekunden
sigma = 0.13*a;
phi = 0.2*pi;
%
am = 0.3*a;
phim = 0.7*pi;
lamm = 0.3*lam;
%
st = a*sin(2*pi*t/lam + phi);
figure(1);
plot( t, s, 'kd', ...
t, st, 'g-.;soll;', ...
t, sf(t,p0), 'b--;Startwerte;', ...
t, sf(t,p1), 'r-;gefittet;');
xlabel('t / s','fontsize', 14);
ylabel('s / V','fontsize', 14);
set(gca,'fontsize',12);
grid on;
```

end

und wir erhalten folgendes Ergebnis

gefittete Werte: $\lambda = 30.0505$; $\phi/\pi = 0.2051$

3.4 Matlabfunktionen zur Berechnung nichtlinearer Fits

Es gibt eine Vielzahl an Bibliotheksfunktionen, die nichtlineare Optimierungsverfahren implementiert haben. So gibt es z.B. z. B. die Matlab-/Octavefunktion

$\mathbf{X} = \text{fminbnd}(\text{FUN}, \mathbf{x1}, \mathbf{x2}, \text{OPTIONS})$. `fminbnd` versucht das lokale Minimum einer Funktion `FUN` mit einer Variablen im Intervall $\mathbf{x1} < \mathbf{X} < \mathbf{x2}$ zu finden. Als ‚OPTIONS‘ können u.a. gewählt werden `TolX` (Toleranz, wann abgebrochen werden soll) oder `MaxIter` (Maximale Anzahl an Iterationen, die durchgeführt werden sollen). Da bei unserem Beispiel mit dem Sinusfit zwei Parameter zu schätzen sind, verwenden wir die Funktion `fminsearch`. Hier wird nur ein Schätzwert (Startwert) $\mathbf{x0}$ angegeben. Die zu minimierende Kostenfunktion Q ist gegeben durch:

$$Q(\mathbf{p}) = \sum_j (a \cdot \sin(2 \cdot \pi \cdot t(j)/p(1) + p(2)) - s(j))^2 \rightarrow \text{Min.} \quad (3.33)$$

Der dazugehörige Matlab-Code lautet:

```
load('pltSS_nonlin_leastsquare_sin_zB2.mat');
a = 5; % Volt
p0 = [34; 0.23*pi];
% Ziel-/ Kostenfunktion:
Q = @(p) sum((a*sin(2*pi*t/p(1) + p(2))-s).^2) ;
tol = 1E-15;
maxit =1000;
p1= fminsearch(Q, p0, optimset('TolX',tol, 'MaxIter', maxit));
disp(['gefittete Werte: lambda =', num2str(p1(1)),'; ', ...
num2str(p1(2)/pi)])
```

Als Ergebnis erhält man:

```
gefittete Werte: lambda =30.0505; phi/pi=0.2051
```

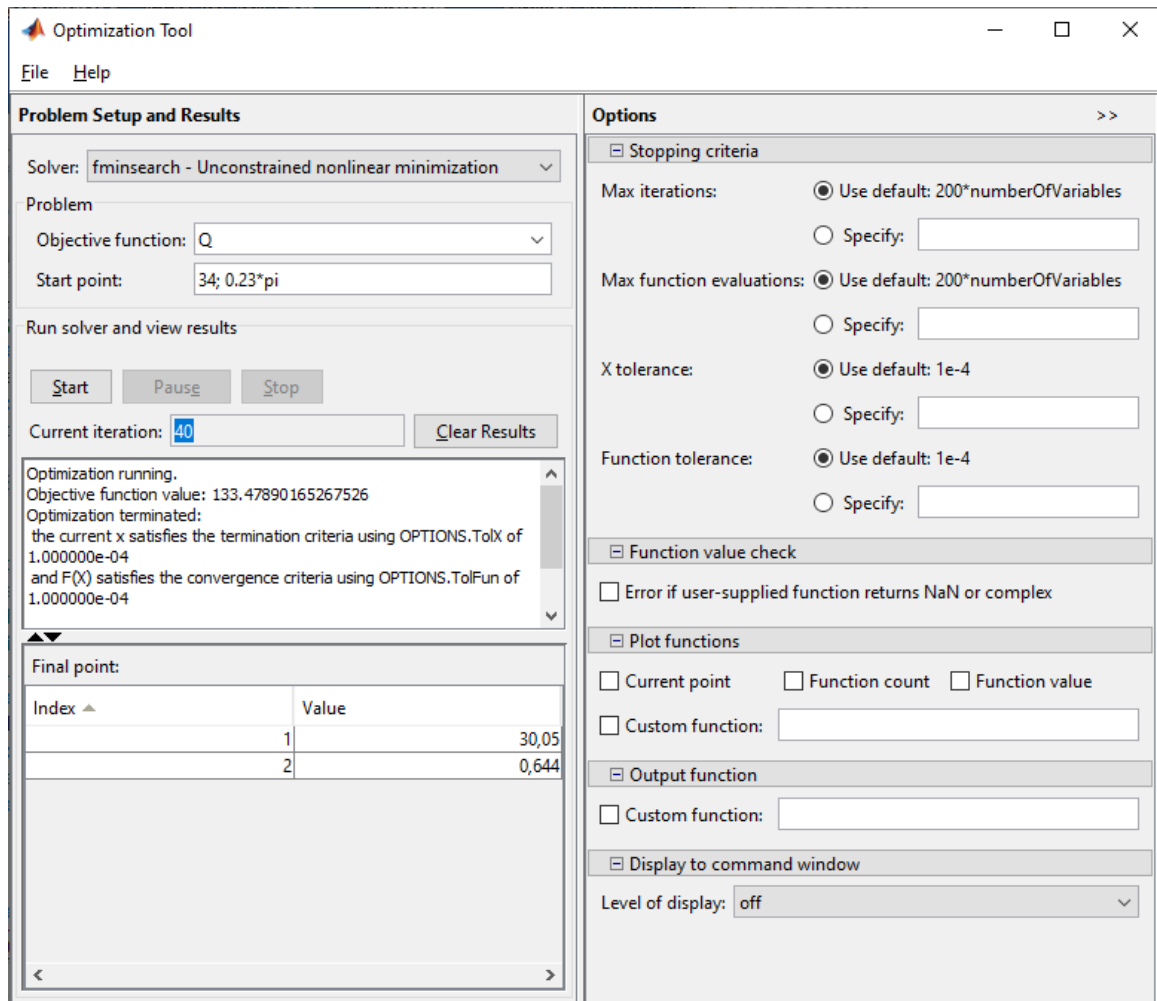



Abbildung 3.7: Optimtool von Matlab: Ergebnis $p(1) = 30,05$ und $p(2) = 0,644$

Tabelle 3.1 gibt einen Überblick über Matlabfunktionen bei Optimierungsproblemen, siehe dazu auch S. S. Rao [Rao09]. **lb** steht für ‚lower bound‘ und **ub** steht für ‚upper bound‘

Bei Matlab gibt es die in Abb. 3.7 gezeigte graphische Oberfläche, das sog. ‚Optimization tool‘ (Funktionsaufruf: ‚optimtool‘), die die in Tab. 3.1 aufgeführten Optimierungsverfahren und noch einige mehr zur Verfügung stellt. Dieses Tool eignet sich ganz gut, wenn man sich einen ersten Überblick über die in Matlab implementierten Optimierungsverfahren verschaffen will.

Tabelle 3.1: Optimierungsverfahren mit Matlab

Typ des Optimierungsproblem	Standardform für die Lösung unter Matlab/Octave	Matlab/Octave Funktion um das Problem zu lösen
Funktion von einer Variablen	Finde x , indem $f(x)$ minimiert wird mit $x_1 < x < x_2$	fminbnd
Minimierung ohne Randbedingung von einer Funktion oder mehreren Variablen	Finde \mathbf{x} , so dass $f(\mathbf{x})$ minimiert wird.	fminunc oder fminsearch
Lineares Problem	Finde \mathbf{x} durch Minimierung von $\mathbf{f}^T \mathbf{x}$ mit $A\mathbf{x} \leq \mathbf{b}$, $A_{eq}\mathbf{x} = \mathbf{b}_{eq}$, $\mathbf{lb} \leq \mathbf{x} \leq \mathbf{ub}$	linprog
Quadratisches Problem	Finde \mathbf{x} durch Minimierung von $\frac{1}{2}\mathbf{x}^T \mathbf{f} \mathbf{x}$ mit $A\mathbf{x} \leq \mathbf{b}$, $A_{eq}\mathbf{x} = \mathbf{b}_{eq}$, $\mathbf{lb} \leq \mathbf{x} \leq \mathbf{ub}$	quadprog
Minimierung von Funktionen von mehreren Variablen mit Nebenbedingungen	Finde \mathbf{x} durch Minimierung von $\frac{1}{2}\mathbf{x}^T \mathbf{f} \mathbf{x}$ mit $\mathbf{c}(\mathbf{x}) \leq \mathbf{0}$, $\mathbf{c}_{eq} = \mathbf{0}$ $A\mathbf{x} \leq \mathbf{b}$, $A_{eq}\mathbf{x} = \mathbf{b}_{eq}$, $\mathbf{lb} \leq \mathbf{x} \leq \mathbf{ub}$	fmincon
Minimax Problem	Finde das \mathbf{x} welches die Maxima von mehreren gegeben Funktionen $F_i(\mathbf{x})$ minimiert, so dass $\mathbf{c}(\mathbf{x}) \leq \mathbf{0}$, $\mathbf{c}_{eq} = \mathbf{0}$ $A\mathbf{x} \leq \mathbf{b}$, $A_{eq}\mathbf{x} = \mathbf{b}_{eq}$, $\mathbf{lb} \leq \mathbf{x} \leq \mathbf{ub}$	fminimax

3.5 Gradientenverfahren mit unterschiedlichen Varianzen

Die Varianzen (und Kovarianzen) der Residuen sind immer auch Funktion der Modellparameter. Als wir die Maximum-Likelihood-Methode eingeführt haben, hatten wir sie ausgeklammert und bei der Bestimmung des Minimums über die Summer der kleinsten Residuenquadrate weggekürzt. Sie lassen sich für den Fall ausklammern, bei dem die Streuung bezüglich des Modells für alle Messpunkte gleich ist, also

$$Q(\mathbf{p}) = \frac{\boldsymbol{\varepsilon}(\mathbf{p})^T \boldsymbol{\varepsilon}(\mathbf{p})}{\sigma^2(\mathbf{p})} \quad (3.34)$$

und

$$\nabla_{\mathbf{p}} Q(\mathbf{p}) = \frac{2}{\sigma^2(\mathbf{p})} \boldsymbol{\varepsilon}^T(\mathbf{p}) \mathbf{J} \stackrel{!}{=} \begin{pmatrix} 0 & \dots & 0 \end{pmatrix} \quad (3.35)$$

was einfach mit $\frac{\sigma^2(\mathbf{p})}{2}$ multipliziert werden kann, so dass wir mit den Residuen wie oben beschreiben weiter arbeiten können, indem wir diese in Taylorreihe entwickeln.

Für den Fall, dass es unterschiedliche Varianzen für Residuen der verschiedenen Beobachtungstupel

$(X_{1,j}, \dots, X_{N,j})$ gibt, ordnen wir $\sqrt{\sigma_j^2} = \sigma_j$ den Residuen zu mit

$$Q(\mathbf{p}) = \begin{pmatrix} \frac{\varepsilon_1(\mathbf{p})}{\sigma_1(\mathbf{p})} & \dots & \frac{\varepsilon_J(\mathbf{p})}{\sigma_J(\mathbf{p})} \end{pmatrix} \begin{pmatrix} \frac{\varepsilon_1(\mathbf{p})}{\sigma_1(\mathbf{p})} \\ \vdots \\ \frac{\varepsilon_J(\mathbf{p})}{\sigma_J(\mathbf{p})} \end{pmatrix} \quad (3.36)$$

d.h. kurz ohne jedesmal das „Funktion von \mathbf{p} “ oder „abhängig von \mathbf{p} “, also (\mathbf{p}) , mitzuschreiben

$$Q(\mathbf{p}) = \boldsymbol{\varepsilon}^\top \begin{pmatrix} \frac{1}{\sigma_1^2} & \dots & 0 \\ & \ddots & \\ 0 & \dots & \frac{1}{\sigma_J^2} \end{pmatrix} \boldsymbol{\varepsilon} \quad (3.37)$$

Je stärker eine Beobachtung j abweicht, also je größer die Varianz σ_j^2 , desto geringer fällt der Summand $\frac{\varepsilon_j^2}{\sigma_j^2}$ in der Kostenfunktion Q ins Gewicht, weshalb man diese auch als Gewichte $w_j = \frac{1}{\sigma_j^2}$ bezeichnet und die Matrix als Gewichtsmatrix \mathbf{W} also

$$Q(\mathbf{p}) = \boldsymbol{\varepsilon}^\top \mathbf{W} \boldsymbol{\varepsilon} \quad (3.38)$$

und mit

$$\frac{1}{2} \nabla_{\mathbf{p}} Q(\mathbf{p}) = \boldsymbol{\varepsilon}^\top(\mathbf{p}) \mathbf{W} \mathbf{J} \quad (3.39)$$

sieht unser Optimierungsproblem wie folgt aus

$$\lim_{\mathbf{p} \rightarrow \hat{\mathbf{p}}} \boldsymbol{\varepsilon}^\top(\mathbf{p}) \mathbf{W}(\mathbf{p}) \mathbf{J}(\mathbf{p}) = \begin{pmatrix} 0 & \dots & 0 \end{pmatrix}. \quad (3.40)$$

Die Gleichungen (3.23) und (3.28) werden damit zu

$$\mathbf{J}^{(\kappa)\top} \mathbf{W}(\mathbf{p}_\kappa) \mathbf{J}^{(\kappa)} \Delta \mathbf{p}_\kappa \stackrel{!}{=} -\mathbf{J}^{(\kappa)\top} \mathbf{W}(\mathbf{p}_\kappa) \boldsymbol{\varepsilon}(\mathbf{p}_\kappa). \quad (3.41)$$

und mit Dämpfungs- bzw. Marquardtparameter

$$\left(\mathbf{J}^{(\kappa)\top} \mathbf{W}(\mathbf{p}_\kappa) \mathbf{J}^{(\kappa)} + \mu \operatorname{diag}(\mathbf{J}^{(\kappa)\top} \mathbf{W}(\mathbf{p}_\kappa) \mathbf{J}^{(\kappa)}) \right) \Delta \mathbf{p}_\kappa \stackrel{!}{=} -\mathbf{J}^{(\kappa)\top} \mathbf{W}(\mathbf{p}_\kappa) \boldsymbol{\varepsilon}(\mathbf{p}_\kappa). \quad (3.42)$$

Es gibt aber auch Fragestellungen mit unterschiedlichen Varianzen σ_i^2 zu den direkten Messgrößen X_i . Ein Beispiel dafür ist eine Gerade in einer Ebene mit den Modellparametern α für den Winkel und d für den Abstand der Geraden vom Koordinatenursprung

$$\varepsilon_j = \begin{pmatrix} X_{1,j} \\ X_{2,j} \end{pmatrix} \cdot \begin{pmatrix} -\sin(\alpha) \\ \cos(\alpha) \end{pmatrix} - d. \quad (3.43)$$

Hier stehen die Residuen ε_j des j -ten Messpunktes $(X_{1,j}, X_{2,j})$ senkrecht auf der Modellgeraden, siehe Abb. 3.8, weil beide Messgrößen X_1 und X_2 gleichermaßen Zufallsgrößen sind und nicht eine davon ein Regressor ist die andere ein Regressand ist. Bei dieser *Least-Square* Methode spricht man auch von *Total Least-Square* Methode, weil „total“ alle Größen streuen können, also Zufallsgrößen sind.

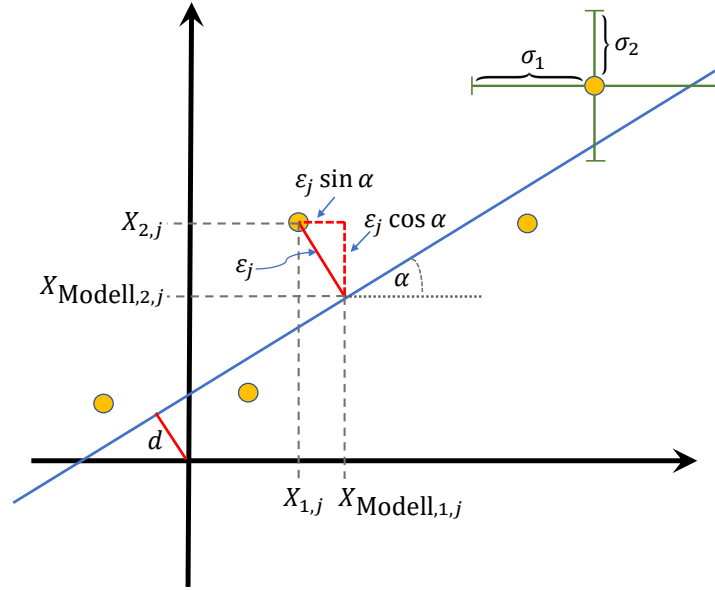


Abbildung 3.8: Veranschaulichung für Schätzung von Geradenparametern durch Minimierung der Summe der Quadrate der Residuen bei verschiedenen Varianzen σ_1^2 und σ_2^2 der beiden direkten Messgrößen X_1 und X_2

$$Q(\alpha, d) = \sum_{j=1}^J \left(\frac{\epsilon_j(\alpha, d) \sin(\alpha)}{\sigma_1} \right)^2 + \sum_{j=1}^J \left(\frac{\epsilon_j(\alpha, d) \cos(\alpha)}{\sigma_2} \right)^2. \quad (3.44)$$

Hier haben wir also die σ_i^2 als Varianzen der Abweichungskomponenten der Messgröße X_i (vektorierte Residuen!) für alle $j = 1, \dots, J$ Beobachtungen und nicht Varianzen σ_j^2 für das skalarwertige Residuum der j -ten Beobachtung. Die allgemeine Form mit Kovarianzen sieht dann also so aus

$$Q(\mathbf{p}) = \sum_{j=1}^J \vec{\epsilon}(X_{1,j}, \dots, X_{N,j}, \mathbf{p})^\top \Sigma(X_{1,j}, \dots, X_{N,j}, \mathbf{p})^{-1} \vec{\epsilon}(X_{1,j}, \dots, X_{N,j}, \mathbf{p})_j. \quad (3.45)$$

In diesem Zusammenhang erfolgt die nichtlineare Optimierung, indem nicht die Residuen in Taylorreihe entwickelt, sondern die Gradientenfunktion ∇Q direkt.

3.6 Robuste Schätzverfahren

Wir haben nun eine Kostenfunktion Q mit Gewichtungsfaktoren kennengelernt, Gl. (3.38), bei der die Gewichte die Kehrwerte der Varianzen sind $w_j = \frac{1}{\sigma_j^2}$, die sich eigentlich aus der Abweichung der Beobachtungen vom Modell ergeben. Wenn kein à priori Wissen über diese Varianzen vorliegt, aber bekannt ist, dass die Verteilung der Residuen aller Beobachtungen nicht gaußverteilt ist, soll das Problem dennoch mit der Maximum-Likelihood-Methode gelöst werden. Es kommt vor, dass einzelne Beobachtungen sehr weit von der Erwartung entfernt sind. Man sagt, dass sie signifikant abweichen. Die Gaußverteilung erlaubt dieses auch, denn sie hat einen Definitionsbereich für die Residuen ϵ , der von minus Unendlich bis plus Unendlich geht, ihre Ausläufer auch *Tails* genannt, sind unbegrenzt. Die

Wahrscheinlichkeit ist aber sehr gering, kleiner als 5 %, dass der Betrag der Residuen ε einen Wert annimmt, der größer als 1.65σ ist, aber er kann auftreten. Liegen deutlich mehr als 5 % der Beobachtungen außerhalb eines Vertrauensintervalls von beispielsweise $[-3\sigma, 3\sigma]$, dann ist dies ein Indiz dafür sein, dass das Modell unzureichend ist. Es kann sein, dass es auch nicht um ein vollständigeres Modell geht, sondern darum, nur die zum betrachteten Modell passenden Beobachtungen berücksichtigen zu wollen. Die übrigen Beobachtungen sollen als Ausreißer, die nicht zum eigentlichen zu untersuchenden Prozess gehören, nicht an dem *Least-Square-Fit* beteiligt werden oder zumindest ihre Beteiligung daran (ihr Gewicht) reduziert wird.

Die Idee für die Realisierung dieser Aufgabe ist, die Verteilung der Residuen durch Ungewichten so zu verändern, dass sie wieder die Gestalt einer Normalverteilung bekommt. Wir führen Gewichtungsfaktoren $w = \delta$ ein, die eine Funktion des Abstands ε einer Beobachtung vom Modell ist, so dass die Wahrscheinlichkeitsdichte der Abweichung $\tilde{\varepsilon}$ einer Normalverteilung folgt. Gesucht ist eine Gewichtsfunktion $\delta: \varepsilon \mapsto \delta(\varepsilon)$ für die gilt

$$\tilde{\varepsilon} = \delta(\varepsilon) \varepsilon \quad \Rightarrow \quad \tilde{\varepsilon} \sim \mathcal{N}(0, \sigma). \quad (3.46)$$

Mit Hilfe der Gewichtsfunktion werden aus den Residuen die Gewichtungsfaktoren $w_j = \delta_j = \delta(\varepsilon_j)$ berechnet. Da man die Residuen erst nach Schätzung der Modellparameter gewinnt, erfordert dies wie bei der nichtlinearen Optimierung auch sonst, einen iterativen Prozess.

Wir betrachten folgende Kostenfunktion

$$Q(\mathbf{p}) = \sum_{j=1}^J \delta_j \varepsilon_j(\mathbf{p})^2. \quad (3.47)$$

Zunächst werden die Parameter ungeachtet möglicher Ausreißer mittels *Least Square*-Methode ohne Gewichte

$$\min_{\mathbf{p}} Q(\mathbf{p}) \quad \text{mit} \quad Q(\mathbf{p}) = \sum_{j=1}^J \varepsilon_j(\mathbf{p})^2. \quad (3.48)$$

geschätzt mit \mathbf{p}_0 .

Wir betrachten wieder ein einfaches Beispiel, in dem es nur einen Parameter y gibt und eine Messgröße X_1 mit J Beobachtungen $(X_{1,1}, \dots, X_{1,J})$. Dabei wird davon ausgegangen, dass die Beobachtungen zu einer Grundgesamtheit, also zu einem Parameter gehören. Die Beobachtungswerte werden histogrammiert. Das Histogramm ist in Abb. 3.9 dargestellt. Für das Histogramm werden $K = 80$ Klassen (engl. *bins*) gewählt, die Anzahl der Beobachtungen umfasst $J = 2200$. Die Klassenbreite ist dann

$$\Delta\xi = \frac{1}{K} (\max\{X_{1,j}\} - \min\{X_{1,j}\}) \quad (3.49)$$

so dass $K + 1$ Klassengrenzen $\xi_{G,k}$ vorliegen, deren Indizes wir von Null bis K zählen:

$$\xi_{G,k} = k \Delta\xi + \min\{X_{1,j}\} \quad k = 0, \dots, K \quad (3.50)$$

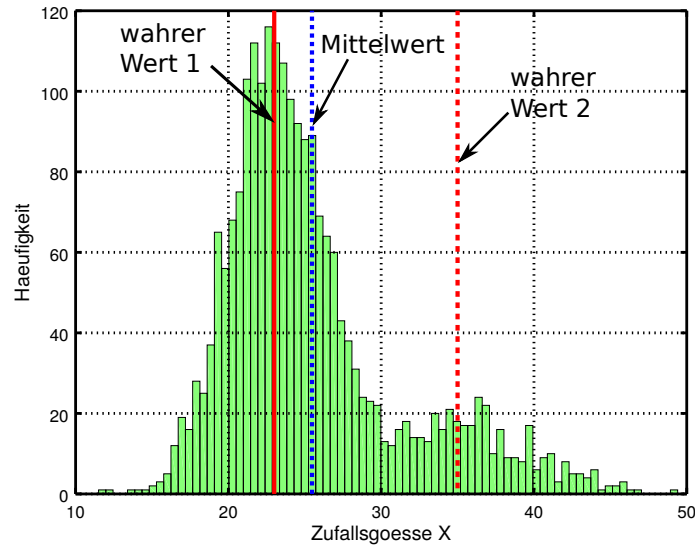


Abbildung 3.9: Beispiel für nicht erwartungsgemäße Beobachtungen

und die Mitte in der jeweiligen Klasse für den Wert der Klasse verwenden

$$\xi_k = k \Delta\xi + \frac{1}{2} \Delta\xi + \min \{X_{1,j}\} \quad k = 1, \dots, K \quad (3.51)$$

wobei die Anzahl der Klassen K ist.

Die Anzahl der Einzelbeobachtungen, für die gilt

$$\xi_{G,k-1} \leq X_{1,j} < \xi_{G,k}, \quad (3.52)$$

nennen wir *Häufigkeit* n_k . Bei der letzten Klasse, also für $k = K$ wird auch bei der rechten Intervallgrenze ein \leq . Damit ist die diskrete Funktion $n(\xi)$ die Häufigkeitsverteilung.

Der Schätzwert für den Parameter ist nach der Maximum-Likelihood / Least-Square-Methode der Mittelwert, der hier den Wert 25.47 annimmt. In Wirklichkeit gehören die Beobachtungswerte gar nicht zu nur einer Grundgesamtheit, sondern 1750 gehören zu dem Parameter $\mu_1 = 23.00$ und 450 zu $\mu_2 = 35.00$. Diese Wirklichkeit kennt aber niemand. Bei Betrachten des Histogramms sieht der Messtechniker jedoch, dass ein Nebenmaximum vorliegt, es keine reine Gaußverteilung ist. Die weiter entfernt liegenden Beobachtungen sollen unterdrückt werden durch entsprechende Wahl zugehöriger Gewichtungsfaktoren δ_j .

Eine mögliche Wahl für die Gewichtungsfaktoren ist die *Tuckey biweight*-Funktion.

$$\delta_j = \begin{cases} \left(1 - \left(\frac{X_{1,j} - y^{(\kappa)}}{c^{(\kappa)}}\right)^2\right)^2 & \text{falls } |X_{1,j} - y^{(\kappa)}| \leq c^{(\kappa)} \\ 0 & \text{sonst} \end{cases} \quad (3.53)$$

wobei κ der Zähler für den Iterationsschritt ist und $c^{(\kappa)}$ ein passend zur Anwendung zu wählender Faktor ist, hier beispielsweise

$$c^{(\kappa)} = \text{median } |X_{1,j} - y^{(\kappa)}| \quad (3.54)$$

Die Kostenfunktion, deren Gewichte von den iterativ bestimmten Residuen und damit von den iterativ

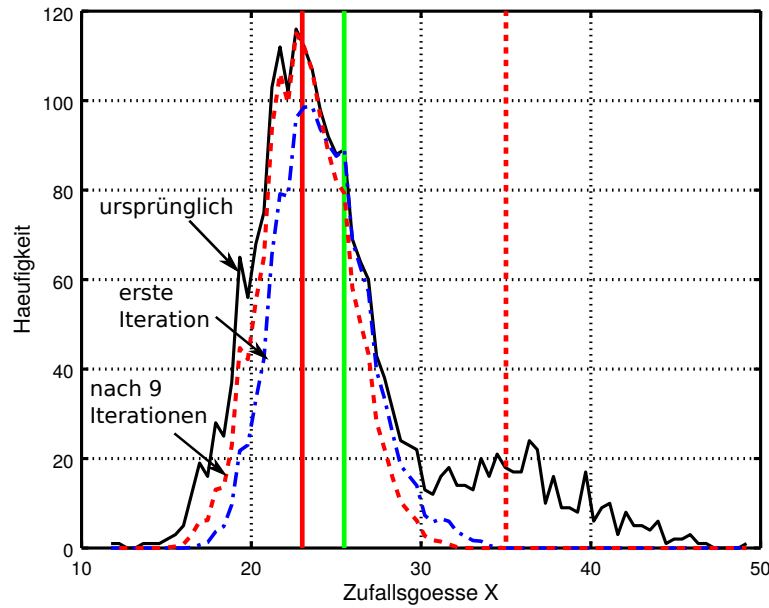


Abbildung 3.10: Umgewichten um die Verteilungsform der Normalverteilung anzupassen

geschätzten Parametern abhängt, stellt ein nichtlineares Optimierungsproblem $\min_{\mathbf{p}} Q(\mathbf{p})$ dar, mit

$$Q(\mathbf{p}) = \sum_{j=1}^J \varepsilon_j^2 \left(1 - \left(\frac{\varepsilon_j(\mathbf{p})}{c(\mathbf{p})} \right)^2 \right)^2 H(|\varepsilon_j(\mathbf{p})| - c(\mathbf{p})) H(c(\mathbf{p}) - |\varepsilon_j(\mathbf{p})|), \quad (3.55)$$

wobei H für *Heaviside*- oder Sprungfunktion steht.

Der Median von den Absolutbeträgen der Residuen $\varepsilon_j = X_{1,j} - y^{(\kappa)}$ wird dadurch gewonnen, dass man die Werte der Größe nach sortiert und dann den mittleren nimmt, also den, auf dem Platz in der Mitte der sortierten Anordnung liegt. Abb. 3.10 stellt die ursprüngliche Häufigkeitsverteilung aus Abb. 3.9 als durchgezogene, schwarze Linie dar. Nach dem ersten Iterationsschritt der Umgewichtung, hat sich die Verteilung so verändert, wie es durch die blaue gestrichpunktete Linie gezeigt wird, zu der ein Mittelwert von 24.20 gehört. Nach mehr als acht Iterationen wurde ein Mittelwert von 23.23 erzielt und die mit rot gestrichelter Verteilungskurve eingezeichnet sind.

Kapitel 4

Konzepte der Statistik für die Messdatenanalyse

4.1 Konzepte zum Lösen inverser Probleme

In den letzten drei Kapiteln (Vorlesungswochen) haben wir einen Einblick in die Vorgehensweise der Maximum-Likelihood-Methode erhalten:

1. Die Idee von *Modell* und *Lösen inverser Probleme* wurde genannt.
2. Zur Lösung inverser Probleme wurde das Schätzen von Parametern, die linear in die Modellgleichung eingehen, durch *lineare Regression* vorgestellt.
3. Ein Einblick in mögliche Verfahren zur Lösung inverser Probleme durch Schätzen von Parametern, die nicht-linear in die Modellgleichung eingehen, wurde gegeben. Hier wurden kurz drei Verfahren zur *nicht-linearen Optimierung* vorgestellt.

Wir haben dabei gesehen, dass es bei der Messdatenanalyse um die Schätzung von physikalischen Größen geht, die indirekt über das Messen anderer Größen, direkter Messgrößen, ermittelt werden. Die Aufgabe der Messtechnik lässt sich auch beschreiben als die Bestimmung des Wertes von physikalischen Größen durch Messverfahren, die auf physikalischen Prinzipien basieren. Die von einem Sensorsystem angezeigten Größen sind dabei im allgemeinen andere Größen als diejenigen, die zu messen sind.

Wir hatten in der ersten Vorlesung, in Abschnitt ??, das Beispiel „Beugung am Gitter“ betrachtet, um die Aufgabenstellung zu verstehen, dass man aus den einem Messsystem direkt zugänglichen Größen die gesuchten physikalischen Größen gewinnen möchte. Dabei wurde der Begriff des *Lösens inverser Probleme* eingeführt. Nachdem wir bereits Methoden zur Ermittlung von indirekten Messgrößen als Modellparameter kennen gelernt haben, haben wir die Chance uns unter den Ideen dieses Konzeptes mehr vorstellen zu können.

Darüber hinaus wurde in Abschnitt ?? schon mal erwähnt, dass es etwas unterschiedliche Konzepte oder Heransgehensweisen für Schätzverfahren gibt. Auf der einen Seite steht der sogenannte *frequentistische* Ansatz, bei dem allein Beobachtungswerte direkter Größen verwendet werden und die Annahmen über die Verteilung der Streuung der direkten Größen impliziert schon ins Schätzverfahren selber eingebaut ist. Der Begriff Frequenz von englischen *frequency* übernommen heißt hier *Häufigkeit* und nicht *Frequenz* wie wir das im Deutschen vielfach mit einer periodischen Bewegung verbunden verstehen. Aber auch die Deutsche Sprache kennt die Sprechweise: „Dies oder jenes Restaurant oder so wird gut frequentiert“ im Sinne von „wird häufig besucht“.

Eine Verallgemeinerung der Schätzmethoden wurde in den letzten Jahrzehnten in die Messdatenanalyse eingeführt, bei der auch die Modellparameter selber als Zufallsgrößen behandelt werden. Auch ihnen wird eine Streuung, also Wahrscheinlichkeitsverteilung zugrunde gelegt. Bei diesen Verfahren lassen sich Schätzwerte und Werte zur Unsicherheit der Modellparameter in die aktuelle Schätzung mit einbeziehen. Diese verallgemeinerten Methoden, die die Modellparameter als Zufallsgrößen betrachten, werden *bayesische* Methoden genannt.

Bevor wir diese beiden Ansätze (frequentistisch und bayesisch) gegenüber stellen und vergleichen, wollen wir folgende zwei Blickrichtungen, messtechnische Aufgaben zu betrachten, beleuchten:

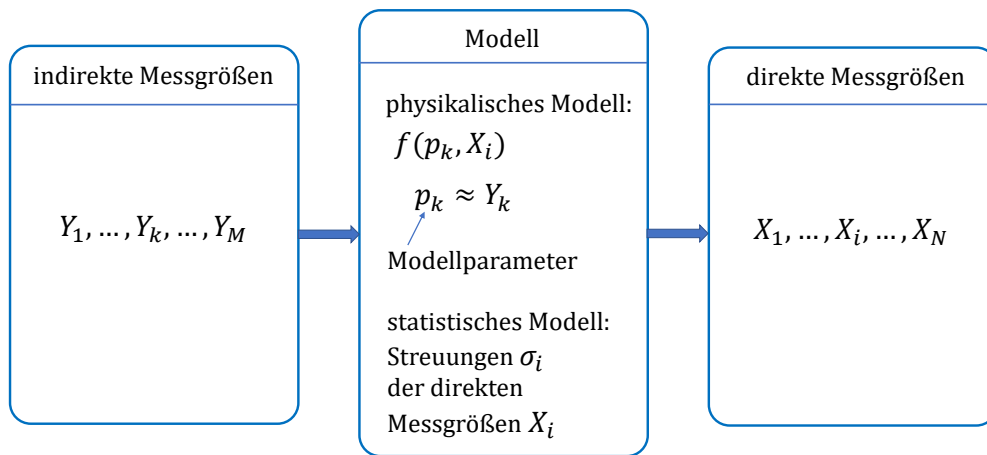


Abbildung 4.1: Modellbildung zur Gewinnung indirekter Messgrößen aus direkten Messgrößen

- als Modell von der Physik des Messprinzips gemeinsam mit dem statistischen Modell, dass die Größen Zufallsgrößen sind,
- als inverses Problem.

Ein Modell wird dargestellt durch eine Vorstellung darüber, wie die indirekten Messgrößen Y_m mit den direkten Messgrößen X_i verknüpft sind:

$$(Y_1, \dots, Y_M) \xrightarrow{\text{Modell}} (X_1, \dots, X_N) \quad (4.1)$$

Ein Messvorgang liefert Beobachtungen zu den direkten Messgrößen. Die wahren Werte der indirekten Größen bleiben verborgen. Die indirekten Größen können nur geschätzt werden, oftmals auch nur approximiert werden, weil die Modelle den physikalischen Sachverhalt nur annähernd beschreiben; denn die Realität ist deutlich komplexer und wird von vielfältigen Einflussfaktoren bestimmt. Die approximierenden und zufällig streuenden Größen, die die indirekten Messgrößen repräsentieren, sind die Modellparameter. Die Abweichungen der Modellparameter entstehen also zum einen durch vereinfachende Modellannahmen und zum anderen durch zufällige Einflüsse wie das thermische Rauschen von Elektronik oder Vibrationen im Laborraum.

Im vorigen Kapitel hatten wir bereits den Bezeichner \mathbf{p} für die Modellparameter verwendet: $\mathbf{p} = (P_1, \dots, P_M)$. Der Buchstabe \mathbf{p} steht hier einfach nur für den Anfangsbuchstaben von dem Begriff Parameter und der Fettdruck weist darauf hin, dass es ein Vektor mit vielen Einzelparametern sein kann und dient ferner dazu, das Symbol von dem für die Wahrscheinlichkeitsdichte zu unterscheiden. Diese wird mit p für den Anfangsbuchstaben von *probability* bezeichnet. Für die einzelnen Parameter P_m verwenden wir hier den Großbuchstaben, um die Unterscheidung zur Wahrscheinlichkeitsdichtefunktion zu haben, sowie in Anlehnung an die Konvention Großbuchstaben X_i und Y_m für die Messgrößen

zu verwenden.

$$P_1 \approx Y_1, \dots, P_M \approx Y_M \quad (4.2)$$

Der Schätzvorgang wird als *Lösen eines inversen Problems* interpretiert.

$$(X_1, \dots, X_N) \xrightarrow{\text{inverses Problem}} (Y_1, \dots, Y_M) \quad (4.3)$$

Dazu haben wir die *Maximum-Likelihood-Methode* kennen gelernt, und gesehen, wie aus ihr die Methode der kleinsten Residuenquadratsumme hervorgeht. Die Wahrscheinlichkeit, dass ein Messwert um den Wert ε von dem geschätzten Modell abweicht, folgt einer Normalverteilung. Mit anderen Worten kann man auch sagen, dass die Verteilung der Residuen eine Gaußverteilung ist. Die Begriffe Gaußverteilung und Normalverteilung verwenden wir synonym.

$$p(\varepsilon | \theta_1, \dots, \theta_M, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\varepsilon}{\sigma}\right)^2} \quad (4.4)$$

Die Symbolik $p(\varepsilon | \theta_1, \dots, \theta_M, \sigma)$ wird gesprochen:

Wahrscheinlichkeit p des Eintretens des Ereignisses ε gegeben die Parameter $\theta_1, \dots, \theta_M$ und σ .

Der Sprachgebrauch „Eintretens des Ereignisses ε “ stellt die Sprechweise der Statistik dar. In den angewandten Wissenschaften, Messtechnik, Elektrotechnik, etc. spricht man von der Wahrscheinlichkeit, dass die Messgröße oder hier die Abweichung der Messgröße vom Modell einen Wert ε annimmt oder dass dieser Wert „beobachtet“ wird.

In Kapitel 2 wurden Sie mit der Regressionsrechnung vertraut gemacht, die eine spezielle Anwendung der Maximum-Likelihood-Methode darstellt, bei der Größen beteiligt sind, die keine Zufallsgrößen darstellen, die Regressoren. Die Residuen und entsprechend die Regressanden sind die Zufallsgrößen, für die die in Gl. (4.4) formulierte Wahrscheinlichkeitsdichte angenommen wird. Diese Modellannahme, dass die *Residuen* als normalverteilt seien, hatten wir in Abschnitt 2.1 wie folgt formuliert:

Die abhängigen Größen Y , d.h. die *Regressanden* streuen und ihre Residuen sind normalverteilt mit Erwartungswert $E(\varepsilon) = 0$ und Varianz $\text{Var}(\varepsilon) = \sigma^2$. Die Residuen sind *unabhängig und identisch verteilt*, kurz u.i.v.

$$\varepsilon \stackrel{\text{u.i.v.}}{\sim} \mathcal{N}(0, \sigma). \quad (4.5)$$

Ein Residuum der linearen Regression mit Y als Regressand und X_i als Regressoren hat folgende Gestalt

$$\varepsilon_j = Y_j - \sum_{i=1}^M \theta_i X_{i,j} \quad (4.6)$$

wobei der Bezeichner Y der Regressionsrechnung die direkte Messgröße ist und die Regressoren ebenfalls zu den direkten Größen gehören, jedoch nicht als Zufallsgrößen mit Streuung, sondern mit vorgegebenen, deterministischen Werten.

Ein Residuum nach der Methode der kleinsten Quadrate für ein Modell mit genau nur einer einzigen Zufallsgröße X mit einem Modellparameter μ ist einfach

$$\varepsilon_j = X_j - \mu. \quad (4.7)$$

Ein Residuum nach der Methode der kleinsten Quadrate für ein Modell ohne Regressoren, also mit allen beteiligten Größen als Zufallsgrößen, lässt sich allgemein nicht mit einem Ansatz darstellen, in dem die Modellparameter linear sind. In Abschnitt 3.5 haben wir gesehen, dass bereits bei einer Geraden, für die sowohl die Abzissengröße X_1 als auch die Ordinatengröße X_2 streut, die Modellparameter nichtlinear in die Kostenfunktion eingehen. Die Modellparameter sind der Steigungswinkel α der Geraden relativ zur Abzisse und der Abstand d der Geraden vom Koordinatenursprung. Das Residuum ε_j stellt den Abstand des j -ten Messpunktes $(X_{1,j}, X_{2,j})$ zum Modell dar, siehe Abb. 3.8.

$$\varepsilon_j = \begin{pmatrix} X_{1,j} \\ X_{2,j} \end{pmatrix} \cdot \begin{pmatrix} -\sin(\alpha) \\ \cos(\alpha) \end{pmatrix} - d. \quad (4.8)$$

Wir haben in vorherigen Kapitel gesehen, dass bei der Lösung von inversen Problemen, bei denen im Gegensatz zur linearen Regression eine nichtlineare Verknüpfung der Modellparameter vorliegt, die Parameter nicht mehr durch Lösen eines linearen Gleichungssystems bestimmen werden können. Die Schätzung von Modellparametern besteht dann darin, die Modellparameter auszuprobieren, d.h. viele unterschiedliche Werte vorzugeben,

$$(P_1, \dots, P_M) \xrightarrow{\text{Modell}} (X_{\text{Modell},1,j}, \dots, X_{\text{Modell},N,j}) \quad (4.9)$$

und das Ergebnis der Modellrechnungen (4.9) mit den beobachteten Werten zu vergleichen, solange, bis das Ergebnis „gut zu den beobachteten Werten passt“, die Kostenfunktion minimal wird. Wir haben dazu im vorigen Kapitel einen Einblick erhalten, wie numerische entsprechende Optimierungsverfahren prinzipiell funktionieren. Im Fall der linearen Regression repräsentiert $\mathbf{p} = (\theta_1, \dots, \theta_M)$, im Fall der Geraden $\mathbf{p} = (\alpha, d)$ und im Fall der Einzelgröße $\mathbf{p} = \mu$.

Die beiden Annahmen, dass - wie wir in Kapitel 2 bereits bei der linearen Regression aufgeschrieben hatten - jede einzelne Beobachtung unabhängig und identisch verteilt (u.i.v) ist, sind Voraussetzung der Maximum-Likelihood-Methode:

- Jeder Messpunkt wurde unabhängig von allen anderen gewonnen.
- Jeder Messpunkt ist identisch verteilt (sie haben die gleiche Streuung).

Die gemeinsame Wahrscheinlichkeitsdichte, engl. *joint probability density*, der Residuen von j voneinander unabhängigen Messungen ist dann das Produkt der Wahrscheinlichkeitsdichten $p(\varepsilon_j|\mathbf{p}, \sigma)$ jedes Residuums:

$$p(\varepsilon_1, \dots, \varepsilon_J|\mathbf{p}, \sigma) = \prod_{j=1}^J p(\varepsilon_j|\mathbf{p}, \sigma) = \frac{1}{(\sigma\sqrt{2\pi})^J} \prod_{j=1}^J e^{-\frac{1}{2}\left(\frac{\varepsilon_j}{\sigma}\right)^2}. \quad (4.10)$$

Anhand des kleinen Beispiels mit dem ohm'schen Gesetz haben wir in Abschnitt ?? gesehen (Gln. (??), (??) und Gl. (??)), wie die Maximum-Likelihood-Methode in die Methode der kleinsten Summe der Residuenquadrate überführt wird: Unter Anwendung des Potenzgesetzes wird das Produkt der

Wahrscheinlichkeiten also

$$p(\varepsilon_1, \dots, \varepsilon_J | \mathbf{p}, \sigma) = \frac{1}{(\sigma\sqrt{2\pi})^J} e^{-\frac{1}{2} \sum_{j=1}^J \left(\frac{\varepsilon_j}{\sigma}\right)^2}. \quad (4.11)$$

Für die Maximum-Likelihood-Methode wird diese Wahrscheinlichkeitsdichte so gesehen, dass die in den Residuen steckenden Beobachtungen vorgegeben sind, und die Parameter \mathbf{p} durch Maximieren der Verteilung gesucht werden und sich daraus die Streuung der Residuen σ ergibt. Die Residuen sind Funktion der Beobachtungen $X_{1,j}, \dots, X_{N,j}$ und der Parameter \mathbf{p} , also $\varepsilon_j = \varepsilon(X_{1,j}, \dots, X_{N,j}, \mathbf{p})$. Deshalb formulieren wir die Wahrscheinlichkeitsdichte $p(\varepsilon_1, \dots, \varepsilon_J | \mathbf{p}, \sigma)$ als *Likelihood* $l(\mathbf{p}, \sigma | \{X_{1,1}, \dots, X_{1,J}\}, \dots, \{X_{N,1}, \dots, X_{N,J}\})$ um und Gl. (4.11) formen wir damit wie folgt um zu

$$l(\mathbf{p}, \sigma | \{X_{1,1}, \dots, X_{1,J}\}, \dots, \{X_{N,1}, \dots, X_{N,J}\}) = \frac{1}{(\sigma\sqrt{2\pi})^J} e^{-\frac{1}{2} \sum_{j=1}^J \left(\frac{\varepsilon(X_{1,j}, \dots, X_{N,j}, \mathbf{p})}{\sigma}\right)^2}. \quad (4.12)$$

Die Maximum-Likelihood-Methode bedeutet: Finde derartige Parameter \mathbf{p} , für die die *Likelihood* maximal wird

$$\max_{\mathbf{p}} \{l(\mathbf{p})\}. \quad (4.13)$$

Wir maximieren die in Gl. (4.12) gegebene *Likelihood*, indem wir ihren Exponenten minimieren

$$\min_{\mathbf{p}} \left\{ \sum_{j=1}^J \left(\frac{\varepsilon(X_{1,j}, \dots, X_{N,j}, \mathbf{p})}{\sigma} \right)^2 \right\}. \quad (4.14)$$

Hierbei fällt die Varianz, die für alle Residuen und Parameter denselben Wert hat, raus und das Optimierungsproblem, d.h. die Suche des Minimums, reduziert sich auf das Variieren der Parameter \mathbf{p}

$$\min_{\mathbf{p}} \left\{ \sum_{j=1}^J \varepsilon(X_{1,j}, \dots, X_{N,j}, \mathbf{p})^2 \right\}. \quad (4.15)$$

Die Aussage, dass jeder Messpunkt identisch verteilt ist, also die gleiche Streuverteilung hat, bedeutet aber nicht unbedingt, dass die Streubreite eine skalare Größe σ sein muss, sondern nur, dass für alle Beobachtungen die identischen Varianzen gelten.

Wenn der Messpunkt wie in Gl. (4.8) ein Vektor aus zwei Größen ist, die nicht dieselbe physikalische Dimension haben oder zwar dieselbe physikalische Dimension, aber dennoch unterschiedlich stark streuen, weil die Messprinzipien sich unterscheiden, dann stellt sich die Darstellung der Residuen komplizierter dar. Ein Beispiel für unterschiedliche physikalische Dimensionen kann wieder das mit der Messung eines elektrischen Widerstandes sein, bei dem X_1 eine Strommessung und X_2 eine Spannungsmessung repräsentieren kann. Ein Beispiel für dieselbe physikalische Dimension mit unterschiedlichem Streuverhalten kann ein Oberflächenmessgerät sein, bei dem die Sensorik in vertikaler Richtung zur Messung der Höhen einer Topographie ein optischer oder taktile Punktsensor sein kann und die horizontale Achse durch einen Schlitten mit einem Wegmesssystem in Form eines Glasmaßstabs sein kann.

Angenommen Größe X_1 zu Gl. (4.8) streue normalverteilt mit σ_1 und X_2 mit σ_2 , so müsste man

beispielsweise die Residuen aus Gl. (4.8) in Vektorkomponenten parallel zu den die beiden Größen repräsentierenden Achsen zerlegen

$$\vec{\varepsilon}_j = \begin{pmatrix} -\varepsilon_j \sin(\alpha) \\ \varepsilon_j \cos(\alpha) \end{pmatrix}, \quad (4.16)$$

so dass wir für die Wahrscheinlichkeitsdichte der j -ten Beobachtung den verschiedenen Streubreiten Rechnung tragen können:

$$p(\varepsilon_j | \alpha, d, \sigma_1, \sigma_2) \propto e^{-\frac{1}{2} \left(\frac{\varepsilon_j \sin(\alpha)}{\sigma_1} \right)^2} e^{-\frac{1}{2} \left(\frac{\varepsilon_j \cos(\alpha)}{\sigma_2} \right)^2} \quad (4.17)$$

In der Literatur werden zur Anpassung einer Geraden an zwei Größen unterschiedlicher Varianz sehr verschiedene Ansätze und Methoden vorgeschlagen.

Die Abweichungen der direkten Messgrößen zu den entsprechenden Werten, also die Residuen, die bei der Modellschätzung Gl. (4.9) errechnet werden, müssen als vektorielle Größen behandelt werden, wenn die verschiedenen direkten Messgrößen X_i eine unterschiedliche Varianz $\sigma_i^2 = \sigma_{i,i}$ oder noch allgemeiner auch Kovarianzen $\sigma_{i,k}$ aufweisen, die als Kovarianzmatrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \dots & \sigma_{1,N} \\ & \ddots & \\ \sigma_{N,1} & \dots & \sigma_N^2 \end{pmatrix} \quad (4.18)$$

zusammengefasst wird.

Die Wahrscheinlichkeitsdichteverteilung für den Vektor der Residuen der j -ten Beobachtung ist dann

$$p(\vec{\varepsilon}_j | \mathbf{p}, \Sigma) \propto e^{-\frac{1}{2} \vec{\varepsilon}_j^\top \Sigma^{-1} \vec{\varepsilon}_j} \quad (4.19)$$

wobei

$$\vec{\varepsilon}_j = \begin{pmatrix} X_{1,j} \\ \vdots \\ X_{N,j} \end{pmatrix} - \begin{pmatrix} X_{\text{Modell},1,j} \\ \vdots \\ X_{\text{Modell},N,j} \end{pmatrix}. \quad (4.20)$$

Die *joined probability density* ist dann entsprechend

$$p(\vec{\varepsilon}_1, \dots, \vec{\varepsilon}_J | \mathbf{p}, \Sigma) \propto \prod_{j=1}^J e^{-\frac{1}{2} \vec{\varepsilon}_j^\top \Sigma^{-1} \vec{\varepsilon}_j} = e^{-\frac{1}{2} \sum_{j=1}^J \vec{\varepsilon}_j^\top \Sigma^{-1} \vec{\varepsilon}_j}. \quad (4.21)$$

Für die Lösung eines inversen Problems geht es darum, die Parameter \mathbf{p} für gegebene Beobachtungen $X_{i,j}$ zu finden, die die Wahrscheinlichkeitsdichte in Gl. (4.21) maximieren. Wir schreiben deshalb die Residuenvektoren $\vec{\varepsilon}_j$ als Funktion der Beobachtungen $\vec{\varepsilon}(X_{1,j}, \dots, X_{N,j}, \mathbf{p})$ auf und verändern die Schreibweise für die Wahrscheinlichkeit so, dass wir als Likelihood l , die eine Funktion der Variablen \mathbf{p} und Σ für gegebene Beobachtungen $(X_{1,j}, \dots, X_{N,j})$ ist, ausdrücken. Gl. (4.21) bekommt dann folgende Gestalt

$$l(\mathbf{p}, \Sigma | \{X_{1,1}, \dots, X_{1,J}\}, \dots, \{X_{N,1}, \dots, X_{N,J}\}) \propto e^{-\frac{1}{2} \sum_{j=1}^J \vec{\varepsilon}^\top(X_{1,j}, \dots, X_{N,j}, \mathbf{p}) \Sigma^{-1} \vec{\varepsilon}(X_{1,j}, \dots, X_{N,j}, \mathbf{p})}. \quad (4.22)$$

Die Kovarianzmatrix ist wie die Residuen Funktion der Parameter und der Beobachtungen, also $\Sigma(X_{1,j}, \dots, X_{N,j}, \mathbf{p})$ und variiert werden die Parameter \mathbf{p} . Zur Maximierung der in Gl. (4.22) gegebenen *Likelihood* wird der Exponent minimiert, was in die Methode der kleinsten Residuenquadratsumme übergeht, also

$$\min_{\mathbf{p}} \left\{ \sum_{j=1}^J \tilde{\varepsilon}(X_{1,j}, \dots, X_{N,j}, \mathbf{p})^T \Sigma(X_{1,j}, \dots, X_{N,j}, \mathbf{p})^{-1} \tilde{\varepsilon}(X_{1,j}, \dots, X_{N,j}, \mathbf{p}) \right\}. \quad (4.23)$$

4.2 Konzept der bayesischen Verfahren zum Lösen inverser Probleme

Für komplexere Fragestellungen hinsichtlich der Varianzen und hinsichtlich der Möglichkeit, sukzessive neue Informationen durch weitere Beobachtungen zu bereits geschätzten Modellparametern hinzuzufügen und damit die Werte von interessierenden indirekten Messgrößen immer weiter zu verbessern, stoßen die gängigen Ansätze der Maximum-Likelihood-Methode an gewisse Grenzen. Insbesondere die Möglichkeit, bereits durch vergangene Optimierungsprozesse gewonnene Parameter durch Gewinnen neuer Beobachtungen zu verändern, ist nicht Gegenstand der „herkömmlichen“ Statistik. Mit „herkömmlich“ ist an dieser Stelle die *frequentistische* Statistik gemeint, bei der empirisch aus Beobachtungen Häufigkeitsverteilungen gewonnen werden, empirische Wahrscheinlichkeitsverteilungen (Likelihood) berechnet werden, und entsprechend eine Schätzung von Modellparametern erfolgt. Die lineare Regression ist ein Teilgebiet der *frequentistischen* Statistik

In der Statistik gibt es dazu zweierlei Blickrichtungen:

1. In der **frequentistischen Statistik** wird angenommen, dass der Wert einer indirekten Messgröße unbekannt, aber konstant/ fest ist und deshalb auch *wahrer Wert* genannt wird. In Bezug auf die direkt messbaren Größen wird angenommen, dass sie aufgrund des Mangels an Erkenntnis über alle möglichen Einflüsse und Abläufe im Messprozess streuen, so dass sie als Zufallsgrößen behandelt werden. Die Behandlung als Zufallsgröße bedeutet dabei, dass der Wahrscheinlichkeit für eine Beobachtung einen konkreten Wert anzunehmen eine Wahrscheinlichkeitsdichteverteilung zugrunde gelegt wird.

$$\underbrace{(Y_1, \dots, Y_M)}_{\text{unbek, konst, wahr}} \xrightarrow{\text{Messprozess}} \underbrace{(X_1, \dots, X_N)}_{\text{Zufallsgroessen}} \quad (4.24)$$

2. Umgekehrt ist die Weise, wie das Zustandekommen der Streuung von Größen gesehen wird, in der Statistik, die den Satz von Bayes zum Berechnen bedingter Wahrscheinlichkeiten verwendet. Dieses Gebiet der Statistik wird wegen der Verwendung des Satzes von Bayes auch **bayesische Statistik** genannt.

Hier werden die Beobachtungen der direkten Größen, als konkrete, feste Werte eingesetzt, um die Erkenntnis über die zu bestimmende indirekte Größe zu überprüfen bzw. zu vermehren. Die Vorstellung ist, dass durch den Mangel an Erkenntnis die indirekten Größen als Zufallsgrößen zu behandeln sind. Es wird nun also nicht mehr nur eine Likelihood-Verteilung berechnet, deren Maximum gesucht wird. Sondern es wird eine Kombination der Likelihood mit einer Wahrscheinlichkeitsdichteverteilung, die eine *à priori* Annahme über die indirekten Messgrößen darstellt,

berechnet und der Erwartungswert (das erste statistische Moment) und die Varianz (das zweite statistische Moment) dieser kombinierten Verteilungsdichte ermittelt.

Dabei repräsentieren die Wahrscheinlichkeiten für die zu erwartenden Beobachtungen eines Parameters (einer indirekten Messgröße), den Grad der Erkenntnis, vernünftiger Glaubwürdigkeit, über die indirekte Messgröße. In der englischsprachigen Literatur wird dies *Degree of Belief* genannt.

Vielleicht kann man sich diese Denkweise so vorstellen wie die Vorgehensweise eines Detektivs oder Kriminalkommissars beim Sammeln von immer mehr Indizien zum Aufklären eines Falls.

$$\underbrace{(X_1, \dots, X_N)}_{\text{Beobachtungen}} \xrightarrow{\text{inverses Problem}} \underbrace{(Y_1, \dots, Y_M)}_{\text{Zufallsgroessen}} \quad (4.25)$$

Die Erkenntnis über die indirekten Größen (Modellparameter Y_1, \dots, Y_M) wird mit jeder neuen Messkampagne κ revidiert

$$\left. \begin{array}{l} \underbrace{(X_1, \dots, X_N)}_{\text{neue Beobachtungen}} \\ \underbrace{(Y_1, \dots, Y_M)_{\kappa-1}}_{\text{Zufallsgroessen, vorher}} \end{array} \right\} \xrightarrow{\text{inverses Problem}} \underbrace{(Y_1, \dots, Y_M)_\kappa}_{\text{Zufallsgroessen}} \quad (4.26)$$

Da wir wie vorher detailliert dargelegt die indirekten Größen durch approximative Modellparameter ersetzen, also $Y_m \approx P_m$, sind es die Schätzungen zu den Parametern $\mathbf{p} = (P_1, \dots, P_M)$, die sozusagen „updated“ werden.

$$\left. \begin{array}{l} \underbrace{(X_1, \dots, X_N)}_{\text{neue Beobachtung}} \\ \underbrace{(P_1, \sigma_1^2, \dots, P_M, \sigma_M^2)_{\kappa-1}}_{\text{Zufallsgroesse, vorher}} \end{array} \right\} \xrightarrow{\text{inverses Problem}} \underbrace{(P_1, \sigma_1^2, \dots, P_M, \sigma_M^2)_\kappa}_{\text{Zufallsgroesse}} \quad (4.27)$$

Wir betrachten wie in Gl. (4.7) den einfachsten Fall einer Größe, die wir wie in Gl. (4.7) auch hier als Modellparameter $P = \mu$, der eine indirekte Messgröße Y repräsentieren soll, aufschreiben. Die Größe μ ist aus einer Stichprobe zu schätzen. Als à priori Information seien folgende Angaben bekannt:

- Der à priori Schätzwert der Modellgröße μ sei y_0 .
- Der à priori Schätzwert seiner Varianz σ^2 sei s_0^2 .
- Diese Größe sei normalverteilt, also verwenden wir als Verteilungsdichtefunktion p die Gaußverteilung.

Die neue Stichprobe $\{X_{1,1}, \dots, X_{1,J}\}$ sei vom Umfang deutlich kleiner, so dass für die Streuung der Varianz die χ^2 -Verteilung $p_{\chi^2, \nu}$ für $\nu = J - 1$ Freiheitsgrade verwendet wird. Hierzu müssen wir im Stoff vorgreifen. Diese Verteilungsdichtefunktion werden wir in der 5. Vorlesung behandeln. Sie wird verwendet als Verteilungsdichte für Varianzen und ist eine schiefsymmetrische Verteilung mit längerem

Ausläufer zu größeren Werten. Je kleiner der Stichprobenumfang, desto schiefer die Verteilung und ausgeprägter die Ausläufer.

Die gesamte Wahrscheinlichkeitsdichteverteilung $p(Y, \sigma)$ für die indirekte Messgröße Y bzw. deren Approximation als Modellparameter $P = \mu$ mit Streuung σ wird als Produkt folgender Wahrscheinlichkeiten berechnet,

- der Likelihood p_L ,
- der Verteilungsdichte $p_N(y|y_0, s_0)$ der Größe μ aus den à priori Informationen,
- der Verteilungsdichte $p_{\chi^2, \nu}(s)$ der Varianz σ^2 dieser Größe.

Die zu ermittelnde, bedingte Wahrscheinlichkeitsdichte wird so interpretiert, dass die Verteilung Funktion der Variablen μ und σ ist, gegeben die Beobachtungswerte der Stichprobe. Die Likelihood wird dabei betrachtet als die bedingte Wahrscheinlichkeitsdichte p_L der direkten Messgröße X_1 gegeben die Modellparameter μ und σ :

$$p_L(\{X_{1,1}, \dots, X_{1,J}\}|\mu, \sigma) = \prod_{j=1}^J \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{X_{1,j}-\mu}{\sigma}\right)^2} = l(\mu, \sigma|\{X_{1,1}, \dots, X_{1,J}\}). \quad (4.28)$$

Die Varianz σ^2 wird gemäß der χ^2 -Verteilung variiert. Das Variieren der Varianz bzw. deren Wurzel σ symbolisieren wir durch den Gebrauch der Variablen s . Das Variieren des Modellparameters μ symbolisieren wir durch den Gebrauch der Variablen y . Der Parameter μ wird ebenfalls variiert gemäß der Gaußverteilung, wir verwenden folgende Likelihood:

$$p_L(\{X_{1,1}, \dots, X_{1,J}\}|y, s) = \prod_{j=1}^J \frac{1}{\sqrt{2\pi}s} e^{-\frac{1}{2}\left(\frac{X_{1,j}-y}{s}\right)^2} = l(y, s|\{X_{1,1}, \dots, X_{1,J}\}). \quad (4.29)$$

Das Produkt ist die Wahrscheinlichkeit des Modellparameters und dessen Varianz gegeben die Stichprobe $\{X_{1,1}, \dots, X_{1,J}\}$ und die à priori Informationen y_0, s_0

$$p(y, s|\{X_{1,1}, \dots, X_{1,J}\}, y_0, s_0) = C p_L(\{X_{1,1}, \dots, X_{1,J}\}|y, s) p_N(y|y_0, s_0) p_{\chi^2, \nu}(s) \quad (4.30)$$

mit C als Normierungsfaktor derart zu wählen, dass die integrierte Wahrscheinlichkeit Eins ist, d.h.

$$\int_{-\infty}^{\infty} \int_0^{\infty} p(y, s|\{X_{1,1}, \dots, X_{1,J}\}, y_0, s_0) ds dy = 1.$$

Die Verteilungsdichte $p(y, s|\{X_{1,1}, \dots, X_{1,J}\}, y_0, s_0)$ ist Funktion der Variablen y und s mit vorgegebenen Parametern (festen Werten) $\{X_{1,1}, \dots, X_{1,J}\}, y_0, s_0$, wie es beispielhaft in Abb. 4.2 dargestellt wird.

Die Wahrscheinlichkeitsdichteverteilung $p_N(y|y_0, s_0)$ wird *Prior* genannt und die Wahrscheinlichkeitsdichteverteilung $p(y, s|\{X_{1,1}, \dots, X_{1,J}\}, y_0, s_0)$ wird *Posterior* genannt.

Wir integrieren Gl. (4.30) über s um eine Wahrscheinlichkeitsdichteverteilung zu gewinnen, die nur

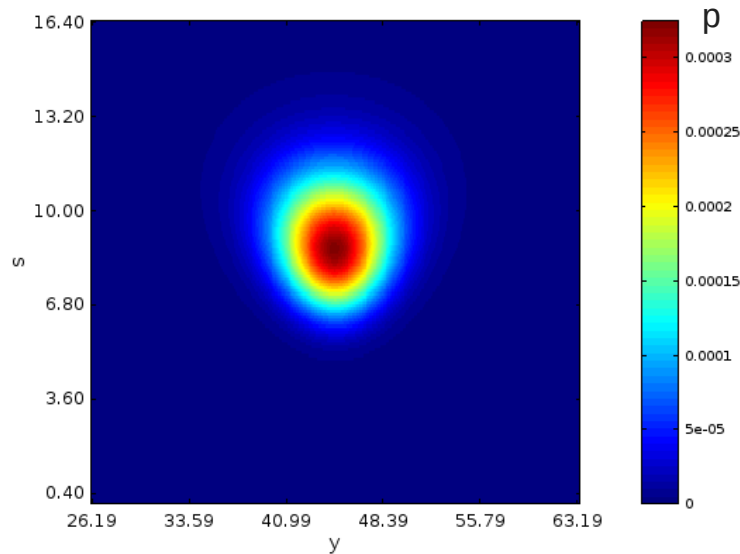


Abbildung 4.2: Posterior-Wahrscheinlichkeitsdichte als Funktion der beiden zu schätzenden Größen y und s .

noch Funktion von y ist

$$p(y|\{X_{1,1}, \dots, X_{1,J}\}, y_0, s_0) = \int_0^\infty p(y, s|\{X_{1,1}, \dots, X_{1,J}\}, y_0, s_0) \, ds \quad (4.31)$$

Diese Verteilung wird *Randverteilung* oder *Marginalverteilung* genannt. Allgemeiner gilt: Gegeben zwei Zufallsgrößen A und B mit gemeinsamer Verteilungsdichte $p(A, B)$; dann heißen die Verteilungen der einzelnen Zufallsgrößen A und B die Randverteilungen des Zufallsvektors (A, B) .

Als Schätzwert für Y berechnen wir den Erwartungswert des Modellparameters μ durch Integration über alle Werte y gewichtet mit der Wahrscheinlichkeitsdichte $p(y|\{X_{1,1}, \dots, X_{1,J}\}, y_0, s_0)$ aus Gl. (4.31)

$$y_1 = \int_{-\infty}^{\infty} y p(y|\{X_{1,1}, \dots, X_{1,J}\}, y_0, s_0) \, dy \quad (4.32)$$

wobei wir y_1 mit Index 1 schreiben, um anzuzeigen, dass dies die Revision der Schätzung des Modellparameters $\mu \approx Y$ gegenüber vorheriger Kenntnis mit Schätzwert y_0 ist.

Zu einem *Messergebnis* gehört jeweils ein Intervall, das ein Maß für die Unsicherheit des Wertes/ der Werte der Messgröße ist, siehe [VIM 2.9]

Messergebnis

Menge von Größenwerten, die einer Messgröße zugewiesen sind, zusammen mit jeglicher verfügbarer relevanter Information

ANMERKUNG 1 Ein *Messergebnis* enthält im Allgemeinen „relevante Information“ über die Menge der Größenwerte, von denen einige repräsentativer für die Messgröße sein können

als andere. Dies kann in Form einer Wahrscheinlichkeitsdichtefunktion ausgedrückt werden.

ANMERKUNG 2 Ein *Messergebnis* wird im Allgemeinen als ein einziger Messwert und eine *Messunsicherheit* ausgedrückt. Wird die Messunsicherheit für einige Zwecke als vernachlässigbar angesehen, kann das Messergebnis als ein einziger Messwert ausgedrückt werden. In vielen Bereichen ist dies die übliche Art, ein Messergebnis auszudrücken.

ANMERKUNG 3 In der traditionellen Literatur und in der vorhergehenden Ausgabe des VIM war das Messergebnis als ein Wert definiert, der einer Messgröße zugewiesen ist, und so erklärt, dass er eine Anzeige oder ein unkorrigiertes oder ein korrigiertes Ergebnis bedeutet, je nach Kontext.

Mit der in Anmerkung 1 angesprochenen Wahrscheinlichkeitsdichtefunktion sind Verteilungen wie die in den Gln. (4.12) und (4.22) vorgestellten Likelihoods oder der in Gl. (4.31) vorgestellten Posterior mit $p(y|\{X_{1,1}, \dots, X_{1,J}\}, y_0, s_0)$ gemeint. Die Likelihood wird für kleine Stichprobenumfänge ($N < 100$) anstelle durch eine Gaußverteilung durch eine Verteilung repräsentiert, deren Charakteristik im Kurvenverlauf überhöhter und mit länger auslaufenden Rändern ist. Sie wird *t*-Verteilung genannt. Wir werden später in der 5. Vorlesung auf diesen Verteilungstyp zurück kommen.

Allgemein betrachten wir für beide Fälle eine Wahrscheinlichkeitsdichtefunktion $p: y \mapsto p(y)$. Der Definitionsbereich für y reicht von minus bis plus Unendlich. Von Interesse ist der Kernbereich der Dichteverteilung für eine spezifizierte Wahrscheinlichkeit $1 - \alpha$, beispielsweise $1 - \alpha = 0.95$ oder 0.90 oder so. Dies ist die Wahrscheinlichkeit einer Größe, einen Messwert im Bereich (Intervall) $[y_{\min}, y_{\max}]$, anzunehmen

$$1 - \alpha = \int_{y_{\min}}^{y_{\max}} p(y) \, dy. \quad (4.33)$$

Die Breite des Intervalls gibt dann die *Messunsicherheit* an.

Das Intervall $[y_{\min}, y_{\max}]$, zu dem die Fläche $1 - \alpha$ unter der die Verteilungsdichte $p(y)$ repräsentierenden Kurve gehört, wird *Überdeckungsintervall* genannt. Es ist der verallgemeinerte Begriff aus der Metrologie für das, was in der *frequentistischen* Statistik *Vertrauensintervall*, engl. *Confidence Interval*, und in der *bayesischen* Statistik *Glaubwürdigkeitsintervall*, engl. *Credible Interval*, heißt. Während der Begriff *Vertrauensintervall* in der deutschsprachigen Literatur üblich ist, wird anstelle des Begriffs *Glaubwürdigkeitsintervall* üblicherweise auch in der deutschsprachigen Literatur der Begriff *Credible Interval* verwendet.

Die Verwendung unterschiedlicher Bezeichnungen, *Confidence Interval* und *Credible Interval*, sollen die Unterschiedlichkeit der Vorstellungen über den Charakter einer indirekten Messgröße Y zum Ausdruck bringen. Die Unterschiedlichkeit in der Vorstellung von einer indirekten Messgröße Y besteht darin, dass im Fall der *bayesischen* Statistik die indirekte Messgröße Y als Größe, die intrinsisch eine Zufallsgröße ist, betrachtet wird. Im Fall der *frequentistischen* Statistik wird Y als eine Größe betrachtet, die einen festen, wahren Wert hat, der aber nicht zugänglich ist. Nur der Modellparameter, der eine indirekte Messgröße approximiert und der indirekt aus anderen Zufallsgrößen gewonnen wird, wird als daraus resultierend einer Streuung unterliegend betrachtet.

Die Wahrscheinlichkeit $1 - \alpha$ wird in der *frequentistischen* Statistik *Vertrauensniveau* genannt. Die beiden Flächen unter den Ausläufern der Dichteverteilung stellt eine Wahrscheinlichkeit α dar, die in

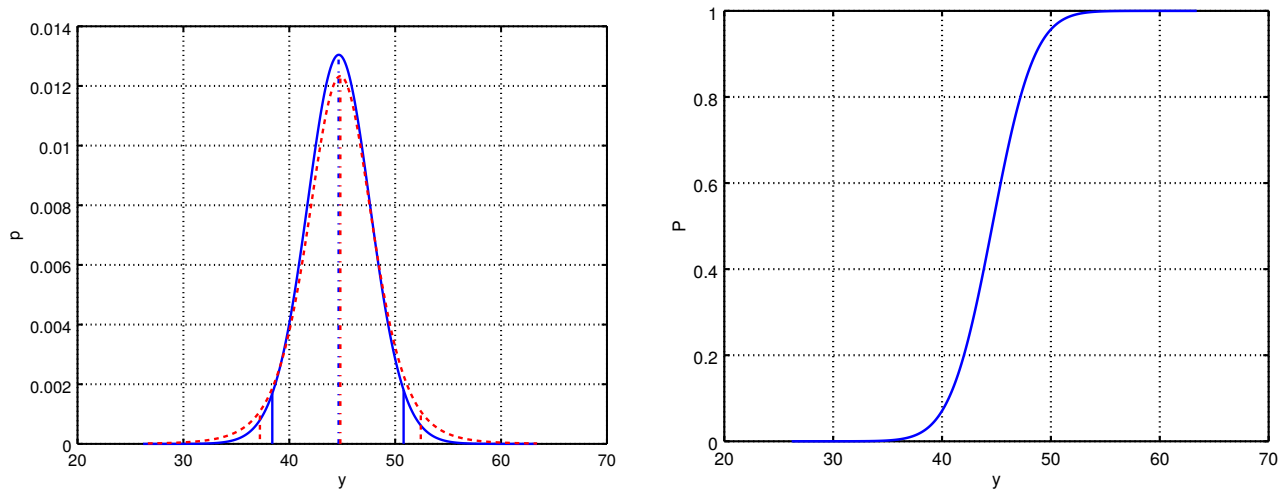


Abbildung 4.3: *Links:* Die blaue Kurve stellt die Randverteilung der Posterior-Wahrscheinlichkeitsdichte als Funktion der beiden zu schätzenden Größen Y mit *Credible Interval* dar und die rot gestrichelte Kurve die Student-t-Verteilung mit Vertrauensintervall. *Rechts:* Kumulierte Randverteilung der Posterior-Wahrscheinlichkeitsdichte, deren Umkehrfunktion verwendet wird, um die Intervallgrenzen des *Credible Intervals* zu ermitteln.

der *frequentistischen* Statistik *Signifikanzniveau* genannt wird.

Im folgenden werden wir sehen, wie eine *Messunsicherheit* im Rahmen der *bayesischen* Statistik gewonnen wird. Die Grenzen des *Credible Interval* $[y_{\min}, y_{\max}]$ bilden die Integrationsgrenzen y_{\min} und y_{\max} der *Posterior*, so dass

$$1 - \alpha = \int_{y_{\min}}^{y_{\max}} p(y|\{X_{1,1}, \dots, X_{1,J}\}, y_0, s_0) dy. \quad (4.34)$$

Die Integrationsgrenzen werden über die Umkehrfunktion der kumulativen Verteilung P berechnet. Die kumulative Verteilung P ist

$$P(y) = \int_{-\infty}^y p(y'|\{X_{1,1}, \dots, X_{1,J}\}, y_0, s_0) dy'. \quad (4.35)$$

Die Umkehrfunktion schreiben wir symbolisch als P^{-1} . Wir berechnen für das *Credible Interval*

$$y_{\min} = P^{-1}\left(\frac{\alpha}{2}\right) \quad \text{und} \quad y_{\max} = P^{-1}\left(\frac{1 - \alpha}{2}\right). \quad (4.36)$$

Abb. 4.3 links zeigt im Vergleich folgende Verteilungsdichten

- die Posterior als Funktion des Parameters y mit einem *Credible Interval* für eine Wahrscheinlichkeit von $1 - \alpha = 0.95$ (blaue, durchgezogene Kurve und blaue, durchgezogene Linien)
- die t -Verteilung (rote, gestrichelte Kurve), die sich für die in dem graphisch dargestellten Beispiel verwendeten Stichprobe mit Stichprobenumfang $J = 7$, folglich $\nu = J - 1 = 6$ Freiheitsgraden ergeben hat, sowie das Vertrauensintervall (rote, gestrichelte Linien), das aus dem Mittelwert der Stichprobe und der empirischen Standardabweichung des Mittelwerts multipliziert mit dem t -Quantil gewonnen wurde. Mit dem Begriff t -Quantil werden die Integrationsgrenzen dieser

Verteilung zu vorgegebenen Wahrscheinlichkeiten $\frac{\alpha}{2}$ und $1 - \frac{\alpha}{2}$ bezeichnet.

Abb. 4.3 rechts zeigt die kumulierte Posterior, aus deren Umkehrfunktion zu den Wahrscheinlichkeiten $\frac{\alpha}{2}$ und $1 - \frac{\alpha}{2}$ die Intervallgrenzen des *Credible Intervals* gewonnen werden.

Die Ermittlung der *Messunsicherheit* im Rahmen der *frequentistischen* Statistik unter Verwendung der *t*-Verteilung wird in den kommenden Wochen detailliert und mit Beispielen dargelegt werden. Sie wird einen großen Teil dieser Vorlesungsreihe ausmachen und Klausurrelevanz haben.

Kapitel 5

Wahrscheinlichkeiten und Hypothesentests

5.1 Konzept der statistischen Erwartung

Das *Signifikanzniveau* α und die dazu komplementäre Wahrscheinlichkeit, das *Vertrauensniveau* $1 - \alpha$ spielen eine zentrale Rolle bei der Bewertung von Messergebnissen. Als wir in Kapitel 4 die Konzepte der statistischen Auswertung behandelt hatten, haben wir diese beiden Begriffe gemeinsam mit dem des *Vertrauensintervalls* bzw. *Credible Interval* eingeführt, mit dem gemeinsamen Oberbegriff *Überdeckungsintervall*. Es ist das Intervall, in dem mit einer Wahrscheinlichkeit von $P(X_1) = 1 - \alpha$ die beobachteten Werte $X_{1,j}$ zur Größe X_1 liegen, das also die „meisten“ Beobachtungen abdeckt.

Die Idee hinter dem Begriff des *Vertrauensintervalls* oder des *Credible Intervals* ist, dass beim Bestimmen einer physikalischen Größe X eine Vorstellung (eine Erwartung) darüber existiert, wie die gemessenen oder beobachteten Werte verteilt sind. Das schließt die Erwartung von Position und Breite des Bereichs, in dem die Werte liegen sollten, ein. Diese Erwartung wird quantifiziert mit einer Angabe über die Wahrscheinlichkeit, dass die zu messende Größe diesen oder jenen Wert annimmt, eher noch in welchem Bereich (Intervall) der Wert zu erwarten ist. Wir *vertrauen* also darauf, dass wir einen Wert messen, der innerhalb eines gewissen Intervalls liegen müsste, oder sagen, dass es plausibel oder *glaubwürdig* (engl. *credible*) ist, dass die Beobachtungen in dem Intervall liegen. Statt zu sagen, dass wir *eine Vorstellung haben*, können wir auch sagen, dass wir *eine Hypothese aufstellen*.

Die Normalverteilung (Gaußverteilung) ist die grundlegende Verteilungsdichtefunktion zur Beschreibung zufälliger Prozesse, beispielsweise Zerfallsprozesse und Diffusionsprozesse. Bei rein stochastischen Prozessen wird von Zufallsgrößen ausgegangen, deren Wahrscheinlichkeitsdichteverteilung eine Normalverteilung ist

$$p(X) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{1}{2} \left(\frac{X-\mu}{\sigma} \right)^2}. \quad (5.1)$$

Verteilungen wie die Poisson-Verteilung und die Binominal-Verteilung gehen für unendlich viele Beobachtungen oder Stichprobenentnahmen in die Gaußverteilung über.

In der Wahrscheinlichkeitstheorie wird diese Tatsache exakt formuliert und ist der *zentrale Grenzwertsatz* (engl. *central limit theorem*) CLT. In einfacheren Worten können wir sagen:

Die Summe einer großen Anzahl von unabhängigen Zufallsgrößen folgt asymptotisch einer stabilen Verteilung.

Bei endlicher und positiver Varianz der Zufallsgrößen ist ihre Summe annähernd normalverteilt, was die Sonderstellung der Normalverteilung erklärt.

Hierzu wird das Ziehen eines Wertes (das Machen einer Beobachtung) als Zufallsgröße interpretiert. Damit betrachten wir eine Stichprobe als Folge von Zufallsgrößen. Sind die Elemente einer Stichprobe Zufallsgrößen, die *unabhängig und identisch verteilt*, kurz u.i.v., sind, so ist auch ihre Summe eine Zufallsgröße, die normalverteilt ist, wenn der Stichprobenumfang beliebig groß ist, also gegen Unendlich konvergiert. In der englischsprachigen Literatur ist *unabhängig und identisch verteilt* unter dem Begriff *independent and identically distributed*, kurz i.i.d., zu finden.

Liegt nun eine Einzelbeobachtung innerhalb der flachen Ausläufer (*tails*) der Wahrscheinlichkeitsdichteverteilung, also außerhalb des Vertrauensintervalls, so spricht man dann davon, dass der Wert

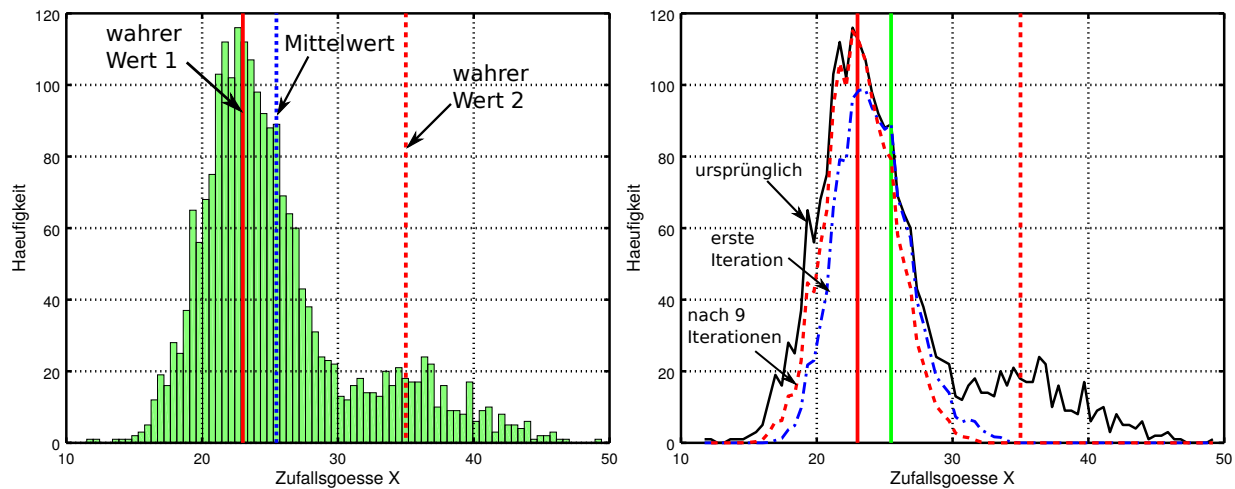


Abbildung 5.1: Beispiel für nicht erwartungsgemäße Beobachtungen (*rechts*) und Umgewichten um die Verteilungsform der Normalverteilung anzupassen (*links*)

(die Einzelbeobachtung) signifikant von dem, was zu erwarten ist, abweicht. Das heißt, dass ein Ereignis eintritt, das *nicht der Erwartung entspricht*. Somit *vermuten wir ein Problem*. Das Aufspüren des Problems erfolgt durch *Hypothesentests*. Da die Normalverteilung lange Ausläufer hat, gibt es eine kleine Wahrscheinlichkeit α dafür, dass es weiter abseits liegende Werte gibt. Deshalb gilt es zu untersuchen, ob solch ein Wert in der Tat im Rahmen der Normalverteilung weiter weg liegt oder aber ob er auf Basis eines anderen Effektes gewonnen wurde, ob er - statistisch formuliert - zu einer anderen Grundgesamtheit gehören könnte.

Der *Kolmogorow-Smirnow-Test* ist ein möglicher Test zum Prüfen, ob eine aus Beobachtungen empirisch ermittelte Verteilung zu einer für die statistische Analyse der Beobachtungen zugrundegelegten (erwarteten) Wahrscheinlichkeitsdichtefunktion passt.

5.2 Kolmogorow-Smirnow-Test

Das in Abb. 5.1 dargestellte Beispiel zeigt dieselben Diagramme und Daten, die wir in Abschnitt 3.6 bereits zur Veranschaulichung robuster Schätzverfahren behandelt hatten. Den robusten Schätzverfahren liegt die im folgenden erläuterte statistische Betrachtung zugrunde: Es liegt eine Annahme (Hypothese) darüber vor, welcher Gestalt die Wahrscheinlichkeitsverteilung der Beobachtungen ist und ob eine Reihe von Beobachtung zu einer gemeinsamen Verteilung gehören oder zu unterschiedlichen Verteilungen. Technisch gesehen stellen sich die Fragen, die statistisch formuliert lauten „gehören Beobachtungen zu einer Verteilung“ oder „gehören Beobachtungen zu mehreren verschiedenen Verteilungen“, als

- „stammen die Messwerte aus einem Prozess“ oder „aus unterschiedlichen Prozessen“,
- „werden alle Messwerte von demselben Effekt beeinflusst“ oder „gibt es für verschiedene Gruppen aus der Messreihe unterschiedliche beeinflussende Effekte“.

Die Annahmen bzw. Hypothesen bezüglich der Verteilungen zur Wahrscheinlichkeit von Beobachtungswerten werden geprüft. Diese Art der Prüfung wird *Hypothesentest* genannt. Es gibt unter-

schiedliche Typen von Hypothesentests, einer davon ist der *Kolmogorow-Smirnow-Test*. Er ist ein statistischer Test auf Übereinstimmung zweier Wahrscheinlichkeitsverteilungen.

Der Test verwendet die kumulierten Wahrscheinlichkeitsfunktionen

$$P(X) = \int_{-\infty}^X p(X') dX'. \quad (5.2)$$

Das hat den Vorteil, dass die Verteilung der empirischen Häufigkeiten ohne Histogrammierung, d.h. ohne Einteilung in Klassen, erfolgen kann.

Wir wollen damit prüfen, ob die Beobachtungen des in Abschnitt 3.6 zur robusten Schätzung betrachteten Beispiels der Annahme folgt, zu der Grundgesamtheit *genau einer* Zufallsgröße X_1 zu gehören. Ferner nehmen wir an, dass ihre Ausprägungen (Beobachtungswerte) unabhängig und identisch der Normalverteilung folgen.

Die Nullhypothese lautet:

Die Beobachtungen zu X_1 gehören zu einer normalverteilten Grundgesamtheit.

Wir prüfen, ob die Stichprobe $(X_{1,1}, X_{1,2}, \dots, X_{1,J})$ bezüglich des Mittelwertes $y = \frac{1}{J} \sum_{j=1}^J X_{1,j}$ und der empirischen Varianz $s^2 = \frac{1}{J-1} \sum_{j=1}^J (X_{1,j} - y)^2$ normalverteilt ist. Die kumulierte Normalverteilung, mit der wir die kumulierte Verteilung der relativen Häufigkeiten vergleichen, ist

$$P(X) = \frac{1}{\sqrt{2\pi}s} \int_{-\infty}^X e^{-\frac{1}{2}\left(\frac{X'-y}{s}\right)^2} dX'. \quad (5.3)$$

Die kumulierte Verteilung der relativen Häufigkeiten gewinnen wir darüber, dass wir die Werte der Stichprobe in aufsteigender Reihenfolge sortieren

$$X_{1,k_1} \leq X_{1,k_2} \leq \dots \leq X_{1,k_J} \quad (5.4)$$

Die Teilstichprobe mit der Beobachtung mit dem kleinsten Wert (X_{1,k_1}) hat die relative kumulierte Häufigkeit $\frac{1}{J}$, die Teilstichprobe mit den zwei kleinsten Beobachtungen (X_{1,k_1}, X_{1,k_2}) hat die relative kumulierte Häufigkeit $\frac{2}{J}$. Die Teilstichprobe

$$(X_{1,k_1}, X_{1,k_2}, \dots, X_{1,k_j}) \quad \text{die rel. kum. Häufigk.} \quad H(X_{1,k_j}) = \frac{j}{J} \quad (5.5)$$

Abb. 5.2 stellt zum Vergleich die Wahrscheinlichkeitsfunktion (kumulierte Verteilungsdichte) $P(X)$ gemäß Gl. (5.3) als rote, gestrichelte Kurve und die kumulierte, relative Häufigkeit $H(X_{1,k_j}) = \frac{j}{J}$ als schwarze, durchgezogene gezeichnete Kurve in einem Diagramm dar.

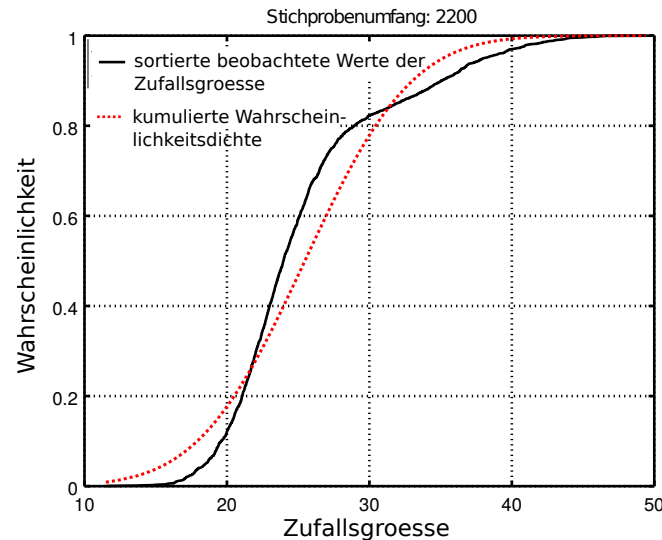


Abbildung 5.2: Vergleich der Wahrscheinlichkeiten (kumulierte Verteilungsdichten) aus der empirischen Häufigkeit der beobachteten Werte der Zufallsgröße (schwarze, durchgezogene Kurve) mit der kumulierten Normalverteilung, die sich auf die empirischen Erwartungswerte, den Mittelwert und die empirische Varianz, bezieht (rote, gestrichelte Kurve).

Die Prüfgröße des Kolmogorow-Smirnow-Tests ist

$$\sup_X |H(X_{1,k_j}) - P(X)| = \max_{k_j} \left\{ \max \left\{ |H(X_{1,k_j}) - P(X_{1,k_j})|, \lim_{X \rightarrow X_{1,k_j}} |H(X) - P(X_{1,k_j} - 1)| \right\} \right\}. \quad (5.6)$$

Dabei steht \sup für Supremum, was *kleinste, obere Schranke* heißt. Zu jeder Ausprägung der Stichprobe wird der Betrag der beiden Wahrscheinlichkeitsverteilungen gebildet. Die Berechnung des Grenzwertes $\lim_{X \rightarrow X_{1,k_j}} |H(X) - P(X_{1,k_j} - 1)|$ für das Supremum ist erforderlich, wenn die kumulierte relative Häufigkeit (Summenhäufigkeit) durch Kumulieren eines Histogramms, also einer in Klassen quantisierten Stichprobe, berechnet wurde. Dabei soll die Schreibweise mit dem Grenzwert $\lim_{X \rightarrow X_{1,k_j}} |H(X) - P(X_{1,k_j} - 1)|$ zum Ausdruck bringen, dass man infinitesimal dicht aber doch prinzipiell rechts von X_{1,k_j} an der Stufe ist, praktisch wird aber direkt $|H(X_{1,k_j}) - P(X_{1,k_j} - 1)|$ berechnet. Für die nichtklassierte, sortierte Stichprobe berechnen wir einfach nur

$$\max_{k_j} \{|H(X_{1,k_j}) - P(X_{1,k_j})|\}. \quad (5.7)$$

Die Wahrscheinlichkeit, dass die Beobachtungen nicht zur in der Hypothese angenommenen Verteilung passt, ist die Wahrscheinlichkeit, dass die Beobachtung in den Ausläufern (*Tails*) der Dichteverteilung liegt, weshalb wir die durch die Fläche in den Tails repräsentierte Wahrscheinlichkeit mit *Signifikanzniveau* α bezeichnen.

Beim Kolmogorow-Smirnow-Test wird die Nullhypothese verworfen, wenn die empirische relative Häufigkeit einen Grenzwert überschreitet, der vom Signifikanzniveau α und dem Stichprobenumfang

J abhängt

$$\sup_X |H(X_{1,k_j}) - P(X)| > K_{\alpha,J} \quad (5.8)$$

wobei es zu Schwellwerten $K_{\alpha,J}$ für $J < 35$ Tabellen in den gängigen Formelsammlungen gibt und für $J \geq 35$ folgende Näherung verwendet wird:

$$K_{\alpha,J} = \sqrt{-\frac{1}{2J} \ln\left(\frac{\alpha}{2}\right)} \quad (5.9)$$

5.3 Wahrscheinlichkeitsdichtefunktionen und ihre Parameter

Bei vielen beobachteten Stichproben, die signifikant von dem Modell, das besagt, dass sie zu genau einer normalverteilten Grundgesamtheit gehören, abweichen, handelt es sich um Beobachtungen aus Prozessen, denen mehrere überlagerte stochastische Anteile zugrunde liegen. Bei dem in Abb. 5.1 illustrierten Beispiel waren dies zwei normalverteilte Grundgesamtheiten mit verschiedenen Erwartungswerten μ_1 und μ_2 und mit unterschiedlichen Varianzen σ_1^2 und σ_2^2 .

Für eine Bewertung von gemessenen Werten einer Größe, d.h. von Beobachtungen, werden statistische Modelle zugrunde gelegt. Ein statistisches Modell zu beobachtbaren Größen umfasst die Vorstellung von Zufallsgrößen und der Verteilung ihrer Beobachtungswerte. Wahrscheinlichkeitsdichtefunktionen werden parametrisiert über ihre statistischen Momente. Diese charakterisieren den Kurvenverlauf der Funktion hinsichtlich ihrer Lage, Breite, Symmetrie und so weiter.

Der *Erwartungswert einer Zufallsgröße* X ist das erste statistische Moment der Verteilungsdichte p ihrer Grundgesamtheit

$$E(X) = \int_{-\infty}^{\infty} X p(X) dX =: \mu^{(1)}(X) \quad (5.10)$$

das ihren Schwerpunkt angibt. Der Erwartungswert der *Varianz* einer Zufallsgröße X entspricht dem zweiten statistischen Moment der Verteilungsdichte p . Das zweite statistische Moment ist definiert durch

$$\int_{-\infty}^{\infty} X^2 p(X) dX =: \mu^{(2)}(X) \quad (5.11)$$

und die Varianz ist das zweite Moment der um den Erwartungswert verschobenen Zufallsgröße

$$\text{Var}(X) = \int_{-\infty}^{\infty} (X - E(X))^2 p(X) dX \quad (5.12)$$

deren Wurzel

$$\sigma = \sqrt{\text{Var}(X)} \quad (5.13)$$

ein Maß für die Breite von p darstellt. Die Schiefe einer Verteilungsdichte wird durch das dritte statistische Moment

$$\int_{-\infty}^{\infty} X^3 p(X) dX =: \mu^{(3)}(X) \quad (5.14)$$

gewonnen, bei dem die Zufallsgröße um ihren Erwartungswert verschoben und auf die Wurzel ihrer Varianz normiert

$$\text{Skew}(X) = \int_{-\infty}^{\infty} \left(\frac{X - E(X)}{\sigma} \right)^3 p(X) \, dX \quad (5.15)$$

wird. Das dritte statistische Moment der normierten Zufallsgröße heißt *Skewness*.

Ein Maß für die Überhöhung oder Abflachung relativ zur Normalverteilung wird durch das vierte statistische Moment

$$\int_{-\infty}^{\infty} X^4 p(X) \, dX =: \mu^{(4)}(X) \quad (5.16)$$

gewonnen wird, bei dem die Zufallsgröße um ihren Erwartungswert verschoben und auf die Wurzel ihrer Varianz normiert

$$\text{Kurt}(X) = \int_{-\infty}^{\infty} \left(\frac{X - E(X)}{\sigma} \right)^4 p(X) \, dX \quad (5.17)$$

wird. Das vierte statistische Moment der normierten Zufallsgröße heißt *Kurtosis*.

5.4 Vertrauensintervall und t-Verteilung

Wir hatten bereits in Kapitel 4 angesprochen, wie *Vertrauensintervall* $[x_1, x_2]$, *Wahrscheinlichkeitsdichtefunktion* p , *Signifikanzniveau* α und *Vertrauensniveau* $1 - \alpha$ zusammen hängen. Das Vertrauensintervall gibt den Bereich an, in dem ein Ereignis mit Wahrscheinlichkeit $1 - \alpha$ beobachtet wird bzw. die Messgröße einen Wert in diesem Bereich annimmt, also

$$\int_{x_1}^{x_2} p(X) \, dX = 1 - \alpha. \quad (5.18)$$

Eine Wahrscheinlichkeitsdichteverteilung $p(X)$ wird immer so normiert, dass ihre Fläche 1 ist:

$$\int_{-\infty}^{\infty} p(X) \, dX = 1 \quad (5.19)$$

Das heißt mit anderen Worten: Unsere Größe nimmt mit Wahrscheinlichkeit 1 irgend einen beliebigen Wert zwischen minus und plus Unendlich an.

Auch hatten wir in Kapitel 4 durchgenommen, wie man aus unterschiedlichen vorliegenden Informationen eine Wahrscheinlichkeitsdichteverteilung berechnet, um den Erwartungswert

$$E(X) = \int_{-\infty}^{\infty} X p(X) \, dX \quad (5.20)$$

zu bestimmen und um aus der Umkehrfunktion P^{-1} der kumulativen Verteilung das Überdeckungsintervall

$$[P^{-1}(\frac{\alpha}{2}), P^{-1}(1 - \frac{\alpha}{2})] \quad (5.21)$$

zu ermitteln.

Man kann zur Berechnung eines Überdeckungsintervalls rechenzeiteffizient auf vorhandene tabellierte Werte für die Umkehrfunktion der Normalverteilung zurück greifen, wenn folgendes der Fall ist:

- Über die reinen Beobachtungen $X_{1,j}$ hinaus liegen keine Informationen vor, also keine vorherigen Ergebnisse zu den zu schätzenden Modellparameter (kein *Prior*) und keine weiteren beeinflussenden Verteilungen.
- Es kann davon ausgegangen werden, dass die Streuung der einzelnen Beobachtungswerte um das Modell normalverteilt sind.

Für kleine Stichprobenumfänge wird eine der Normalverteilung ähnliche Verteilung herangezogen, die je kleiner der Stichprobenumfang ist, um so stärker ausgeprägte Ausläufer (*Tails*) hat, die *t*-Verteilung. Diese werden wir im Laufe dieses Kapitels beleuchten.

Zunächst erörtern wir die Berechnung der Vertrauensintervalle auf Basis tabellierter Werte von Integrationsgrenzen zu ausgesuchten Vertrauensniveaus der Normalverteilung. Eine solche Integrationsgrenze wird *Quantil* genannt. Die Tabellen beziehen sich auf normierte Zufallsgrößen. Für die Gaußverteilung normieren wir die Zufallsgröße, so dass der Erwartungswert $\mu = 0$ ist und die Varianz $\sigma^2 = 1$, d.h. Z ist *standardnormalverteilt*

$$Z \sim \mathcal{N}(0, 1). \quad (5.22)$$

Die Wahrscheinlichkeitsdichtefunktion $p(Z)$ ist die *Standardnormalverteilung*

$$p(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2} \quad \text{mit} \quad Z = \frac{X - \mu}{\sigma}. \quad (5.23)$$

Die in den Formelsammlungen und Tabellenwerken aufgelisteten Grenzen sind als obere Integrationsgrenze definiert

$$\int_{-\infty}^{z_\alpha} p(Z) dZ = \alpha. \quad (5.24)$$

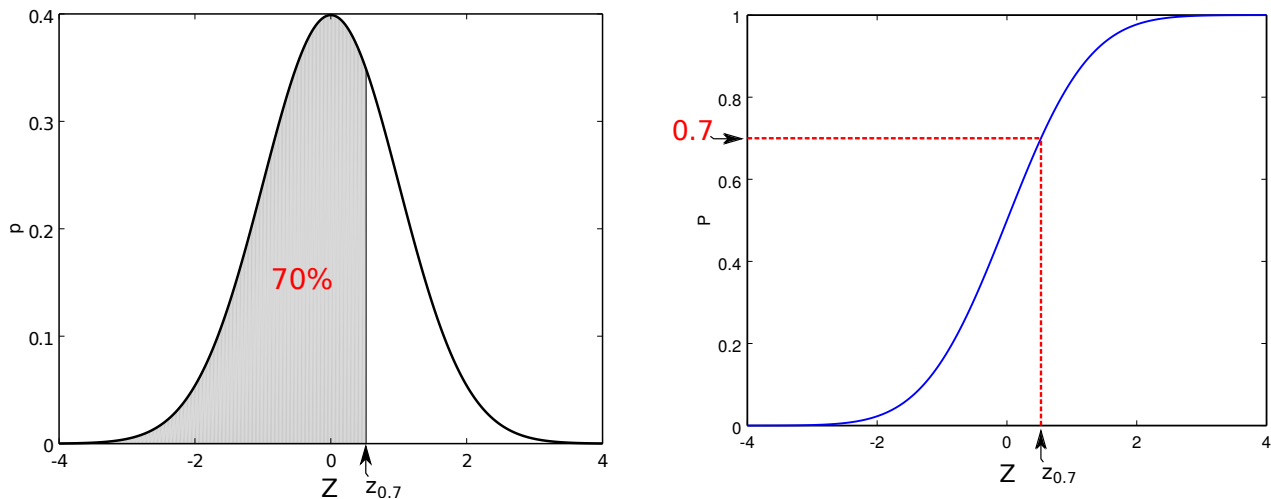
In Abb. 5.3 wird dies anhand des Quantils $z_\alpha = 0.5244$ zu $\alpha = 0.7$ also 70% Wahrscheinlichkeit beispielhaft gezeigt.

Dies lesen wir so:

Die Ereignisse Werte zu Z zu beobachten, die kleiner sind als z_α , können mit einer Wahrscheinlichkeit von α eintreten.

Die in den Tabellenwerken gelisteten Wahrscheinlichkeiten α werden aus der Integration der Wahrscheinlichkeitsdichte der normierten Zufallsgröße von minus Unendlich bis zu einer endlichen oberen Integrationsgrenze z_α gewonnen. Die zu einem Wert eines Quantils gehörende Wahrscheinlichkeit wird durch den tiefgestellten Index bei z_α gekennzeichnet, in unserem Beispiel aus Abb. 5.3 ist dies dann

$$z_{0.7} = 0.5244$$


 Abbildung 5.3: Quantil $z_\alpha = 0.5244$ und Wahrscheinlichkeit $\alpha = 0.7$ der Standardnormalverteilung.

Als weiteres Beispiel betrachten wir die Frage, wie hoch die Wahrscheinlichkeit ist, Werte zu beobachten, die in den Ausläufern einer Verteilungsdichte liegen. Die Ausläufer werden im Englischen als *Tails* (Schwänze) bezeichnet. Dies sollen Beobachtungen zu der normierten Zufallsgröße Z mit Werten, die zum einen kleiner oder gleich -1.960 sind, sein

$$z_{0.025} = -1.960$$

und zum anderen, die Beobachtungen mit Werten größer oder gleich 1.960 , also die in dem rechten Ausläufer (*Tail*) der Standardnormalverteilung liegen. Die Wahrscheinlichkeit für Beobachtungen im linken *Tail* ist

$$\int_{-\infty}^{-1.960} p(Z) dZ = 0.025.$$

Aufgrund der Symmetrie der Standardnormalverteilung ist die Wahrscheinlichkeit für das Liegen von Werten im rechten *Tail* ebenfalls 0.025 . Wir rechnen also zusammen, dass mit einer Wahrscheinlichkeit von $0.05 = 0.025 + 0.025$ Werte in den beiden *Tails* außerhalb des Intervalls $[-1.960, 1.960]$ liegen können. Somit ist die Wahrscheinlichkeit, dass sie innerhalb des Intervalls liegen $1 - 0.05 = 0.95$

$$\int_{-1.96}^{1.96} p(Z) dZ = 0.95 \quad (5.25)$$

was uns zeigt, wie wir die Tabellen zu verwenden haben. Wenn wir ein zweiseitiges Intervall betrachten wollen, wie hier zu einem Vertrauensniveau von 95% , so müssen wir im Tabellenwerk nach dem Quantil, das für 97.5% gelistet ist, schauen

$$z_{0.975} = 1.960$$

weil

$$0.95 = \underbrace{\int_{-\infty}^{z_{0.975}} p(Z) dZ}_{0.975} - \underbrace{\int_{-\infty}^{z_{0.025}} p(Z) dZ}_{0.025}$$

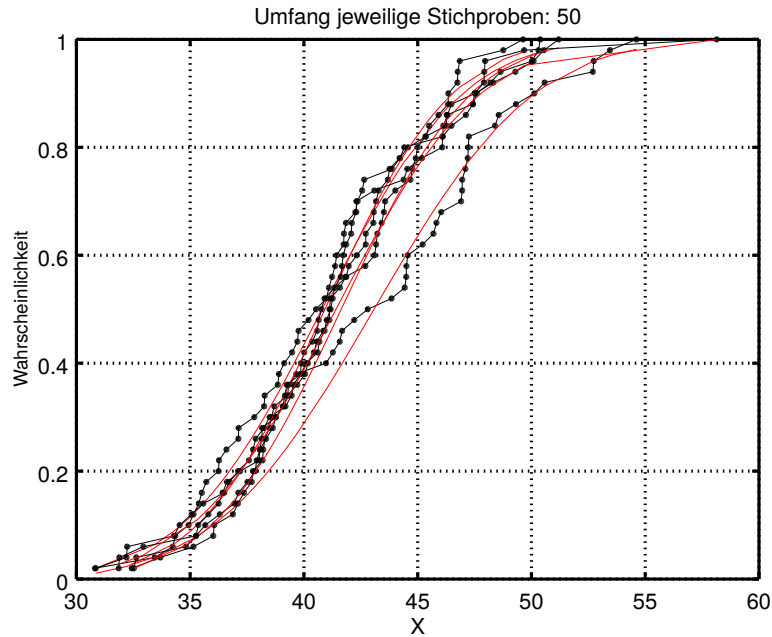


Abbildung 5.4: Kumulative relative Wahrscheinlichkeitsfunktionen und relative Summenhäufigkeiten von verschiedenen Stichproben mit jeweils $J = 50$ Werten, entnommen aus einer Grundgesamtheit mit $\mu = 42$ und $\sigma = 5$.

Aufgrund der Achssymmetrie der Normalverteilung gilt

$$z_{\frac{1}{2}\alpha} = -z_{1-\frac{1}{2}\alpha}.$$

Da sich die Quantile zum jeweiligen Vertrauensniveau $1 - \alpha$ auf die normierte Zufallsgröße beziehen, muss für die Berechnung des Vertrauensintervall zurückgerechnet werden

$$Z = \frac{X - \mu}{\sigma} \quad \Leftrightarrow \quad X = \mu + Z\sigma \quad (5.26)$$

also

$$[z_{\frac{1}{2}\alpha}, z_{1-\frac{1}{2}\alpha}] = [-z_{1-\frac{1}{2}\alpha}, z_{1-\frac{1}{2}\alpha}] \quad \Leftrightarrow \quad [\mu - z_{1-\frac{1}{2}\alpha}\sigma, \mu + z_{1-\frac{1}{2}\alpha}\sigma] \quad (5.27)$$

Mit α sind die 5% und mit $1 - \alpha$ die 95% Vertrauensniveau gemeint.

Die Likelihood wird maximal für

$$y = \frac{1}{J} \sum_{j=1}^J X_{1,j}, \quad (5.28)$$

wobei y der Schätzwert für den Modellparameter Y ist. Der Schätzwert s^2 für die Varianz σ^2 wird berechnet mit

$$s^2 = \frac{1}{J-1} \sum_{j=1}^J (X_{1,j} - y)^2. \quad (5.29)$$

Nachdem wir gesehen haben, wie wir für eine normalverteilte Zufallsgröße das Vertrauensintervall ermitteln können, wenden wir uns jetzt der Thematik zu, wenn wir zu einer Zufallsgröße nur relativ wenige Beobachtungen haben.

Wenn wir aus einer sehr großen Stichprobe, die einer Grundgesamtheit entnommen wurde, disjunkte Teilstichproben von kleinerem Umfang ($J < 100$) in Grüppchen unterteilen, so können wir beobachten, dass sowohl die Mittelwerte streuen als auch die Varianzen. Diese Grüppchen oder Teilstichproben können wir als mehrere Stichproben aus einer normalverteilten Grundgesamtheit ansehen. Abb. 5.4 illustriert wie unterschiedlich die Verteilungsfunktionen von Stichproben mit relativ kleinem Umfang ($J = 50$) aussehen können, obwohl sie zu derselben Grundgesamtheit gehören. Dieses Beispiel wurde wie folgt mit Gnu-octave generiert:

```
function example_t()
    J1 = 50;
    mu1 = 42;
    sigma1 = 5;
    samples = [];
    for kappa = 1:5
        samples = [samples (mu1 + sigma1*randn(J1,1))];
    end
    figure(3000); hold on;
    [J1, nS] = size(samples);
    for kappa = 1:5
        [P, H, xsort, xbar, s] = cumulatives(samples(:,kappa));
        plot( xsort, H, 'k.-');
        plot( xsort, P, 'r-');
    end
    xlabel('X', 'fontsize', 14);
    ylabel('Wahrscheinlichkeit', 'fontsize', 12);
    title(['Umfang jeweilige Stichproben: ' num2str(J1)], 'fontsize', 14);
    grid on;
    set(gca, 'fontsize', 14, 'linewidth', 2);
    hold off;
end
function [P, H, xsort, xbar, s] = cumulatives(x)
    xbar = mean(x);
    s = std(x);
    J1 = length(x);
    [xsort, isort] = sort(x, 'ascend');
    H = [1:J1]'/J1;
    P = normcdf( xsort, xbar, s);
end
```

Zur Bewertung der Flachheit bzw. Ausgeprägtheit der Ausläufer einer Verteilung berechnen wir ihre *Kurtosis*. Die *Kurtosis* der Normalverteilung hat den Wert 3.

Die empirischen Schätzer für den Erwartungswert, die Varianz, die Skewness und Kurtosis berechnen wir aus einer Stichprobe mit Beobachtungswerten $X_{1,1}, \dots, X_{1,J}$ wie folgt:

Der Mittelwert wird berechnet mit

$$y = \frac{1}{J} \sum_{j=1}^J X_{1,j}. \quad (5.30)$$

Die empirische Standardabweichung wird berechnet mit

$$s = \sqrt{\frac{1}{J-1} \sum_{j=1}^J (X_{1,j} - y)^2}. \quad (5.31)$$

Zu bemerken ist, dass jetzt durch $J-1$ und nicht durch J geteilt wird, weil das y aus den Werten der Stichprobe $X_{1,1}, \dots, X_{1,J}$ gewonnen wurde und somit die Anzahl der Freiheitsgrade ν um Eins verringert wird, also $\nu = J - 1$.

Die empirische Skewness wird berechnet mit

$$\text{skew} = \frac{1}{J} \sum_{j=1}^J \left(\frac{X_{1,j} - y}{s} \right)^3. \quad (5.32)$$

Die empirische Kurtosis wird berechnet mit

$$\text{kurt} = \frac{1}{J} \sum_{j=1}^J \left(\frac{X_{1,j} - y}{s} \right)^4. \quad (5.33)$$

Bei dem in Abb. 5.4 gezeigten Beispiel mit mehreren Stichprobenentnahmen aus einer Grundgesamtheit erhalten wir folgende aus den Momenten abgeleiteten empirischen Parameter der Verteilungen:

Tabelle 1:

Einzelstichproben, Umfang jeweils $J = 50$				
ltd. Nr.	y	s	skew	kurt
1	40.79	4.83	0.33	2.50
2	43.08	5.52	0.16	2.16
3	40.92	4.38	-0.26	2.57
4	41.64	4.46	0.09	2.44
5	41.41	4.98	0.59	4.16
alle Stichproben zusammen, Umfang $J = 5 \cdot 50 = 250$				
	41.57	4.88	0.28	2.89
wahre Grundgesamtheit				
	μ	σ	Skew	Kurt
Normalvert.	42.00	5.00	0.00	3.00

Man kann der Beschreibung einer kleinen Stichprobe besser gerecht werden, wenn die Verteilung flacher ist als eine Standardnormalverteilung. Das bedeutet, dass ihre Ausläufer ausgeprägter sind. Eine Verteilungsdichtefunktion, die der Gaußverteilung ähnlich ist, aber unter Berücksichtigung des

begrenzten Stichprobenumfangs entsprechend ausgeprägtere Tails hat, ist die im folgenden dargestellte *Student-t-Verteilung*, kurz auch einfach *t-Verteilung* genannt.

Diese Wahrscheinlichkeitsdichteverteilung wurde anfang des 20. Jahrhunderts von William Sealy Gosset entwickelt. Sie wird der Tatsache gerecht, dass die Varianz zunimmt mit kleiner werdendem Stichprobenumfang. Gosset veröffentlichte 1908 erstmals zu dem Thema während er als Chemiker für die Guinness-Brauerei in Dublin arbeitete. Er entwickelte einen Test zum Vergleich von Mittelwerten einzelner Stichproben als eine billige Art und Weise, die Qualität des Stout zu überwachen. Dieser Test wird entsprechend der zugrunde gelegten Wahrscheinlichkeitsdichtefunktion *t-Test* genannt. Da sein Arbeitgeber die Veröffentlichung nicht gestattete, veröffentlichte Gosset sie unter dem Pseudonym Student. Die zugehörige Theorie wurde erst durch die Arbeiten von R. A. Fisher belegt, der die Verteilung *Student-Verteilung* (engl. *Student's distribution*) nannte.

Die *t-Verteilung* bezieht sich auf eine standardnormalverteilte Zufallsgröße Z und sieht wie folgt aus

$$p_\nu(Z) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi} \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{Z^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad (5.34)$$

wobei ν die Anzahl der Freiheitsgrade ist (was für den Fall, dass Z die Summe der Stichprobenelemente darstellt, um den Mittelwert der Stichprobe mit Umfang J zu repräsentieren, dann $\nu = J - 1$ ist), und wobei die Gamma-Funktion für natürliche Zahlen $\nu \in \mathbb{N}$ dividiert durch Zwei wie folgt definiert ist:

$$\Gamma\left(\frac{\nu}{2}\right) = \sqrt{\pi} \frac{(\nu-2)!!}{2^{\frac{\nu-1}{2}}} \quad \nu \in \mathbb{N} \quad (5.35)$$

mit $\nu!! = \nu(\nu-2)(\nu-4) \dots 4 \cdot 2$. Wir werden auf den theoretischen Hintergrund aus der Wahrscheinlichkeitstheorie nicht genau eingehen, also ist die Definition dieser Verteilung auch kein Klausurstoff. Für die Klausur werden die dazugehörigen Quantiltabellen zur Verfügung gestellt, d.h. den Aufgabenstellungen beigelegt.

Je kleiner der Stichprobenumfang ist desto flacher wird die Verteilung, mit entsprechend längeren Ausläufern, *Tails*, siehe Abb. 5.5. Um zu verdeutlichen, wie sich die Gaußverteilung und die Student-t-Verteilung hinsichtlich Breite und lange Ausläufer unterscheiden, haben wir hier dies im Vergleich berechnet. Dabei haben wir bewusst ganz extrem wenig Freiheitsgrade gewählt, nämlich nur 5. Wir verwenden jetzt mal den empirischen Mittelwert und die empirische Standardabweichung, die wir aus der Gesamtheit aller Stichproben gewonnen hatten, also $y = 41.57$ und $s = 4.88$. Daraus rechnen wir sowohl die Normalverteilung als auch die Student-t-Verteilung für nur fünf Freiheitsgrade und erhalten eine größere Standardabweichung und eine größere Kurtosis aus der Student-t-Verteilung:

Verteilung	σ	Kurt
Gauß	4.88	2.99
Student-t	5.86	3.59

Dementsprechend sind auch die Quantile größer.

Die Kurtosis der Student-t Verteilungen nimmt Werte an, die um so größer werden, je flacher die Verteilungsdichte ist, also je ausgeprägter die *Tails* sind, also je kleiner die Stichprobenumfänge bzw. die Anzahl der Freiheitsgrade sind.

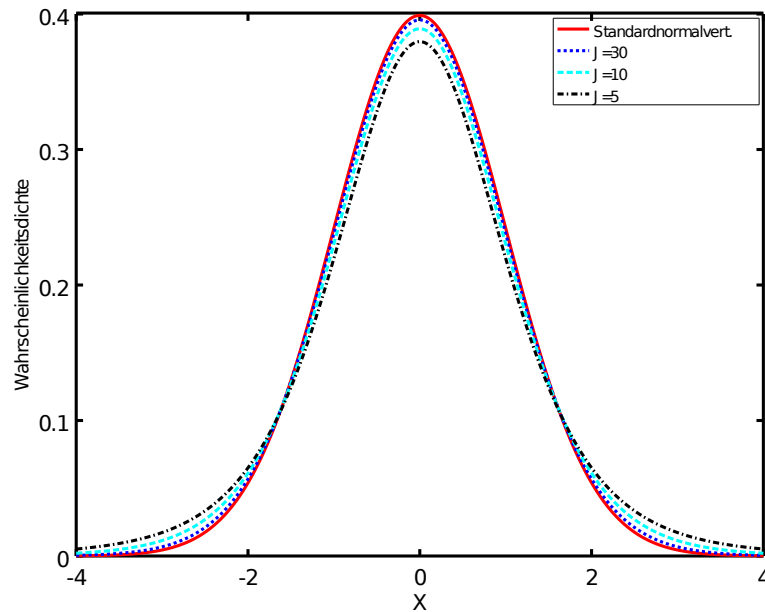


Abbildung 5.5: Student-t Verteilungen für unterschiedliche Stichprobenumfänge.

Das Vertrauensintervall zu einem Vertrauensniveau von $1 - \alpha$ für eine endlich große Stichprobe mit Stichprobenumfang J schätzen wir mit dem Quantil der Student-t-Verteilung ab. Die Anzahl der Freiheitsgrade entspricht dem Stichprobenumfang abzüglich der Anzahl der zu schätzenden Modellparameter, was in diesem Fall einer ist, also $\nu = J - 1$.

Während für die Standardnormalverteilung das Quantil $z_{0.975} = 1.96$ beträgt, haben wir aus der Student-t-Verteilung zu demselben Vertrauensniveau für $\nu = 50$ Freiheitsgrade $t_{0.975,50} = 2.01$ und für $\nu = 5$ Freiheitsgrade $t_{0.975,5} = 2.57$.

Für das zweiseitige, symmetrische Intervall verwenden wir das Quantil $t_{1-\alpha/2,\nu}$, so dass wir folgendes Vertrauensintervall erhalten:

$$[y - t_{1-\alpha/2,\nu}s, y + t_{1-\alpha/2,\nu}s]$$

Die Breite wird durch die Anzahl der Freiheitsgrade beeinflusst, je kleiner die Anzahl der Freiheitsgrade ist, desto breiter wird das Intervall. Die beiden Abbn. 5.5 und 5.6 zeigen, dass die *Tails* einer Student-t-Verteilung für wenige Freiheitsgrade deutlich breiter sind als bei der Standardnormalverteilung, folglich auch das Vertrauensintervall breiter ist. In Abb. 5.6 sind die Quantile für ein Vertrauensniveau von $1 - \alpha = 0.95$ eingezeichnet.

Das (*vollständige*) *Messergebnis* geben wir an als Schätzwert zur Größe Y gemeinsam mit dem Vertrauensintervall. Eine gebräuchliche Schreibweise ist

$$Y = y \pm t_{1-\alpha/2,\nu}s. \quad (5.36)$$

Als nächstes befassen wir uns damit, das Vertrauensintervall des Mittelwertes abzuschätzen. Dazu betrachten wir ein Beispiel, bei dem aus einer Grundgesamtheit mit den wahren Werten $Y = 42$ und $\sigma = 7$ mehrere Stichproben genommen werden, beispielsweise $K = 15$, wobei jede Stichprobe einen Umfang von $J = 25$ habe. Zu jeder der Stichproben $k = 1 \dots K$ berechnen wir den Mittelwert y_k und

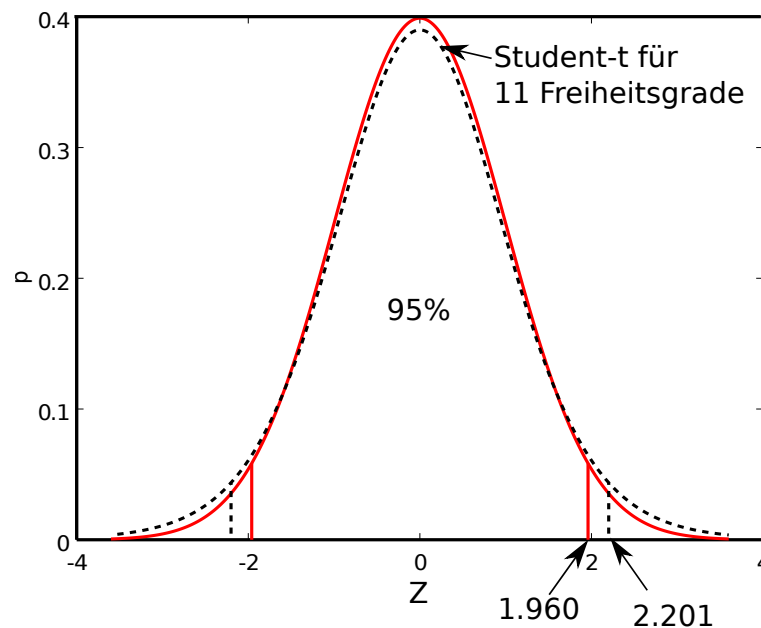


Abbildung 5.6: Vertrauensintervalle in Vergleich für die Standardnormalverteilung (rote Kurve) und die Student-t-Verteilung für $\nu = 11$ Freiheitsgrade (schwarze gestrichelte Kurve).

die empirische Standardabweichung s_k .

k	1	2	3	4	5	6	7	8
y_k	41.813	41.854	39.084	42.117	41.532	40.428	42.717	39.727
s_k	6.476	7.706	6.996	7.729	6.431	7.516	7.505	7.439

k	9	10	11	12	13	14	15
y_k	40.858	39.670	44.351	42.011	41.935	44.509	39.921
s_k	8.568	7.649	6.082	7.661	7.084	7.878	7.084

Der Mittelwert \bar{y} der Mittelwerte liefert dann

$$\bar{y} = \frac{1}{K} \sum_{k=1}^K y_k = 41.502$$

und die empirische Standardabweichung der Mittelwerte y_k liefert

$$\bar{s} = \sqrt{\frac{1}{K-1} \sum_{k=1}^K (y_k - \bar{y})^2} = 1.606$$

Das wiederholte Ziehen von Stichproben kann für so manche Anwendung kostspielig sein, so dass es gilt die empirische Standardabweichung des Mittelwerts aus nur einer Stichprobe abzuschätzen. Dazu betrachten wir die Likelihood

$$l(Y, \sigma | \{X_{1,1}, \dots, X_{1,J}\}) = \prod_{j=1}^J \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \left(\frac{(X_{1,j} - Y)}{\sigma} \right)^2} \quad (5.37)$$

und setzen den Schätzer y der Größe Y als nahrhafte Null ein

$$l(Y, \sigma | \{X_{1,1}, \dots, X_{1,J}\}) = \prod_{j=1}^J \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{1}{2} \left(\frac{(X_{1,j} - y) - (Y - y)}{\sigma} \right)^2} \quad (5.38)$$

d.h.

$$l(Y, \sigma | \{X_{1,1}, \dots, X_{1,J}\}) = \frac{1}{(\sqrt{2\pi} \sigma)^J} e^{-\frac{1}{2} \sum_{j=1}^J \left(\frac{(X_{1,j} - y) - (Y - y)}{\sigma} \right)^2}. \quad (5.39)$$

mit folgender Nebenrechnung für die Summe im Exponenten

$$\begin{aligned} \sum_{j=1}^J ((X_{1,j} - y) - (Y - y))^2 &= \sum_{j=1}^J ((X_{1,j} - y)^2 + (Y - y)^2 - 2(X_{1,j} - y)(Y - y)) \\ &= \left(\sum_{j=1}^J (Y - y)^2 \right) + \left(\sum_{j=1}^J (X_{1,j} - y)^2 \right) - 2 \sum_{j=1}^J (X_{1,j} - y)(Y - y) \\ &= J(Y - y)^2 + \left(\sum_{j=1}^J (X_{1,j} - y)^2 \right) \end{aligned}$$

und $\sum_{j=1}^J (X_{1,j} - y)(Y - y) = 0$ wegen $Jy = \sum_{j=1}^J X_{1,j}$ gilt für die Likelihood folgende Proportionalität

$$l(Y, \sigma | \{X_{1,1}, \dots, X_{1,J}\}) \propto e^{-\frac{1}{2} J \left(\frac{(Y - y)}{\sigma} \right)^2} \quad (5.40)$$

und wir definieren folgende Größe

$$\bar{\sigma}^2 := \frac{\sigma^2}{J} \quad (5.41)$$

als *Varianz des Mittelwertes*.

Damit lässt sich Gl. (5.40) umschreiben in

$$l(Y, \sigma | \{X_{1,1}, \dots, X_{1,J}\}) \propto e^{-\frac{1}{2} \left(\frac{(Y - y)}{\bar{\sigma}} \right)^2}. \quad (5.42)$$

Nun setzen wir in die Definitionsgleichung der Varianz des Mittelwertes Gl. (5.41) bzw. die Wurzel daraus die empirisch ermittelten Werte unseres Beispiels ein:

$$\bar{s} = \frac{s}{\sqrt{J}} \quad (5.43)$$

k	1	2	3	4	5	6	7	8
\bar{s}_k	1.295	1.541	1.399	1.546	1.286	1.503	1.501	1.488

k	9	10	11	12	13	14	15
\bar{s}_k	1.714	1.530	1.216	1.532	1.417	1.576	1.417

Jede einzelne Stichprobe liefert eine etwas unterschiedliche Abschätzung für die Standardabweichung des Mittelwertes. Die Werte für \bar{s} liegen mit 1.3 bis 1.5 knapp unter 1.6. Es wird die Standardabweichung des Mittelwertes leicht unterschätzt, weil der einzelne Stichprobenumfang mit $J = 25$ größer

ist als die Anzahl der Stichproben mit $K = 15$.

Das *Vertrauensintervall für die Schätzung des Mittelwertes* im Gegensatz zur Schätzung des Einzelwertes ist damit

$$[y - t_{1-\alpha/2, \nu} \bar{s}, y + t_{1-\alpha/2, \nu} \bar{s}]$$

und in der Schreibweise als *vollständiges Messergebnis*

$$Y = y \pm t_{1-\alpha/2, \nu} \bar{s}. \quad (5.44)$$

5.5 t-Test - Mittelwerttest

Nicht nur um die Qualität von schwarzem Bier (Stout) zu überprüfen, sondern ganz allgemein, wird der t-Test eingesetzt, um zu testen, ob

1. eine Stichprobe zu einer Grundgesamtheit gehört, deren Erwartungswert bekannt ist;
2. zwei Stichproben zur gleichen Grundgesamtheit gehören, deren Erwartungswert nicht bekannt ist, sondern deren jeweilige Mittelwerte miteinander verglichen werden.

Sei μ_1 der Erwartungswert der Grundgesamtheit zu Stichprobe 1 mit Beobachtungen $(X_{1,1}, \dots, X_{1,J_1})$ und sei μ_2 der Erwartungswert der zu Stichprobe 2 mit Beobachtungen $(X_{2,1}, \dots, X_{2,J_2})$. Die Nullhypothese H_0 lautet, dass beide zu derselben Grundgesamtheit gehören, was soviel bedeutet wie $\mu_1 = \mu_2$. Die Alternativhypothese H_a lautet, dass sie zu unterschiedlichen Grundgesamtheiten gehören, was soviel bedeutet wie $\mu_1 \neq \mu_2$.

H_0	$\mu_1 = \mu_2$	beide Stichproben gehören derselben Grundgesamtheit an
H_a	$\mu_1 \neq \mu_2$	beide Stichproben gehören unterschiedlichen Grundgesamtheiten an

Die Prüfgröße ist die Differenz der empirisch berechneten Mittelwerte normiert auf die Standardabweichungen der Mittelwerte. Mit Gl. (5.31) wird die empirische Standardabweichung einer Stichprobe berechnet. Das in Tabelle 1 aufgeführte Beispiel zeigt, dass die Mittelwerte deutlich weniger streuen, als die Werte innerhalb einer einzelnen Stichprobe. Die Varianz des Mittelwertes einer Stichprobe wird gemäß Gl. (5.43) darüber abgeschätzt, dass sie durch den Stichprobenumfang geteilt wird:

$$\bar{s}_i^2 = \frac{1}{J_i(J_i - 1)} \sum_{j=1}^{J_i} (X_{i,j} - y_i)^2 \quad (5.45)$$

Die Prüfgröße des t-Tests für den Vergleich der beiden Stichproben $i = 1, 2$ ist wie folgt definiert

$$T = \frac{y_1 - y_2}{\sqrt{\bar{s}_1^2 + \bar{s}_2^2}} \quad (5.46)$$

mit

$$y_i = \frac{1}{J_i} \sum_{j=1}^{J_i} X_{i,j} \quad \text{mit} \quad i = 1, 2 \quad (5.47)$$

Die Nullhypothese wird mit einem Signifikanzniveau von $\alpha = 0.05$ abgelehnt, wenn der Betrag der Prüfgröße größer als das entsprechende Quantil der t -Verteilung ist

$$|T| > t_{1-\frac{1}{2}\alpha, \nu} \quad \Rightarrow \quad H_0 \text{ ablehnen} \quad (5.48)$$

Für den t -Test mit zwei Stichproben wird unterschieden nach *gepoolten* und nicht *gepoolten* Stichproben. Unter *gepoolten* Stichproben (*Samples*) versteht man diejenigen, deren Varianzen im wesentlichen als gleich zu betrachten sind. Die gemeinsame Anzahl der Freiheitsgrade für *gepoolte* Samples, die für die Wahl des t -Quantils gebraucht wird, ist

$$\nu = J_1 + J_2 - 2 \quad (5.49)$$

und für NICHT *gepoolte*

$$\nu = \frac{\left(\frac{s_1^2}{J_1} + \frac{s_2^2}{J_2}\right)^2}{\frac{\left(\frac{s_1^2}{J_1}\right)^2}{J_1-1} + \frac{\left(\frac{s_2^2}{J_2}\right)^2}{J_2-1}}. \quad (5.50)$$

Die Herleitung zur Berechnung der Anzahl der Freiheitsgrade für nicht gepoolte Stichproben wird in Kapitel 7 dran kommen. Gl. (5.50) wird auch Satterthwaite'sche Gleichung genannt.

Bei $\nu = J_1 + J_2 - 2$ Freiheitsgraden, bei dem hier betrachteten Beispiel ist dies $\nu = 98$ ist $t_{1-\frac{1}{2}\alpha, \nu} = t_{0.975, 98} = 1.985$. Der Vergleich der ersten mit der zweiten Stichprobe aus Tabelle 1 liefert

$$|T| = \frac{|40.79 - 43.08|}{\sqrt{\frac{1}{50}(4.83)^2 + \frac{1}{50}(5.52)^2}} = 2.21 > 1.99$$

dass die Mittelwerte der beiden Stichproben signifikant, auf einem Signifikanzniveau von $\alpha = 0.05$, von einander abweichen und die Nullhypothese verworfen wird.

Betrachtet man jedoch einen größeren Vertrauensbereich, also anstelle von 95% einen Bereich von 98%, nimmt man also mehr aus dem Bereich der *Tails* hinzu, so wird die Nullhypothese nicht verworfen. Auf einem Signifikanzniveau von nur $\alpha = 0.02$ hat das Vertrauensintervall die Grenzen $-t_{1-\frac{1}{2}\alpha, \nu} = -t_{0.99, 98} = -2.365$ und $t_{1-\frac{1}{2}\alpha, \nu} = t_{0.99, 98} = 2.365$. Die Nullhypothese wird für diese Wahl des Signifikanzniveaus nicht verworfen, weil die Prüfgröße innerhalb des breiteren Intervalls mit $|T| = 2.21 < 2.37$ liegt.

<http://www.itl.nist.gov/div898/handbook/eda/section3/eda353.htm>

5.6 Chi2-Verteilung und Varianztest

Eine Verteilungsdichtefunktion, die die Verteilung der Quadrate einer normierten Zufallsgröße Z , also die Verteilung der Varianzen einer Zufallsgröße X , beschreibt, ist die *Chi-Quadrat*-Verteilung.

Bisher haben wir uns damit befasst, wie eine Zufallsgröße X streut. Jetzt betrachten wir, wie die Streuung ihrerseits streut. Wir betrachten eine Zufallsgröße, die zu einer normalverteilten Grundgesamtheit

gehört.

X sei eine normalverteilte Zufallsgröße mit Erwartungswert μ und Varianz σ^2

d.h.

$$X \sim \mathcal{N}(\mu, \sigma). \quad (5.51)$$

Werden einzelne kleinere Stichproben X_i aus der Grundgesamtheit von X entnommen, die jeweils einen Stichprobenumfang J_i haben und werden die jeweiligen Mittelwerte und empirischen Varianzen berechnet

$$y_i = \frac{1}{J_i} \sum_{j=1}^{J_i} X_{i,j} \quad s_i^2 = \frac{1}{J_i - 1} \sum_{j=1}^{J_i} (X_{i,j} - y_i)^2 \quad (5.52)$$

so haben wir anhand der zuvor behandelten Beispiele festgestellt, dass nicht nur die Mittelwerte y_i sondern auch die Varianzen s_i^2 streuen. Die Charakteristik der Verteilungsdichte der Varianzen hängt vom Stichprobenumfang J_i ab. Der Stichprobenumfang entspricht der Anzahl der Freiheitsgrade.

Gegeben sei eine Stichprobe Z_1, \dots, Z_J unabhängiger Beobachtungen einer standardnormalverteilten Zufallsgröße Z : Dann ist die Summe der Quadrate der Beobachtungen Q

$$Q = \sum_{j=1}^J Z_j^2 \quad (5.53)$$

gemäß folgender Wahrscheinlichkeitsdichtefunktion verteilt

$$p(Q, J) = \frac{Q^{\frac{J}{2}-1} e^{-\frac{Q}{2}}}{2^{\frac{J}{2}} \Gamma(\frac{J}{2})} \quad Q > 0 \quad (5.54)$$

mit der Gammafunktion für natürliche Zahlen $\nu = J$, deren Definition im Abschnitt zuvor als Gl. (5.35) angegeben wurde.

Eine Schreibweise für die Aussage

Q ist Chi2-verteilt

ist

$$Q \sim \chi^2. \quad (5.55)$$

Die Definition der χ^2 -Verteilungsdichtefunktion braucht nicht auswendig gelernt zu werden. Der Umgang mit den Quantiltabellen ist aber zu üben. Wichtig zu wissen ist, dass die χ^2 -Verteilungsdichtefunktion für einen positiven Wertebereich gilt, sie ihr Maximum in der Nähe von $Q = J$ hat, sie um so schiefer und breiter ist, je kleiner J ist und die Verteilungsdichte der Varianzen

$$s_{\mu,i}^2 = \frac{1}{J_i} \sum_{j=1}^{J_i} (X_{i,j} - \mu)^2 \quad (5.56)$$

$$s_{\mu}^2 = \frac{1}{J} Q \sigma^2 \quad \Leftrightarrow \quad Q = J \left(\frac{s_{\mu}}{\sigma} \right)^2 \quad (5.57)$$

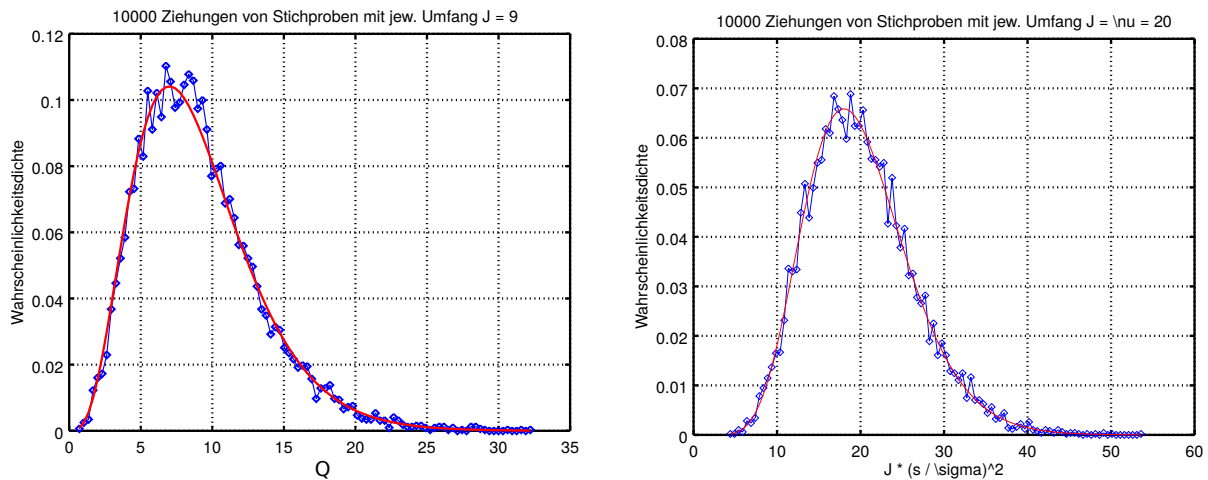


Abbildung 5.7: χ^2 -Verteilung der normierten Zufallsgröße $Q = J \left(\frac{s_\mu}{\sigma}\right)^2$ für $s_\mu^2 = \frac{1}{J} \sum_{j=1}^J (X_{\cdot,j} - \mu)^2$ für $J = 9$ (links) und $J = 20$ (rechts).

liefert. Was in der Praxis hinsichtlich des Statistiktests, dem Chi2-Test, gebraucht wird, ist das Verständnis wie die Tabellen verwendet werden und für Methoden der Bayesischen Statistik erforderlichenfalls auch wie die Funktionen der jeweils eingesetzten Numerik-/Statistikbibliotheken zu benutzen sind.

Abb. 5.7 zeigt zwei Beispiele für die Verteilungsdichte der Varianzen, zum einen für $J = 9$ und zum anderen für $J = 20$. Mit

```
function understandchi2()
    Jz = 9;
    nbin = 100;
    n = 10000;
    d = zeros(n,1);
    for k=1:n
        x = randn(Jz,1);
        d(k) = sum(x.^2);
    end
    [haeuf, bin] = hist(d,nbin);
    Deltabin = bin(2)-bin(1);
    figure(1000);
    plot(bin,haeuf/(n*Deltabin),'bd-',bin,chi2pdf(bin,Jz),'r-');
    xlabel('J * (s / \sigma)^2', 'fontsize', 14);
    ylabel('Wahrscheinlichkeitsdichte', 'fontsize', 14);
    title([num2str(n) 'Stichprobenumfang J = \nu = ' num2str(Jz)], 'fontsize', 14);
    grid on;
    set(gca, 'fontsize', 14, 'linewidth', 2);
    print(1000,['understandchi2_df' num2str(Jz) '_solid.svg'],'-dsvg');
```

wurden die Diagramme erzeugt. Die Verteilung der Grundgesamtheit ist die Standardnormalverteilung mit $\mu = 0$ und $\sigma = 1$. Eine Stichprobenentnahme wird mit $\mathbf{x} = \text{randn}(Jz,1)$; simuliert.

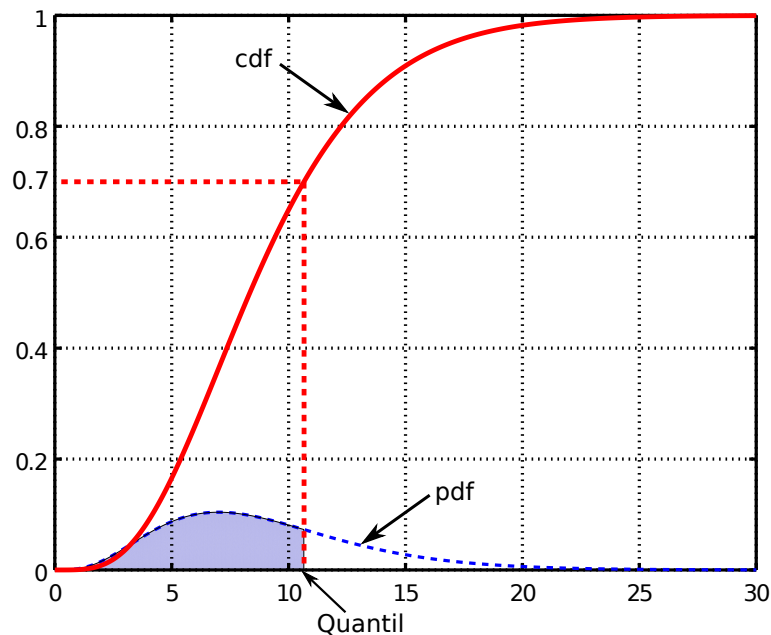


Abbildung 5.8: Wahrscheinlichkeitsverteilungsdichte (pdf) und kumulierte Wahrscheinlichkeitsfunktion (cdf) der normierten Varianzen, das Quantil wird χ^2 genannt. Hier für $\nu = 9$ Freiheitsgrade und $1 - \alpha = 0.7$, somit $\chi^2_{\nu, 1-\alpha} = 10.66$ dargestellt.

Wie bei der Quantildefinition für die Standardnormalverteilung und für die t -Verteilung ist das Quantil der χ^2 -Verteilung die obere Integrationsgrenze zur Gewinnung des Flächeninhalts der Verteilungsdichte. Dies quantifiziert die Wahrscheinlichkeit, mit der für die Größe Q Werte beobachtet werden, die kleiner sind als das Quantil. Hier ist die untere Integrationsgrenze Null und nicht minus Unendlich, weil Q per Definitionem nur positive Werte haben kann, siehe Abb. 5.8.

$$P(Q) = \int_0^{\chi^2_{1-\alpha, \nu}} p(Q', \nu) \, dQ' = 1 - \alpha \quad (5.58)$$

Für den χ^2 -Test wird, weil der Erwartungswert der Grundgesamtheit nicht bekannt ist, die Varianz durch die empirische Varianz

$$s_{\mu, i}^2 = \frac{1}{J_i} \sum_{j=1}^{J_i} (X_{i,j} - \mu)^2 \approx \frac{1}{J_i - 1} \sum_{j=1}^{J_i} (X_{i,j} - y_i)^2 = s_i^2 \quad (5.59)$$

approximiert.

```
for k=1:n
    x = randn(Jz,1);
    y = mean(x);
    d(k) = sum((x-y).^2);
end
```

Dann ist die Anzahl der Freiheitsgrade $\nu = J - 1$ um einen vermindert, weil der Mittelwert y_i von den Werten $X_{i,j}$ abhängt. Die χ^2 -Verteilung ist mit der Anzahl der Freiheitsgrade und nicht mit dem

Stichprobenumfang zu verwenden

$$s^2 = \frac{1}{\nu} Q \sigma^2 \quad \Leftrightarrow \quad Q = \nu \left(\frac{s}{\sigma} \right)^2, \quad (5.60)$$

so dass die χ^2 -Verteilungsdichte, hier mit dem Funktionsaufruf `chi2pdf` für $\nu = J - 1$ zu verwenden ist:

```
[haeuf, bin] = hist(d,nbin);
Deltabin = bin(2)-bin(1);
plot(bin,haeuf/(n*Deltabin),'bd-',bin,chi2pdf(bin,Jz-1),'r-');
xlabel('(J-1) * (s / \sigma)^2', 'fontsize', 14);
```

Für den Fall, den wir weiter oben betrachtet hatten, nämlich dass für den Erwartungswert μ nicht sein Schätzer (der Mittelwert) verwendet wird, ist die Anzahl der Freiheitsgrade gleich dem Stichprobenumfang.

Beim Chi-Quadrat-Test wird geprüft, ob die Varianz einer Stichprobe gleich einer spezifizierten Varianz σ_0^2 ist. Wir formulieren die Hypothesen

$$\begin{array}{l|l} H_0 & \sigma^2 = \sigma_0^2 \quad \text{die Stichprobe gehört zu einer Grundgesamtheit mit der Varianz } \sigma_0^2 \\ H_a & \sigma^2 \neq \sigma_0^2 \quad \text{die Stichprobe gehört nicht zu einer Grundgesamtheit mit der Varianz } \sigma_0^2 \end{array}$$

Die Testgröße ist

$$T = \nu \left(\frac{s}{\sigma_0} \right)^2 \quad (5.61)$$

Die Nullhypothese wird mit einem Signifikanzniveau von α verworfen, falls die Testgröße außerhalb des Intervalls liegt, das durch die in Gl. (5.58) definierten Quantile begrenzt wird, d.h.

$$T < \chi_{\alpha,\nu}^2 \quad \text{oder} \quad T > \chi_{1-\alpha,\nu}^2. \quad (5.62)$$

<http://www.itl.nist.gov/div898/handbook/eda/section3/eda358.htm>

5.7 Anwendung von Hypothesentests

Anwendung finden die Hypothesentests im Bereich der Qualitätssicherung in der Produktion (siehe beispielsweise zur Prüfung des Stout in der Guinnessbrauerei schon vor 100 Jahren), oder im Bereich der Naturwissenschaften (Biologie, Chemie, Pharmazie, Medizinen) zum Nachweis von Substanzen. Es geht darum, dass zu bewerten ist, mit welcher Wahrscheinlichkeit Beobachtungen zu welcher Verteilung gehören. Einer Entscheidung liegen Regeln zugrunde. Mit dem t-Test haben wir die Regeln kennengelernt, über die wir entscheiden, ob zwei Datensätze zu der gleichen gaußverteilten Grundgesamtheit - der gleichen Normalverteilung mit demselben Erwartungswert gehören.

Als nächstes wollen wir untersuchen, welche Fehler mit welcher Wahrscheinlichkeit dabei entstehen können. Anschließend wollen wir uns ansehen, wie eine Stichprobe mit einer Toleranzvorgabe ver-

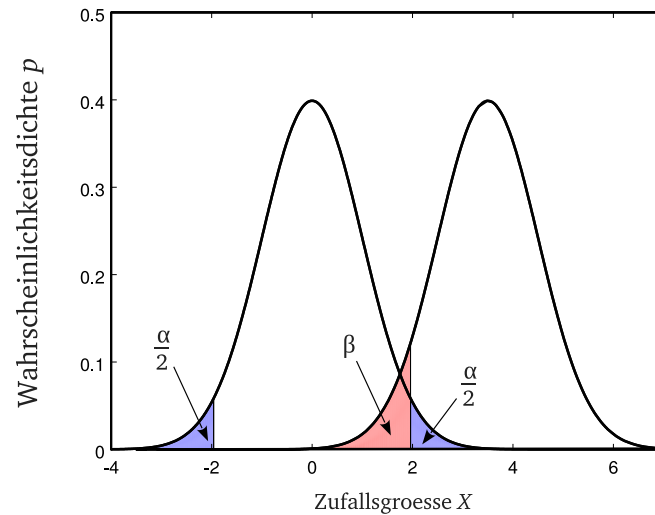


Abbildung 5.9: Die Ereignisse (beobachtete Werte) aus der linken Wahrscheinlichkeitsdichteverteilung, die in deren *Tails* liegen, werden mit einem Signifikanzniveau von α verworfen. Die Ereignisse (beobachtete Werte) aus der rechten Wahrscheinlichkeitsdichteverteilung, die in dessen linkem *Tail* liegen, werden mit Wahrscheinlichkeit β als der linken Wahrscheinlichkeit zugehörig erachtet.

glichen wird. Die Toleranzvorgabe liefert ein Intervall, dessen Lage mit der Lage der Wahrscheinlichkeitsverteilung der zu untersuchenden Stichprobe verglichen wird.

5.7.1 Entscheidung bei Vergleich zweier Stichproben

Den Hypothesentests, die wir uns angesehen haben, ist gemeinsam, dass

1. eine Nullhypothese H_0 (evtl. auch eine Alternativhypothese H_a) aufgestellt wurde,
2. ein Signifikanzniveau α spezifiziert wurde,
3. eine Stichprobe mit einer Anzahl Freiheitsgrade ν genommen wurde und
4. anhand der vorher aufgestellten Entscheidungsregeln (Vergleich einer Prüfgröße mit dem Quantil einer Wahrscheinlichkeitsdichteverteilung) die Hypothese verworfen oder angenommen wurde.

In der Entscheidungsregel werden durch Vorgabe eines Signifikanzniveaus Verwerfungsbereich und Annahmebereich festgelegt. Das Signifikanzniveau ist dabei Komplementärwahrscheinlichkeit (Gegenwahrscheinlichkeit) zum Vertrauensniveau, das ist die Wahrscheinlichkeit dafür wie sicher die Hypothese ist.

Die Wahrscheinlichkeit, mit einem Hypothesentest eine falsche Entscheidung zu treffen wird in zwei Klassen von Fehlern aufgeteilt, den α -Fehler oder den β -Fehler

	H_0 annehmen	H_0 ablehnen
H_0 ist wahr	richtige Entscheidung, Wahrscheinlichkeit: $1 - \alpha$	Fehler 1. Art, Wahrscheinlichkeit α
H_0 ist falsch	Fehler 2. Art, Wahrscheinlichkeit: β	richtige Entscheidung, Wahrscheinlichkeit: $1 - \beta$

Abb. 5.9 soll illustrieren, dass Beobachtungswerte, die zu einer Grundgesamtheit mit einer Wahrscheinlichkeitsdichteverteilung mit Erwartungswert $\mu_a = 3.5$ gehören, per Hypothesentest einer Grundgesamtheit mit $\mu_0 = 0$ zugeordnet werden. Die Wahrscheinlichkeit für das Auftreten solcher Beobachtungen ist dann β . Das Diagramm wurde mit folgendem Octaveskript erzeugt

```
function plot_entscheidungsfehler()
    dz = 0.02;
    lim = 7;
    z = [-lim:dz:lim];
    p1 = normpdf(z,0,1);
    t = 1.96;
    zleft = [-lim:dz:-t];
    pleft = normpdf(zleft,0,1);
    zright = [t:dz:lim];
    pright = normpdf(zright,0,1);
    mu2 = 3.5;
    p2 = normpdf(z+mu2,mu2,1);
    zbeta = [mu2-lim:dz:t];
    pbeta = normpdf(zbeta,mu2,1);
    figure(200); hold on;
    area( zleft, pleft, 'Facecolor', [0.5 0.5 1]);
    area( zright, pright, 'Facecolor', [0.5 0.5 1]);
    area( zbeta, pbeta, 'Facecolor', [1 0.5 0.5]);
    plot( z, p1, 'k-', 'linewidth', 2);
    plot( z+mu2, p2, 'k--', 'linewidth', 2);
    xlabel('X', 'fontsize', 14);
    ylabel('p', 'fontsize', 14);
    axis([-4 7 0 0.5]);
    set(gca, 'fontsize', 14);
    hold off;
    print(200, 'entscheidungsfehler.svg', '-dsvg');
end
```

5.7.2 Konformitätswahrscheinlichkeiten

Nachdem wir gesehen haben wie wir zwei Hypothesen über die Lage zweier Wahrscheinlichkeitsdichteverteilungen verglichen haben, befassen wir uns jetzt mit dem Vergleich eines Intervalls, einem

Toleranzintervall, mit einer Wahrscheinlichkeitsdichteverteilung. Für die Fertigung werden in den Konstruktionszeichnungen zu den Merkmalen von Bauteilen die Toleranzen eingetragen, innerhalb derer die Merkmale des Werkstücks liegen müssen. Wie nun bewertet man nach Fertigstellung des Werkstücks, ob ein Merkmal innerhalb einer Toleranzvorgabe liegt? Ein Merkmal wird durch einen Messvorgang geprüft. Der Messvorgang liefert ein Ergebnis: Einen Zahlenwert für das Merkmal und ein Überdeckungsintervall zusammen mit einem Vertrauensniveau. Anstelle von Überdeckungsintervall und Vertrauensniveau kann auch direkt eine Wahrscheinlichkeitsdichteverteilung angegeben werden, ist aber sehr unüblich in der industriellen Praxis. Für die Bewertung, ob das gefertigte Werkstück mit der Vorgabe der Konstruktion übereinstimmt, konform ist, werden Messergebnis und Toleranzvorgabe verglichen.

Als Ergänzung zum „Guide of Uncertainty“ gibt es das Dokument *Evaluation of measurement data - The role of measurement uncertainty in conformity assessment*, das darlegt, wie Konformitätsbewertungen vorzunehmen sind.

https://www.bipm.org/utils/common/documents/jcgm/JCGM_106_2012_E.pdf

Dort heißt es in Paragraph 7.1.1

An item conforms to a specified requirement if the true value of its associated property Y lies in the tolerance interval. Knowledge of Y is conveyed by a probability density function $p(y|X_1, \dots, X_J)$ so that a statement of conformity is always an inference, with some probability of being true. Denoting the set of permissible (conforming) values Y by C , the conformance probability, denoted by p_c , is given by

$$p_c = P(Y \in C | X_1, \dots, X_J) = \int_C p(y | X_1, \dots, X_J) dy. \quad (5.63)$$

Im Originaldokument werden Sie eine etwas andere Schreibweise für die Bezeichner in der Formel vorfinden. Hier ist es so aufgeschrieben, wie es zu der Notation innerhalb dieses Vorlesungsskriptes, insbesondere zu Kapitel 4, passt. Für die Toleranz wird in der Richtlinie zur Konformitätsbewertung zur Wahrung der Allgemeingültigkeit eine Menge C angegeben. Falls die Größe Y eine skalarwertige Größe wie bei dem Beispiel, anhand dessen wir in Kapitel 4 das Prinzip der bayesischen Statistik illustriert haben, so steht C für ein Intervall $C = [y_L, y_U]$. Dabei sollen der Index L für „lower limit“ und der Index U für „upper limit“ stehen.

Die Wahrscheinlichkeit p_c , dass die Beobachtungen X_1, \dots, X_J innerhalb des Toleranzintervalls $C = [y_L, y_U]$ liegen, ist

$$p_c = \int_{y_L}^{y_U} p(y | X_1, \dots, X_J) dy. \quad (5.64)$$

Für den Fall, dass von einer normalverteilten Stichprobe ausgegangen werden kann und dass eine Einzelgröße vorliegt, wird auch der Stichprobe X_1, \dots, X_J der Schätzer \bar{y} aus einfacher Mittelwertbildung gewonnen ($\bar{y} = \sum_{j=1}^J X_j$) und die Varianz als empirische Varianz s aus $s = \frac{1}{J-1} \sum (X_j - \bar{y})^2$ und

für die Wahrscheinlichkeitsdichte p die Gaußverteilung:

$$p_c = \frac{1}{\sqrt{2\pi}s} \int_{y_L}^{y_U} e^{-\frac{1}{2}\left(\frac{y-\bar{y}}{s}\right)^2} dy \quad (5.65)$$

Die Größe $\frac{y-\bar{y}}{s}$ ist eine normierte Zufallsgröße. Die Integrationsgrenzen können gleichfalls normiert werden: $z = \frac{y-\bar{y}}{s}$, $z_L = \frac{y_L-\bar{y}}{s}$ und $z_U = \frac{y_U-\bar{y}}{s}$, so dass die Tabellenwerke oder Bibliotheksfunktion für die gauß'sche Fehlerfunktion (*error function*) wie folgt verwendet werden kann:

$$p_c = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_U} e^{-\frac{1}{2}z^2} dz - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_L} e^{-\frac{1}{2}z^2} dz \quad (5.66)$$

d.h.

$$p_c = \text{erf}(z_U) - \text{erf}(z_L). \quad (5.67)$$

5.8 Übungsaufgaben zum Selbststudium

Aufgabe 1

Probieren Sie anhand von Beobachtungen, die Sie mit Hilfe eines Generators, der normalverteilte Zufallszahlen liefert, den Kolmogorow-Smirnow-Test, kurz KS-Test, aus.

In Matlab/Gnu-Octave könnten Sie dies beispielsweise wie folgt realisieren:

```
J1 = 2000; % Stichprobenumfang
mu1 = 23; % Erwartungswert der Zufallsgroesse
s1 = 3; % Wurzel aus dem Erwartungswert fuer die Varianz
Xarray = s1*randn(J1,1) + mu1;
```

Das Sortieren mit Matlab/Gnu-Octave können Sie mit

```
[xsort, isort] = sort(Xarray, 'ascend');
```

und die Wahrscheinlichkeiten mit

```
h = [1:J1]' / J1;
```

berechnen.

1. Verwenden Sie nicht die à priori eingesetzten Erwartungswerte, sondern Mittelwert $\bar{y} = \text{mean}(Xarray)$ und empirische Standardabweichung $s = \text{std}(Xarray)$, um die Funktion $P(X)$ aus Gl. (5.3) zu

realisieren. Berechnen Sie die relativen Häufigkeiten H gemäß Gl. (5.5), sowie den in Gl. (5.9) definierten Schwellwert

$$K_{\alpha,J}$$

zu einem Signifikanzniveau von $\alpha = 0.05$, wobei J der Stichprobenumfang ist. Führen Sie den Test mit den von Ihnen erzeugten Zufallswerten durch.

2. Erzeugen Sie einen zusätzlichen Satz von Zufallszahlen, beispielsweise mit den Werten $\mu_2 = 35$ und $s_2 = 5$. Wählen Sie für diese einen deutlich kleineren Stichprobenumfang. Führen Sie den KS-Test für die vereinigten Zufallszahlenarrays durch und tun Sie so, als ob Sie nicht wüssten, dass dies keine gemeinsame Grundgesamtheit ist.

Aufgabe 2

Gegeben seien zwei Stichproben

Stichprobe 1:

21	33	19	39	7
----	----	----	----	---

Stichprobe 2:

53	69	63	47	49	44	47	44	38
----	----	----	----	----	----	----	----	----

- a) Geben Sie zu jeder der beiden Stichproben die Mittelwerte und Standardabweichungen an.
- b) Prüfen Sie auf einem Signifikanzniveau von $\alpha = 0.05$ die Hypothese H_0 , ob beide Stichproben zu einer Grundgesamtheit mit demselben Erwartungswert μ gehören. Geben Sie dazu die Formel und den Wert der Testgröße an und vergleichen Sie diese mit dem entsprechenden Quantil der für diesen Test zu verwendenden Verteilung. Mit welcher Verteilung ist dieser Test durchzuführen?
- c) Prüfen Sie auf einem Signifikanzniveau von $\alpha = 0.05$ die Hypothese H_0 , ob die zweite Stichprobe zu einer Grundgesamtheit mit der Standardabweichung $\sigma_0 = 6$ gehört. Geben Sie dazu die Formel und den Wert der Testgröße an und vergleichen Sie diese mit dem entsprechenden Quantil der für diesen Test zu verwendenden Verteilung. Mit welcher Verteilung ist dieser Test durchzuführen? Führen Sie nur den einseitigen Test durch, das heißt prüfen Sie nur bezüglich des rechten *Tail* der Verteilungsdichte.
- d) Führen Sie denselben Test wie in (c) durch, aber dieses Mal für $\sigma_0 = 9$.

Verwenden Sie die auf den folgenden Seiten abgedruckten Quantiltabellen.

5.9 Quantiltabellen

5.9.1 Quantile der Student-t Verteilung

d.o.f	Wahrscheinlichkeit				
ν	0.995	0.990	0.975	0.950	0.800
2	9.925	6.965	4.303	2.920	1.061
3	5.841	4.541	3.182	2.353	0.978
4	4.604	3.747	2.776	2.132	0.941
5	4.032	3.365	2.571	2.015	0.920
6	3.707	3.143	2.447	1.943	0.906
7	3.499	2.998	2.365	1.895	0.896
8	3.355	2.896	2.306	1.860	0.889
9	3.250	2.821	2.262	1.833	0.883
10	3.169	2.764	2.228	1.812	0.879
11	3.106	2.718	2.201	1.796	0.876
12	3.055	2.681	2.179	1.782	0.873
13	3.012	2.650	2.160	1.771	0.870
14	2.977	2.624	2.145	1.761	0.868
15	2.947	2.602	2.131	1.753	0.866
16	2.921	2.583	2.120	1.746	0.865
17	2.898	2.567	2.110	1.740	0.863
18	2.878	2.552	2.101	1.734	0.862
19	2.861	2.539	2.093	1.729	0.861
20	2.845	2.528	2.086	1.725	0.860
21	2.831	2.518	2.080	1.721	0.859
22	2.819	2.508	2.074	1.717	0.858
23	2.807	2.500	2.069	1.714	0.858
24	2.797	2.492	2.064	1.711	0.857
25	2.787	2.485	2.060	1.708	0.856

d.o.f	Wahrscheinlichkeit				
ν	0.995	0.990	0.975	0.950	0.800
26	2.779	2.479	2.056	1.706	0.856
27	2.771	2.473	2.052	1.703	0.855
28	2.763	2.467	2.048	1.701	0.855
29	2.756	2.462	2.045	1.699	0.854
30	2.750	2.457	2.042	1.697	0.854
31	2.744	2.453	2.040	1.696	0.853
32	2.738	2.449	2.037	1.694	0.853
33	2.733	2.445	2.035	1.692	0.853
34	2.728	2.441	2.032	1.691	0.852
35	2.724	2.438	2.030	1.690	0.852
36	2.719	2.434	2.028	1.688	0.852
37	2.715	2.431	2.026	1.687	0.851
38	2.712	2.429	2.024	1.686	0.851
39	2.708	2.426	2.023	1.685	0.851
40	2.704	2.423	2.021	1.684	0.851
41	2.701	2.421	2.020	1.683	0.850
42	2.698	2.418	2.018	1.682	0.850
43	2.695	2.416	2.017	1.681	0.850
44	2.692	2.414	2.015	1.680	0.850
45	2.690	2.412	2.014	1.679	0.850
46	2.687	2.410	2.013	1.679	0.850
47	2.685	2.408	2.012	1.678	0.849
48	2.682	2.407	2.011	1.677	0.849
49	2.680	2.405	2.010	1.677	0.849
50	2.678	2.403	2.009	1.676	0.849

5.9.2 Quantile der χ^2 -Verteilung

	Wahrscheinlichkeit				
ν	0.995	0.990	0.975	0.950	0.800
2	10.597	9.210	7.378	5.991	3.219
3	12.838	11.345	9.348	7.815	4.642
4	14.860	13.277	11.143	9.488	5.989
5	16.750	15.086	12.833	11.070	7.289
6	18.548	16.812	14.449	12.592	8.558
7	20.278	18.475	16.013	14.067	9.803
8	21.955	20.090	17.535	15.507	11.030
9	23.589	21.666	19.023	16.919	12.242
10	25.188	23.209	20.483	18.307	13.442
11	26.757	24.725	21.920	19.675	14.631
12	28.300	26.217	23.337	21.026	15.812
13	29.819	27.688	24.736	22.362	16.985
14	31.319	29.141	26.119	23.685	18.151
15	32.801	30.578	27.488	24.996	19.311
16	34.267	32.000	28.845	26.296	20.465
17	35.718	33.409	30.191	27.587	21.615
18	37.156	34.805	31.526	28.869	22.760
19	38.582	36.191	32.852	30.144	23.900
20	39.997	37.566	34.170	31.410	25.038
21	41.401	38.932	35.479	32.671	26.171
22	42.796	40.289	36.781	33.924	27.301
23	44.181	41.638	38.076	35.172	28.429
24	45.559	42.980	39.364	36.415	29.553
25	46.928	44.314	40.646	37.652	30.675

	Wahrscheinlichkeit				
ν	0.995	0.990	0.975	0.950	0.800
26	48.290	45.642	41.923	38.885	31.795
27	49.645	46.963	43.195	40.113	32.912
28	50.993	48.278	44.461	41.337	34.027
29	52.336	49.588	45.722	42.557	35.139
30	53.672	50.892	46.979	43.773	36.250
31	55.003	52.191	48.232	44.985	37.359
32	56.328	53.486	49.480	46.194	38.466
33	57.648	54.776	50.725	47.400	39.572
34	58.964	56.061	51.966	48.602	40.676
35	60.275	57.342	53.203	49.802	41.778
36	61.581	58.619	54.437	50.998	42.879
37	62.883	59.893	55.668	52.192	43.978
38	64.181	61.162	56.896	53.384	45.076
39	65.476	62.428	58.120	54.572	46.173
40	66.766	63.691	59.342	55.758	47.269
41	68.053	64.950	60.561	56.942	48.363
42	69.336	66.206	61.777	58.124	49.456
43	70.616	67.459	62.990	59.304	50.548
44	71.893	68.710	64.201	60.481	51.639
45	73.166	69.957	65.410	61.656	52.729
46	74.437	71.201	66.617	62.830	53.818
47	75.704	72.443	67.821	64.001	54.906
48	76.969	73.683	69.023	65.171	55.993
49	78.231	74.919	70.222	66.339	57.079
50	79.490	76.154	71.420	67.505	58.164

Kapitel 6

Auswertung von Mess- und Ringvergleichen

6.1 Auswertung von Messvergleichen mit Referenzlabor

In dieser Vorlesung wird die Anwendung des t-Tests und des Chi2-Tests auf eines der zentralen Themen der Metrologie gezeigt, welches die Durchführung von Ringvergleichen ist. Wesen der Metrologie ist, dass Messgrößen auf die SI-Einheiten zurückgeführt werden, damit sie weltweit vergleichbar sind. Um die Vergleichbarkeit von Ergebnissen unterschiedlicher gesetzlich geprüfter Laboratorien, seien dies Staatsinstitute verschiedener Länder oder Kalibrierlaboratorien innerhalb eines Landes, zu gewährleisten, werden Ringvergleiche durchgeführt. Dazu wird beispielsweise ein Messobjekt (Prüfkörper) rumgeschickt und jedes beteiligte Labor misst an demselben Objekt eine genau spezifizierte Messgröße nach einem vorgegebenen Verfahren. Um die Ergebnisse miteinander zu vergleichen, werden die statistischen Verfahren der Hypothesentests auf Gleichheit der Mittelwerte und der Standardabweichungen eingesetzt.

Es wird bei der Verwendung des t-Tests danach unterschieden, ob die Nullhypothese getestet wird, dass Mittelwerte der Laboratorien mit einem Erwartungswert (dem Referenzwert) übereinstimmen. Der Referenzwert wird beispielsweise von einem Referenzlabor zur Verfügung gestellt, das die Möglichkeit hatte, mit deutlich mehr Aufwand und genaueren Geräten, messen zu können. Für den Vergleich wird die zu vergleichende Messgröße normiert, wie wir es in der letzten Vorlesung schon kennengelernt hatten.

Wir betrachten die Messergebnisse mit Größenwerten x_i für $i = 1, \dots, N$ und Unsicherheiten u_i von N Laboratorien (Partner des Ringvergleichs).

Die standardnormalverteilten Zufallsgrößen, die die Differenz zwischen dem Ergebnis eines Labors bezogen auf den Referenzwert repräsentieren, werden auch Z-Scores oder Z-Werte genannt. Sie werden genutzt, wenn z. B. das Messergebnis eines Partners i mit dem Referenzwert (Erwartungswert) verglichen werden soll. Liegt das Messergebnis des Partners i oberhalb des Referenzwertes, so ist der Z-Wert positiv. Liegt das Messergebnis unterhalb des Erwartungswertes, so ist der Z-Wert negativ. Um den Z-Wert zu bestimmen, müssen der Erwartungswert μ und die Standardabweichung σ der zu Grunde liegenden Verteilung bekannt sein. Es reicht nicht aus nur die empirischen Werte (empirischer Erwartungswert, d.h. Mittelwert, und empirische Standardabweichung) aus den Stichproben zu schätzen. Ist X eine Zufallsvariable mit dem Erwartungswert $E(X) = \mu$ und der Varianz $\text{Var}(X) = \sigma^2$ erhält man die zugehörige normierte Zufallsgröße Z durch:

$$Z = \frac{X - \mu}{\sigma} \quad (6.1)$$

Für den Erwartungswert und die Varianz von Z gilt:

- $E(Z) = E\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma}(E(X) - \mu) = 0$
- $\text{Var}(Z) = \text{Var}\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma^2}\text{Var}(X) = 1$

In Abb.6.1 ist eine Standardnormalverteilung mit den Z-Scores (Z-Werten) dargestellt. Z-Scores können nur berechnet werden, wenn die zugrunde liegende Verteilung bekannt ist. Im Falle von t-Tests haben wir hingegen auch die Möglichkeit auf die aus der Stichproben geschätzten Parameter für den Erwartungswert und die Varianz zurückzugreifen. In Abb.6.1 ist zusätzlich zu der Größe Z-Score ein

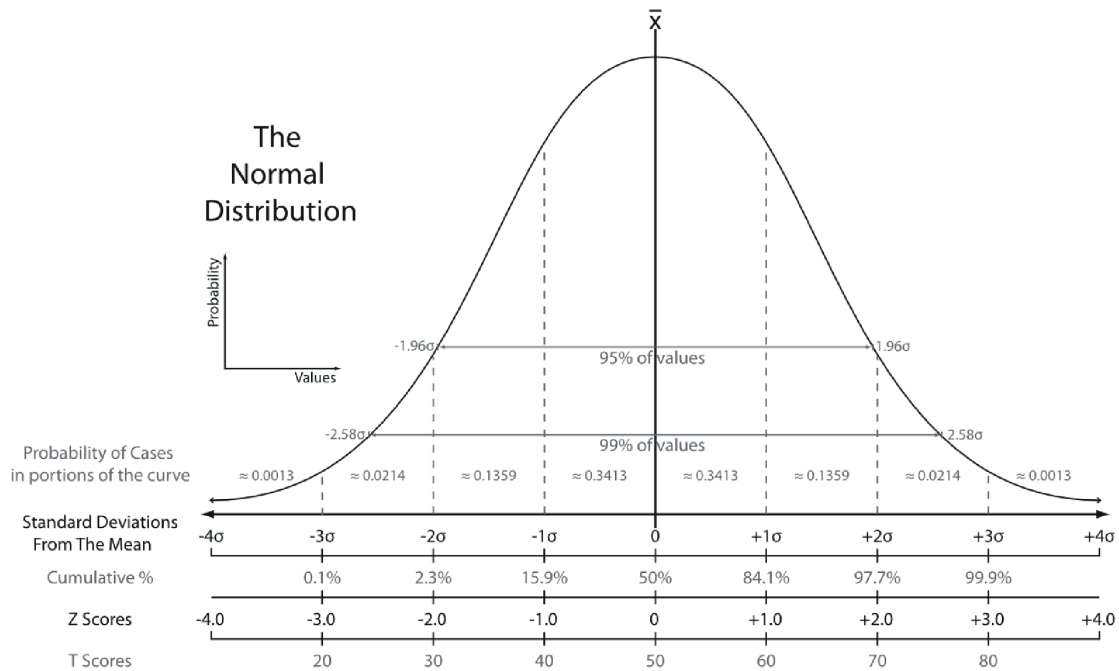


Abbildung 6.1: Die Standard-Normalverteilung $\mathcal{N}(0, \sigma^2)$ mit den entsprechenden Z-Werten.

Quelle: https://en.wikipedia.org/wiki/Standard_score

weitere weitere Größe aufgetragen: T-Score. Diese ist direkt proportional zu den Z-Scores nur mit einer anderen Normierung, derart skaliert, dass ihr Erwartungswert 50 ist und die Standardabweichung den Wert 10 annimmt.

Entsprechend der internationalen Norm [ISO13528] kann der Z-Wert für den Vergleich eines Messwertes x_i eines Labors i mit dem Referenzwert x_{ref} , der eine Standardabweichung (Standardunsicherheit) σ_{ref} hat, verglichen werden. Es wird folgender Z-Score berechnet und anschließend bewertet:

$$z_i = \frac{x_i - x_{\text{ref}}}{\sigma_{\text{ref}}} \quad (6.2)$$

Nach [ISO13528] gibt es die folgende übliche Interpretation für die Bewertung:

- Ein Ergebnis mit $|z_i| \leq 2.0$ ist noch akzeptabel.
- Ein Ergebnis mit $2.0 < |z_i| < 3.0$ gibt ein Warnsignal.
- Ein Ergebnis mit $|z_i| \geq 3.0$ wird als nicht akzeptierbar bewertet.

Sind die erweiterten Unsicherheiten sowohl des Referenzlabors U_{ref} als auch des Teilnehmerlabors des i -ten Partners U_i für das Überdeckungsintervall $[-1,96u; 1,96u]$ zum 97,5% Vertrauensniveau bekannt, so kann der sogenannte **En-Wert** berechnet werden.

In der vorigen Vorlesung hatten wir die Prüfgröße $T = \frac{x_1 - x_2}{\sqrt{s_1^2 + s_2^2}}$, die prüft ob zwei Stichproben mit ihren Erwartungswerten x_1 und x_2 zur selben Grundgesamtheit gehören. Dieser Hypothesentest prüft diese Größe auf $|T| > k$. Der Parameter k steht hier allgemein für das Quantil, also für $z_{1-\alpha/2}$ im Falle

der Verwendung der Standardnormalverteilung wie sie hier in Abb. 6.1 dargestellt ist. Er steht für das Quantil $t_{1-\alpha/2,\nu}$ im Falle der Verwendung der t-Verteilung (α : Signifikanzniveau, ν : Freiheitsgrad). Ferner steht der Parameter k auch für das Quantil einer Posteriorverteilung, also den Faktor eines *Credible* Intervalls, weshalb er allgemein *Erweiterungsfaktor* genannt wird.

Wenn wir eine Unsicherheit u_i eines Labors vorliegen haben, die beispielsweise eine Standardabweichung s_i sein kann, und haben wir ein symmetrisches Überdeckungsintervall, wie es bei normalverteilten oder t-verteilten Zufallsgrößen der Fall ist, so ist die halbe Breite des Überdeckungsintervalls $U_i = ku_i$. Dabei wird U_i die *erweiterte Messunsicherheit* genannt.

Jetzt dividieren wir die Prüfgröße T durch das Quantil bzw. den Erweiterungsfaktor k

$$\frac{T}{k} = \frac{x_1 - x_2}{k\sqrt{s_1^2 + s_2^2}} = \frac{x_1 - x_2}{\sqrt{(ks_1)^2 + (ks_2)^2}}$$

und führen die bei Ringvergleichen allgemein verwendete Bezeichnung $E_n := T/k$ ein, so dass die Prüfgröße mit $U_i = ku_i$ bzw. $U_i = ks_i$ wie folgt aussieht:

$$E_n = \frac{x_i - x_{\text{ref}}}{\sqrt{U_i^2 + U_{\text{ref}}^2}} \quad (6.3)$$

Der **En-Wert** gibt an, wie gut der Laborwert x_i mit dem Referenzwert x_{ref} übereinstimmt. Es werden hier die erweiterten Messunsicherheiten U für $k = 2$ kombiniert. Da hier mit den erweiterten Unsicherheiten gerechnet wird, liegen die Grenzen der akzeptablen Werte nicht bei 2 sondern bei 1. $|E_n|$ ist gemäß t-Test zu bewerten (siehe auch [ISO13528]):

- $|E_n| < 1$: Die Nullhypothese wird angenommen und gilt als Indikator für eine gute Übereinstimmung
- $|E_n| \geq 1$: Die Nullhypothese wird abgelehnt und gilt als Indikator, dass die Messdaten nicht konsistent zueinander sind.

6.2 Auswertung von Ringvergleichen ohne Referenzlabor

6.2.1 Vorgehensweise

Ringvergleiche zwischen den NMIs (*National Metrology Institutes*), sog. **Key-Comparisons** dienen u.a. dazu einen Referenzwert (KCRV: *key comparison reference value*) mit einer zugeordneten Unsicherheit festzulegen. Der Grad der Übereinstimmung der Messdaten von teilnehmenden Instituts i zum Referenzwert ist hier gesucht. Es gibt keinen Referenzwert, weil man a priori nicht davon ausgehen kann, dass es ein Staatsinstitut gibt, das signifikant besser messen kann als alle anderen. Ein mögliches Auswerteverfahren, das häufig Anwendung findet, ist bei [Cox02] beschrieben. Bei diesem Auswerteverfahren müssen folgende 3 Voraussetzungen erfüllt sein:

- Jeder teilnehmende Partner i ($i = 1, \dots, N$) stellt Messdaten x_i mit beigeordneter Messunsicherheit $u(x_i)$ eines Prüflings bereit. Der Prüfling muss eine gute Stabilität - auch während des Transportes - aufweisen.

- Jeder Partner stellt Messdaten zur Verfügung, die unabhängig von den anderen Partnern sind. D.h. es darf keine Messung eines Partners i von einem anderen Partner j mit $i \neq j$ abhängen.
- Die Messdaten jedes Partners (Instituts) sind normalverteilt.

Gegeben sind N teilnehmende Institutionen mit $i = 1, \dots, N$. Jeder Partner/Institut i stellt ein Messergebnis der Messgröße zur Verfügung, also einen Schätzwert x_i und eine zugeordnete Standardmessunsicherheit $u(x_i)$. Der Ablauf der Auswertung sieht wie folgt aus:

• **Schritt 1:**

Da kein Referenzwert vorliegt, wird in diesem Fall ein gewichteter Mittelwert \tilde{x} aus den Messergebnissen aller Partner bestimmt. Benutze dazu das inverse der Quadrate der zugeordneten Standardmessunsicherheiten als Gewichte:

$$\tilde{x} = \frac{x_1/u^2(x_1) + \dots + x_N/u^2(x_N)}{1/u^2(x_1) + \dots + 1/u^2(x_N)} = \frac{\sum_{i=1}^N x_i/u^2(x_i)}{\sum_{i=1}^N u^{-2}(x_i)} \quad (6.4)$$

• **Schritt 2:**

Bestimme die Standardabweichung des gewichteten Mittelwertes $u(\tilde{x}) = \sigma(\tilde{x})$:

$$\frac{1}{u^2(\tilde{x})} = \frac{1}{u^2(x_1)} + \dots + \frac{1}{u^2(x_N)} \quad (6.5)$$

• **Schritt 3:**

Führe einen Konsistenzcheck durch, ob die angegebenen Messergebnisse konsistent zueinander sind. Führe dazu den χ^2 -Test durch mit der χ^2 -Variablen (Testgröße T):

$$T := \chi_{\text{obs}}^2 = \frac{(x_1 - \tilde{x})^2}{u^2(x_1)} + \dots + \frac{(x_N - \tilde{x})^2}{u^2(x_N)} \quad (6.6)$$

Diese Testgröße weist eher einen kleinen Wert auf, wenn alle Partner recht dicht am Mittelpartner liegen und natürlich umgekehrt. Vergleiche diese Testgröße mit der $\chi^2(\nu)$ -Verteilung mit Freiheitsgrad $\nu = N - 1$ (N Anzahl der Partner), die am Messvergleich teilgenommen haben. Der Konsistenzcheck schlägt fehl, wenn die folgende Bedingung erfüllt ist:

$$\Pr(\chi^2(\nu) > T) < 0.05 \quad (6.7)$$

„Pr“ steht für „Probability (Wahrscheinlichkeit)“. Der χ^2 -Test verlangt als Voraussetzung, dass die Messergebnisse normalverteilt sind.

• **Schritt 4:**

Falls der Konsistenzcheck nicht fehlschlägt, wird der Wert \tilde{x} als Referenzwert (KCRV: key comparison reference value) x_{ref} akzeptiert mit der Unsicherheit $u(x_{\text{ref}}) = u(\tilde{x})$. Nun kann der **Grad der Übereinstimmung** $d_i = x_i - x_{\text{ref}}$ der Partner $i = 1, \dots, N$ mit dem Referenzwert x_{ref} wie folgt bestimmt werden. Wir definieren die normierten Gewichte \tilde{w}_i

$$\tilde{w}_i := \frac{u^2(x_{\text{ref}})}{u^2(x_i)} \quad (6.8)$$

mit

$$\sum_{i=1}^N \tilde{w}_i = 1. \quad (6.9)$$

Der Referenzwert $x_{\text{ref}} \equiv \tilde{x}$ aus Gl. (6.4) sieht damit wie folgt aus:

$$x_{\text{ref}} = \sum_{i=1}^N \tilde{w}_i x_i \quad (6.10)$$

Für den Grad der Übereinstimmung erhalten wir:

$$d_i = x_i - x_{\text{ref}} = x_i - \sum_{j=1}^N \tilde{w}_j x_j = x_i - \tilde{w}_i x_i - \sum_{\substack{j=1 \\ j \neq i}}^N \tilde{w}_j x_j = (1 - \tilde{w}_i) x_i - \sum_{\substack{j=1 \\ j \neq i}}^N \tilde{w}_j x_j \quad (6.11)$$

Wenn die Messungen nicht gegenseitig voneinander abhängen, so kann zur Berechnung der Unsicherheit von d_i das Gesetz der Fortpflanzung der Unsicherheiten angewendet werden. (Hinweis: Sind ein Modell mit $y = f(x_1, x_2)$ und die dazugehörigen Unsicherheiten $u(x_1)$ sowie $u(x_2)$ gegeben, so ergibt sich die Unsicherheit von y bei Nichtkorrelation von x_1 und x_2 nach dem Gesetz der Fehlerfortpflanzung zu: $u^2(y) = \left(\frac{\partial f}{\partial x_1}\right)^2 u^2(x_1) + \left(\frac{\partial f}{\partial x_2}\right)^2 u^2(x_2)$)

Für die Unsicherheit der Differenz d_i ergibt sich somit:

$$u^2(d_i) = \left(\frac{\partial d_i}{\partial x_i}\right)^2 \cdot u^2(x_i) + \sum_{\substack{j=1 \\ j \neq i}}^N \left(\frac{\partial d_i}{\partial x_j}\right)^2 \cdot u^2(x_j) \quad (6.12)$$

$$= (1 - \tilde{w}_i)^2 u^2(x_i) + \sum_{\substack{j=1 \\ j \neq i}}^N \tilde{w}_j^2 u^2(x_j) \quad (6.13)$$

$$= ((1 - \tilde{w}_i)^2 - \tilde{w}_i^2) u^2(x_i) + \sum_{j=1}^N \tilde{w}_j^2 u^2(x_j) \quad (6.14)$$

Mit Gl.(6.8) ergibt sich:

$$u^2(d_i) = (1 - 2\tilde{w}_i^2) u^2(x_i) + \sum_{j=1}^N \tilde{w}_j u^2(x_{\text{ref}}) \quad (6.15)$$

Mit Gl.(6.8) und Gl.(6.9) ergibt sich daraus:

$$\begin{aligned} u^2(d_i) &= u^2(x_i) - 2u^2(x_{\text{ref}}) + u^2(x_{\text{ref}}) \sum_{j=1}^N \tilde{w}_j \\ &= u^2(x_i) - u^2(x_{\text{ref}}) \end{aligned} \quad (6.16)$$

Auch hier kann nun wieder ein En-Wert angegeben werden. Da alle Größen normalverteilt sind, sind die erweiterten Unsicherheiten für einen Erweiterungsfaktor von $k = 2$, $U(x_i) = k \cdot u(x_i) = 2 \cdot u(x_i)$. Damit lässt sich der En-Wert mit Gl.(6.3) bestimmen, der den Grad der Übereinstimmung

des beteiligten Partners i mit dem Referenzwert angibt.

$$E_n(x_i) = \frac{d_i}{2\sqrt{u^2(d_i)}} \quad (6.17)$$

Es gelten wieder die Aussagen von Kapitel 1:

- $|E_n| < 1$ ist ein Indikator für eine gute Übereinstimmung.
- $|E_n| \geq 1$ ist ein Indikator, dass die Messdaten nicht konsistent zueinander sind.

6.2.2 Beispiel

5 Partner messen ein Massestück von ca. 100 g. Die Ergebnisse der Partner sind wie folgt (alle Angaben sind in g):

i	1	2	3	4	5
x_i/g	99.82	99.05	99.17	99.20	99.38
$u(x_i)/\text{g}$	0.80	0.63	0.86	0.82	0.98

Als Ergebnis erhalten wir: $x_{\text{ref}} = 99.29$ und $u_{\text{ref}} = 0.35$. Mit der Testgröße $T = 0.6239$ ergibt sich die Wahrscheinlichkeit $\Pr\{\chi^2(\nu) > T\} = 0.8858$. Da die Wahrscheinlichkeit größer als 0.05 ist, schlägt der Konsistenzcheck nicht fehl, d.h. die Daten der 5 Messpartner sind konsistent.

Für die Abweichungen d_i erhalten wir:

$$[0.7172 \ 0.5209 \ 0.7836 \ 0.7395 \ 0.9137].$$

Für die En-Werte erhalten wir:

$$[0.3676 \ -0.2329 \ -0.0783 \ -0.0626 \ 0.0478].$$

Das heisst der Partner 5 ist in bester Übereinstimmung mit dem Referenzwert.

Der Octave-/Matlab-Code dazu lautet (Hinweis zu Octave: Evtl. ist die Funktion `chi2pdf` in Octave noch nicht vorhanden. Dann zuerst das Paket statistics installieren, `pkg install -forge package_name` mit `package_name = statistics`):

```
% clear all;
x = [99.82, 99.05, 99.17 , 99.20 ,99.38];
u_x = [0.80 0.63, 0.86, 0.82, 0.98];
U_x = 2*u_x;

% Anzahl der Freiheitsgrade
nu = 4;

% Schritt 1:
```

```

x_ref = sum(x./(u_x).^2) ./ sum(1./(u_x).^2)

% Schritt 2:
u_ref = sqrt(1./sum(1./(u_x).^2))

% Schritt 3:
T = sum((x-x_ref).^2./(u_x.^2))

% Bestimme die Wahrscheinlichkeit Pr der Chi^2-Verteilung für Chi^2(\nu) > T:
Pr = 1 - chi2pdf(T,nu)

% Schritt 4: Grad der Übereinstimmung
d = x-x_ref;
u_d = sqrt(u_x.^2 - u_ref.^2)
En = d./(2*sqrt(u_d.^2))

```

6.2.3 Identifikation von Ausreißern und Konsistenzcheck

Wir wollen uns hier noch etwas genauer anschauen, wie man feststellt, ob die Messdaten eines Ringvergleiches konsistent sind und wie man mit Inkonsistenten umgeht. Als **Ausreißer** werden häufig Messwerte deklariert, deren Abweichungen größer als 3 mal die erweiterte Messunsicherheit des Referenzwertes [GuideKey] sind. Diese Ausreißer werden in Abstimmung mit dem Partner gelöscht. Anschließend werden die verbliebenen Daten auf statistische Konsistenz überprüft. Wir haben dazu bereits in Gl.(6.6) und Gl.(6.7) eine Formel angegeben, mit der überprüft werden kann, ob die angegebenen Messergebnisse (Messwerte mit Unsicherheiten) konsistent zu dem Referenzwert und dessen Unsicherheit ist. Die Testgröße T basiert auf dem Birge-Test, welche ebenso ein Test auf die Konsistenz der Messdaten in einem Ringvergleich ist. Der Birge-Test ist wie der χ^2 -Test nur unter folgenden Bedingungen gültig:

- Die gemessenen Messwerte x_i der N Institute sind unkorreliert
- Man benötigt Kenntnis über die Verteilungsdichtefunktionen. Es wird vorausgesetzt, dass die Größen normalverteilt mit den Unsicherheiten $u(x_i)$ bzw. den Varianzen σ_i sind.

Das Birge-Verhältnis ist -bis auf die Anzahl der Freiheitsgrade- die Prüfgröße des Chi2-Tests (siehe vorherige Vorlesung, Gl. 5.61). Es wird folgendes geprüft (H_0 : Nullhypothese, H_a : Alternativhypothese):

H_0	$u_{\text{partner}}^2 = u_{\text{ref}}^2$ Die Stichprobe gehört zu einer Grundgesamtheit mit Varianz u_{ref}^2
H_a	$u_{\text{partner}}^2 \neq u_{\text{ref}}^2$ Die Stichprobe gehört nicht zu einer Grundgesamtheit mit Varianz u_{ref}^2

Das **Birge-Verhältnis** ist definiert als das Verhältnis der Streuungen der Unsicherheiten aller Partner (bezeichnen wir mit u_{partner}) zu der Unsicherheit des Referenzwertes u_{ref} . Es wird analog der Testgröße

des χ^2 -Tests, $T = \nu \left(\frac{s}{\sigma_0} \right)^2$ Gl. 5.61, wie folgt definiert:

$$R_B := \frac{u_{\text{partner}}}{u_{\text{ref}}} \quad \text{mit} \quad \frac{u_{\text{partner}}^2}{u_{\text{ref}}^2} \propto \nu \left(\frac{s}{\sigma_0} \right)^2. \quad (6.18)$$

Die Varianz aller N Partner ist wie folgt definiert:

$$u_{\text{partner}}^2 := \frac{\frac{1}{N-1} \sum_{i=1}^N w_i (x_i - x_{\text{ref}})^2}{\sum_{i=1}^N w_i} \quad \text{mit} \quad w_i = \frac{1}{u^2(x_i)}. \quad (6.19)$$

Entsprechend Gl.(6.5) ist die Unsicherheit des Referenzwertes wie folgt gegeben

$$u_{\text{ref}} = \left(\frac{1}{u^2(x_1)} + \cdots + \frac{1}{u^2(x_N)} \right)^{-\frac{1}{2}}, \quad (6.20)$$

so dass für das Birge-Verhältnis gilt

$$R_B = \sqrt{\frac{1}{N-1} \sum_{i=1}^N \left(\frac{x_i - x_{\text{ref}}}{u(x_i)} \right)^2}. \quad (6.21)$$

Bei dem Konsistenzcheck mit dem Birge-Verhältnis wird untersucht, ob

- $R_B \leq 1$: Konsistenzcheck ok.
- $R_B > 1$: Konsistenzcheck schlägt fehl.

Wenn $R_B > 1$ ist, bedeutet dies, dass entsprechend Gl.(6.21) die Unsicherheiten der Partner $u(x_i)$ nicht zu der Unsicherheit des Referenzwertes $u(x_{\text{ref}})$ passen. Die gemessenen Daten x_i der Partner sind inkonsistent. Für unser Beispiel Abschnitt 6.2.2 erhalten wir $R_B = 0.39494$. Damit schlägt der Birge-Test nicht fehl und die Daten sind danach konsistent. Für die Auswertung eines Ringvergleichs ist es empfehlenswert sowohl das Birge-Verhältnis als auch den χ^2 -Test (siehe Gl.(6.7)) durchzuführen und zu prüfen, ob beide Tests ok sind.

Häufig wird das Birge-Verhältnis in Gl.(6.21) umgeschrieben und in folgender Form dargestellt:

$$R_B^2 = \sum_{i=1}^N \frac{w_i (x_i - x_{\text{ref}})^2}{N-1} \quad (6.22)$$

mit den Gewichten $w_i = 1/\sigma_i^2$ für $i = 1, 2, \dots, N$ und dem Referenzwert (gewichteten Mittelwert) $x_{\text{ref}} = \sum_i w_i x_i / \sum_i w_i$. Der Birge Test kann auch bei korrelierten Messdaten x_1, \dots, x_N , wenn die Kovarianzen $\sigma_{12}, \dots, \sigma_{(N-1)N}$ gegeben sind, durchgeführt werden, siehe [Kac08].

6.2.4 Paule-Mandel Methode zur Anpassung von Gewichtungsfaktoren

Bei Ringvergleichen kommt es oftmals vor, dass jeder am Vergleich beteiligte Partner eine Unsicherheit U angibt die kleiner ist als der Abstand des Messwertes zu anderen Partnern, so dass der En-Wert bzw. des Birge-Verhältnisses unter Verwendung der Gewichte gemäß Gl. (6.8) größer als Eins wird. Durch eine Anpassung der Gewichtungsfaktoren durch Hinzufügen einer zusätzlichen Komponente, die sich aus der Streuung der Ergebnisse von Partner zu Partner ergibt, kann eine bessere Konsistenz erzielt werden.

Der Leitfaden zum Erstellen eines *Key Comparison Reports* [GuideKey] empfiehlt als Anpassungsverfahren die **Paule-Mandel Methode**, bei der die Varianzen durch Additions einer Streuung zwischen Laboratorien erhöht werden. Dadurch werden die Gewichtungsfaktoren w_i verkleinert und damit Referenzwert x_{ref} so angepasst, so dass das Birge-Verhältnis kleiner als Eins werden kann. Ein Beispiel dazu ist im Anhang B der Richtlinie für Ringvergleiche [GuideKey] zu finden.

Im folgenden schauen wir uns das Prinzip von Mandel und Paule an, das in der Publikation [Pau82] zu finden. Es werden zwei eher künstliche Beispiele gewählt, um das Prinzip besser zu verstehen. In Beispiel I wird eine Messgröße mit zwei verschiedenen Methoden (bzw. von zwei verschiedenen Partnern) A und B gemessen. Die gemessenen Einzelwerte von jedem der beiden Partner sind in Tab. 6.1 angegeben.

Tabelle 6.1: Messdaten des Beispiels I

Methode/Partner	A			B		
Gemessener Wert x_i	1.1	1.9	1.5	16	25	3

Der gleichgewichtete Mittelwert \bar{x}_I über alle Werte beider Partner gemeinsam ist:

$$\bar{x}_I = \frac{1}{N} \sum_{i=1}^N x_i = \frac{1}{6} (1.1 + 1.9 + 1.5 + 16 + 25 + 3) \approx 8.1 \quad (6.23)$$

Intuitiv würden wir jedoch sagen, dass wir den Messdaten mit der Methode A mehr vertrauen schenken würden, als den Messdaten mit der Methode B, da die Messdaten der Methode A weniger streuen als die Messdaten mit der Methode B.

Als zweites Beispiel betrachten wir, dass die Methoden (bzw. Partner) A und B ähnlich genau messen, jedoch werden von Partner A einen größeren Stichprobenumfang als von Partner B:

Tabelle 6.2: Messdaten des Beispiels II

Methode/Partner	A						B	
Gemessener Wert x_i	2.0	1.0	1.5	1.8	1.2	1.7	16.3	16.8

Für den Mittelwert über alle Werte gemeinsam erhalten wir:

$$\bar{x}_{II} = \frac{1}{N} \sum_{i=1}^N x_i = \frac{1}{8} (2.0 + 1.0 + 1.5 + 1.8 + 1.2 + 1.7 + 16.3 + 16.8) = 5.2875 \quad (6.24)$$

Es fällt auf, dass die beiden Werte von Partner B zueinander passen, sich aber deutlich von denen von Partner A unterscheiden.

Die beiden Mittelwerte der jeweiligen Partner A und B sind

$$\bar{x}_{II,A} = \frac{1}{6} (2.0 + 1.0 + 1.5 + 1.8 + 1.2 + 1.7) = 1.533, \quad \bar{x}_{II,B} = \frac{1}{2} (16.3 + 16.8) = 16.550 \quad (6.25)$$

Dieses würde ich hier an dieser Stelle gar nicht bringen, auch wenn es in Pau82 so präsentiert wird, weil Du ja hier in diesem Skript schon mit den Gewichten als Kehrwert der Varianzen für den Referenzwert, der aus dem gewichteten Mittel der Resultate aller Partner gewonnen wurde, um En-Werte für gleiche Partner zu bestimmen.

zunächst hatte ich entlang am Pau82 die Sachen so nachgerechnet und verstanden, wie Paule und Mandel sie argumentiert hatte, dann wurde mir langsam klar, dass wir in unseren Vorlesungen aber längst an dem Punkt angekommen sind, dass man die Kehrwerte der Varianzen nimmt und nicht anfängt die Mittelwerte einfach zu mitteln \Rightarrow alles Blaue hiernach hier wegnehmen

und der Mittelwert dieser beiden Mittelwerte ist

$$\bar{x}_{II,AB} = \frac{1}{2} (1.533 + 16.550) = 9.0417 \quad (6.26)$$

Dadurch dass wir Mittelwerte mitteln erhalten wir eine andere Wichtung: $\frac{\sum w_i x_i}{\sum w_i}$:

$$\frac{1}{2} \left(\frac{1}{6} \left(\sum_{i=1}^6 x_i \right) + \frac{1}{2} \left(\sum_{i=7}^8 x_i \right) \right) = \frac{1}{12} \left(\sum_{i=1}^6 x_i \right) + \frac{1}{4} \left(\sum_{i=7}^8 x_i \right)$$

also

$$\frac{w_i}{\sum_{i=1}^8 w_i} = \begin{cases} \frac{1}{12} & \text{für } i = 1, \dots, 6 \\ \frac{1}{4} & \text{für } i = 7, 8 \end{cases}$$

Diese Art der Wichtung, die Einzelwerte zu gewichten, führt dazu dass Stichprobe B sehr viel stärker gewichtet wird, in diesem Beispiel um einen Faktor 3 mit $\frac{1}{4} = \frac{3}{12}$. Es macht jedoch mehr Sinn, die Gewichte danach zu richten, wie genau die Größen sind, im Sinne von der Frage wie breit oder schmal die Wahrscheinlichkeitsdichteverteilung ist.

Wir haben im Laufe dieser Vorlesungsreihe bereits gesehen, dass Messergebnisse mit größerer Unsicherheit sinnvollerweise mit geringerer Wichtung zu berücksichtigen sind, so dass Gewichte deshalb aus dem Kehrwert der Varianzen bestimmt werden. Der Grund dafür liegt darin, dass die Varianz des gewichteten Mittelwertes x_{ref} minimiert wird, wenn die Gewichte als Reziprokwert der Varianzen der Einzelwerte berechnet werden. In Gl. (??) bzw. Gl. (6.10) haben wir deshalb zur Ermittlung des

Referenzwertes x_{ref} beim Ringvergleich mit gleichwertigen Partnern gerechnet mit

$$w_i = \frac{1}{u_i^2}$$

und

$$x_{\text{ref}} = \frac{\sum_{i=1}^N w_i x_i}{\sum_{i=1}^N w_i}.$$

wahrscheinlich macht es keinen Sinn mit der empirischen Varianz anzufangen, sondern direkt mit der empirischen Var des Mittelwertes, also nehme ich diesen Teil wieder zurück Verwenden wir für die Unsicherheit u_i die empirische Varianz $u_i = \text{Var}(x_i)$ für Beispiel II folgende Werte

$$\text{Var}(x_{II,A}) = \frac{1}{6-1} \sum_{k=1}^6 (x_{II,k} - \bar{x}_{II,A})^2 = 0.14267$$

und

$$\text{Var}(x_{II,B}) = \frac{1}{2-1} \sum_{k=7}^8 (x_{II,k} - \bar{x}_{II,B})^2 = 0.12500$$

so erhalten wir für die beiden Gewichte $w_A = 7.0093$ und $w_B = 8.0000$, die beide fast gleich. Der gewichtete Mittelwert, der als Referenzwert für En-Tests und das Birge-Verhältnis genutzt wird ist damit

$$x_{\text{ref}} = \frac{7.0093 \cdot 1.533 + 8.0000 \cdot 16.550}{7.0093 + 8.0000} = 9.5372$$

Nach Gl. (6.21) das Birge-Verhältnis hierfür

$$R_B^2 = \frac{1}{N-1} \sum_{i=1}^N \left(\frac{x_i - x_{\text{ref}}}{u^2(x_i)} \right)^2 = \frac{1}{2-1} \left(\frac{(1.5333 - 9.5372)^2}{0.14267} + \frac{(16.550 - 9.5372)^2}{0.12500} \right) = 842.47 \gg 1$$

–nix!

Da es sich um den Vergleich der Mittelwerte der beiden Stichproben A und B handelt, verwenden wir für die Unsicherheit u_i die Varianz des Mittelwertes, also $u_i = \text{Var}(\bar{x}_i)$ verwenden. Für Beispiel II sind dies folgende Werte:

$$\text{Var}(\bar{x}_{II,A}) = \frac{1}{6(6-1)} \sum_{k=1}^6 (x_{II,k} - \bar{x}_{II,A})^2 = 0.023778$$

und

$$\text{Var}(\bar{x}_{II,B}) = \frac{1}{2(2-1)} \sum_{k=7}^8 (x_{II,k} - \bar{x}_{II,B})^2 = 0.062500$$

so erhalten wir für die beiden Gewichte $w_A = 42.056$ und $w_B = 16.000$. Der gewichtete Mittelwert, der als Referenzwert für En-Tests und das Birge-Verhältnis genutzt wird, ist damit

$$x_{\text{ref}} = \frac{42.056 \cdot 1.533 + 16.000 \cdot 16.550}{42.056 + 16.000} = 5.6716$$

Nach Gl. (6.21) das Birge-Verhältnis hierfür

$$R_B^2 = \frac{1}{N-1} \sum_{i=1}^N \left(\frac{x_i - x_{\text{ref}}}{u(x_i)} \right)^2 = \frac{1}{2-1} \left(\frac{(1.5333 - 5.6716)^2}{0.023778} + \frac{(16.550 - 5.6716)^2}{0.0625} \right) = 2613.7$$

also

$$R_B = \sqrt{2613.7} = 51.1 \gg 1$$

Wir sehen, dass die Varianzen der Mittelwerte der beiden Partner A und B im Verhältnis zur Differenz zwischen den beiden Partnern sehr klein ist, oder umgekehrt gesagt liegen die Mittelwerte der Methoden A und B sehr weit auseinander im Verhältnis zu den Varianzen. Die Varianzen $\text{Var}(\bar{x}_A)$ und $\text{Var}(\bar{x}_B)$ beschreiben nur die interne Streuung, also nur die Streuung der jeweiligen Stichproben A und B. Die Wahrscheinlichkeitsverteilungen werden durch die beiden Parameter Erwartungswert, also hier Mittelwert, und Varianz charakterisiert und wir hatten bei den Hypothesentests gelernt, auf beides zu prüfen. Wir haben gesehen, dass Stichproben einer gemeinsamen Grundgesamtheit angehören, wenn beides, Mittelwert und Varianz zueinander passen.

Bei den Ringvergleichen, bei denen von verschiedenen Laboratorien oder Instituten mit unterschiedlichen Methoden oder jedenfalls unabhängigen experimentellen Aufbauten die Ergebnisse verglichen werden, kann beobachtet werden, dass entweder die Schätzer (empirischen Erwartungswerte) oder die empirischen Varianzen oder sogar beides nicht zusammen passen. In der Praxis kommt es vor, dass die jeweiligen Labore oder Institute ihre Methode immer weiter optimiert haben und ihre Genauigkeit erhöht haben, dass jedoch verbleibende, unerkannte systematische Effekte vorhanden sind. Dass es Unterschiede gibt, tritt erst bei dem Vergleich zutage und es lässt sich im Rahmen der gesetzten Zeit für das Projekt des Vergleichens nicht aufklären. In solchen Fällen lebt man dann damit, dass es auch zwischen den unterschiedlichen Methoden bzw. Laboren eine Streuung gibt.

Statistisch wird dies damit ausgedrückt, dass die Stichproben auf den verschiedenen Laboren zu verschiedenen Grundgesamtheiten gehören. Konkret bedeutet das, dass für die Gewichte zusätzlich zur Varianz des Mittelwert zur jeweiligen Methode/ des jeweiligen Laboratoriums (Streuung innerhalb einer Gruppe, *within group variability*) auch eine Varianz der Methoden/Laboratorien (Streuung zwischen Gruppen, *between group variability*) eingeführt. Die Streuung zwischen den Gruppen, die Varianz s_b^2 wird iterativ geschätzt. Für die verschiedenen Partner oder Stichproben A und B verwenden wir jetzt die Notation S_i , hier also $S_1 = A$ und $S_2 = B$, um die Gleichungen allgemein aufschreiben zu können.

$$\text{Var}(\bar{x}_{S_i,c}) = \text{Var}(\bar{x}_{S_i}) + s_b^2 = \frac{1}{M_i(M_i-1)} \sum_{k \in S_i} (x_k - \bar{x}_{S_i})^2 + s_b^2$$

mit M_i für den Stichprobenumfang zu S_i - oder kürzer aufgeschrieben mit i als Index für die Nummerierung der Partner und für die Varianzen der Mittelwerte bzw. Messergebnisse der Partner $u_i^2 = \bar{s}_i^2 = \text{Var}(\bar{x}_{S_i})$ bzw. für die Konsensusvarianzen $\bar{s}_{i,c}^2 = \text{Var}(\bar{x}_{S_i,c})$

$$\bar{s}_{i,c}^2 = \bar{s}_i^2 + s_b^2 \quad (6.27)$$

wobei der Index b für *between* und c für *consensus* stehen. Auch die Mittelwerte bzw. allgemein die Messergebnisse der jeweiligen Partner schreiben wir jetzt in der Form $\bar{x}_{S_i} = \bar{x}_i$.

Die Gewichte für den *Consensus Value*, dem besser übereinstimmenden Wert für die Messgröße, sind damit dann

$$w_{i,c} = \frac{1}{\text{Var}(\bar{x}_{S_i,c})} = \left(\frac{1}{M_i(M_i - 1)} \sum_{k \in S_i} (x_k - \bar{x}_{S_i})^2 + s_b^2 \right)^{-1}$$

mit M_i für den Stichprobenumfang zu S_i und dasselbe in der kürzeren Notation geschrieben

$$w_{i,c} = \frac{1}{\bar{s}_{i,c}^2} = \frac{1}{\bar{s}_j^2 + s_b^2}. \quad (6.28)$$

Der Konsenzreferenzwert aus den Werten aller N Partner ist dann

$$x_{\text{ref},c} = \frac{\sum_{i=1}^N w_{i,c} x_i}{\sum_{i=1}^N w_{i,c}}. \quad (6.29)$$

Das Birge-Verhältnis für diese Konsensuswerte ist somit

$$R_B = \sqrt{\frac{1}{N-1} \sum_{i=1}^N \frac{1}{\bar{s}_i^2 + s_b^2} (\bar{x}_i - x_{\text{ref},c})^2}. \quad (6.30)$$

Schätzen der Zwischengruppenvarianz s_b^2

Geschätzt wird die Zwischengruppenvarianz s_b^2 gemäß Paule und Mandel [Pau82] als diejenige Varianz, für die die Ergebnisse der Partner konsistent sein sollen, also für die das Birge-Verhältnis gegen Eins streben soll: $R_B \rightarrow 1$.

Die Optimierungsaufgabe formulieren wir für das Quadrat des Birge-Verhältnisses, um eine Kostenfunktion zu haben, die sich einfacher linearisieren lässt. Dies geht ganz gut, weil $1^2 = 1$ ist, also $R_B^2 \rightarrow 1$, d.h.

$$\frac{1}{N-1} \sum_{i=1}^N \frac{1}{\bar{s}_i^2 + s_b^2} (\bar{x}_i - x_{\text{ref},c})^2 \rightarrow 1$$

oder genauer ausgedrückt

$$\min_{s_b^2} \left\{ \frac{1}{N-1} \left(\sum_{i=1}^N \frac{1}{\bar{s}_i^2 + s_b^2} (\bar{x}_i - x_{\text{ref},c})^2 \right) - 1 \right\}. \quad (6.31)$$

Diese Optimierungsaufgabe hat eine Kostenfunktion $Q(s_b^2) = \frac{1}{N-1} \left(\sum_{i=1}^N \frac{1}{\bar{s}_i^2 + s_b^2} (\bar{x}_i - x_{\text{ref},c})^2 \right) - 1$ die von dem zu schätzenden Parameter, der Zwischengruppenvarianz $v := s_b^2$ nicht-linear abhängt, so dass sie iterativ zu lösen ist, es handelt sich also nichtlineare Optimierung wie wir es in Kapitel 3 besprochen haben.

Wir entwickeln die Kostenfunktion in eine Taylorreihe bis zum linearen Term

$$Q(v_0 + dv_0) \approx Q_0 + \left(\frac{\partial Q}{\partial v} \right) \bigg|_{v=v_0} dv_0 \quad (6.32)$$

Mit $Q(v) \rightarrow 0$ setzen wir für den ersten Iterationsschritt

$$Q_0 + \left(\frac{\partial Q}{\partial v} \right) \Big|_{v=v_0} dv_0 = 0 \quad (6.33)$$

und lösen dies nach dv_0 auf zu

$$dv_0 = - \left(\frac{Q_0}{\frac{\partial Q}{\partial v}} \right) \Big|_{v=v_0} \quad (6.34)$$

Wir beginnen mit dem Startwert für v_0 , indem wir ihn auf $v_0 = 0$ setzen, so dass wir mit dem sehr großen Birge-Verhältnis beginnen und iterativ die Zwischengruppenvarianz erhöhen, um das Birge-Verhältnis solange zu reduzieren bis es Eins wird. Im nächsten Schritt setzen wir $v_1 = v_0 + dv_0$. Dies setzen wir fort mit

$$v_{k+1} = v_k + dv_k \quad (6.35)$$

bis Q zu einem Minimum möglichst nahe bei Null konvergiert.

Die Ableitung der Kostenfunktion Q nach der Zwischengruppenvarianz ist

$$\frac{\partial Q}{\partial v} = \frac{1}{N-1} \left(\sum_{i=1}^N \frac{\partial}{\partial v} \frac{1}{\bar{s}_i^2 + v} (\bar{x}_i - x_{\text{ref},c})^2 \right) - 0$$

d.h.

$$\frac{\partial Q}{\partial v} = \frac{1}{N-1} \sum_{j=1}^N (\bar{x}_j - x_{\text{ref},c})^2 \frac{\partial}{\partial v} \frac{1}{\bar{s}_i^2 + v}$$

d.h.

$$\frac{\partial Q}{\partial v} = \frac{1}{N-1} \sum_{i=1}^N (\bar{x}_i - x_{\text{ref},c})^2 \frac{-1}{(\bar{s}_i^2 + v)^2} \quad (6.36)$$

Die Iterationen laufen, solange R_B größer als Eins ist und eine maximale Anzahl Iterationsschritte einen Wert von 50 nicht überschreitet. In dem hier behandelten Beispiel werden 17 Iterationsschritte gebraucht, bis die Optimierung konvergiert und folgendes Ergebnis für die Zwischengruppenvarianz $v = s_b^2$ liefert:

$$v = 112.707 \quad \text{d.h.} \quad s_b = 10.6$$

ich werde hier noch einen Plot erzeugen, der die Gaußverteilungen der beiden Einzelstichproben zeigt (vielleicht außerdem noch die t-Verteilungen) sowie die Gaußverteilung $\mathcal{N}(x_{\text{ref}} = 9.040, s_b = 10.6)$ für die Zwischengruppenverteilung.

Als konsistenter Referenzwert ergibt sich:

$$x_{\text{ref}} = 9.040 \quad \text{mit einer gepoolten Varianz von } \text{Var}(x_{\text{ref}}) = 0.1397 \quad (6.37)$$

wie hast Du die gepoolte Varianz definiert? Ich finde sie nicht in Deinem Matlabskript auch nicht in dem Text des Vorl-Skripts ist das

$$\text{Var}(x_{\text{ref}}) = \left(\sum_{i=1}^N \frac{1}{\bar{s}_i^2 + s_b^2} \right)^{-1} = 56.375 \quad \sqrt{\text{Var}(x_{\text{ref}})} = 7.5$$

ich muss das Octave-Skript ausprobieren, und auch die gepoolte Var

Der dazu gehörige Quellcode für Gnu-Octave/Matlab sieht wie folgt aus:

```
function consensus_fit()
    x_A = [2.0,1.0,1.5,1.8,1.2,1.7];
    x_B = [16.3,16.8];

    % jeweiliger Stichprobenumfang
    N_A = length(x_A);
    N_B = length(x_B);

    % Mittelwerte
    bar_x_A = mean(x_A);
    bar_x_B = mean(x_B);
    fprintf(stdout,'Mittelwerte: %1.2f und %1.2f\n', ...
        bar_x_A, bar_x_B);

    % Varianzen der Mittelwerte
    Var_A_mittel = var(x_A)/N_A;
    Var_B_mittel = var(x_B)/N_B;
    fprintf(stdout, ...
        'Varianzen der Mittelwerte innerhalb jeder Stichprobe:\n %1.3f und %1.3f\n', ...
        Var_A_mittel, Var_B_mittel);

    % Berechnung des Birge ratio fuer den Startwert v = 0.0
    v = 0.0;
    [RB, x_ref, dv] = eval_Birge_dv([Var_A_mittel, Var_B_mittel], ...
        [bar_x_A, bar_x_B], v);
    %
    max_num_iterations = 50;
    k = 0;
    while (RB > 1) && (k < max_num_iterations)
        fprintf(stdout,'x_ref: %1.3f, RB: %1.2f - dv: %e ==> v = %1.3f\n', ...
            x_ref, RB, dv, v);
        v = v + dv;
        [RB, x_ref, dv] = eval_Birge_dv([Var_A_mittel, Var_B_mittel], ...
            [bar_x_A, bar_x_B], v);
        k = k + 1;
    end
    fprintf(stdout,'x_ref: %1.3f, RB: %1.2f - dv: %e ==> v = %1.3f\n', ...
        x_ref, RB, dv, v);
    fprintf(stdout,'Standardabweichung zwischen den Partnern:\n');
    fprintf(stdout,'s_b = %1.1f nach %d Iterationen\n', ...
        sqrt(v), k);
```

```

end
%
%
function [RB, x_ref, dv] = eval_Birge_dv(var_barx, bar_x, var_b)
    n = length(bar_x);
    weight = 1./(var_barx + var_b);
    x_ref = (1/(n-1)) * sum(weight.*bar_x) / sum(weight);
    RBSq = (1/(n-1)) * sum(weight .* (bar_x-x_ref).^2);
    RB = sqrt(RBSq);
%
    Q = RBSq - 1;
    dQ_dv = (-1/(n-1)) * sum(weight.^2 .* (bar_x-x_ref).^2);
    dv = -Q / dQ_dv;
end

```

mit dem Output:

```

>> consensus_fit()
Mittelwerte: 1.53 und 16.55
Varianzen der Mittelwerte innerhalb jeder Stichprobe:
    0.024 und 0.062
x_ref: 5.672, RB: 51.12 - dv: 4.312238e-02 ==> v = 0.000
x_ref: 7.356, RB: 36.15 - dv: 8.619528e-02 ==> v = 0.043
x_ref: 8.199, RB: 25.57 - dv: 1.721928e-01 ==> v = 0.129
x_ref: 8.620, RB: 18.09 - dv: 3.435958e-01 ==> v = 0.302
x_ref: 8.830, RB: 12.80 - dv: 6.840440e-01 ==> v = 0.645
x_ref: 8.936, RB: 9.06 - dv: 1.355587e+00 ==> v = 1.329
x_ref: 8.988, RB: 6.43 - dv: 2.661878e+00 ==> v = 2.685
x_ref: 9.015, RB: 4.57 - dv: 5.132109e+00 ==> v = 5.347
x_ref: 9.028, RB: 3.27 - dv: 9.539961e+00 ==> v = 10.479
x_ref: 9.034, RB: 2.37 - dv: 1.649219e+01 ==> v = 20.019
x_ref: 9.038, RB: 1.76 - dv: 2.470307e+01 ==> v = 36.511
x_ref: 9.039, RB: 1.36 - dv: 2.797615e+01 ==> v = 61.214
x_ref: 9.040, RB: 1.12 - dv: 1.861186e+01 ==> v = 89.190
x_ref: 9.040, RB: 1.02 - dv: 4.691663e+00 ==> v = 107.802
x_ref: 9.040, RB: 1.00 - dv: 2.129841e-01 ==> v = 112.494
x_ref: 9.040, RB: 1.00 - dv: 4.038510e-04 ==> v = 112.707
x_ref: 9.040, RB: 1.00 - dv: 1.446530e-09 ==> v = 112.707
x_ref: 9.040, RB: 1.00 - dv: 0.000000e+00 ==> v = 112.707
Standardabweichung zwischen den Partnern:
s_b = 10.6 nach 17 Iterationen

```


Kapitel 7

Messunsicherheitsfortpflanzung bei linearen Modelle

7.1 Konzepte der Messunsicherheitsfortpflanzung (M.U.F.)

In der Messtechnik geht es darum, indirekte Messgrößen über Beobachtungen von direkten Messgrößen und über ein Schätzverfahren auf Basis eines Modells zu ermitteln. In der ersten Vorlesung hatten wir dazu hervorgehoben, dass zu dem Modell zum einen die mathematische Beschreibung des physikalischen Sachverhaltes gehört und zum anderen die statistische Beschreibung von nicht deterministischem Verhalten, das in erster Linie von nicht vorhersehbaren, kleinen Störeinflüssen und Auflösungsbegrenzungen beim Beobachtungsvorgang verursacht wird, aber auch von dem physikalischen Vorgang, der zu untersuchen ist, selber stammt, wie beispielsweise bei radioaktiven Zerfällen, Streuprozesse atomarer Vorgänge oder bei biologischen Prozessen. Das nicht deterministische Verhalten, die Stochastik, eines Prozesses wird mittels entsprechender Wahrscheinlichkeitsverteilungen modelliert.

Ein Messergebnis einer Größe X wird deshalb statistisch mit Wahrscheinlichkeitsverteilungen ausgedrückt

- in Form einer Wahrscheinlichkeitsdichteverteilung

$$p(X)$$

oder

- in Form der statistischen Momente, Erwartungswert x und Varianz (bzw. deren Wurzel σ_X), einer Wahrscheinlichkeitsdichteverteilung, zusammen mit einer Wahrscheinlichkeit $1 - \alpha$ bzw. einem Quantil k

$$x = \int X p(X) \, dX, \quad \sigma_X = \sqrt{\int (X - x)^2 p(X) \, dX}, \quad k = P^{-1}(1 - \alpha/2).$$

Dreh- und Angelpunkt bei der Quantifizierung von Messgrößen ist die Vorgehensweise, wie ein entsprechendes Messergebnis durch Modellbildung, Parameterschätzung und Bestimmung der Wahrscheinlichkeitsdichtefunktion der Parameter gewonnen wird. Ein Messergebnis einer Größe Y oder eines Größenvektors \mathbf{Y} , die bzw. der *explizit* oder *implizit* von direkten Messgrößen \mathbf{X} abhängt, soll entsprechend ausgedrückt werden durch eine Wahrscheinlichkeitsdichteverteilung $p(Y)$ bzw. durch den entsprechenden Erwartungswert y und die Wurzel aus der Varianz σ_Y mit einem Quantil k , dem Erweiterungsfaktor, für eine spezifizierte Wahrscheinlichkeit $1 - \alpha$.

Wir kategorisieren die Modelle für indirekte Messgrößen danach, ob sie univariat oder multivariat und ob sie explizit oder implizit sind:

	univariat: 1 indirekte Größe	multivariat: $M > 1$ indirekte Größen
explizit	$Y = f(\mathbf{X})$	$\mathbf{Y} = \vec{f}(\mathbf{X})$
implizit	$f(Y, \mathbf{X}) = 0$	$f(\mathbf{Y}, \mathbf{X}) = 0$ oder $\vec{f}(\mathbf{Y}, \mathbf{X}) = 0$

mit $\mathbf{X} = (X_1, \dots, X_N)^\top$ und $\mathbf{Y} = (Y_1, \dots, Y_M)^\top$.

Bei der linearen Regression beispielsweise haben wir, wie wir es in der zweiten Vorlesung gelernt haben, die Regressoren und Regressanden als

1. gemeinsam ausgelesene (z.B. durch Triggerung zugeordnete) Beobachtungstupel

$$(Y_{\text{Regr},j}, \mathbf{X}_j) = (Y_{\text{Regr},j}, X_{1,j}, \dots, X_{M,j}) \quad (7.1)$$

vorliegen mit $j = 1, \dots, J$ und J für den Stichprobenumfang, so dass gemäß einem zeitlichen Ablauf eine Veränderung aller Größen vorliegen kann,

2. ansonsten können die direkten Messgrößen in voneinander unabhängigen Experimenten und damit im allgemeinen auch mit verschiedenen Stichprobenumfängen gewonnen werden.

Der erstere Fall repräsentiert als direkte Größen sowohl die Regressoren \mathbf{X} als auch die Regressanden Y_{Regr} und als indirekte Messgrößen, die Modellparameter $\boldsymbol{\theta} \equiv \mathbf{Y}$. Hier handelt es sich also hinsichtlich der Unsicherheitsfortpflanzung um ein implizites, multivariates Problem, auch bei der univariaten, linearen Regression.

$$0 = f(\mathbf{Y}, \mathbf{X}) \equiv f(\boldsymbol{\theta}, \mathbf{X}) = Y_{\text{Regr}} - \sum_{i=1}^M \theta_i X_i \quad (7.2)$$

Das „Multivariate“ für die Unsicherheitsberechnung betrifft die indirekten Messgrößen $\boldsymbol{\theta} \equiv \mathbf{Y}$, das „Univariate“ für die Regression betrifft den Regressanden Y_{Regr} .

Fall zwei kann im allgemeinen so geartet sein, dass jede der direkten Größen in jeweils unabhängigen Messvorgängen gewonnen wird, bei denen im allgemeinen unterschiedliche Stichprobenumfänge J_i, J_k vorliegen können und dann keine paarweise Zuordnung der einzelnen Beobachtungswerte $X_{i,j}$ zu $X_{k,l}$ für $i \neq k$ vorliegt.

Bei dem expliziten, univariaten Fall

$$Y = f(\mathbf{X}) \quad (7.3)$$

können die direkten Messgrößen als unkorrelierte unterschiedliche Stichproben vorliegen (Fall 2) oder auch als gemeinsam getriggerte Tupel

$$\mathbf{X}_j = (X_{1,j}, \dots, X_{N,j}) \quad (7.4)$$

wie in Fall 1, aber dann nicht mit Regressionskoeffizienten als Modellparameter, sondern mit der indirekten Größe explizit als Funktion von (X_1, \dots, X_N) gemäß Gl. (7.3).

Für eine internationale Vergleichbarkeit und den internationalen Handel ist es erforderlich, dass die Verfahren zur Bestimmung von Messergebnissen möglichst einheitlich sind. Für das gesetzliche Messwesen ist deshalb die international übereingekommene Richtlinie zur Bestimmung von Messunsicherheiten bindend. Die internationale Richtlinie zur Berechnung von Messunsicherheiten umfasst mehrere Dokumente:

- JCGM 100:2008 GUM 1995 with minor corrections *Evaluation of measurement data - Guide to the expression of uncertainty in measurement*; oftmals einfach mit „GUM“ bezeichnet, betrifft univariate und explizite Modelle,

- die linear sind

$$Y = f(\mathbf{X}) = \sum_{i=1}^N c_i X_i \quad (7.5)$$

oder für die eine Linearisierung zulässig ist

$$Y = f(\mathbf{X}) = f(\mathbf{X})|_{\bar{\mathbf{x}}} + \sum_{i=1}^N \underbrace{\frac{\partial f}{\partial X_i} \Big|_{\bar{\mathbf{x}}}}_{c_i} \Delta X_i \quad (7.6)$$

- die Zufallsgrößen betreffen, deren Streuung (*dispersion*) normalverteilt oder *t*-verteilt ist.

Die indirekte Messgröße Y ist damit Linearkombination der direkten Messgrößen X_i , d.h. $Y = \sum X_i$, so dass für die Varianz $\text{Var}(Y) = \text{Var}(\sum X_i)$ gemäß Gl. (??) aus Abschnitt ?? gilt

$$\text{Var}(Y) = \sum_{i=1}^N \sum_{k=1}^N c_i c_k \text{Cov}(X_i, X_k) \quad (7.7)$$

Da sich viele physikalischen Zusammenhänge für definierte Bereiche linearisieren lassen findet Gl. (7.7) in großen Teilen der Messdatenanalyse Anwendung und ist in der Literatur als das *Gesetz zur Fortpflanzung von Messunsicherheiten* bekannt.

- JCGM 101:2008 *Evaluation of measurement data - Supplement 1 to the Guide to the expression of uncertainty in measurement - Propagation of distributions using a Monte Carlo method*; kurz mit „GUM - supplement 1“ bezeichnet, betrifft explizite und implizite, univariate Modelle,
 - die sich nicht einfach linearisieren lassen
 - und/oder die sich nicht in einer geschlossenen analytischen Form darstellen lassen.

Hier wird zu jeder direkten Messgröße eine Wahrscheinlichkeitsdichteverteilung vorgegeben. Gemäß den jeweiligen Verteilungen wird eine große Stichprobe (ein großes *Sample*) $\mathbf{x}_1 = (x_{1,1}, \dots, x_{N,1})^T, \dots, \mathbf{x}_J = (x_{1,J}, \dots, x_{N,J})^T$ für die direkten Größen \mathbf{X} per Zufallszahlengenerator erzeugt. Diese wird in das Modell f gesteckt, um eine Stichprobe

$$y_1 = f(x_{1,1}, \dots, x_{N,1}), \dots, y_J = f(x_{1,J}, \dots, x_{N,J}) \quad (7.8)$$

der indirekten Größe Y zu gewinnen. Das Histogramm der Stichprobe der indirekten Größe liefert somit die Wahrscheinlichkeitsdichteverteilung. Aus der sortierten Stichprobe nach der Art wie wir sie für den Kolmogoroff-Smirnow-Test berechnet haben, gewinnen wir die kumulierte Wahrscheinlichkeitsverteilung aus deren inverser Funktion das Überdeckungsintervall gewonnen werden kann.

→ Vorlesung 10 und 11

- JCGM 102:2011 *Evaluation of measurement data – Supplement 2 to the Guide to the expression of uncertainty in measurement - Extension to any number of output quantities* liefert die Verallgemeinerung der beiden Richtliniendokumente JCGM 100 und JCGM 101 für den multivariaten Fall mit mehreren $M > 1$ indirekten Größen, also einem Größenvektor \mathbf{Y} . Hier wird dargelegt, wie man die Unsicherheit

- bei linearen oder linearisierbaren Modellen mittels der Berechnung der Kovarianzen (in der Art wie wir es in Vorlesung 2 für die lineare Regression kennen gelernt haben)
- oder im Fall der nicht linearisierbaren und/oder komplexeren (oftmals impliziten) Modelle via Monte-Carlo-Berechnungen analog zu GUM-supplement 1

$$\mathbf{y}_1 = \vec{f}(x_{1,1}, \dots, x_{N,1}), \dots, \mathbf{y}_J = \vec{f}(x_{1,J}, \dots, x_{N,J}) \quad (7.9)$$

- sowie für alle Fälle, ob uni- oder multivariat, ob explizit oder implizit, ob analytisch oder via Monte-Carlo-Verfahren unter Einbeziehung von à priori-Information mittels Bayesischen Methoden
→ Vorlesung 9

ermittelt.

- JCGM 103 CD 2018-10-04 *Guide to the expression of uncertainty in measurement - Developing and using measurement models* behandelt in umfassender Weise die Problematik der Modellentwicklung. Es gibt den Bereich physikalischer, deterministischer Prozesse, den Bereich der nicht-deterministischen, physikalischen oder biologischen oder soziologischen Prozesse. Ferner gibt es den Bereich der statistischen Modelle, die dazu dienen, nicht-deterministische Anteile von Prozessen zu behandeln. Das JCGM 103 Dokument soll diese Dinge konzeptionell für die Metrologie beleuchten. Es ist bisher ein Entwurf; das Kürzel CD steht für *committee draft*.
- JCGM 104:2009 *Guide to the expression of uncertainty in measurement - An introduction to the Guide to the expression of uncertainty in measurement and related documents* liefert die Konzepte und Hintergründe zur Unsicherheitsbestimmung. Dieses Dokument soll ein Verständnis für die Konzepte der Wahrscheinlichkeitsdichteverteilungen im Zusammenhang mit der Bestimmung von Messunsicherheiten liefern.
- JCGM 106:2012 *Evaluation of measurement data - The role of measurement uncertainty in conformity assessment*. Im Rahmen der Vorlesungen zu Hypothesentests und Ringvergleichen haben wir bereits gelernt, dass Messergebnisse zu vergleichen sind, und wie dies gehandhabt wird. In der Qualitätssicherung ist die zentrale Aufgabe, gefertigte Bauteile mit den vorgegebenen Daten der technischen Konstruktionszeichnung zu vergleichen hinsichtlich der Übereinstimmung, also der Einhaltung von Toleranzgrenzen. Dies wird Konformitätsbewertung genannt und wird im JCGM 106-Dokument behandelt.

Das historisch älteste Dokument JCGM 100, das für linearisierbare, explizite, univariate, skalare und gauß- oder t -verteilte indirekte Messgrößen gilt, wurde zunächst erweitert durch das JCGM 101, mit dem nicht nur die Unsicherheit expliziter, univarter, linearisierbarer Messgrößen abschätzbar ist, sondern auch die Unsicherheit für implizite, multivariate Messgrößen, deren direkte Eingangsgrößen beliebig verteilt sein können auch mit von Null verschiedener Skewness, oder mit U-Verteilung, oder anderen Verteilungen gehorchen, so dass die Wahrscheinlichkeitsverteilungen der indirekten Messgrößen entsprechend einen irgendwie gearteten Verlauf aufweisen können. Die nächste Erweiterung liefert dann das JCGM 102-Dokument, das eine vollständige Verallgemeinerung darstellt. Demgemäß werden auch Verfahren betrachtet, die die Berücksichtigung der Unsicherheit des Modells an sich zulassen (die indirekte Messgröße aufgrund von Mangel an Information als Zufallsgröße behandelt) sowie die Behandlung von vorherigen Informationen über die indirekten Messgrößen (Bayesische Statistik).

Die Dokumente JCGM 100, 101, 102, 104 und 106 sind über die Webseiten des *Bureau International des Poids et Mesure*, abgekürzt BIPM,

<https://www.bipm.org/en/publications/guides/#gum>

erhältlich.

Die wesentlichen Komponenten der Messunsicherheit sind

- die Begrenztheit des Messvorgangs bedingt durch
 - endliche Auflösung der Geräte,
 - Messbereichsgrenzen,
 - Störeinflüsse von außerhalb und innerhalb der Geräte,
 - subjektive Komponenten durch den Operateur,
- die Begrenztheit der Modellbildung bedingt durch
 - Vereinfachungen (Parsimonie = „Sparsamkeit“) und Grenzen durch Rechenkapazität zeitlich (Rechenzeit/Rechenleistung) und kostenmäßig (Speicherkapazität und Maschinengenauigkeit, numerische Stabilität, numerische Zuverlässigkeit),
 - Vereinfachungen mangels Kenntniss über quantitative Details zu Einflussgrößen oder genaueren Details der Physik innerhalb des betreffenden physikalischen Effektes,
 - mangels Informationen zu genaueren Werten von Einflussgrößen der physikalischen Prozesse.

Bei der Schätzung von Modellparametern (Quantifizierung der indirekten Messgrößen) spielt vielfach die numerische Zuverlässigkeit des Optimierungsalgorithmus eine wichtige Rolle. Wenn wir das Kostenfunktionsbeispiel der Abbildungen 2/ 3 mit dem der Abbildung 5 der vierten Vorlesung vergleichen, erkennen wir, dass die Modellparameter bei Abb. 2/3 in etwa die gleiche Skalierung aufweisen, während die in Abb. 5 stark unterschiedlich sind. In Abb. 2/3 sehen wir eine schöne, runde Kuhle und in Abb. 5 einen schmalen Graben. Im letzteren ist das Modell mit seinen Parametern nicht so gut konditioniert, so dass das Minimum nicht klar erkennbar ist. Nicht so gut konditionierte Modelle führen eine größere Unsicherheit bei der Schätzung seiner Parameter mit sich.

Für die Wahl des Modellansatzes gilt es nicht nur den physikalischen Zusammenhang geeignet wieder zugeben, sondern auch das Verhalten der Gleitkommaarithmetik zu berücksichtigen. Wenn zwei sehr große Zahlen voneinander zu subtrahieren sind, muss berücksichtigt werden, wie hoch die Maschinengenauigkeit ist, also wie groß die Mantisse ist. Werden die Zahlen so groß, dass die relevanten Nachkommastellen, die nach Subtraktion gewonnen werden sollen, nicht vorhanden sind, liefert der Algorithmus kein brauchbares Ergebnis.

Doing floating point operations is like moving piles of sand: every time you move a pile, you lose a little sand and pick up a little dirt!

Ein Algorithmus ist deshalb derart zu implementieren, dass hohe Potenzen oder vielfache Multiplikationen vor Additionen oder Subtraktionen zu vermeiden, zu umgehen sind. Dazu gehört es, zu überlegen, ob man und wie man das Modell formuliert, beispielsweise ob man Polynome in der einfachen Darstellung, als Legendre oder als Tschebycheff-Polynome definiert. Dazu gehört auch, zu prüfen, ob man und wie man ein Modell zunächst in einer einfacheren Approximation aufbaut und erforderlichenfalls Effekte höherer Ordnung in einem nächsten Schritt hinzufügt.

Ein Modell ist so zu wählen, dass es der entsprechenden Anwendung gerecht wird hinsichtlich der Anforderungen an Kosten vs. Genauigkeit. Wie gut ein Modell einen physikalischen Sachverhalt beschreibt, wird in der Modellunsicherheit ausgedrückt. Ein kleines Beispiel hatten wir in der zweiten Vorlesung gesehen, in der ein Vergleich zwischen einer Regressionsgeraden und einem Polynom 6. Grades gemacht wurde, siehe dort Beispiel 2.2 mit Abb. 3.

7.2 M.U.F. für univariate, explizite indirekte Größe, linear von direkten Größen abhängig

In der ersten Vorlesung, in Abschnitt ??, wurden die Varianz und Kovarianz von Zufallsgrößen eingeführt. In dem Zusammenhang haben wir die Rechenregeln für Erwartungswerte, Summen, Produkte und Skalierungen dazu eingeführt. Auf Grundlage dieser Rechenregeln haben wir die Varianz einer Zufallsgröße, die Linearkombination von mehreren Zufallsgrößen X_i ist, als Funktion der Kovarianzen der Zufallsgrößen X_i bestimmt. Wir wiederholen hier noch mal kurz die Herleitung zu Gl. (??) und damit also Gl. (7.7) und zeigen, weshalb die Varianz einer indirekten Größe, die linear von den direkten Messgrößen gemäß Gl. (7.5) abhängt, Funktion der Kovarianzen und Varianzen der direkten Messgrößen ist.

$$\text{Var}(Y) = \text{Var}\left(\sum_{i=1}^N c_i X_i\right) = \sum_{i,k=1}^N \text{Cov}(c_i X_i, c_k X_k), \quad (7.10)$$

Mit

$$\text{Var}\left(\sum_{i=1}^N c_i X_i\right) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\left(\sum_{i=1}^N c_i X_i \right) - E\left(\sum_{i=1}^N c_i X_i\right) \right]^2 p(X_1, \dots, X_N) \, dX_1 \dots dX_N.$$

und

$$\left(\sum_{i=1}^N c_i X_i \right) - E\left(\sum_{i=1}^N c_i X_i\right) = \left(\sum_{i=1}^N c_i X_i \right) - \sum_{i=1}^N c_i E(X_i)$$

gilt

$$\text{Var}\left(\sum_{i=1}^N c_i X_i\right) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\sum_{i=1}^N (c_i X_i - E(c_i X_i)) \right]^2 p(X_1, \dots, X_N) \, dX_1 \dots dX_N. \quad (7.11)$$

Mit Anwendung des Assoziativgesetzes gilt

$$\text{Var} \left(\sum_{i=1}^N c_i X_i \right) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \sum_{i=1}^N \sum_{k=1}^N c_i (X_i - E(X_i)) c_k (X_k - E(X_k)) p(X_1, \dots, X_N) dX_1 \dots dX_N.$$

Durch Berechnung der Marginalverteilungen und weil $\int p(X_j) dX_j = 1$ für alle j , die weder i noch k sind erhalten wir

$$\text{Var} \left(\sum_{i=1}^N c_i X_i \right) = \sum_{i=1}^N \sum_{k=1}^N \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} c_i (X_i - E(X_i)) c_k (X_k - E(X_k)) p(X_i, X_k) dX_i dX_k. \quad (7.12)$$

Der Term auf der rechten Seite ist genau die Kovarianz der beiden Größen X_i und X_k .

Mit der Definition der Kovarianz

$$\text{Cov}(X_i, X_k) := \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (X_i - E(X_i)) (X_k - E(X_k)) p(X_i, X_k) dX_i dX_k$$

erhalten wir das Fortpflanzungsgesetz in der Form, wie wir es aus Abschnitt ?? Gl. (??) bereits kennen gelernt haben.

$$\begin{aligned} \text{Var}(Y) &= \text{Var} \left(\sum_{i=1}^N c_i X_i \right) \\ &= \sum_{i=1}^N \sum_{k=1}^N \text{Cov}(c_i X_i, c_k X_k) \\ &= \sum_{i=1}^N c_i^2 \text{Var}(X_i) + \sum_{i,k=1, i \neq k}^N c_i c_k \text{Cov}(X_i, X_k) \\ &= \sum_{i=1}^N c_i^2 \text{Var}(X_i) + 2 \sum_{i=1}^{N-1} \sum_{k=i+1}^N c_i c_k \text{Cov}(X_i, X_k). \end{aligned} \quad (7.13)$$

Bei getrennter Betrachtung der beiden Terme mit den Varianzen $\text{Var}(X_i)$ und mit den Kovarianzen $\text{Cov}(X_i, X_k)$, also in der Notation $\sum c_i^2 \text{Var}(X_i) + 2 \sum \sum c_i c_k \text{Cov}(X_i, X_k)$, wird deutlich, ob die direkten Messgrößen korrelieren oder nicht.

Es gibt Anwendungen, bei denen die Korrelationskoeffizienten, definiert in Abschnitt ?? Gl. (??) $\rho_{i,k} = \frac{\text{Cov}(X_i, X_k)}{\sqrt{\text{Var}(X_i) \text{Var}(X_k)}}$, angegeben wird und die Standardabweichungen $\sigma_i = \sqrt{\text{Var}(X_i)}$, so dass das Unsicherheitsfortpflanzungsgesetz auch in dieser Form aufgeschrieben wird

$$\text{Var}(Y) = \sigma_Y^2 = \sum_{i=1}^N c_i^2 \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{k=i+1}^N c_i c_k \rho_{i,k} \sigma_i \sigma_k, \quad (7.14)$$

und σ_Y dabei dann *kombinierte Standardabweichung* genannt wird.

Oftmals, so auch im *Guide* für Messunsicherheit GUM - JCGM 100, wird anstelle der Bezeichnung $\rho_{i,k}$ die Bezeichnung $r(X_i, X_k)$ verwendet, meint aber dasselbe.

Alternativ, insbesondere im GUM - JCGM 100, gibt es die Schreibweise mit u anstelle der Standard-

abweichungen σ als Kürzel für *uncertainty* oder *Unsicherheit*, und heißt in diesem Zusammenhang *Standardunsicherheit* (engl. *standard uncertainty*).

$$\begin{aligned}
 u^2(Y) &= \sum_{i=1}^N c_i^2 \text{Var}(X_i) + 2 \sum_{i=1}^{N-1} \sum_{k=i+1}^N c_i c_k \text{Cov}(X_i, X_k) \\
 &= \underbrace{\sum_{i=1}^N c_i^2 u^2(X_i)}_{\text{unkorrelierter Fall}} + 2 \underbrace{\sum_{i=1}^{N-1} \sum_{k=i+1}^N c_i c_k \rho_{i,k} u(X_i) u(X_k)}_{\text{Mischterme: korrelierter Fall}}.
 \end{aligned} \tag{7.15}$$

mit

$$\text{Cov}(X_i, X_i) \equiv \text{Var}(X_i) \equiv \sigma_i^2 \equiv u^2(X_i) \tag{7.16}$$

und

$$\text{Cov}(X_i, X_k) \equiv \sigma_{i,k} \equiv u(X_i, X_k) \tag{7.17}$$

Hier sind keine Tippfehler bezüglich der Quadrate, es gilt tatsächlich $u^2(X_i) = u(X_i, X_i)$ sowie $\sigma^2(X_i) = \sigma(X_i, X_i)$.

Im JCGM 100-Dokument des *Guide* für Messunsicherheit GUM in Absatz 5.2.4 werden als typische Ursachen für die Korrelation von Eingangsgrößen folgende aufgeführt:

- Verwendung desselben Normals (Prüfkörpers) oder Messgeräts;
- Verwendung mehrerer Normale, die in derselben Vorrichtung kalibriert wurden (z.B. gestückelte Masse-Normale bei der Kalibrierung einer Waage);
- Verwendung des gleichen Referenzwertes
- Eine Eingangsgröße hängt direkt von einer weiteren ab (z.B. Luftdruck von Umgebungstemperatur);
- Zwei oder mehr Eingangsgrößen sind von demselben Effekt beeinflusst (z.B. Stromstärke und Spannung in einem Messkreis, so dass beide von Schwankungen der Stromquelle beeinflusst werden).

Bei gemeinsam aufgenommenen Tupeln direkter Messgrößen (z.B. Stromstärke und Spannung in einem Messkreis), die miteinander korreliert sein können, sieht die Kovarianz aufgrund der Zuordnung mit gemeinsamer Beobachtung (Ziehung) jeweils eines Tupels X_j wie folgt aus:

$$\text{Cov}(X_i, X_k) = \frac{1}{J-1} \sum_{j=1}^J (X_{i,j} - \bar{x}_i)(X_{k,j} - \bar{x}_k) \quad \text{und mit} \quad \bar{x}_i = \frac{1}{J} \sum_{j=1}^J X_{i,j}$$

für $i, k = 1, \dots, N$.

Bei den aufgelisteten Fällen, bei denen es keine direkte Zuordnung von Beobachtungen der verschiedenen direkten Messgrößen gibt, wie es bei der Verwendung desselben Normals oder Messgeräts sein kann, erfordert die Abschätzung eines Korrelationskoeffizienten ρ oder r eine geeignete Modellbildung.

7.3 M.U.F. für univariate, explizite indirekte Größe, Abhängigkeit von direkten Größen linearisierbar

Zunächst beleuchten wir ein Beispiel, bei dem ein Modellparameter Y berechnet wird, der sich aus dem Produkt zweier direkter Messgrößen X_1 und X_2 ergibt. Wir beschäftigen uns zunächst mit dem Fall, dass die beiden direkten Größen nicht von einander abhängen. Sie sollen zwei unterschiedlichen und unabhängigen Grundgesamtheiten angehören, d.h. unkorreliert sein.

Das Modell sei

$$Y = X_1 X_2 \quad (7.18)$$

und es wurden zu jeder der beiden direkten Größen zu unterschiedlichen Zeiten unterschiedlich große Stichproben gezogen.

Die naheliegende Vorgehensweise ist nun, die Mittelwerte der beiden Stichproben zu berechnen und diese miteinander zu multiplizieren

$$y = \left(\frac{1}{J_1} \sum_{j=1}^{J_1} X_{1,j} \right) \left(\frac{1}{J_2} \sum_{j=1}^{J_2} X_{2,j} \right) \quad (7.19)$$

Für die Berechnung der kombinierten Standardabweichung von Y bestimmen wir die Standardabweichungen der beiden Stichproben $i = 1, 2$ zu den Größen X_1 und X_2 .

$$s_i = \sqrt{\frac{1}{J_i - 1} \sum_{k=1}^{J_i} \left(X_{i,k} - \frac{1}{J_i} \sum_{j=1}^{J_i} X_{i,j} \right)^2} \quad (7.20)$$

Wir betrachten die Empfindlichkeit der Modellfunktion, wie stark die Größe Y auf Abweichungen/Veränderungen der Größen X_1 und X_2 reagiert. In diesem Beispiel ist es sehr einfach, wenn die Größe X_1 um ΔX_1 verschoben wird

$$Y + \Delta Y = (X_1 + \Delta X_1) X_2 = X_1 X_2 + \Delta X_1 X_2 \quad (7.21)$$

dann ist

$$\Delta Y = \Delta X_1 X_2 \quad (7.22)$$

weil Y linear von X_1 abhängt und es gilt

$$c_1 := \frac{\Delta Y}{\Delta X_1} = X_2 \quad (7.23)$$

was die Empfindlichkeit, d.h. Sensitivität c_1 , der Größe Y auf Änderungen der Größe X_1 genannt wird. Da die beiden Größen X_1 und X_2 zwei unabhängige Dimensionen repräsentieren, werden ihre Standardabweichungen nach dem Satz des Pythagoras addiert, eine Veranschaulichung dieses Sachverhalts zeigt Abb. 7.1. Da die Größen über die Modellgleichung miteinander verknüpft sind und die gesuchte Standardabweichung die Dimension der Größe Y trägt, müssen die Standardabweichungen der Größen

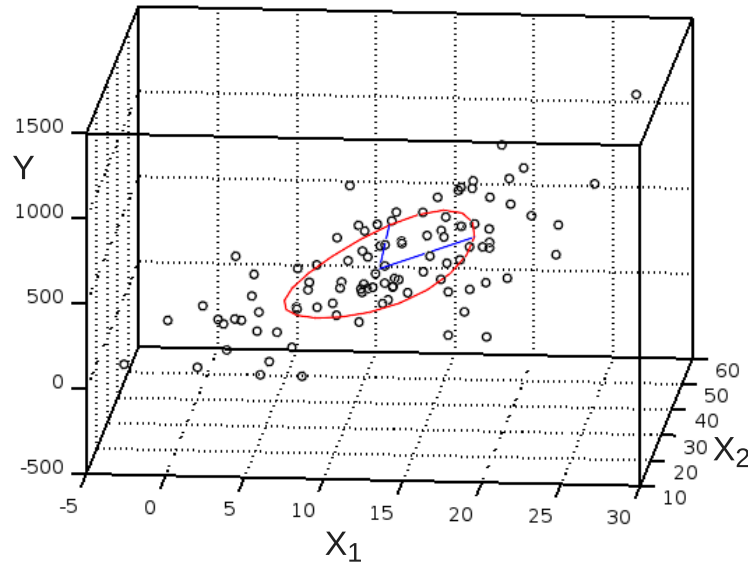


Abbildung 7.1: Eine anschauliche Deutung für die pythagoräische Addition der Standardabweichungen liefert die Betrachtung der mit roter Kurve dargestellten Ellipse. Die blauen Geraden repräsentieren die Abstände vom Mittelpunkt als Halbmesser der Ellipse, die aus den Standardabweichungen jeweils von X_1 und X_2 berechnet wurden.

X_1 und X_2 mit den entsprechenden Sensitivitäten verknüpft werden, so dass

$$s_y^2 = (c_1 s_1)^2 + (c_2 s_2)^2. \quad (7.24)$$

Diese Varianz s_y^2 ist also wieder eine *kombinierte Varianz* und deren Wurzel s_y eine *kombinierte Standardabweichung*.

Für den allgemeinen Fall mit einer Linearisierung durch die Taylorreihenentwicklung Gl. (7.6) stellen die partiellen Ableitungen, also der Gradient, an der Stelle der Schätzer der direkten Größen die Sensitivitäten $c_i = \left. \frac{\partial f}{\partial X_i} \right|_{\bar{\mathbf{x}}}$. Sie drücken aus, wie stark sich die indirekte Größe Y gemäß der Steigung (Gradienten) des Kurvenverlaufs des Modells ändert, wenn es kleine Änderungen der direkten Größen gibt. Mit anderen Worten besagt dies, wie empfindlich die indirekte Größe auf Änderungen der direkten Größen reagiert, d.h. wie sensitiv das Modell an der Position $\bar{\mathbf{x}}$ ist.

Als nächstes wollen wir wieder das Fortpflanzungsgesetz für die Messunsicherheit in der Gestalt von Gl. (7.13) erhalten, wobei jetzt die Koeffizienten c_i die Steigungen der Tangenten $\left. \frac{\partial f}{\partial X_i} \right|_{\bar{\mathbf{x}}}$ an der Stelle der geschätzten Erwartungswerte sind.

Die Varianz der gewichteten Summe der Zufallsgrößen ΔX_i ist

$$\text{Var}(Y) = \text{Var}(f(\mathbf{X})|_{\bar{\mathbf{x}}}) + \text{Var}\left(\sum_{i=1}^N \left. \frac{\partial f}{\partial X_i} \right|_{\bar{\mathbf{x}}} \Delta X_i\right). \quad (7.25)$$

Dabei sind $f(\mathbf{X})|_{\bar{x}}$ und $c_i = \frac{\partial f}{\partial X_i} \Big|_{\bar{x}}$ Konstanten, so dass daraus $\text{Var}(f(\mathbf{X})|_{\bar{x}}) = 0$ und

$$\text{Var}(Y) = \text{Var}\left(\sum_{i=1}^N c_i \Delta X_i\right) \quad (7.26)$$

wird, was also die Form der letzten Zeile von Gl. (7.13) annimmt, so dass gilt

$$\text{Var}(Y) = \sum_{i=1}^N c_i^2 \text{Var}(\Delta X_i) + 2 \sum_{i=1}^{N-1} \sum_{k=i+1}^N c_i c_k \text{Cov}(\Delta X_i, \Delta X_k) \quad (7.27)$$

mit $\text{Var}(\Delta X_i) = s_i^2$ für die Varianz in der Umgebung vom Erwartungswert \bar{x}_i und $\text{Cov}(\Delta X_i, \Delta X_k) = s_{i,k}$ für die Kovarianz in der Umgebung von den Erwartungswerten \bar{x}_i und \bar{x}_k .

Zu Kapitel 5.2 im JCGM 100-Dokument des *Guide* für Messunsicherheit GUM sei angemerkt, dass für die Sensitivitäten eine verkürzte, unmathematische Schreibweise für die Sensitivitäten gewählt wurde, indem dort geschrieben steht $\frac{\partial f}{\partial x_i}$ anstelle von $\frac{\partial f}{\partial X_i} \Big|_{\bar{x}_1, \dots, \bar{x}_N}$. Im GUM wird angemerkt dass $\frac{\partial f}{\partial x_i}$ für $\frac{\partial f}{\partial X_i}$ unter Verwendung der Schätzwerte x_i zu den X_i . In unserer Vorlesungsreihe sowie im GUM steht das Symbol mit dem großgeschriebenen X_i für die direkten Messgrößen. Das im GUM verwendete x_i als Kleinbuchstabe steht für den Schätzwert, was wir in der Vorlesung mit \bar{x}_i bezeichnen. Da der Schätzwert eine Konstante ist, kann man nach dieser natürlich nicht ableiten.

7.4 Überdeckungsintervall für indirekte Messgrößen

In Vorlesung 5 haben wir das Überdeckungsintervall behandelt, als *Credible intervall* und als *Vertrauensintervall*. Die Intervallgrenzen des Vertrauensintervall haben wir auch als Produkt aus einem Quantil und einer Standardabweichung repräsentiert, im Fall, dass eine Normalverteilung zugrunde gelegt wird in der Form

$$\left[\bar{x} - z_{1-\frac{1}{2}\alpha} \sigma_X, \bar{x} + z_{1-\frac{1}{2}\alpha} \sigma_X \right] \quad (7.28)$$

und im Fall, dass eine t -Verteilung zugrunde gelegt wird

$$\left[\bar{x} - t_{1-\frac{1}{2}\alpha, \nu} \sigma_X, \bar{x} + t_{1-\frac{1}{2}\alpha, \nu} \sigma_X \right] \quad (7.29)$$

Bei dem Fall, dass eine t -Verteilung zugrunde gelegt wird, und zwar, wenn es um kleinere Stichprobenumfänge geht hängt das Quantil von der Anzahl der Freiheitsgrade $\nu_i = J_i - 1$ ab, die jedoch im allgemeinen für die beiden Größen X_i unterschiedlich sein kann. Es gilt also, ein gemeinsames Quantil zu finden.

Mit der bayesischen Methode, die wir zuvor betrachtet haben, konnten wir das *Credible Interval* aus der kumulierten Posterior gewinnen. Bei dem zuvor erörterten Beispiel Gl. (7.18) $Y = X_1 X_2$ haben wir statt der einen Likelihood drei Verteilungsdichten: Für jede Größe X_i haben wir je eine Likelihood

$$p_{L,i}(\{X_{i,1}, \dots, X_{i,J}\} | X_i, s_i) = \prod_{j=1}^J \frac{1}{\sqrt{2\pi} s_i} e^{-\frac{1}{2} \left(\frac{X_{i,j} - X_i}{s_i} \right)^2} = l(X_i, s_i | \{X_{i,1}, \dots, X_{i,J}\}) \quad (7.30)$$

was zwei Verteilungen liefert, und als dritte Verteilung noch

$$p_{L,Y}(X_1, X_2|Y, s) = \frac{1}{\sqrt{2\pi} s} e^{-\frac{1}{2} \left(\frac{Y - X_1 X_2}{s} \right)^2} \quad (7.31)$$

wobei dann nicht nur über s , sondern auch über X_1 und X_2 integriert wird, um die Posterior als Marginalverteilung, die nur noch von Y abhängt, zu erhalten.

Für relativ einfache analytische Zusammenhänge, also für explizite Modelle und für einfachere implizite Modellansätze wie die lineare Regression, ist es nicht erforderlich die numerisch aufwendigere kolmogoroffschen Wahrscheinlichkeitsrechnung (Bildung der Produkte der Verteilungen) anzuwenden. Für explizite linearisierbare Modelle

$$Y = f(\mathbf{X}) \quad \mathbf{Y} = \vec{f}(\mathbf{X})$$

mit $\vec{f} = (f_1, \dots, f_l, \dots, f_M)^\top$ ist das Fortpflanzungsgesetz gemäß Gl. (7.15) wie folgt

$$u^2(Y_l) = \sum_{i=1}^N \left(\left. \frac{\partial f_l(\mathbf{X})}{\partial X_i} \right|_{\bar{\mathbf{x}}} \right)^2 u^2(X_i) + 2 \sum_{i=1}^{N-1} \sum_{k=i+1}^N \left. \frac{\partial f_l(\mathbf{X})}{\partial X_i} \right|_{\bar{\mathbf{x}}} \left. \frac{\partial f_l(\mathbf{X})}{\partial X_k} \right|_{\bar{\mathbf{x}}} \rho_{i,k} u(X_i) u(X_k) \quad (7.32)$$

verwendbar. Für die lineare Regression

$$Y_{\text{Regr}} = \sum_{l=1}^M \theta_l X_l$$

die bzgl. der indirekten Messgrößen $\theta_l = Y_l$ ein implizites Modell ist, gilt für die Unsicherheitsfortpflanzung Gl. (87) mit Gl. (88) aus Vorlesung 2. Mit einer Stichprobe mit Tupeln $(Y_{\text{Regr},j}, X_{1,j}, \dots, X_{M,j})$ und mit $\mathbf{Y}_{\text{Regr}} = (Y_{\text{Regr},1}, \dots, Y_{\text{Regr},J})^\top$ und der Regressormatrix $\mathbf{X} = (X_{l,j})$ liefert das Fortpflanzungsgesetz die Kovarianz der Modellparameter

$$\text{Cov}(\boldsymbol{\theta}) = \left(\mathbf{X}^\top \mathbf{X} \right)^{-1} \text{Var}(\boldsymbol{\varepsilon}) \quad (7.33)$$

mit J für den Stichprobenumfang und M für die Anzahl der Modellparameter und mit

$$\text{Var}(\boldsymbol{\varepsilon}) = \frac{1}{J - M} \left(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\theta}} \right)^\top \left(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\theta}} \right) \quad (7.34)$$

Die Hauptdiagonale der Kovarianz der Modellparameter enthält die Varianzen

$$\sigma_{\theta_l}^2 = u^2(\theta_l \equiv Y_l)$$

der jeweiligen Modellparameter.

Für die Ermittlung des Überdeckungsintervalls $[y_l - U(Y_l), y_l + U_l]$ wird ein entsprechender Erweiterungsfaktor k für $U(Y_l) = k u(Y_l)$ gebraucht. Wird als Erweiterungsfaktor das Quantil der t -Verteilung eingesetzt, so wird eine *effektive Anzahl von Freiheitsgraden* ν_{eff} approximiert. In Abschnitt G.4 des Dokuments JCGM 100 GUM:2008 wird auf die Berechnungsmethode für ν_{eff} von Satterthwaite aus dem Jahr 1941 [Sat41] zurückgegriffen.

Satterthwaite betrachtet die Varianz $\text{Var}(s_i^2)$ der Varianz s_i^2 . Die s_i^2 sind unabhängige Zufallsgrößen,

also als Größen zu betrachten, die unkorreliert sind, so dass

$$\text{Var}(s_y^2) = \text{Var}(c_1^2 s_1^2) + \text{Var}(c_2^2 s_2^2) \quad (7.35)$$

gilt. Die empirischen Varianzen s_1^2 und s_2^2 der Größen X_1 und X_2 sind χ^2 -verteilt, d.h. für die normierte Größe gilt

$$Q_i = \nu_i \left(\frac{s_i}{\sigma_i} \right)^2 \sim \chi^2(\nu_i). \quad (7.36)$$

Der Erwartungswert für Q_i ist

$$\text{E}(Q_i) = \int_0^\infty Q p_{\chi^2}(Q) \, dQ = \nu_i. \quad (7.37)$$

Die Varianz $\text{Var}(Q_i)$ ist

$$\text{Var}(Q_i) = \int_0^\infty (Q - \nu_i)^2 p_{\chi^2}(Q) \, dQ = 2\nu_i. \quad (7.38)$$

Die Herleitungen zu den beiden Gln. (7.37) und (7.38) befinden sich in Anhang 7.7.

Ferner gilt allgemein für die Varianz einer Zufallszahl R multipliziert mit einem konstanten Faktor a

$$\begin{aligned} \text{Var}(aR) &= \int_0^\infty (aR - \text{E}(aR))^2 p(R) \, dR \\ &= \int_0^\infty a^2 (R - \text{E}(R))^2 p(R) \, dR \\ &= a^2 \int_0^\infty (R - \text{E}(R))^2 p(R) \, dR \\ &= a^2 \text{Var}(R) \end{aligned} \quad (7.39)$$

so dass

$$\text{Var}(Q_i) = \text{Var} \left(\nu_i \left(\frac{s_i}{\sigma_i} \right)^2 \right) = \left(\frac{\nu_i}{\sigma_i^2} \right)^2 \text{Var}(s_i^2) \quad (7.40)$$

mit Gl. (7.38) zu

$$2\nu_i = \left(\frac{\nu_i}{\sigma_i^2} \right)^2 \text{Var}(s_i^2) \quad (7.41)$$

wird. Das heißt

$$2\nu_i = \frac{\nu_i^2}{\sigma_i^4} \text{Var}(s_i^2) \quad (7.42)$$

also

$$\text{Var}(s_i^2) = \frac{2\sigma_i^4}{\nu_i}. \quad (7.43)$$

so dass nach Satterthwaite aus Gl. (7.35) die *effektive Anzahl der Freiheitsgrade* ν_{eff} für die Größe Y , also $\nu_{\text{eff}} = \nu_y$ abgeschätzt wird mit

$$\frac{2\sigma_y^4}{\nu_y} = \frac{2c_1^4\sigma_1^4}{\nu_1} + \frac{2c_2^4\sigma_2^4}{\nu_2}$$

mit $\text{Var}(s_y^2) = \frac{2\sigma_y^4}{\nu_y}$ und nach Rauskürzen des Faktors 2

$$\frac{\sigma_y^4}{\nu_y} = \frac{c_1^4 \sigma_1^4}{\nu_1} + \frac{c_2^4 \sigma_2^4}{\nu_2} \quad (7.44)$$

so dass

$$\nu_y = \frac{\sigma_y^4}{\frac{(c_1 \sigma_1)^4}{\nu_1} + \frac{(c_2 \sigma_2)^4}{\nu_2}}. \quad (7.45)$$

Wenn mit Gleichung (7.37) der Erwartungswert von Q gleich der Anzahl der Freiheitsgrade ist, also $E(Q) = \nu$ für $Q = \nu \left(\frac{s}{\sigma}\right)^2$, dann ist der Erwartungswert der Varianz s^2

$$E(s^2) = E\left(\frac{\sigma^2}{\nu} Q\right) \quad (7.46)$$

und mit $\frac{\sigma^2}{\nu}$ als konstanter Faktor bezüglich der Wahrscheinlichkeitsdichte und mit Gl. (7.37) erhalten wir

$$E(s^2) = \frac{\sigma^2}{\nu} E(Q) = \frac{\sigma^2}{\nu} \nu = \sigma^2 \quad (7.47)$$

und wir setzen in Gl. (7.45) die Erwartungswerte $E(s_y^2)$, $E(s_1^2)$ und $E(s_2^2)$ ein, deren Schätzer die empirisch ermittelten Varianzen sind.

Das t -Quantil wird damit für $\nu_y = \nu_{\text{eff}}$ berechnet und das vollständige Messergebnis wie folgt angegeben

$$Y = y \pm t_{1-\alpha/2, \nu_y} s_y. \quad (7.48)$$

Verallgemeinert für N direkte Größen ist dies

$$\nu_{\text{eff}} = \frac{\sigma_y^4}{\sum_{i=1}^N \frac{(c_i \sigma_i)^4}{\nu_i}}. \quad (7.49)$$

Aufgrund der Verallgemeinerung, dass *Überdeckungsintervalle* sowohl klassisch ermittelte Vertrauensintervalle als auch bayesisch ermittelte *Credible Intervals* sein können, werden die zu einer gewissen Wahrscheinlichkeit $1 - \alpha$ gehörenden Faktoren *Erweiterungsfaktor* genannt. Sei $[z_{\min}, z_{\max}]$ das *Credible Interval* einer Größe Y mit Posterior $p(Y|\dots)$ deren Erwartungswert

$$\bar{Y} = \int_{-\infty}^{\infty} Y p(Y|\dots) dY \quad (7.50)$$

und deren Varianz

$$\text{Var}(Y) = \int_{-\infty}^{\infty} (Y - \bar{Y})^2 p(Y|\dots) dY \quad (7.51)$$

ist, so ist der *Erweiterungsfaktor* k für symmetrische Verteilungen mit $z_{\max} - \bar{Y} = \bar{Y} - z_{\min}$

$$k = \frac{z_{\max} - \bar{Y}}{\sqrt{\text{Var}(Y)}} \quad (7.52)$$

was für klassische Berechnungsverfahren dem t-Quantil $t_{1-\alpha/2, \nu_{\text{eff}}}$ entspricht

$$k = t_{1-\alpha/2, \nu_{\text{eff}}} \quad (7.53)$$

Für symmetrische Verteilungen ist die halbe Breite des Überdeckungsintervalls gleich der *erweiterten Messunsicherheit* der Größe Y . Die *erweiterte Messunsicherheit* kann das Produkt aus Erweiterungsfaktor und empirischer Standardabweichung sein.

Während mit *kleinem* Buchstaben u die Standardabweichungen bezeichnet werden, werden mit *großem* Buchstaben U die erweiterten Unsicherheiten geschrieben, die das Produkt aus dem Erweiterungsfaktor k , der in der klassischen Statistik das t -Quantil ist, und der Unsicherheit u ist, sind.

7.5 Vergleich der klassischen mit der bayesischen Methode

Als nächstes wollen wir zeigen, dass für die „gutmütigen“ Fälle, für die das Gesetz der Messunsicherheitsfortpflanzung anwendbar ist, die Berechnung des vollständigen Ergebnisses unter Verwendung der Methode der bayesischen Statistik im wesentlichen dasselbe Resultat liefert wie unter Verwendung des Fortpflanzungsgesetzes.

Wir betrachten ein Beispiel, bei dem irgendeine physikalische Größe zu messen ist. Dabei werde ein Gerät bzw. Sensor verwendet, dessen Funktionsprinzip auf einem physikalischen Effekt beruht und dadurch die Größe so erfasst, dass als direkte Messgröße eine Spannung in Volt angezeigt wird. Dies kann beispielsweise die Messung einer Temperatur sein, in der das Phänomen, dass sich ein elektrischer Widerstand proportional zur Tempertur verändert und die Widerstandsänderung über die Änderung der elektrischen Spannung, die über dem Widerstand abfällt, bestimmt wird. Dies kann beispielsweise eine Stufenhöhe sein, die mit Hilfe eines induktiven Wegaufnehmers gemessen wird, der auf dem Phänomen beruht, dass sich die Induktivität einer Spule in Abhängigkeit von der Position ihres Ferritkerns verändert. Es sind viele Beispiel denkbar. Sensoren sind oft so konzipiert, dass sie ein physikalisches Phänomen nutzen und zur elektronischen Erfassung der zu messenden Größe als direkte Größe ein Messsignal in Form einer elektrischen Spannung liefern.

Wir wollen im folgenden ganz allgemein die indirekte Messgröße mit Y bezeichnen und uns auf keine physikalische Einheit festlegen, sondern diese allgemein nur „Einheit“ nennen. Das Modell, das wir betrachten, ist wie folgt: Das rohe Messsignal, das als Spannung U bzw. direkte Größe X_M vorliegt, ist über einen Kalibrierfaktor K bzw. eine direkte Größe X_K in die physikalische Einheit der indirekten physikalischen Größe Y umzurechnen. Der Index M steht hier für Messung und der Index K für Kalibrierfaktor.

Die indirekte Größe Y habe irgendeine physikalische Einheit E

$$Y = f(X_K, X_M) = X_K X_M \quad (7.54)$$

und die beiden direkten Größen X_M und X_K sollen die Einheiten Volt V und $\frac{E}{V}$ haben.

Die Stichprobe der Beobachtungen zum rohen Messsignal habe einen recht kleinen Umfang $J_M = 9$:

X_M/V	479.58	526.47	516.77	522.01	506.61	497.99	481.71	484.90	491.41
---------	--------	--------	--------	--------	--------	--------	--------	--------	--------

Als vollständiges Messergebnis zum Kalibrierfaktor X_K liegen uns folgende Angaben vor

$$X_K = (K_0 \pm U_K) \frac{E}{V} = (0.0925 \pm 0.0180) \frac{E}{V} \quad \text{mit} \quad k = 2 \quad \text{und} \quad \nu_K = 45$$

Der Erweiterungsfaktor $k = 2$ ist der gerundete Wert für das t-Quantil für 95 % Vertrauensniveau $\nu_K = 45$ Freiheitsgrade. Der genauere Wert wäre $k = 2.0141$, die Rundung ist hier zulässig.

Wir beleuchten den Fall, dass wir die dem Sensor intrinsische Unsicherheit nicht kennen, aber davon ausgehen, dass die Unsicherheit σ_{winzig} aufgrund der Modellapproximation für die Sensorik, hier also u.a. dass vorausgesetzt wird, dass der Sensor linear ist, wesentlich kleiner ist als die Streuung der Messvorgänge beim Erfassen der Rohdaten und der Sensorkalibrierung:

$$p(Y, X_K, X_M | (X_{M,1}, \dots, X_{M,J}), K_0, s_K) \propto e^{-\frac{1}{2} \left(\frac{Y - X_K X_M}{\sigma_{\text{winzig}}} \right)^2} e^{-\frac{1}{2} \left(\frac{X_K - K_0}{s_K} \right)^2} \prod_{j=1}^J e^{-\frac{1}{2} \left(\frac{X_M - X_{M,j}}{s_M} \right)^2} \quad (7.55)$$

mit $\sigma_{\text{winzig}} = 0.003 \cdot s_{KM}$, mit $s_{KM} = K_0 s_M$ und mit $s_K = \frac{U_K}{2} = 0.0090$. Wir verwenden zur Berechnung der Likelihood die empirische Standardabweichung der Daten zu X_M :

$$\bar{x}_M = \frac{1}{9} \sum_{j=1}^9 X_{M,j} = 500.83 \text{ V}$$

und

$$s_M = \sqrt{\frac{1}{\nu_M} \sum_{j=1}^9 (X_{M,j} - \bar{x}_M)^2} = 17.8927 \text{ V} \approx 17.89 \text{ V}$$

mit $\nu_M = 8$ Freiheitsgraden, damit also $K_0 s_M = s_{KM} = 1.655 \text{ E}$ und entsprechend für $\sigma_{\text{winzig}} = 0.00497$. Da die Stichprobenwerte der Größe X_M in der Tabelle mit 2 Nachkommastellen angegeben werden, wird das Ergebnis auch auf 2 Stellen gerundet.

Wir bestimmen den Schätzer und das Überdeckungsintervall im folgenden zum einen aus der Unsicherheitsfortpflanzung gemäß (7.15) und zum anderen aus der Marginalverteilung zu Gl. (7.55):

$$p(Y | (X_{M,1}, \dots, X_{M,J}), K_0, s_K) = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2} \left(\frac{Y - X_K X_M}{\sigma_{\text{winzig}}} \right)^2} e^{-\frac{1}{2} \left(\frac{X_K - K_0}{s_K} \right)^2} \prod_{j=1}^J e^{-\frac{1}{2} \left(\frac{X_M - X_{M,j}}{s_M} \right)^2} dX_M dX_K}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2} \left(\frac{Y - X_K X_M}{\sigma_{\text{winzig}}} \right)^2} e^{-\frac{1}{2} \left(\frac{X_K - K_0}{s_K} \right)^2} \prod_{j=1}^J e^{-\frac{1}{2} \left(\frac{X_M - X_{M,j}}{s_M} \right)^2} dX_M dX_K dY} \quad (7.56)$$

Wir erhalten den Schätzwert \bar{y} aus

$$\bar{y} = \int_{-\infty}^{\infty} Y p(Y | (X_{M,1}, \dots, X_{M,J}), K_0, s_K) dY \quad (7.57)$$

und für die Bestimmung der Intervallgrenzen y_1, y_2 für das Überdeckungsintervall berechnen wir die kumulierte Verteilung

$$P(Y) = \int_{-\infty}^Y p(Y' | (X_{M,1}, \dots, X_{M,J}), K_0, s_K) dY' \quad (7.58)$$

und dann für 95% Wahrscheinlichkeit

$$P(y_1) = 0.025 \Leftrightarrow y_1 = P^{-1}(0.025) \quad \text{und} \quad P(y_2) = 0.975 \Leftrightarrow y_2 = P^{-1}(0.975). \quad (7.59)$$

Durch numerische Integration der Gl. (7.57) erhalten wir

$$\bar{y} = 46.278 \text{ E}$$

und für das *Credible Interval* durch numerische Integration von Gl. (7.58)

$$[36.855 \text{ E}, 56.185 \text{ E}] = [\bar{y} - 9.423, \bar{y} + 9.907] \text{ E}$$

und vergleichen dieses Ergebnis mit dem Ergebnis, das wir auf dem klassischen Wege erhalten. Zunächst berechnen wir das Produkt der beiden Werte $\bar{x}_M = 500.83 \text{ V}$ und $K_0 = 0.0925 \frac{\text{E}}{\text{V}}$. Wir erhalten mit

$$\bar{y} = K_0 \bar{x}_M = 46.327 \text{ E}$$

einen Schätzwert für Y , der um einen Wert von 0.049 E also um 1 Promille von dem aus der bayesischen Berechnung differiert.

In Anhang 7.8 befindet sich das Gnu-Octave/Matlab-Skript, mit dem die Berechnungen zu diesem Beispiel durchgeführt wurden,.

Als nächstes vergleichen wir das *Credible Interval* mit dem Vertrauensintervall. Dazu bestimmen wir die Unsicherheit mit dem Fortpflanzungsgesetz für unkorrelierte direkte Messgrößen.

Wir berechnen für $\rho_{M,K} = 0$

$$u^2(Y) = \left(\frac{\partial}{\partial X_M} X_M X_K \right)_{\bar{x}_M, K_0}^2 s_M^2 + \left(\frac{\partial}{\partial X_K} X_M X_K \right)_{\bar{x}_M, K_0}^2 s_K^2$$

d.h.

$$u(Y) = \sqrt{K_0^2 s_M^2 + \bar{x}_M^2 s_K^2} = 4.802 \text{ E.}$$

Die Anzahl der Freiheitsgrade berechnen wir gemäß der Satterthwaite'schen Gleichung (7.49), also

$$\frac{u(Y)^4}{\nu_y} = \frac{K_0^4 s_M^4}{\nu_M} + \frac{\bar{x}_M^4 s_K^4}{\nu_K}$$

d.h. mit $\nu_K = 45$, $\nu_M = 8$ und mit $K_0 = 0.0925 \frac{\text{E}}{\text{V}}$, $s_K = 0.009 \frac{\text{E}}{\text{V}}$, $\bar{x}_M = 500.83 \text{ V}$, $s_M = 17.89 \text{ V}$

$$\nu_y = u(Y)^4 \left(\frac{K_0^4 s_M^4}{\nu_M} + \frac{\bar{x}_M^4 s_K^4}{\nu_K} \right)^{-1} = 52.58 \approx 52$$

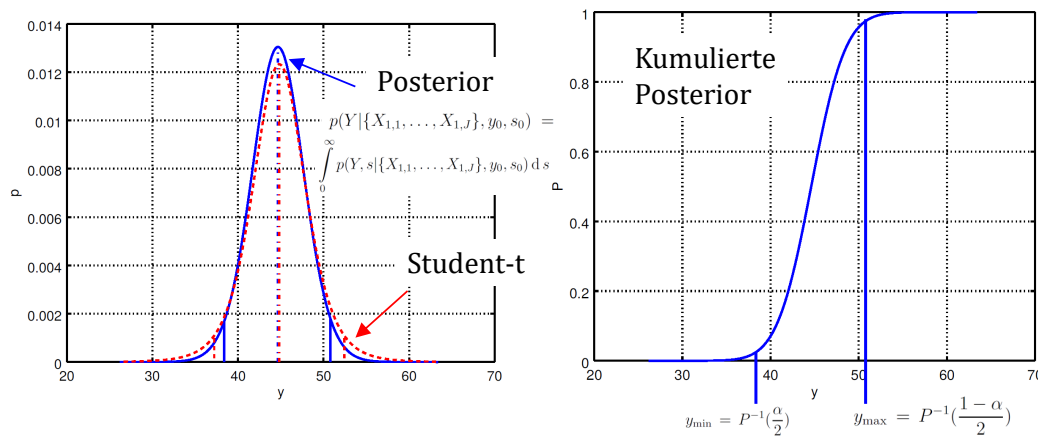


Abbildung 7.2: Wahrscheinlichkeitsdichtefunktionen und kumulierte Posterior für die Bestimmung des Überdeckungsintervalls einer indirekten Messgröße

Die Anzahl der Freiheitsgrade kann auf ganzzahlige Werte abgerundet werden, das heißt die Nachkommastellen unterdrückt werden, siehe GUM JCGM 100:2008, Seite 73:

NOTE 1 If the value of ν_{eff} obtained from Equation (G.2b) is not an integer, which will usually be the case in practice, the corresponding value of t_p may be found from Table G.2 by interpolation or by truncating ν_{eff} to the next lower integer.

```
>> tinv(0.975,52)
ans = 2.0066
>> tinv(0.975,52.58)
ans = 2.0061
```

Wir verwenden für das t-Quantil (den Erweiterungsfaktor) den Wert $k = 2.006$, in der Praxis nimmt man dann auch einfach $k = 2$. Gemäß GUM kann die Formel für die effektive Anzahl von Freiheitsgraden also auch wie folgt geschrieben werden

$$\nu_y = \lfloor u(Y)^4 \left(\sum_{i=1}^N \frac{(c_i s_i)^4}{\nu_i} \right)^{-1} \rfloor \quad (7.60)$$

wobei die beiden Symbole $\lfloor r \rfloor$ um die Variable r geschrieben heißen, dass die Nachkommastellen von r abzuschneiden sind, was in vielen Programmiersprachen mit der Funktion `floor` erfolgt.

Das Vertrauensintervall wird damit schließlich

$$[36.691 \text{ E}, 55.962 \text{ E}]$$

und das vollständige Messergebnis

$$Y = (46.327 \pm 9.635) \text{ E.}$$

Der Wert für die erweiterte Unsicherheit $U = 9.635$ E entspricht also in etwa dem Mittelwert aus den Berechnungen aus dem bayesischen Ansatz $\frac{1}{2}(9.423 + 9.907) = 9.665$. Abb. 7.2 zeigt das Prinzip anhand eines ähnlichen Beispiels, die eingezeichneten Werte liegen weiter auseinander, um die Intervalle besser einzeichnen zu können.

Für Größen, deren Messunsicherheit nicht unmittelbar aus vorliegenden Stichproben gewonnen wurde, so dass Stichprobenumfang und Verteilungen nicht ermittelt werden können, wird auf Basis heuristischer Vorstellungen eine Abschätzung der Anzahl der Freiheitsgrade vorgenommen.

In Fällen, bei denen es um Präzisionsmessungen geht, kann vielfach angenommen werden, dass die Anzahl der Freiheitsgrade gegen unendlich konvergiert.

$$\lim_{\nu_i \rightarrow \infty} \left\{ \frac{(c_i s_i)^4}{\nu_i} \right\} = 0 \quad (7.61)$$

Dies ist in Anhang G, Absatz G.4.3 des GUM JCGM 100:2008 nachzulesen. Lässt sich nicht voraussetzen, dass die Anzahl der Freiheitsgrade sehr, sehr groß ist, so wird diese abgeschätzt. Wie bei der Satterthwaite-Gleichung selbst, wird auch hier die Varianz der Varianz

$$\nu \sim \frac{\sigma_i^2}{\text{Var}(s_i)} \quad (7.62)$$

betrachtet, siehe Vorl 4, Gl (38) und GUM Anhang G, Absatz G.4.2. Die Zuverlässigkeit der als à-priori Information mitgeteilten Unsicherheit muss abgeschätzt werden mit einem bestimmten Prozentsatz $(1 - \alpha) \cdot 100\%$. Daraus wird heuristisch in Anlehnung an den Zusammenhang (7.62) die Anzahl der Freiheitsgrade abgeschätzt mit

$$\nu \approx \frac{1}{2\alpha^2}. \quad (7.63)$$

Der Quotient α wird als relative Unsicherheit $\frac{\Delta u}{u}$ der Unsicherheit u interpretiert.

Nachdem wir gesehen haben, wie handlich die Berechnungen mit der klassischen Fortpflanzung gegenüber den Methoden mit den Wahrscheinlichkeitsdichteverteilungen ist, fragt man sich natürlich, weshalb wir den numerischen Aufwand mit dem Berechnen der Integrale der Wahrscheinlichkeitsdichteverteilungen betreiben. Für dieses kleine, handliche und anschauliche Beispiel ist dies selbstverständlich nicht gerechtfertigt. Es dient lediglich dazu, das Grundprinzip des Verfahrens zu verdeutlichen.

Bei den Aufgaben, für die die Voraussetzungen zur Berechnung der Unsicherheit einer indirekten Messgröße gemäß Gl. (7.15) nicht mehr gegeben sind, werden die Verfahren mit Berechnung von Wahrscheinlichkeitsdichteverteilungen eingesetzt. Dies sind Verfahren die entweder auf der Multiplikation von Likelihoodverteilungen - letztlich auf Monte-Carlo-Berechnungen gemäß dem JCGM 101-Dokument der GUM-Reihe - oder auf der bayesischen Statistik basieren. Die Multiplikation von Likelihoodverteilungen setzt im Gegensatz zur bayesischen Statistik eine *scharfe* „Verteilung“ für die Modellfunktion ein. Dabei wird hier mit dem Begriff *scharfe* Verteilung eine Wahrscheinlichkeitsdichte bezeichnet, die als Normalverteilung mit dem Grenzübergang für $\sigma_{\text{winzig}} \rightarrow 0$ in eine Dirac'schen Deltafunktion übergeht. Den Begriff „Verteilung“ setzen wir hier in Anführungsstriche, weil so eine Dirac'sche Deltafunktion der Grenzfall ist, bei dem die Verteilung keine Verteilung mehr ist, weil die Werte scharf und nicht mehr verteilt sind.

Bei dem Beispiel (mit Gun-Octave/matalb-Skript in Anhang 7.8) haben wir die numerischen Inte-

grationen in sehr schlichter Weise dadurch realisiert, dass wir die betreffenden Größen in ein Raster äquidistanter Stützstellen diskretisiert haben und dann einfach summiert haben. Es sollte hiermit nur das Grundprinzip der Verfahren der *kolmogoroffschen Wahrscheinlichkeitsrechnung* aufzeigen, die auch die Basis für die Methoden der *bayesischen Statistik* bildet.

In vielen Fällen ist es aber nicht angesagt, beim numerischen Integrieren so zu verfahren. Das nächste Kapitel soll deshalb einen kurzen Überblick über unterschiedliche Verfahren zur numerischen Integration liefern.

7.6 Aufgabe zum Selbststudium

Es sollen Abstände auf einem Werkstück mit Hilfe eines induktiven Wegaufnehmers gemessen werden. Das Messprinzip eines induktiven Wegaufnehmers beruht darauf, dass eine Wechselspannung ein Spulensystem im Sensor anregt. Ein bewegliches ferro-magnetisches Teil am Sensor beeinflusst die Induktivität in den Spulen. Diese - in den Spulenteilen unterschiedliche - Induktivitätsveränderung wird vom Messverstärker ausgewertet und in ein positions-proportionales Gleichspannungssignal umgewandelt.

Um die Abstände in der physikalischen Einheit Mikrometer zu erhalten, muss mit Hilfe eines Bezugsnormals aus dem Spannungssignal ein Weg mit der Dimension einer Länge berechnet werden. Es ist bekannt, dass in dem für die Messung relevanten Messbereich (Hub des Sensors) die Abhängigkeit zwischen Spannungssignal und Auslenkung des ferro-magnetischen Kerns linear ist. Das Bezugsnormal ist eine Stufenhöhe.

Zu dem Bezugsnormal gibt es einen Kalibrierschein, der folgenden Höhenwert für das Normal angibt:

$$d = (4.997 \pm 0.011) \mu\text{m} \quad \text{mit} \quad k = 2.1, \nu = 26 \quad (7.64)$$

Mit dem induktiven Wegaufnehmer wurden auf dem Bezugsnormal folgende Stufenhöhen in der Dimension der elektrischen Spannung mit der Einheit Millivolt gemessen

Tabelle A5.1:

U_B/mV	201.3	187.3	196.5	200.4	193.6	174.2	197.2	185.4	194.4	202.5	205.2
-----------------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

Tabelle A5.2:

U_W/mV	176.5	184.1	180.5	193.6	176.0	194.5	160.9
-----------------	-------	-------	-------	-------	-------	-------	-------

Der Zusammenhang zwischen dem Abstand auf dem Werkstück in Mikrometern d_W , der Stufenhöhe des Bezugsnormals gemäß Kalibrierschein d , dem gemessenen Spannungssignal an der Stufenhöhe U_B und dem gemessenen Spannungssignal U_W am Werkstück ist folgender

$$d_W = \frac{d}{U_B} U_W \quad (7.65)$$

Die Änderungen aufgrund der statistischen Streuung der Spannungswerte U_B sind so klein, dass für die Sensitivität c_B von d_W bezüglich Änderungen von U_B mit dem hyperbolischen Zusammenhang

$d_W \sim \frac{1}{\bar{U}_B}$ die Steigung der Tangenten an die Hyperbel verwendet wird. Es wird die Tangente verwendet, die an dem Punkt, der sich aus den Mittelwerten ergibt, anliegt.

$$c_B = \left. \frac{\partial}{\partial U_B} d_W \right|_{\bar{U}_B, \bar{d}, \bar{U}_W} \quad (7.66)$$

Die Sensitivitäten c_d und c_W von d_W bezüglich Änderungen von d und U_W sind aufgrund des linearen Zusammenhangs genau

$$c_d = \left. \frac{U_W}{U_B} \right|_{\bar{U}_B, \bar{U}_W} \quad c_W = \left. \frac{d}{U_B} \right|_{\bar{U}_B, \bar{d}} \quad (7.67)$$

definiert. Die kombinierte Varianz unter der Voraussetzung, dass die gemessenen Größen und die Angabe aus dem Kalbrierschein unkorreliert sind, ist

$$s_{dW}^2 = (c_d s_d)^2 + (c_B s_B)^2 + (c_W s_W)^2 \quad (7.68)$$

mit s_B für die Standardabweichung der in Tabelle A5.1 aufgelisteten Werte, s_W für die Standardabweichung der in Tabelle A5.2 aufgelisteten Werte und s_d für die Standardabweichung der Angabe aus dem Kalbrierschein.

- Ermitteln Sie aus den Angaben des Kalbrierscheins (hier Gl. (9.6)) die Standardabweichung s_d , indem Sie den dort genannten Erweiterungsfaktor verwenden.
- Ermitteln Sie die beiden Mittelwerte \bar{U}_B und \bar{U}_W , sowie die beiden empirischen Standardabweichungen s_B und s_W aus den beiden Tabellen A5.1 und A5.2.
- Bestimmen Sie die Sensitivitäten c_d , c_W , c_B durch Ausführen der partiellen Ableitung.
- Berechnen Sie die kombinierte empirische Standardabweichung s_{dW} .
- Berechnen Sie die effektive Anzahl der Freiheitsgrade.
- Berechnen Sie den Erweiterungsfaktor k , der hier das t-Quantil für ein zweiseitiges Vertrauensniveau von $1 - \alpha = 0.95\%$ ist.
- Bestimmen Sie das vollständige Messergebnis.

7.7 Anhang: Erwartungswerte zur χ^2 -Verteilung

Die χ^2 -Verteilung ist definiert durch

$$p(Q, \nu) := \frac{Q^{\frac{\nu}{2}-1} e^{-\frac{Q}{2}}}{\Gamma\left(\frac{\nu}{2}\right) 2^{\frac{\nu}{2}}} \quad \nu \in \mathbb{N} \quad (7.69)$$

mit

$$\Gamma\left(\frac{\nu}{2}\right) = \sqrt{\pi} \frac{(\nu-2)!!}{2^{\frac{\nu-1}{2}}} \quad (7.70)$$

und

$$\nu!! = \nu(\nu-2)(\nu-4)\dots 4 \cdot 2. \quad (7.71)$$

Der Erwartungswert für die Zufallsgröße Q ist

$$E(Q) = \int_0^\infty Q' p(Q', \nu) dQ' \quad (7.72)$$

und nach Integration mit Computeralgebrasoftware (beispielsweise Maxima)

$$E(Q) = \frac{2\Gamma\left(\frac{\nu}{2} + 1\right)}{\Gamma\left(\frac{\nu}{2}\right)} \quad (7.73)$$

und mit

$$\Gamma\left(\frac{\nu}{2} + 1\right) = \Gamma\left(\frac{\nu}{2} + \frac{2}{2}\right) = \Gamma\left(\frac{\nu+2}{2}\right)$$

und

$$\Gamma\left(\frac{\nu+2}{2}\right) = \sqrt{\pi} \frac{(\nu+2-2)!!}{2^{(\nu+2-1)/2}}$$

erhalten wir

$$\begin{aligned} 2 \frac{\Gamma(\nu/2+1)}{\Gamma(\nu/2)} &= 2 \frac{(\nu)!!}{(\nu-2)!!} \frac{2^{(\nu-1)/2}}{2^{(\nu-1)/2+1}} \\ &= 2 \frac{(\nu)!!}{(\nu-2)!!} \frac{1}{2} \\ &= \nu \end{aligned} \quad (7.74)$$

Der Erwartungswert für die Varianz $\text{Var}(Q)$ von Q ist

$$\text{Var}(Q) = \int_0^\infty (Q' - \nu)^2 p(Q', \nu) dQ' \quad (7.75)$$

und nach Integration mit dem Computeralgebraprogramm Maxima

$$\text{Var}(Q) = \nu^2 - 2\nu \frac{\Gamma(\nu/2+1)}{\Gamma(\nu/2)} + 2 \frac{\Gamma(\nu/2+2)}{\Gamma(\nu/2)} \quad (7.76)$$

d.h.

$$\text{Var}(Q) = \nu^2 - 2\nu^2 + \nu(\nu+2) = 2\nu \quad (7.77)$$

7.8 Anhang: Beispiel indirekte Größe aus Produkt zweier direkter Größen

```

function bayes_indirect_quantity_product()
% Kalibrierfaktor
    K_faktor_0 = 0.0925;
    sigma_K = 0.0090;
%
%
    JM = 9;
    sig = 12;
    mue = 500;
    data_M = [479.58; 526.47; 516.77; 522.01; 506.61; 497.99; 481.71; 484.90; 491.41];
%
%
    xK = [-4*sigma_K:0.005:4*sigma_K] + K_faktor_0;
    nK = length(xK)
    xM = [-4*sig:0.05:4*sig] + mue;
    nM = length(xM)
    y = [25:0.005:65];
    nY = length(y)
%
    std_M = std(data_M);
    std_KM = K_faktor_0 * std_M;
    printf('s_M = %1.4f, s_KM = %1.4f\n', std_M, std_KM);
%
% Messungen in Volt
    XM = data_M * ones(1,nM) - ones(JM,1) * xM;
%
% Das Modell: y = f(xK, xM)
% für jedes xK und jedes xM kombiniert
%  $y_{\{KM,i,j\}} = xK_i * xM_j$  für alle  $i=1,\dots,nK$  und  $j=1,\dots,nM$ 
    y_KM_matrix = xK' * xM;
% in einen langen Spaltenvektor gebracht
    y_KM = y_KM_matrix(:);
    nKM = length(y_KM);
    delta_y = y_KM * ones(1,nY) - ones(nKM, 1)*y;
%
% direkte Groessen
% Kalibrierfaktor
    p_K = exp(-0.5 * ( (xK - K_faktor_0)/sigma_K ).^2 );
    p_K = p_K / sum(p_K);
% Messungen: Likelihood
    p_M_matrix = exp(-0.5 * ( XM/std_M ).^2 );
% Summe über alle Messungen

```

```

p_M_2 = sum( p_M_matrix );
p_M_2 = p_M_2 / sum(p_M_2);
%
p_KM_matrix = p_K' * p_M_2;
% in einen langen Spaltenvektor gebracht
p_KM_2 = p_KM_matrix(:);
%
% Modellprior
p_y2 = exp(-0.5 * ( delta_y/(std_KM*0.003) ).^2 );
% Posterior
post_matrix = p_y2 .* (p_KM_2 * ones(1,nY));
% Normierung des Posteriors
sumpost = sum(post_matrix(:));
post_matrix = post_matrix/sumpost;
%
% Marginalverteilung durch Summation über xM, xK
posterior3 = sum(post_matrix);
%
% kumulative Marginale-PDF also die CDF
cdf3 = cumsum(posterior3);
%
figure(200);
plot( y, cdf3, 'b-', 'linewidth', 2);
xlabel('indirekte Groesse / Einheit', 'fontsize', 14);
ylabel('cdf', 'fontsize', 14);
set(gca, 'fontsize', 12);
%
% Ergebnis
y_bar = sum( y.*posterior3 );
[dmy, imin] = min( abs(cdf3-0.025) )
y1 = y(imin);
[dmy, imin] = min( abs(cdf3-0.975) )
y2 = y(imin);
fprintf('Credible interval = [%1.4f, %1.4f]\n', y1, y2);
fprintf('y_bar = %1.4f + (%1.4f) + (%1.4f)\n', y_bar, y1-y_bar, y2-y_bar);
% ---
% Vergleich mit Fortpflanzungsgesetz
% ---
xMbar = mean(data_M);
fprintf('mean(X_M): %1.4f\n', xMbar);
fprintf('s_M = %1.4f, s_KM = %1.4f\n', std_M, std_KM);
u = sqrt((K_faktor_0*std_M)^2 + (xMbar*sigma_K)^2);
fprintf('u(Y) = %1.4f\n',u);
hlp = (K_faktor_0*std_M)^4/8 + (xMbar*sigma_K)^4/45;

```

```
nu_eff = u^4/hlp
t = tinv(0.975,floor(nu_eff))
printf('U = %1.5f \n', t*u);
printf('ymean = %1.3f, [%1.3f,%1.3f]\n', ...
    K_faktor_0*xMbar, K_faktor_0*xMbar-t*u,K_faktor_0*xMbar+t*u);
end
```


Kapitel 8

Numerische Werkzeuge für Messunsicherheitsberechnungen

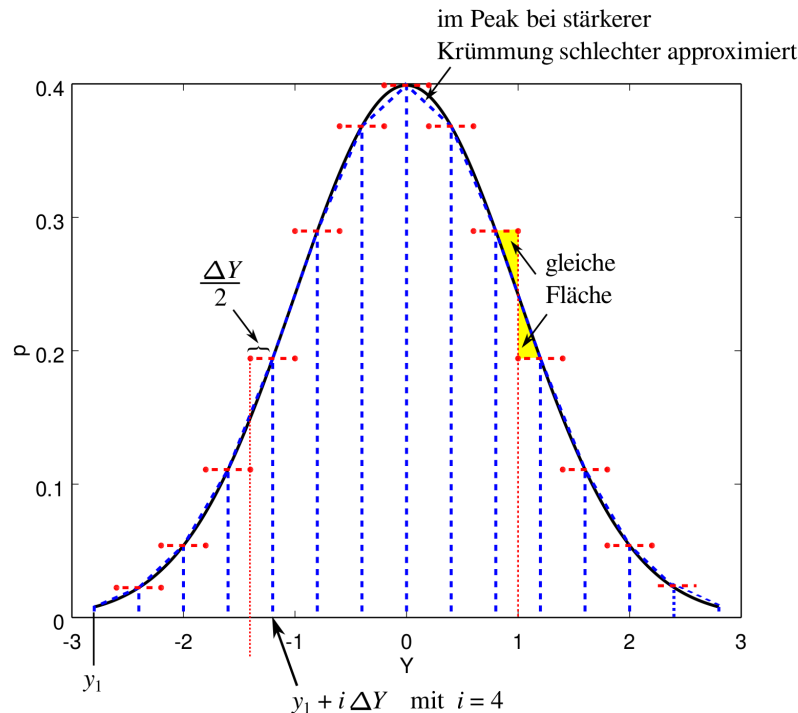


Abbildung 8.1: Trapezapproximation für die Berechnung der Fläche unter einer Kurve zur numerischen Integration.

8.1 Numerische Integrationsverfahren

Sei p eine Funktion $p: Y \rightarrow p(Y)$, die numerisch über ein Intervall $[y_1, y_2]$ integriert werden soll.

$$P(y_1, y_2) = \int_{y_1}^{y_2} p(Y) \, dY \quad (8.1)$$

Integrieren bedeutet, die Fläche unter der Kurve zu berechnen. Wir approximieren die Fläche unter der Kurve durch Rechtecke einer festen Breite ΔY und einer Höhe $p(Y_i)$ in der in Abb. 8.1 dargestellten Weise. Die Flächen der mit rot, gestrichpunkteter Linie gekennzeichneten Rechtecke sind genauso groß wie die mit blau gestrichelter Umrandung markierten Trapezstücke. Diese Flächen werden summiert, um näherungsweise die Fläche unter der Kurve zu bestimmen. Das erste und das letzte Flächenstückchen sind nur halb so groß. Das Integrationsintervall wird in n Flächen unterteilt, so dass für die Breite der Flächenstücke gelte

$$\Delta Y = \frac{y_2 - y_1}{n}. \quad (8.2)$$

Es wird nun so aufgeteilt, dass $n - 1$ Stücke betrachtet werden gemäß den rot gestrichpunkteten Rechtecken mit der Breite ΔY und zwei weitere Stücke je der Hälfte dieser Breite und je eines ganz am Anfang und das andere als letztes Stückchen ganz am Ende. Das erste Stückchen hat die halbe Breite und die Fläche

$$\underbrace{\frac{\Delta Y}{2} p(y_1)}_{\text{Rechteck}} + \underbrace{\frac{\Delta Y}{4} (p(y_1 + \frac{\Delta Y}{2}) - p(y_1))}_{\text{Dreieck}} = \frac{\Delta Y}{4} (p(y_1 + \frac{\Delta Y}{2}) + p(y_1))$$

das letzte Stückchen analog

$$\frac{\Delta Y}{4} \left(p(y_2 - \frac{\Delta Y}{2}) + p(y_2) \right)$$

so dass für die gesamte Fläche gilt

$$P(y_1, y_2) = \varepsilon + \frac{\Delta Y}{4} \left(p(y_1 + \frac{\Delta Y}{2}) + p(y_1) + p(y_2 - \frac{\Delta Y}{2}) + p(y_2) \right) + \sum_{i=1}^{n-1} p(y_1 + i \Delta Y) \Delta Y \quad (8.3)$$

mit ε für die Abweichung von der Fläche, die unter der Kurve liegt. Dabei gilt es eine geeignete Wahl für die Breite ΔY der Intervalle oder andersrum ausgedrückt für die Anzahl n der Intervalle zu treffen.

Eine Möglichkeit ist, sich iterativ zu einer sinnvollen Anzahl von Intervallen vorzuarbeiten, indem die Intervallbreite ΔY mit jedem Iterationsschritt halbiert wird. Dabei arbeitet man sich solange vor bis sich die approximierte Fläche durch weiteres Halbieren der Intervallbreiten nicht mehr signifikant ändert. Diese Methode wird *Romberg-Verfahren* genannt. Werner Romberg war ein in Berlin geborener Mathematiker, der in Heidelberg und München studierte und promovierte. Kurz nach dem Rigorosum im Jahr 1933 musste er als Kritiker des Nationalsozialismus Deutschland verlassen. 1949 wurde Werner Romberg Dozent für Physik am Norwegian Institute of Technology (NTH) in Trondheim und richtete dort den Studiengang für Mathematische Physik ein. 1955 publizierte er dieses numerische Integrationsverfahren.

Ein Quellcode-Beispiel in C ist auf der Wikipediaseite zum Romberg-Verfahren zu finden:

https://en.wikipedia.org/wiki/Romberg%27s_method

Bei jedem Iterationsschritt wird ΔY halbiert. Der Iterationsschritt $n = 1$ bedeutet $\Delta Y_1 = y_2 - y_1$, Schritt $n = 2$ bedeutet $\Delta Y_2 = (y_2 - y_1)/2 = \Delta Y_1/2$ etc.

$$P_n(y_1, y_2) = \frac{\Delta Y_n}{4} \left(p(y_1 + \frac{\Delta Y_n}{2}) + p(y_1) + p(y_2 - \frac{\Delta Y_n}{2}) + p(y_2) \right) + \sum_{i=1}^{n-1} p(y_1 + i \Delta Y_n) \Delta Y_n \quad (8.4)$$

mit

$$\Delta Y_n = \frac{1}{2} \Delta Y_{n-1}.$$

Die Iteration endet, sobald der Betrag der Differenz der approximierten Flächen

$$| P_n(y_1, y_2) - P_{n-1}(y_1, y_2) | < \varepsilon_P \quad (8.5)$$

einen Wert annimmt, der kleiner als ein zuvor definierter sehr kleiner Wert ε_P ist.

Dabei werden die Flächenstücke durch Halbierung der Intervallbreite gleichförmig schmäler gemacht. Die einzelnen Abweichungen zu jedem der Intervalle können für unterschiedlich stark gekrümmte Segmente der unterschiedlichen Stückchen verschieden groß sein, so dass es für manche Anwendungen Sinn machen kann dadurch Rechenzeit zu sparen, dass nur ausgewählte Bereiche verfeinert werden. Hier ist mehr Aufwand bei der Implementierung der Verästelung des Baumes, der für unterschiedliche Integrationsbereiche unterschiedliche Iterationstiefen verwaltet, erforderlich. Wir nennen die Grenzen

der Flächenstückchen, die bei Romberg gleichförmig die Positionen

$$Y_i = y_1 + (i - 1)\Delta Y$$

annehmen, *Samples*, für die im allgemeinen gelten kann

$$Y_{i+1} - Y_i \neq Y_{k+1} - Y_k \quad \text{für } i \neq k. \quad (8.6)$$

Insbesondere bei der Integration mehrdimensionaler Funktionen, wie wir sie in unseren Beispielen hatten, bevor wir die Marginalverteilung erworben hatten, also soetwas wie bei folgender Integration

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(Y, X_1, \dots, X_N | \{X_{i,1}, \dots, X_{i,J_i}\}, \bar{x}_k, s_k, \dots) dX_1 \dots dX_N$$

gilt es die *Samples* für die Hypervolumenstückchengrenzen optimal zu wählen. Die Berechnungen sollen in minimaler Zeit fertig werden und dennoch soll das gesamte Hypervolumen möglichst genau approximiert werden.

Betrachten wir konkret die Aufgabe, Wahrscheinlichkeitsdichteverteilungen p integrieren zu wollen, welche dadurch gekennzeichnet sind, dass ihre Funktionswerte positiv sind und dass sie aus einem oder vielleicht auch ganz wenig mehr als einem breiteren *Peak* (Glocke) bestehen. Diese Glocke muss dabei nicht unbedingt so schön symmetrisch sein wie bei der Normalverteilung oder bei der Student-t-Verteilung, sondern kann eine gewisse Schiefe haben wie beispielsweise bei der χ^2 -Verteilung für wenige Freiheitsgrade. Wenn wir Abb. 8.1 genauer anschauen, sehen wir, dass die Sekanten im Zentrum des *Peak*, der Glocke, stärker von der Kurve der Glocke abweichen als in den *Tails*. Die Sample-Stützstellen wollen wir also im Bereich des Glockenmaximums, bzw. in den Bereichen der Glockenmaxima, enger legen als in den *Tails*. Dies ist äquivalent zu dem, die Dichte der Stützstellen gemäß der Verteilung p selbst zu sampeln. Mit anderen Worten, es wird eine nach p verteilte Stichprobe der Größen X_i entnommen. Die Verteilung p wird verwendet, um gleichverteilte Pseudozufallszahlen auf nach p verteilte Zahlen abzubilden. Dieser Ansatz liegt den sogenannten *Importance Sampling*-Algorithmen zugrunde.

Wenn eine Verteilung p stark fluktuiert, also innerhalb eines Intervalls (des i -ten Intervalls) auf der X -Achse sehr unterschiedliche Funktionswerte $p(X_i)$ annimmt, so bedeutet das, dass $p(X_i)$ eine große Varianz hat.

$$\text{Var}(p(X_i)) = \frac{1}{J-1} \sum_{j=1}^J (p(X_{i,j}) - \bar{p}(X_i))^2. \quad (8.7)$$

Die Varianz bezüglich unterschiedlicher Intervalle, bzw. Regionen, wird als Maß verwendet, das der Entscheidung zugrunde gelegt wird, ob weitere Stützstellen in dieser Region gesampelt werden, die Region in kleinere Regionen zu unterteilen, um je Region kleinere Varianzen zu erhalten. Das Sampeln innerhalb der verschiedenen Regionen geschieht dabei voneinander unabhängig. Die Regionen werden auch Gruppen oder Schichten genannt. Ein nach diesem Verfahren gebildetes Sample heißt *geschichtete Zufallsstichprobe* oder *stratifizierte Zufallsstichprobe* (engl. *stratified Sample*).

8.2 Zufallszahlen und Wahrscheinlichkeitsverteilungen

Die Thematik des Samplings nach Methoden unter Verwendung von Zufallszahlen ist ein weites Feld in der Statistik. Große Bedeutung haben zufallszahlenbasierte Integrationsmethoden die auf Erkenntnissen aus der Thermodynamik, d.h. der statistischen Physik beruhen. Sie wurden aus Beobachtung der Brown'schen Molekularbewegung abgeleitet, also aus der Zufälligkeit und Art und Weise der Bewegung von Molekülen eines Fluids. Der zurückgelegte Pfad eines Moleküls, *Random Walk*, ändert zufällig die Richtung.

Die Abfolge der Transformationen, die eine aktuelle Position eines Gegenstands oder den aktuellen Zustand eines Prozesses in die nachfolgende Position bzw. den nachfolgenden Zustand überführt, und daraus die danach nachfolgende Position (den daraus folgenden Zustand) und so weiter, heißt *Markow-Kette*. Kennzeichen einer Markow-Kette ist, dass die Positionsänderung des Moleküls nur von der aktuellen Position, nicht aber von den Positionen davor, abhängt. Allgemein lassen sich viele statistische Prozesse, nicht nur Pfade von Fluidpartikeln oder Abfolgen der Werte von Zufallssampeln, mit Hilfe von Markow-Ketten beschreiben und es gilt:

Der zukünftige Zustand des Prozesses ist nur durch den aktuellen Zustand bedingt und wird nicht durch vergangene Zustände beeinflusst.

Mit dem Begriff Pseudozufallszahlen ist hier gemeint, dass mittels eines deterministischen Algorithmus eine Zahlenfolge berechnet wird, die den Anschein hat, dass die Zahlenwerte zufällig verteilt sind und nicht deterministisch.

Prinzipiell ist ein solcher Algorithmus, oftmals Zufallszahlengenerator genannt, wie folgt aufgebaut.

1. Setzen eines Startwertes, der *Seed* genannt wird, was ein englischsprachiger Begriff ist und Saat, Samenkorn heißt.
2. Wechselweise Multiplikationen und Restklassenoperationen mit sehr großen natürlichen Zahlen, wobei bei der Restklassenoperation als Teiler eine möglichst große Primzahl, beispielsweise 16807, zum Einsatz kommt.

Mit Restklassenoperation ist hier gemeint, dass eine Division mit natürlichen Zahlen durchgeführt wird, deren Rest als Ergebnis in die weiteren Berechnungsschritte eingeht.

Die heutigen Bibliotheken für Compiler und Programmierertools wie Matlab bieten nicht nur Zufallszahlengeneratoren, die gleichverteilte Zahlenfolgen liefern. Sie stellen schon fertige, schnelle, effektive Algorithmen zur Verfügung, die die Transformation auf die üblichen Verteilungsdichtefunktionen p zur Verfügung stellen. So ist `randn` der Name der Funktion in Matlab und Gnu-Octave zur Erzeugung von normalverteilten Zufallszahlen. `rande` liefert exponentiell verteilte Zufallszahlen, `randg` Zufallszahlen, die nach der Gammafunktion verteilt sind und `randp` poissonverteilte Zufallszahlen.

Wir können aber auch selber gleichverteilte Zufallszahlen abbilden auf eine andere Verteilung p , wenn wir die Umkehrfunktion der kumulierten Verteilung P^{-1} dazu zur Verfügung haben. Sei $R =$

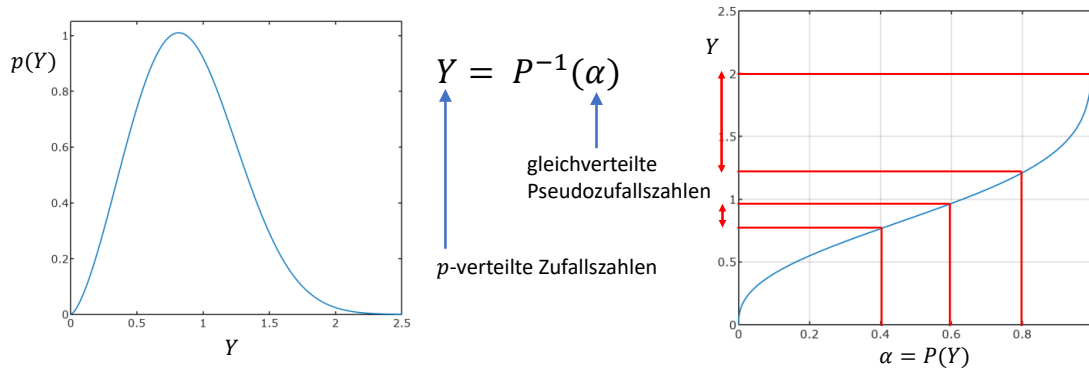


Abbildung 8.2: Eine gleichverteilte Größe wie beispielsweise eine Wahrscheinlichkeit α wird über die inverse Funktion P^{-1} der kumulierten Wahrscheinlichkeitsverteilung P , also der aufintegrierten Verteilung der Wahrscheinlichkeitsdichte p , abgebildet auf die Größe Y , die dann p -verteilt ist.

$\{R_1, \dots, R_J\}$ ein Sample, eine Stichprobe, gleichverteilter Zufallszahlen mit $R \in [0, 1)$, so ist die Menge der mit P^{-1} transformierten Zufallszahlen

$$Z = P^{-1}(R)$$

gemäß p verteilt, siehe Abb. 8.2. Als Beispiel wollen wir in Matlab/Gnu-octave 1000 Zufallszahlen generieren, die verteilt sind gemäß einer t -Verteilung für 45 Freiheitsgrade:

```
Jz = 1000;
nu = 45;
R = rand( Jz, 1);
Z = tinv( R, nu);
```

8.3 Monte-Carlo-Verfahren gemäß GUM-supplement 1 JCGM 101

In der zehnten Vorlesung werden wir das Monte-Carlo-Verfahren gemäß GUM-supplement 1 JCGM 101 erläutern, bei dem wie oben bereits erwähnt, gemäß den jeweiligen Verteilungen der unterschiedlichen direkten Messgrößen eine große Stichprobe (ein großes *Sample*) $\mathbf{x}_1 = (x_{1,1}, \dots, x_{N,1})^\top, \dots, \mathbf{x}_J = (x_{1,J}, \dots, x_{N,J})^\top$ für die direkten Größen \mathbf{X} per Zufallszahlengenerator erzeugt wird. Diese wird in das Modell f gesteckt, um eine Stichprobe

$$y_1 = f(x_{1,1}, \dots, x_{N,1}), \dots, y_J = f(x_{1,J}, \dots, x_{N,J}) \quad (8.8)$$

der indirekten Größe Y zu gewinnen. Entweder wird aus der inversen Funktion der kumulierten Wahrscheinlichkeitsverteilung das Überdeckungsintervall gewonnen oder es werden aus dem Sample y_1, \dots, y_J der Erwartungswert \bar{y} und die Standardabweichung σ_Y gemäß

$$\bar{y} = \frac{1}{J} \sum_{j=1}^J y_j \quad \sigma_Y = \sqrt{\frac{1}{J-1} \sum_{j=1}^J (y_j - \bar{y})^2}$$

berechnet.

Dieses Verfahren entspricht gemäß [Wue08] in etwa den Berechnungen wie in Gl. (7.55), nämlich wenn wir für die Standardabweichung σ_{winzig} den Grenzübergang $\sigma_{\text{winzig}} \rightarrow 0$ machen, sodass die zu integrierende Verteilung in die folgende übergeht:

$$p(Y, X_K, X_M | (X_{M,1}, \dots, X_{M,J}), K_0, s_K) \propto \delta(Y - X_K X_M) e^{-\frac{1}{2} \left(\frac{X_K - K_0}{s_K} \right)^2} \prod_{j=1}^J e^{-\frac{1}{2} \left(\frac{X_M - X_{M,j}}{s_M} \right)^2} \quad (8.9)$$

Zu den Größen X_K und X_M werden dann zufällige Samples gewählt.

Im folgenden werden wir den Grenzübergang der Wahrscheinlichkeitsdichte für das Modell

$$\lim_{\sigma \rightarrow 0} \left\{ e^{-\frac{1}{2} \left(\frac{Y - f(\mathbf{X})}{\sigma} \right)^2} \right\} = \delta(Y - f(\mathbf{X})) \quad (8.10)$$

in eine Dirac'sche Distribution als *scharfen Modellprior* bezeichnen.

Für univariate, explizite Modellfunktionen und für normalverteilte Größen X_i sieht die Posterior mit *scharfen Modellprior* dann wie folgt aus

$$p(Y, X_1, \dots, X_N | (X_{1,1}, \dots, X_{1,J_1}), \dots, (X_{K,1}, \dots, X_{K,J_K}), \bar{x}_{K+1}, s_{K+1}, \dots, \bar{x}_N, s_N) \propto \delta(Y - f(X_1, \dots, X_N)) \left(\prod_{i=1}^K \prod_{j=1}^{J_i} e^{-\frac{1}{2} \left(\frac{X_i - X_{i,j}}{s_i} \right)^2} \right) \left(\prod_{i=K+1}^N e^{-\frac{1}{2} \left(\frac{X_i - \bar{x}_i}{s_i} \right)^2} \right). \quad (8.11)$$

In der vorigen Vorlesung hatten wir bereits angesprochen, dass die Monte-Carlo-Verfahren auch insbesondere für Messaufgaben eingesetzt werden, bei denen mindestens ein Teil der direkten Messgrößen nicht normalverteilt ist. Dies kann vielfach sein, dass t -Verteilungen zu betrachten sind. Es kommen auch immer wieder Messaufbauten vor, bei denen die U-Verteilung eine Rolle spielt, also keine Glocke, sondern eine Verteilung, die ihre starken *Peaks* an den Rändern des Bereichs hat. Dies ist beispielsweise der Fall für Größen, deren Streuung durch Vibrationen verursacht wird. Die U-förmige Gestalt der Verteilung entspricht dann dem Quadrat eines Arkussinus.

Anstelle der Gaußfunktionen in Gl. (8.11) schreiben wir jetzt ganz allgemein p_i

$$p(Y, X_1, \dots, X_N | (X_{1,1}, \dots, X_{1,J_1}), \dots, (X_{K,1}, \dots, X_{K,J_K}), \theta_{K+1,1}, \dots, \theta_{N,M_K}) \propto \underbrace{\delta(Y - f(X_1, \dots, X_N))}_{\text{scharfer Modellprior}} \underbrace{\left(\prod_{i=1}^K \prod_{j=1}^{J_i} p_i(X_{i,j} | X_i) \right)}_{\text{Likelihood}} \underbrace{\left(\prod_{i=K+1}^N p_i(X_i | \theta_{i,1}, \dots, \theta_{i,M_i}) \right)}_{\text{Prior}}. \quad (8.12)$$

mit der Symmetrieeigenschaft $p(X_{i,j} | X_i) = p(X_i | X_{i,j}) \equiv l(X_i | X_{i,j})$. Dabei stellen $\theta_{K+1,1}, \dots, \theta_{N,M_K}$ Parameter aus à priori Informationen dar, wie beispielsweise $\theta_{i,1} = \bar{x}_i$ und $\theta_{i,2} = s_i$.

Wie in Abb. 8.3 dargestellt wird zur Erzeugung jeder der Verteilungen $p_i(X_{i,j} | X_i)$ bzw. $p_i(X_i | \theta_{i,1},$

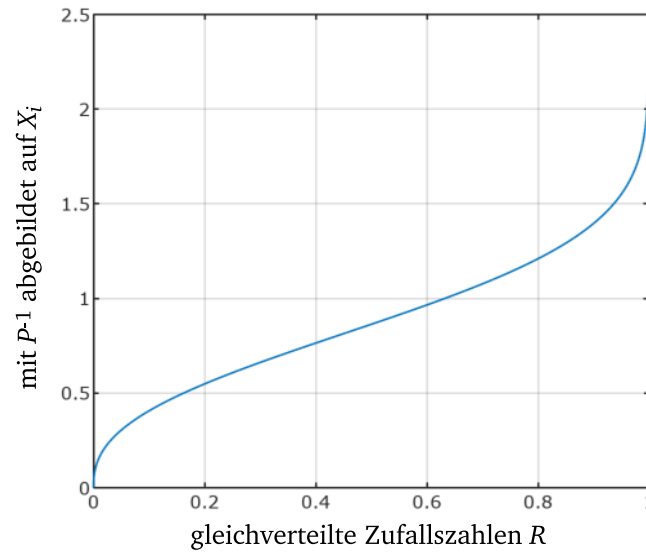


Abbildung 8.3: Ein Sample gleichverteilter Zufallszahlen $\mathbf{R} = \{R_1, \dots, R_L\}$ wird über die inverse Funktion P^{-1} der kumulierten Wahrscheinlichkeitsverteilung P abgebildet auf ein Sample $\mathbf{X} = \{X_1, \dots, X_L\}$, deren Werte dann p -verteilt sind.

$\dots, \theta_{i,M_i})$ der jeweiligen Größe X_i je ein Sample gleichverteilter Zufallszahlen $\mathbf{R}_i = \{R_{i,1}, \dots, R_{i,L}\}$ generiert, das dann mit P_i^{-1} abgebildet wird auf $\mathbf{X}_i = \{X_{i,1}, \dots, X_{i,L}\}$. Wir wissen, dass die $X_{i,j}$ beobachtete Werte der Größe sind, also Konstanten, die die Funktion P_i^{-1} parametrisieren. Entsprechend sind auch die $\theta_{i,1}, \dots, \theta_{i,M_i}$ Konstanten, die die dazugehörige Funktion P_i^{-1} parametrisieren. $\theta_{i,m}$ sind beispielsweise Mittelwert und Standardabweichung, oder andere Modellparameter, die aus Beobachtungen (Messdaten) gewonnen wurden. Zu jeder der Funktionsvariablen X_i wird also wie soeben beschrieben jeweils ein Sample p_i -verteilter Zufallsgrößen generiert. Für die Samples wird jeweils derselbe Stichprobenumfang L gebraucht, damit sich Tupel $(X_{1,l}, \dots, X_{N,l})$ mit $l = 1, \dots, L$ zusammenstellen lassen, aus denen über die Modellfunktion f jeweils der Wert Y_l für die indirekte Messgröße Y berechnet wird

$$Y_l = f(X_{1,l}, \dots, X_{N,l}) \quad (8.13)$$

mit L eine sehr große natürliche Zahl.

Aus der Verteilung der Werte Y_l lässt sich dann die Messunsicherheit bestimmen. Dieses Verfahren wird angewendet, wenn sich keine Sensitivitätskoeffizienten zu f berechnen lassen oder wenn die Sensitivitätskoeffizienten in der Nähe von Null liegen, weil f in der Umgebung der Schätzwerte ganz flach ist, also ein Extremum oder Sattelpunkt hat. Das bedeutet, dass f nicht linearisierbar ist.

Insbesondere wird das Verfahren mit den großen Zufallszahlensamples dann angewendet, wenn es keine analytisch geschlossene Darstellung zu dem Modell f gibt. Numerische Verfahren unter Anwendung von Zufallszahlen werden häufig in Anlehnung an den Zufall des Würfel- oder Roulettespiels Monte-Carlo-Verfahren genannt, so auch hier bei der Ermittlung von Messunsicherheiten.

Kapitel 9

Lösungen zu den Aufgaben

9.1 Lösungen zu den Aufgaben aus Vorl 2

9.1.1 Lösung zur 1. Aufgabe: Lineare Regression

siehe 2.8.1

zu (a) Stellen Sie das lineare Gleichungssystem auf, das sich durch partielles Ableiten für das Optimierungsproblem

$$\min_{a,c,h} \left\{ \sum_{j=1}^{J_T} \varepsilon_j^2 \right\}$$

ergibt, mit $J_T = 21$.

Die Modellgleichung lautet

$$z_j = a x_j + c + h \delta_{j \in C} + \varepsilon_j \quad (9.1)$$

In Vektorschreibweise mit \mathbf{x} und \mathbf{z} als Spaltenvektoren sieht dies mit allen in der oben aufgeführten Werten wie folgt aus

$$\mathbf{x} = \begin{pmatrix} 0 \\ 0.25 \\ 0.50 \\ \vdots \\ 2.25 \\ 2.75 \\ \vdots \\ 5.00 \end{pmatrix} \quad \mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \mathbf{x}_0 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \quad \mathbf{z} = \begin{pmatrix} -54.08 \\ -55.63 \\ -44.65 \\ \vdots \\ -60.77 \\ 36.85 \\ \vdots \\ 23.19 \end{pmatrix}$$

Für die Realisierung der Kroneckersymbole $\delta_{j \in C}$ wurde der Spaltenvektor \mathbf{x}_1 definiert, der als erste $J_C = 10$ Vektorkomponenten Einsen enthält und als weitere Vektorkomponenten Nullen. Der Vektor \mathbf{x}_0 besteht nur aus Einsen.

Als nächstes ist es wichtig, die Größen auf dieselbe Dimension zu bringen, so dass wir eine Steigung berechnen wollen, also rechnen wir die x-Werte auch in Mikrometern

$$\mathbf{x}_2 = 1000 \mathbf{x}$$

Somit sieht die Modellgleichung (9.1) wie folgt aus

$$\mathbf{z} = c \mathbf{x}_0 + h \mathbf{x}_1 + a \mathbf{x}_2 + \boldsymbol{\varepsilon}$$

nun schreiben wir alle Spaltenvektoren mit \mathbf{x} in eine gemeinsame Matrix, die dann 3 Spalten hat und $J_T = 21$ Zeilen

$$\mathbf{X} = (\mathbf{x}_0 \mathbf{x}_1 \mathbf{x}_2) \quad (9.2)$$

so dass die Modellgleichung wie folgt aussieht

$$\varepsilon = \mathbf{z} - \mathbf{X} \begin{pmatrix} c \\ h \\ a \end{pmatrix}.$$

Wenn man diesen Ansatz in dieser Form hat, kann man einfach Gl. (2.78) verwenden und alles einsetzen. Für die Klausur kommt es also nur darauf an, die Regressormatrix \mathbf{X} , das ist Gl. (9.2) aufstellen zu können und Gl. (2.78) aus dieser Vorlesung zu kennen:

$$(\mathbf{X}^T \mathbf{X}) \begin{pmatrix} c \\ h \\ a \end{pmatrix} = \mathbf{X}^T \mathbf{z} \quad (9.3)$$

wobei \mathbf{X}^T die transponierte Matrix ist, die in der ersten Zeile alles Einsen hat, in der zweiten Zeile an den ersten $J_C = 10$ Spaltenpositionen Einsen und an den letzten 11 Nullen hat und in der dritten und letzten Zeile die Werte 0.00 0.25 0.50 0.75 ... 5.00 aus der Tabelle hat. Hier ist dieser Teil der Aufgabenstellung fertig.

zu (b) Schreiben Sie die Gleichung für die Stufenhöhe d als Funktion von h und a auf.

$$d = \frac{h}{\sqrt{1 + a^2}} \quad (9.4)$$

Wer zuvor die x-Werte in Millimetern gerechnet hat, muss spätestens hier die Steigung a entsprechend umrechnen und durch 1000 teilen.

zu (c) Schreiben Sie die Formel für die Varianz der Residuen auf.

$$\sigma_\varepsilon^2 = \frac{1}{J_T - 3} \varepsilon^T \varepsilon$$

zu (d) Schreiben Sie die Formel für die Kovarianzmatrix der Modellparameter a, c, h auf.

Wir verwenden dazu Gl. (2.67) dieser Vorlesung

$$\Sigma = (\mathbf{X}^T \mathbf{X})^{-1} \sigma_\varepsilon^2 \quad (9.5)$$

mit

zu (e) Verwenden Sie eine Programmierumgebung Ihrer Wahl, Matlab, Octave, Python, R, ... die Ihnen einen Solver für lineare Gleichungssysteme zur Verfügung stellt, sowie eine Routine zur Matrixinversion und berechnen Sie die Zahlenwerte für die Modellparameter sowie deren Kovarianzmatrix.

```
function Lsg_Uebung_step()
%
% Teil (a)
x = [0.00  0.25  0.50  0.75  1.00  1.25  1.50 ...
      1.75  2.00  2.25  2.50  2.75  3.00  3.25  3.50 ...
```

```

    3.75  4.00  4.25  4.50  4.75  5.00]';
z = [-54.08 -55.63 -44.65 -51.44 -52.21 -58.01 ...
    -50.76 -56.01 -54.86 -60.77 36.85 38.02 31.71 36.21 ...
    23.39 29.01 30.11 29.35 20.81 33.27 23.19]';
x2 = 1e3*x;
J_T = 21;
J_C = 10;
x0 = ones(J_T,1);
x1 = [ones(J_C,1);zeros(J_T-J_C,1)];
X = [x0 x1 x2];
XTX = X' * X;
b = X' * z;
theta = XTX \ b;
printf('c = %1.4f um\n', theta(1));
printf('h = %1.4f um\n', theta(2));
printf('a = %1.7f um/um\n', theta(3));

%
% Teil (b)
d = theta(2) / sqrt(1 + theta(3)^2);
printf('d = %1.4f um\n', d);

%
% Teil (c)
epsilon = z - X * theta;
var_eps = (epsilon' * epsilon) / (J_T-3);

%
% Teil (e)
Kovarianz = inv(XTX) * var_eps;
printf(' %e %e %e \n', Kovarianz(1,1), Kovarianz(1,2), Kovarianz(1,3));
printf(' %e %e %e \n', Kovarianz(2,1), Kovarianz(2,2), Kovarianz(2,3));
printf(' %e %e %e \n', Kovarianz(3,1), Kovarianz(3,2), Kovarianz(3,3));

```

Ergebnisse:

```

c = 44.5660 um
h = -94.0905 um
a = -0.0038377 um/um
d = -94.0899 um

```

$\Sigma =$

```
2.369818e+01 -1.710178e+01 -5.863466e-03
-1.710178e+01 1.436549e+01 4.104426e-03
-5.863466e-03 4.104426e-03 1.563591e-06
```

Für die Unermüdlichen, die lieber in C programmieren, hier die Routine zum Lösen des linearen Gleichungssystems, haben wir eine Routine zum Lösen von linearen Gleichungssystemen aus den Numerical Recipes abgedruckt:

B. P. Flannery, W. H. Press, S. A. Teukolsky, W. T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, 2. Auflage (1992-2002)

Wichtig, diese Routine überschreibt den Vektor mit der Inhomogenität des Gleichungssystems mit der Lösung, also den gewünschten Modellparametern und die Matrix **a** mit den Summen aus den Regressoren mit deren Inversen \mathbf{a}^{-1} , die Sie dann für die Kovarianz der Modellparameter brauchen.

```
void gaussjordan(double **a, int n, double **b, int m)
/*
 * Linear equation solution by Gauss-Jordan elimination, equation (2.1.1) in
 * Numerical Recipes.
 * a[0..n-1][0..n-1] is the input matrix. b[0..m-1][0..n-1] is input
 * containing the m right-hand side vectors.
 * For most applications we have m = 1
 * On output, a is replaced by its matrix inverse,
 * and b is replaced by the corresponding set of solution vectors.
 */
int *indx, *indxr, *ipiv;
int i, icol, irow, j, k, l, ll;
double big, tmp, pivinv;

// The integer arrays ipiv, indxr, and indx are
// used for bookkeeping on the pivoting.
indx = (int*)calloc( n, sizeof(int));
indxr = (int*)calloc( n, sizeof(int));

ipiv = (int*)calloc( n, sizeof(int));

/* for (j=1;j<=n;j++) ipiv[j]=0; */
/* for (j=1;j<=n;j++) ipiv[j]=0; */
/* calloc initializes to zero:
Allocates a block of memory for an array of num elements,
each of them size bytes long, and initializes all its bits to zero.
*/
```

```
/*
 * This is the main loop over the columns to be reduced.
 */
for (i=0; i<n; i++) {
    big=0.0;

    // This is the outer loop of the search for a pivot element.
    for (j=0; j<n; j++)
        if (ipiv[j] != 1)
            for (k=0; k<n; k++) {
                if (ipiv[k] == 0) {
                    if (fabs(a[k][j]) >= big) {
                        big=fabs(a[k][j]);
                        irow=j;
                        icol=k;
                    }
                }
            }
        ++(ipiv[icol]);

    /*
     * We now have the pivot element, so we interchange rows, if needed,
     * to put the pivot element on the diagonal. The columns are not
     * physically interchanged, only relabeled:
     * indxc[i], the column of the ith pivot element,
     *          is the ith column that is reduced, while
     * indxr[i] is the row in which that pivot element was originally located.
     *          If indxr[i] !=
     * indxc[i] there is an implied column interchange.
     * With this form of bookkeeping, the solution b-s will end up in the
     * correct order, and the inverse matrix will be scrambled by columns.
     */

    if (irow != icol) {
        for (l=0; l<n; l++) SWAP(a[l][irow],a[l][icol])
        for (l=0; l<m; l++) SWAP(b[l][irow],b[l][icol])
    }

    /*
     * We are now ready to divide the pivot row by the
     * pivot element, located at irow and icol.
     */

    indxr[i]=irow;
```

```
indxc[i]=icol;
if (a[icol][icol] == 0.0) printf("gaussj: Singular Matrix\n");
pivinv=1.0/a[icol][icol];
a[icol][icol]=1.0;
for (l=0; l<n; l++) a[l][icol] *= pivinv;
for (l=0; l<m; l++) b[l][icol] *= pivinv;

/*
 * Next, we reduce the rows
 * except, if (ll != icol), for the pivot one, of course.
 */
for (ll=0; ll<n; ll++)
if (ll != icol) {
tmp=a[icol][ll];
a[icol][ll]=0.0;
for (l=0; l<n; l++) a[l][ll] -= a[l][icol]*tmp;
for (l=0; l<m; l++) b[l][ll] -= b[l][icol]*tmp;
}
}

/*
 * This is the end of the main loop over columns of the reduction.
 */

/*
 * It only remains to unscramble the solution in view of the column interchanges.
 * We do this by interchanging pairs of columns in the reverse order
 * that the permutation was built up.
 */
for (l=n-1; l>=0; l--) {
if (indxr[l] != indxc[l])
for (k=0; k<n; k++) SWAP(a[indxr[l]][k],a[indxc[l]][k]);
}
// And we are done.
free(ipiv);
free(indxr);
free(indxc);
}
```

9.2 Hilfe zu den Übungsaufgaben 5.8 zum Selbststudium

Hilfestellung Aufgabe 1

Ein Gnu-Octave/Matlab Skript für KS-Tests verschiedener Stichproben kann beispielsweise so aussehen:

```
function learn_KS(flagnew)
    N1 = 2000; mu1 = 23; s1 = 3;
    N2 = 450; mu2 = 35; s2 = 5;
    if (flagnew==1)
        x1 = s1*randn(N1,1) + mu1;
        x2 = s2*randn(N2,1) + mu2;
        save 'learn_robust.dat' x1 x2
    else
        load('learn_robust.dat');
    end
    x = [x1; x2];
    KStest(x1, 10);
    KStest(x2, 20);
    KStest(x, 0);
    di = abs(x - mean(x));
    medd = median(di);
    iuse = find( di < medd );
    for itest = 1:4
        KStest(x(iuse), itest);
        di = abs(x(iuse) - mean(x));
        medd = median(di);
        iuse2 = find( di < medd );
        iuse = iuse(iuse2);
    end
end

function KStest(x, iplt)
    alpha = 0.05;
    Ntot = length(x)
    [xsort, isort] = sort(x, 'ascend');
    h = [1:Ntot]'/Ntot;
    xbar = mean(x)
    sbar = std(x)
    g = normcdf( xsort, xbar, sbar);
    K = sqrt(-0.5*log(alpha/2))/sqrt(Ntot);
    figure(3000+iplt);
    plot( xsort, h, 'k-;{\fontsize{14}empirisches Histogramm};', ...
        xsort, g, 'r-;{\fontsize{14}Vert. aus Mittelw./ empir. Var.};');
```



```

xlabel('X', 'fontsize', 14);
ylabel('Wahrscheinlichkeitsdichte', 'fontsize', 12);
title(['Stichprobenumfang: ' num2str(Ntot)], 'fontsize', 14);
grid on;
legend(gca, 'location', 'northwest');
set(gca, 'fontsize', 14, 'linewidth', 2);
%
figure(3100+iplt);
plot( xsort, abs(g - h), 'b.;\fontsize{14}abs. Diff.};', ...
      [xsort(1); xsort(Ntot)], [K; K], 'r-;\fontsize{14}K_{\alpha,J}');
xlabel('X', 'fontsize', 14);
ylabel('abs. Differenz d. Wahrscheinlichkeitsdichte', 'fontsize', 12);
title(['Stichprobenumfang J = ' num2str(Ntot)], 'fontsize', 14);
legend(gca, 'location', 'northwest');
grid on;
set(gca, 'fontsize', 14, 'linewidth', 2);
end

```

Lösung zu Aufgabe 2

Gegeben seien zwei Stichproben

Stichprobe 1:

21	33	19	39	7
----	----	----	----	---

Stichprobe 2:

53	69	63	47	49	44	47	44	38
----	----	----	----	----	----	----	----	----

a) Zu jeder der beiden Stichproben die Mittelwerte und Standardabweichungen:

Stichprobe 1: Mittelwert $\bar{x}_1 = (21 + 33 + 19 + 39 + 7)/5 = 23.8$, Standardabweichung $s_1 = \sqrt{((21 - 23.8)^2 + (33 - 23.8)^2 + (19 - 23.8)^2 + (39 - 23.8)^2 + (7 - 23.8)^2)/(5 - 1)} = 12.54$

Stichprobe 2: Mittelwert $\bar{x}_2 = 50.44$, Standardabweichung $s_2 = 9.83$

b) Nullhypothese $H_0: \mu_1 = \mu_2$

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\bar{s}_1^2 + \bar{s}_2^2}}$$

Varianzen der Mittelwerte:

$$\bar{s}_1^2 = \frac{s_1^2}{5} = \frac{(12.54)^2}{5} = 31.44 \quad \bar{s}_2^2 = \frac{s_2^2}{9} = \frac{(9.83)^2}{9} = 10.73$$

damit

$$T = \frac{23.8 - 50.44}{\sqrt{31.44 + 10.73}} = -5.84$$

t-Quantil für Signifikanzniveau $\alpha = 0.05$ für zweiseitige Verteilung: $t_{1-\alpha/2, \nu} = t_{0.975, 5+9-2} = 2.18$

Ergebnis: $|T| = 5.84 > 2.18$ also Nullhypothese ablehnen

c) Nullhypothese $H_0: s_2^2 = \sigma_0^2$ mit $\sigma_0 = 6$

$$T = \nu \left(\frac{s_2}{\sigma_0} \right)^2$$

mit $\nu = 9 - 1$ für Stichprobe 2

$$T = 8 \left(\frac{9.83}{6} \right)^2 = 21.45$$

χ^2 -Quantile für Signifikanzniveau $\alpha = 0.05$ sind $\chi_{\alpha/2, \nu}^2 = \chi_{0.025, 8}^2 = 2.180$ und $\chi_{1-\alpha/2, \nu}^2 = \chi_{0.975, 8}^2 = 17.535$

Gilt $\chi_{\alpha/2, \nu}^2 < T < \chi_{1-\alpha/2, \nu}^2$?

Ergebnis: $2.180 < 21.45$ aber 21.45 ist nicht kleiner als 17.535 , also H_0 ablehnen

c) Nullhypothese $H_0: s_2^2 = \sigma_0^2$ mit $\sigma_0 = 9$

$$T = 8 \left(\frac{9.83}{9} \right)^2 = 9.53$$

Ergebnis: T liegt zwischen den Quantilen $2.180 < 9.53 < 17.535$, also H_0 annehmen

9.3 Lösung zur Aufgabe 7.6

Es sollen Abstände auf einem Werkstück mit Hilfe eines induktiven Wegaufnehmers gemessen werden. Das Messprinzip eines induktiven Wegaufnehmers beruht darauf, dass eine Wechselspannung ein Spulensystem im Sensor anregt. Ein bewegliches ferro-magnetisches Teil am Sensor beeinflusst die Induktivität in den Spulen. Diese - in den Spulenteilen unterschiedliche - Induktivitätsveränderung wird vom Messverstärker ausgewertet und in ein positions-proportionales Gleichspannungssignal umgewandelt.

Um die Abstände in der physikalischen Einheit Mikrometer zu erhalten, muss mit Hilfe eines Bezugsnormals aus dem Spannungssignal ein Weg mit der Dimension einer Länge berechnet werden. Es ist bekannt, dass in dem für die Messung relevanten Messbereich (Hub des Sensors) die Abhängigkeit zwischen Spannungssignal und Auslenkung des ferro-magnetischen Kerns linear ist. Das Bezugsnormal ist eine Stufenhöhe.

Zu dem Bezugsnormal gibt es einen Kalibrierschein, der folgenden Höhenwert für das Normal angibt:

$$d = (4.997 \pm 0.011) \mu\text{m} \quad \text{mit} \quad k = 2.1, \nu_d = 26 \quad (9.6)$$

Mit dem induktiven Wegaufnehmer wurden auf dem Bezugsnormal folgende Stufenhöhen in der Dimension der elektrischen Spannung mit der Einheit Millivolt gemessen

Tabelle A5.1:

U_B/mV	201.3	187.3	196.5	200.4	193.6	174.2	197.2	185.4	194.4	202.5	205.2
-----------------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

Tabelle A5.2:

U_W/mV	176.5	184.1	180.5	193.6	176.0	194.5	160.9
-----------------	-------	-------	-------	-------	-------	-------	-------

Der Zusammenhang zwischen dem Abstand auf dem Werkstück in Mikrometern d_W , der Stufenhöhe des Bezugsnormals gemäß Kalibrierschein d , dem gemessenen Spannungssignal an der Stufenhöhe U_B und dem gemessenen Spannungssignal U_W am Werkstück ist folgender

$$d_W = \frac{d}{U_B} U_W \quad (9.7)$$

Die Änderungen aufgrund der statistischen Streuung der Spannungswerte U_B sind so klein, dass für die Sensitivität c_B von d_W bezüglich Änderungen von U_B mit dem hyperbolischen Zusammenhang $d_W \sim \frac{1}{U_B}$ die Steigung der Tangenten an die Hyperbel verwendet wird. Es wird die Tangente verwendet, die an dem Punkt, der sich aus den Mittelwerten ergibt, anliegt.

$$c_B = \left. \frac{\partial}{\partial U_B} d_W \right|_{\bar{U}_B, \bar{d}, \bar{U}_W} \quad (9.8)$$

Die Sensitivitäten c_d und c_W von d_W bezüglich Änderungen von d und U_W sind aufgrund des linearen Zusammenhangs genau

$$c_d = \left. \frac{U_W}{U_B} \right|_{\bar{U}_B, \bar{U}_W} \quad c_W = \left. \frac{d}{U_B} \right|_{\bar{U}_B, \bar{d}} \quad (9.9)$$

definiert. Die kombinierte Varianz unter der Voraussetzung, dass die gemessenen Größen und die Angabe aus dem Kalbrierschein unkorreliert sind, ist

$$s_{dW}^2 = (c_d s_d)^2 + (c_B s_B)^2 + (c_W s_W)^2 \quad (9.10)$$

mit s_B für die Standardabweichung der in Tabelle A5.1 aufgelisteten Werte, s_W für die Standardabweichung der in Tabelle A5.2 aufgelisteten Werte und s_d für die Standardabweichung der Angabe aus dem Kalibrierschein.

- a) Ermitteln Sie aus den Angaben des Kalibrierscheins (hier Gl. (9.6)) die Standardabweichung s_d , indem Sie den dort genannten Erweiterungsfaktor verwenden.

$$s_d = 0.00524 \mu\text{m}$$

- b) Ermitteln Sie die beiden Mittelwerte \bar{U}_B und \bar{U}_W , sowie die beiden empirischen Standardabweichungen s_B und s_W aus den beiden Tabellen A5.1 und A5.2.

$$\bar{U}_B = 194.364 \text{ mV}, s_B = 9.045 \text{ mV}, \bar{U}_W = 180.871 \text{ mV}, s_W = 11.547 \text{ mV}$$

- c) Bestimmen Sie die Sensitivitäten c_d , c_W , c_B durch Ausführen der partiellen Ableitung.

$$c_d = \frac{\partial d_W}{\partial d} = \frac{U_W}{U_B} = 0.9306$$

$$c_W = \frac{\partial d_W}{\partial U_W} = \frac{d}{U_B} = 0.0257 \frac{\mu\text{m}}{\text{mV}}$$

$$c_B = \frac{\partial d_W}{\partial U_W} = -\frac{dU_W}{U_B^2} = -0.0239 \mu\text{m}$$

- d) Berechnen Sie die kombinierte empirische Standardabweichung $s_{d,W}$.

$$s_{d,W} = \sqrt{(c_d s_d)^2 + (c_W s_W)^2 + (c_B s_B)^2} = 0.3674 \mu\text{m}$$

- e) Berechnen Sie die effektive Anzahl der Freiheitsgrade.

$$\nu_{\text{eff}} = \frac{s_{d,W}^4}{(c_d s_d)^4/\nu_d + (c_W s_W)^4/(J_W - 1) + (c_B s_B)^4/(J_B - 1)} = 12.04$$

- f) Berechnen Sie den Erweiterungsfaktor k , der hier das t-Quantil für ein zweiseitiges Vertrauensniveau von $1 - \alpha = 0.95\%$ ist.

$$k = t_{0.975, 12.04} = 2.178$$

- g) Bestimmen Sie das vollständige Messergebnis.

$$d_W = \frac{d\bar{U}_B}{\bar{U}_W} \pm k s_{d,W} = (4.6501 \pm 0.8002) \mu\text{m}$$

Literaturverzeichnis

- [Fla02] B. P. Flannery, W. H. Press, S. A. Teukolsky, and W. T. Vetterling. Numerical Recipes in C. Cambridge University Press, 2. edition (1992-2002)
- [Wue08] Gerd Wübbeler, Michael Krystek and Clemens Elster: Evaluation of measurement uncertainty and its numerical calculation by a Monte Carlo method *Meas. Sci. Technol.*, 19 (084009), 4pp, 2008. doi:10.1088/0957-0233/19/8/084009
- [Rao09] S.S. Rao: Engineering optimization: theory and practice, John Wiley & Sons, 2009
- [Wiki01] https://en.wikipedia.org/wiki/Non-linear_least_squares
- [Wiki02] https://en.wikipedia.org/wiki/Gauss-Newton_algorithm
- [Sat41] F.E. Satterthwaite: Synthesis of variance. *Psychometrika*, 6(5), 1941.
- [Cox02] M. G. Cox: The evaluation of key comparison data, *Metrologia* 39, 589-595 (2002)
- [ISO13528] ISO: Statistical methods for use in proficiency testing by interlaboratory comparison, Second edition 2015-08-01, corrected version 2016-10-15
- [DIN1319] DIN 1319 Teil 1, Ausgabe Januar 1995 Titel: Grundlagen der Messtechnik – Grundbegriffe. Herausgeber: DIN Deutsches Institut für Normung e.V. Beuth-Verlag GmbH, Berlin–Wien–Zürich.
- [Kac08] R.N.Kacker, A.B.Forgbers R. N. Kacker, A. B. Forbes, R. Kessel and K. Sommer Bayesian posterior predictive p-value of statistical consistency in interlaboratory evaluations *Metrologia* 45 512-523 (2008)
- [GuideKey] Guidelines for CCPR Key Comparison Report Preparation, <https://www.bipm.org/utis/common/pdf/CC/CCPR/CCPR-G2.pdf>
- [ISO17011] ISO/IEC 17011:2005
- [Pau82] R.C. Paupe, J. Mandel: Consensus Values and Weighting Factors, *J. Res. NBS*, Vol.87, No.5,377-385 (1982)