

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

30. Oktober 2020

Inhaltsverzeichnis

1 Einführung: Messtechnik und Statistik	1
1.1 Grundlegende Begriffe	2
1.2 Modelle und Statistik	7
1.3 Ko-Varianzen von Zufallsgrößen	14
1.4 Beispiel zu Methode der kleinsten Quadrate zum Selbststudium	19
2 Lineare Regression (Ausgleichsrechnung)	21
2.1 Die Idee der linearen Regression	22
2.2 Lineare Regression als Beispiel zur Methode der kleinsten Abweichungsquadrate	24
2.3 Beispiel zur linearen Regression	27
2.4 Varianz der Regressionsparameter	29
2.5 Matrixformalismus zur allgemeinen Lösung der linearen Regression	30
2.5.1 Matrixformalismus anhand der Regressionsgeraden	30
2.5.2 Matrixansatz als allgemeiner Ansatz für linearen Regression	35
2.6 Das Bestimmtheitsmaß, empirischer Korrelationskoeffizient	38
2.7 Beispiele zur linearer Regression	40
2.7.1 Fit einer Geraden mittels linearer Regression	40
2.7.2 Fit einer kubischen Funktion (Polynom 3. Ordnung) mittels linearer Regression	46
2.8 Aufgaben zum Selbststudium	49
2.8.1 1. Aufgabe zur Linearen Regression	49
3 Optimierung nichtlinearer Modelle	51

3.1 Konzepte von Optimierungsverfahren für nichtlineare Modelle	52
3.2 Gradientenverfahren für nichtlineare Zielfunktionale	55
3.3 Beispiel zu Gradienten-Verfahren nach Gauß-Newton für den Fall gleicher Varianzen .	58
3.4 Levenberg-Marquardt-Verfahren	63
3.5 Matlabfunktionen zur Berechnung nichtlinearer Fits	73
3.6 Gradientenverfahren mit unterschiedlichen Varianzen	75
3.7 Robuste Schätzverfahren	77
4 Konzepte der Statistik für die Messdatenanalyse	81
4.1 Konzepte zum Lösen inverser Probleme	82
4.2 Konzept der bayesischen Verfahren zum Lösen inverser Probleme	88
5 Wahrscheinlichkeiten und Hypothesentests	95
5.1 Konzept der statistischen Erwartung	96
5.2 Kolmogorow-Smirnow-Test	97
5.3 Wahrscheinlichkeitsdichtefunktionen und ihre Parameter	102
5.4 Vertrauensintervall und t-Verteilung	103
5.5 t-Test - Mittelwerttest	113
5.6 Chi2-Verteilung und Test einer Varianz	114
5.7 Fisherverteilung und Vergleich zweier empirischer Varianzen	118
5.8 Anwendung von Hypothesentests	120
5.8.1 Entscheidung bei Vergleich zweier Stichproben	121
5.8.2 Konformitätsprüfung	123
5.9 Übungsaufgaben zum Selbststudium	124
6 Auswertung von Mess- und Ringvergleichen	127
6.1 Auswertung von Messvergleichen mit Referenzlabor	128
6.2 Auswertung von Ringvergleichen ohne Referenzlabor	130
6.2.1 Vorgehensweise	130

6.2.2 Beispiel	133
6.2.3 Identifikation von Ausreißern und Konsistenzcheck	134
6.2.4 Paule-Mandel Methode zur Anpassung von Gewichtsfaktoren	136
7 Messunsicherheitsfortpflanzung bei linearen Modelle	147
7.1 Konzepte der Messunsicherheitsfortpflanzung (M.U.F.)	148
7.2 M.U.F. für univariate, explizite indirekte Größe, linear von direkten Größen abhängig	153
7.3 M.U.F. für univariate, explizite indirekte Größe, Abhängigkeit von direkten Größen linearisierbar	156
7.4 Überdeckungsintervall für indirekte Messgrößen	159
7.5 Vergleich der klassischen mit der bayesischen Methode	163
7.6 Aufgabe zum Selbststudium	168
7.7 Anhang: Erwartungswerte zur χ^2 -Verteilung	171
7.8 Anhang: Beispiel indirekte Größe aus Produkt zweier direkter Größen	172
8 Numerische Werkzeuge für Messunsicherheitsberechnungen	175
8.1 Numerische Integrationsverfahren	176
8.2 Zufallszahlen und Wahrscheinlichkeitsverteilungen	179
8.3 Monte-Carlo-Verfahren gemäß GUM-supplement 1 JCGM 101	181
9 Messunsicherheitsberechnung mit bayesischer Statistik	185
9.1 Bayes-Theorem	186
9.1.1 Einleitung: Klassische und Bayessche Statistik	186
9.1.2 Das Bayes-Theorem	186
9.2 Parameterschätzung mit Bayesscher Statistik	188
9.2.1 Einführung	188
9.2.2 Beispiel: Schätzung einer physikalischen Konstante	188
9.2.3 Priorverteilungen	197
9.2.4 Posteriorverteilungen von Modellparametern	200
9.3 MU-Bestimmung mittels Posterior aus Bayesansatz	201

9.3.1 Bayessche Intervallschätzung	201
9.3.2 MU-Bestimmung mit der Bayesschen Statistik	202
9.4 Regression mit Vorkenntnis (Prior)	204
9.5 Median einer Verteilung	207
9.6 Übungsaufgaben	212
10 Monte-Carlo Methoden für Messunsicherheitsberechnungen	215
10.1 Einbeziehen von a-Priori-Wissen, „scharfer“ Model-Prior	216
10.1.1 Kurze Wiederholung: Bayessche Statistik zur MU-Bestimmung	216
10.1.2 Modellgleichung als à-Priori-Wissen → klass. Statistik	223
10.1.3 Analyt. Lösung für lineares Modell und Normalverteilungsdichten	226
10.2 MU-Berechnung mit rechteck- und U-förmigen Messgrößen	228
10.2.1 Messunsicherheitsfortpflanzung linearisierbarer Modelle	228
10.2.2 Wahrscheinlichkeitsdichteveerteilungen für Messgrößen	228
10.2.3 Wahrscheinlichkeitsdichten im Fortpflanzungsgesetz	230
10.3 Anhang	239
10.3.1 A1: Matlab-Skript zum Delta-Peaken mit Bayes	239
10.3.2 A2: Octave-Skript zur Fehlerfortpflanzung mit U- und Rechteck-Verteilungen .	242
10.4 Übungsaufgaben	248
11 Infrastrukturen der Metrologie	251
11.1 Internationale Metrologieorganisationen (BIPM,NMIs)	252
11.2 Akkreditierung, Zertifizierung, DAkkS	252
11.2.1 Messen, Prüfen, Lehren	252
11.2.2 Akkreditierung (Wirtschaft)	252
11.2.3 Zertifizierung	258
11.2.4 Konfirmitätsbewertung	260
11.2.5 Audit	261
11.2.6 DAkkS	262

12 Anhang: Quantiltabellen	267
12.1 Quantile der Normalverteilung	268
12.2 Quantile der t-Verteilung	269
12.3 Quantile der χ^2 -Verteilung	270
13 Lösungen zu den Aufgaben	271
13.1 Lösungen zu den Aufgaben aus Vorl 2	272
13.1.1 Lösung zur 1. Aufgabe: Lineare Regression	272
13.2 Hilfe zu den Übungsaufgaben 5.9 zum Selbststudium	278
13.3 Lösung zur Aufgabe 7.6	281

Kapitel 1

Einführung: Messtechnik und Statistik

1.1 Grundlegende Begriffe

Um über einen Sachverhalt kommunizieren zu können, müssen seine Gegenstände, Inhalt und Vorgänge **bezeichnet** werden. Damit die Kommunikationspartner sich gegenseitig verstehen, muss zuvor festgelegt sein, welcher Begriff / welcher Bezeichner / welche Vokabel welchen Bedeutungsinhalt darstellt. Deshalb müssen wir auch die Begrifflichkeiten der Bereiche Messtechnik und Statistik lernen und ihren Bedeutungsinhalt verstehen.

Die statistischen Begriffe und Sprachkonventionen sind nicht einheitlich. Beim Literaturstudium wird man feststellen, dass verschiedene Fachgebiete dieselben statistischen und systemtheoretischen Sachverhalte unterschiedlich darstellen und auch mit teilweise recht unterschiedlichen Ansätzen behandeln.

Im folgenden werden die Vokabeln der in Messtechnik und Statistik verwendeten Sprache aufgeführt:

1. Messen - vergleichen; Messvorgang

- **Messung** (engl. *measurement*): Prozess, bei dem ein Größenwert oder mehrere Größenwerte, die vernünftigerweise einer Größe zugewiesen werden können, experimentell ermittelt werden. [VIM2.1]
- **Messprinzip** (engl. *measuerement principle*): Phänomen, das als Grundlage einer Messung dient. [VIM2.4]

Beispiele: Länge eines Werkstücks mit Maßstab vergleichen, Temperatur darstellen über die Ausdehnung einer Quecksilbersäule in einem Röhrchen

Die Abkürzung VIM steht für *Vocabulaire international de métrologie* und bezeichnet das internationale Wörterbuch der Metrologie *International vocabulary of metrology - Basic and general concepts and associated terms*, das auf folgender Webseite als zweisprachige Version (Englisch und Französisch) frei erhältlich ist:

<https://www.bipm.org/en/publications/guides/vim.html>

Die nachstehende Zahl, hier 2.1 und 2.4, gibt den jeweiligen Absatz an. Die deutschsprachige Version ist kostenpflichtig im Beuthverlag erschienen.

2. Physikalische Größe und Größengleichung

- Eine **physikalische Größe** ist eine quantitativ bestimmbarer Eigenschaft eines physikalischen Objektes, Vorgangs oder Zustands. Ihr Wert (Größenwert) wird als Produkt aus einem Zahlenwert (der Maßzahl) und einer Maßeinheit angegeben. Vektorgrößen werden durch Größenwert und Richtung angegeben.
- Eine **Größengleichung** ist die mathematische Darstellung eines physikalischen Gesetzes, das Zustände eines physikalischen Systems und deren Änderungen beschreibt. Sie stellt den dabei geltenden Zusammenhang zwischen verschiedenen physikalischen Größen dar, wobei in der Regel für jede dieser Größen ein Formelzeichen steht. Größengleichungen gelten unabhängig von den gewählten Maßeinheiten.
- Diejenigen physikalischen Größen, die als Basis eines Größensystems festgelegt sind, heißen **Basisgrößen**.

- Im Jahr 1960 wurde das Internationale Einheitenystem SI (franz. *Système international d'unités*) basierend auf dem metrischen System von der Generalkonferenz für Maß und Gewicht eingeführt. Dieses liefert die Struktur für die Einheiten im (gesetzlichen) Messwesen.
- Das *Bureau international des poids et mesures* (BIPM) ist eine internationale Organisation, die vom internationalen Metervertrags-Bündnis, der Meterkonvention (*Metre Convention*), eingerichtet wurde. Durch das BIPM können die Mitgliederstaaten gemeinsam in Angelegenheiten des Messwesens und Standardisierung im Messwesen handeln.
- Die Basisgrößen im SI-System sind folgende

1.) **Zeit:** Einheit **Sekunde**, Formelzeichen der Einheit: s

– von 1983 bis 2019:

Die Sekunde ist das $9\ 192\ 631\ 770$ fache der Periodendauer der dem Übergang zwischen den beiden Hyperfeinstrukturniveaus des Grundzustandes von Atomen des Nuklids ^{133}Cs entsprechenden Strahlung.

– seit Mai 2019:

Die Sekunde ist definiert, indem für die Cäsiumfrequenz $\Delta\nu_{\text{Cs}}$, der Frequenz des ungestörten Hyperfeinübergangs des Grundzustandes der Zahlenwert $9\ 192\ 631\ 770$ festgelegt wird, ausgedrückt in der Einheit **Hertz**, die gleich Sekunde^{-1} ist.

2.) **Länge:** Einheit **Meter**, Formelzeichen der Einheit: m

– von 1983 bis 2019:

Der Meter ist die Länge der Strecke, die Licht im Vakuum während der Dauer von $(1/299\ 792\ 458)$ Sekunden durchläuft.

– seit Mai 2019:

Der Meter ist definiert, indem für die Lichtgeschwindigkeit in Vakuum c der Zahlenwert $299\ 792\ 458$ festgelegt wird, ausgedrückt in der Einheit Meter · Sekunde^{-1} , wobei die Sekunde mittels der Cäsiumfrequenz $\Delta\nu_{\text{Cs}}$ definiert ist.

3.) **Masse:** Einheit **Kilogramm**, Formelzeichen der Einheit: kg

– von 1983 bis 2019:

Das Kilogramm ist die Einheit der Masse; es ist gleich der Masse des Internationalen Kilogrammprototyps.

– seit Mai 2019:

Das Kilogramm ist definiert, indem für die Planckkonstante h der Zahlenwert $6,62607015 \cdot 10^{-34}$ festgelegt wird, ausgedrückt in der Einheit Joule · Sekunde, die gleich der Einheit Kilogram · Meter² · Sekunde⁻¹ ist, wobei der Meter mittels der Lichtgeschwindigkeit und die Sekunde mittels der Cäsiumfrequenz definiert sind.

4.) **elektrische Stromstärke:** Einheit **Ampere**, Formelzeichen der Einheit: A

– von 1983 bis 2019:

Das Ampere ist die Stärke eines konstanten elektrischen Stromes, der, durch zwei parallele, geradlinige, unendlich lange und im Vakuum im Abstand von einem Meter voneinander angeordnete Leiter von vernachlässigbar kleinem,

kreisförmigem Querschnitt fließend, zwischen diesen Leitern je einem Meter Leiterlänge die Kraft $2 \cdot 10^{-7}$ Newton hervorrufen würde.

- seit Mai 2019:

Das Ampere ist definiert, indem für die Elementarladung e der Zahlenwert $1,602176634 \cdot 10^{-19}$ festgelegt wird, ausgedrückt in der Einheit Coulomb, gleich dem Produkt Ampere · Sekunde ist, wobei die Sekunde über die Cäsiumfrequenz definiert ist.

5.) Temperatur: Einheit **Kelvin**, Formelzeichen der Einheit: K

- von 1983 bis 2019:

Das Kelvin, die Einheit der thermodynamischen Temperatur, ist der 273,16te Teil der thermodynamischen Temperatur des Tripelpunktes des Wassers. Diese Definition bezieht sich auf Wasser, dessen Isotopenzusammensetzung durch folgende Stoffmengenverhältnisse definiert ist: 0,000 155 76 Mol ^2H pro Mol ^1H , 0,000 379 9 Mol ^{17}O pro Mol ^{16}O und 0,002 005 2 Mol ^{18}O pro Mol ^{16}O

- seit Mai 2019:

Das Kelvin ist die SI Einheit der thermodynamischen Temperatur.

Das Kelvin ist definiert, indem für die Boltzmannkonstante k_B der Zahlenwert $1,380649 \cdot 10^{-23}$ festgelegt wird, ausgedrückt in der Einheit Joule · Kelvin $^{-1}$, die gleich dem Produkt Kilogramm · Meter 2 · Sekunde $^{-2}$ · Kelvin $^{-1}$ ist und wobei das Kilogramm über die Planckkonstante, der Meter über die Lichtgeschwindigkeit und die Sekunde über die Cäsiumfrequenz definiert sind.

6.) Stoffmenge: Einheit **Mol**, Formelzeichen der Einheit: mol

- von 1983 bis 2019:

Das Mol ist die Stoffmenge eines Systems, das aus ebenso vielen Einzelteilchen besteht, wie Atome in 0,012 Kilogramm des Kohlenstoffnuklids ^{12}C enthalten sind. Bei Benutzung des Mol müssen die Einzelteilchen spezifiziert sein und können Atome, Moleküle, Ionen, Elektronen sowie andere Teilchen oder Gruppen solcher Teilchen genau angegebener Zusammensetzung sein.

- seit Mai 2019:

Das Mol ist die SI Einheit für Stoffmenge. Ein mol enthält genau $6,02214076 \cdot 10^{23}$ elementare Entitäten (Einzelteilchen). Dieser Zahlenwert entspricht dem für die Avogadrokonstante N_A festgelegten Zahlenwert, ausgedrückt in der Einheit mol $^{-1}$ und wird als Avogadrozahl bezeichnet.

Die Stoffmenge, Formelzeichen n , eines Systems ist ein Maß für die Anzahl der spezifizierten Einzelteilchen. Bei einem Einzelteilchen kann es sich um ein Atom, ein Molekül, ein Ion, ein Elektron oder irgend eine andere Teilchenart oder eine spezifizierte Gruppe von Teilchen mit genau angegebener Zusammensetzung handeln.

7.) Lichtstärke: Einheit **Candela**, Formelzeichen der Einheit: cd

- von 1983 bis 2019:

Die Candela ist die Lichtstärke in einer bestimmten Richtung einer Strahlungsquelle, die monochromatische Strahlung der Frequenz $540 \cdot 10^{12}$ Hertz aussendet und deren Strahlstärke in dieser Richtung ($1/683$) Watt durch Steradian beträgt.

- seit Mai 2019:

Die Candela ist die SI Einheit der Lichtstärke in einer bestimmten Richtung.

Die Candela ist definiert, indem für das photometrische Strahlungsäquivalent K_{Cd} der monochromatischen Strahlung der Frequenz $540 \cdot 10^{12}$ Hertz der Zahlenwert 683 festgelegt wird, ausgedrückt in der Einheit Kandela · Steradian · Watt⁻¹ oder Kandela · Steradian · Kilogramm⁻¹ · Meter⁻² · Sekunde³, wobei das Kilogramm mittels der Planckkonstanten, der Meter mittels der Lichtgeschwindigkeit und die Sekunde mittels der Cäsiumfrequenz definiert sind.

https://en.wikipedia.org/wiki/2019_redefinition_of_the_SI_base_units

https://www.ptb.de/cms/fileadmin/internet/presse_aktuelles/broschueren/intern_einheitensystem/Die_gesetzlichen_Einheiten.pdf

<https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32019L1258&from=EN>

- Durch Verknüpfung der Basisgrößen gewonnene Größen sind beispielsweise

- **Frequenz**: Einheit **Hertz**, Formelzeichen der Einheit: Hz

Sie ist der Kehrwert der Zeit. Die Größengleichung ist

$$f = \frac{1}{t} \quad (1.1)$$

mit den Formelzeichen t für die Größe Zeit und f für die Frequenz

- **Kraft**: Einheit **Newton**, Formelzeichen der Einheit: N

Sie ist die Verknüpfung von den Größen Masse m , Länge L und Zeit t gemäß folgender Größengleichung (physikalischen Zusammenhangs):

$$F = m \frac{L}{t^2} \quad (1.2)$$

wobei F das Formelzeichen für die Größe Kraft ist.

Anmerkung: Die Größe $\frac{L}{t^2}$ hat einen eigenen Namen, sie heißt Beschleunigung und hat häufig das Formelzeichen a , aber ihre Einheit hat keinen eigenen Namen, die Einheit bleibt eine Verknüpfung aus Basiseinheiten: $\frac{\text{m}}{\text{s}^2}$.

- Die **Werte** von physikalischen **Größen** (Naturkonstanten) und von Umrechnungsfaktoren zwischen unterschiedlichen Einheitensystemen werden vom *Committee on Data for Science and Technology* (CODATA) festgelegt. Die Werte basieren auf Anpassungsberechnungen nach der Methode der kleinsten Quadrate. Der im Bericht von 2015, der über den folgenden Link zugänglich frei zugänglich ist beruht auf Daten, die bis zum 31. Dezember 2014 verfügbar waren.

<https://zenodo.org/record/22826#.W8bufNOzZhE>

Die Daten werden aus internationalen Vergleichsmessungen gewonnen.

3. Messgröße - Zufallsgröße

Beispiel Länge des Werkstücks wiederholt vergleichen
→ Beobachtungen

- Beobachtungen = Messwerte
Erfassen von Werten durch einen Messvorgang
- Streuung der beobachteten Werte
 - Die einzelnen beobachteten Werte streuen aufgrund des Mangels einer genauen Kenntnis des Entstehungsprozesses, weshalb man die Methoden der Statistik anwendet.
 - Die Statistik ist das mathematische Framework für den Umgang mit **Zufallsgrößen**.
- Zufallsgrößen - Wahrscheinlichkeit
Die Beobachtungen zu Größen werden in der Wahrscheinlichkeitstheorie mit dem Begriff **Ereignis** bezeichnet.
 - **Zufallsgrößen** sind dadurch gekennzeichnet, dass sie verschiedene Werte annehmen können, wobei jeder dieser Werte ein zufälliges Ereignis darstellt und mit einer bestimmten **Wahrscheinlichkeit** auftritt.
 - Die **Wahrscheinlichkeit** ist ein Maß für die Möglichkeit (engl. *likelihood*), dass ein Ereignis eintritt.
 - **Wahrscheinlichkeitsverteilung** wird die mathematische Funktion genannt, die jedem Wert einer Zufallsgröße die Wahrscheinlichkeit für sein Eintreten zuordnet.

Andrej Nikolaevic Kolmogorow, russischer Mathematiker, hat in den 1930er Jahren die axiomatische Begründung der Wahrscheinlichkeitstheorie entwickelt und damit den Formalismus wie wir ihn heutzutage verwenden. Der aus der kyrillischen Schrift in die lateinische Schrift umgeschriebene Name ist in der Literatur in unterschiedlicher Schreibweise zu finden, beispielsweise auch als Kolmogorov oder Kolmogoroff. Der kolmogorow'schen Wahrscheinlichkeitstheorie liegen folgende drei Axiome zugrunde:

- (a) Für jedes Ereignis ist die Wahrscheinlichkeit eine reelle Zahl zwischen 0 und 1.
- (b) Das sichere Ereignis (damit ist dasjenige gemeint, das immer eintritt) hat die Wahrscheinlichkeit 1.
- (c) Für Ereignisse, die sich gegenseitig ausschließen (disjunkte Ereignisse), gilt, dass die Summe der Wahrscheinlichkeiten jedes einzelnen dieser Ereignisse gleich der Wahrscheinlichkeit der Summe der Ereignisse ist.

Aus diesen drei Axiomen werden die Regeln, wie man mit Wahrscheinlichkeiten rechnen kann, hergeleitet. Im Formalismus der Mathematik nennt man die aus Axiomen hergeleiteten Folgerungen **Satz**. Dies sind die **Sätze** der kolmogorow'schen Wahrscheinlichkeitstheorie, die wir im Laufe dieser Vorlesungsreihe verwenden werden:

- (a) Satz: Aus Axiom 3, d.h. aus der Additivität der Wahrscheinlichkeiten, folgt, dass komplementäre Ereignisse (Gegenereignisse) komplementäre Wahrscheinlichkeiten (Gegenwahrscheinlichkeiten) haben.

Anmerkung zum Begriff **komplementär**: Das Komplement \bar{A} zu einer Menge A bezüglich des Grundbereichs G ist die Menge aller Elemente aus G , die nicht Elemente von A sind. A und \bar{A} sind Komplementärmengen bezüglich G . (Die Vereinigungsmenge von A und \bar{A} ist dann G und die Schnittmenge ist die leere Menge).

- (b) Satz: Das unmögliche Ereignis (die leere Menge) hat die Wahrscheinlichkeit 0.

- (c) Für die Vereinigung nicht notwendig disjunkter Ereignisse gilt, dass die Wahrscheinlichkeit der Vereinigung nicht notwendig disjunkter Ereignisse gleich der Summe der Wahrscheinlichkeiten der einzelnen Ereignisse abzüglich der Wahrscheinlichkeit der Schnittmenge der nicht notwendig disjunkten Ereignisse ist.

1.2 Modelle und Statistik

Hinter jedem Messvorgang und jeder Messgröße stehen physikalische Prinzipien und Naturgesetze, die als Modell beschrieben werden. Ein Modell wird dargestellt durch eine Vorstellung darüber wie die indirekten Messgrößen und die direkten Messgrößen miteinander verknüpft sind. Liegen Werte für indirekte Messgrößen vor, beispielsweise aus älteren Experimenten oder aus heuristischen Annahmen, so liefert das Modell als Ergebnis modellierte Werte für die direkten Größen (Modellannahmen über die direkten Größen):

$$(Y_1, \dots, Y_M) \xrightarrow{\text{Modell}} (X_1, \dots, X_N) \quad (1.3)$$

Die Modellrechnung ist vorwärts gerichtet und wird direkt berechnet.

In der messtechnischen Realität liegen Beobachtungen zu den direkten Messgrößen vor, aus denen die Werte für die indirekten Messgrößen zu schätzen sind:

$$(X_1, \dots, X_N) \xrightarrow{\text{inverses Problem}} (Y_1, \dots, Y_M) \quad (1.4)$$

Beispiel:

Eine Gitterstruktur wird mit kohärentem Licht beleuchtet. Das transmittierte Licht fällt auf eine CCD-Kamera, die die Intensitäten in Abhängigkeit von der Position auf dem CCD-Chip aufnimmt. Es gibt dabei $N = 3$ direkte Größen: zwei Ortskoordinaten (x, y) auf dem CCD und der dazugehörige Intensitätswert I . Die Anzahl der CCD-Pixel ist die Anzahl J der Beobachtungen, $(x_1, y_1, I_1), \dots, (x_J, y_J, I_J)$.

Die gesuchte Information ist die Gitterkonstante, sie ist die indirekte Größe.

Zum Modell gehören die physikalischen Gesetze der Fraunhofer-Beugung, zudem auch weitere Größen, wie die Wellenlänge der Beleuchtung, der Abstand zwischen CCD-Chip und dem zu untersuchenden Gitters, der groß genug sein muss, damit die Gesetze der Fernfelloptik gelten (der Einfachheit halber mal ohne Linsensysteme gedacht). Um die Abstände der Intensitätsmaxima genau zu lokalisieren, werden die Zusammenhänge für die Wellenoptik des Beugungphänomens gebraucht.

Ein Schätzvorgang besteht darin, die Modellparameter (das sind die indirekten Messgrößen) auszuprobieren, d.h. viele unterschiedliche Werte vorzugeben, und das Ergebnis der Modellrechnungen (4.1) mit den beobachteten Werten zu vergleichen, solange, bis das Ergebnis „gut zu den beobachteten Werten passt“. Die Methode des Variierens und Vergleichens „bis es passt“ nennen wir Schätzverfahren.

Die Methoden der Statistik dienen als Werkzeug zur Auswertung der Daten in der Messtechnik.

- In den Naturwissenschaften, den Ingenieurwissenschaften und der Statistik wird teilweise eine unterschiedliche Nomenklatur verwendet.
- In der Messdatenanalyse im Bereich Metrologie ist von *direkten* und *indirekten* Messgrößen die Rede. Sind die indirekten Messgrößen durch Schätzverfahren zu ermitteln, wird auch von **Daten** und **Modellparametern** gesprochen. In der Systemtheorie wird von **Eingangsgrößen** und **Ausgangsgrößen** und im Zusammenhang mit zeitabhängigen Größen (z.B. in der Elektrotechnik und in der Akustik) wird von **Eingangssignalen** und **Ausgangssignalen** gesprochen.
- Die Regressionsrechnung als Teilgebiet der Statistik kennt Eingangsgrößen als **Regressoren** und Ausgangsgrößen als **Regressanden**. Es handelt sich in beiden Fällen um Daten: Die an Geräten eingestellten Werte für die Eingangsgrößen werden als vorgegebene Werte ohne Streuung behandelt, und als Regressoren bezeichnet. Die mittels Messinstrument beobachteten Werte für die Ausgangsgrößen werden als Beobachtungen behandelt, die einer zufälligen Streuung unterliegen, und als Regressanden bezeichnet. Die Regressionsrechnung schätzt Modellparameter, die in der Systemtheorie auch **Zustandsgrößen** genannt werden können.
- Je nach Betrachtungsweise nehmen die indirekten Messgrößen entweder die Rolle der „Ausgangsgrößen“ (A_1, \dots, A_{n_A}) oder der „Zustandsgrößen“ (Z_1, \dots, Z_{n_Z}) oder der „Eingangsgrößen“ (E_1, \dots, E_{n_E}) ein.

$$(E_1, \dots, E_{n_E}) \xrightarrow{\text{System}(Z_1, \dots, Z_{n_Z})} (A_1, \dots, A_{n_A}) \quad (1.5)$$

Für das inverse Problem (4.3) sind die direkten Messgrößen die Eingangsgrößen und die indirekten Messgrößen die Ausgangsgrößen. Für das Modell (4.1) der indirekten Messgrößen sind die indirekten Größen die Eingangsgrößen und die direkten Messgrößen die Ausgangsgrößen. Das bedeutet, dass in der Messtechnik beides, die Regressoren und Regressanden die direkten Messgrößen sind und die Modellparameter dann die indirekten Messgrößen darstellen.

Gängiges Verfahren der Statistik zum Schätzen von Modellparametern ist die **Maximum-Likelihood**-Methode. Bei diesem Verfahren werden die Modellparameter einzig aus den Beobachtungen der direkten Messgrößen bestimmt. Auf dieses Teilgebiet der Statistik wird auch mit der Bezeichnung *frequentistische* Statistik Bezug genommen, um es von den statistischen Verfahren abzugrenzen, bei denen Vorwissen einbezogen wird. Es geht dabei um Vorwissen über die zu schätzenden Parameter sowie über die Wahrscheinlichkeitsverteilungen der beteiligten Größen. Das Vorwissen über die Wahrscheinlichkeitsverteilungen von direkten Messgrößen wird bei der Modellbildung einer konkreten Problemstellung eingebaut. Im Falle der *frequentistischen* Methoden wird beim Schätzverfahren nach der **Maximum-Likelihood**-Methode bereits vorausgesetzt, dass alle beteiligten Zufallsgrößen, die direkten und die indirekten, normalverteilt sind. Um bei den *frequentistischen* Methoden andere Verteilungen für die Schätzer (indirekten Größen) zulassen zu können wird entweder eine Umgewichtung der Abweichungen vorgenommen oder es wird anstelle der Normalverteilung die Laplaceverteilung verwendet. Darauf kommen wir im Laufe der Vorlesungreihe zurück. Die statistischen Verfahren, bei denen die Wahrscheinlichkeitsdichten und die Vorkenntnisse über Schätzparameter direkt eingearbeitet werden, nennt man *bayesische* Methoden.

Die Idee der Maximum-Likelihood-Methode wollen wir hier anhand eines kleinen Regressionsbeispiels illustrieren:

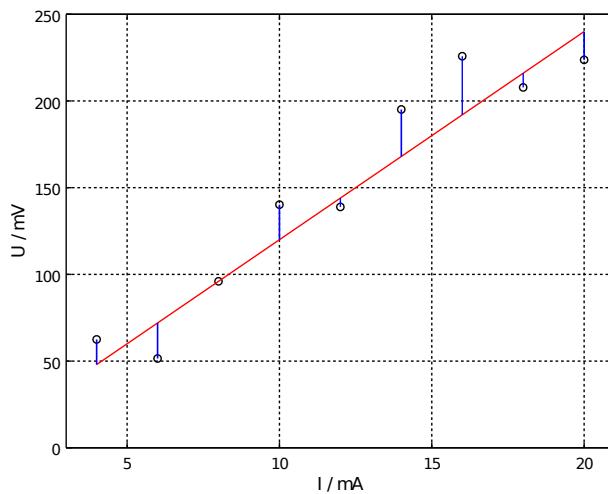


Abbildung 1.1: Regression Beispiel: Modellparameter finden, dessen Wert am besten zu den Daten passt

- Präzisionsstromquelle mit genau einstellbarer Stromstärke I ;
- Messung der Spannung U mit Voltmeter;
- physikalisches Modell: Ohmsches Gesetz: $R = \frac{U}{I}$; Spannung Null, wenn Strom Null (kein Offset!);
- Schätzung des Modellparameters Widerstand R : Suchen eines Wertes für R , für den das Ergebnis der Modellrechnung „gut zu den beobachteten Werten passt“.

Wir tragen die experimentellen Werte in eine Tabelle ein und zeichnen sie in ein Diagramm, Abb. 1.1:

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

Ob die Spannungswerte zu dem Modell „passen“ wird bewertet, wobei das Modell besagt, dass die Spannungswerte gleich dem Widerstand, der genau eine konstante Größe sein soll, multipliziert mit den eingestellten Werten der Stromstärke seien. Das Bewertungskriterium ist „die Treffer-Wahrscheinlichkeit Spannungswert liegt auf Widerstand mal Strom“. Wir gehen davon aus, dass die Abweichungen $\varepsilon_i = U_i - RI_i$ normalverteilt sind, also

$$p(U_i, I_i | R) \propto e^{-\frac{1}{2} \left(\frac{\varepsilon_i}{\sigma_\varepsilon} \right)^2} \quad (1.6)$$

wobei σ_ε ein Maß für die Streuung der Residuen, hier in Millivolt, ist. Wir haben die Werte der Abweichungen ε_i . Man nennt sie **Residuen**. Das Produkt aller Wahrscheinlichkeiten $p(U_i, I_i | R)$ heißt **Likelihood**.

$$L((U_1, \dots, U_J), (I_1, \dots, I_J) | R) = \prod_{i=1}^J p(U_i, I_i | R) \quad (1.7)$$

Unsere Modellparameter bzw. der Modellparameter, die bzw. den wir ermitteln wollen, ist R und für die Streuung der Werte des Parameters R verwenden wir das Formelzeichen σ_R .

In der frequentistischen Statistik wird von der Vorstellung ausgegangen, dass es einen wahren Wert für R gibt, dem man sich nähert, je mehr und genauer man misst, der aber wegen der Streuung der Spannungswerte U nur annähernd bestimmbar ist. Wenn wir die Präzisionsstromquelle unverändert auf einen festen Wert einstellen und nicht durchfahren und dann mehrfach die Spannungen ablesen, so beobachten wir dennoch leicht unterschiedliche Werte für die Spannung. Dies ist der Grund dafür dass die Wertepaare bei durchgefahrem Strom (I_j, U_j) nicht exakt auf einer Geraden liegen.

Am besten passt das ganze für einen Wert für R , für den $L((U_1, \dots, U_J), (I_1, \dots, I_J) | R)$ maximal wird. Die Zielfunktion unseres Optimierungsvorgangs ist also:

$$\max_R L((U_1, \dots, U_J), (I_1, \dots, I_J) | R) \quad (1.8)$$

Solch eine Optimierungsaufgabe heißt **Maximum-Likelihood**-Verfahren.

$$\max_R \prod_{i=1}^J e^{-\frac{1}{2} \left(\frac{\varepsilon_i}{\sigma_\varepsilon} \right)^2} \quad (1.9)$$

d.h. nach den Gesetzen der Potenzrechnung

$$\max_R e^{-\frac{1}{2} \sum_{i=1}^J \left(\frac{\varepsilon_i}{\sigma_\varepsilon} \right)^2}. \quad (1.10)$$

Das Maximum der Gaußfunktion liegt an der Stelle, an der der Exponent minimal wird. Wir maximieren die Likelihood, wenn wir ihren Logarithmus minimieren:

$$\min_R \sum_{i=1}^J \left(\frac{\varepsilon_i}{\sigma_\varepsilon} \right)^2 \quad (1.11)$$

dazu suchen wir die Nullstelle der Ableitung nach dem Modellparameter. Auf die Suche des Minimums wird auch mit dem Terminus *Methode der kleinsten Abweichungsquadrate* oder *Least Mean Square Method* Bezug genommen. Hätten wir nicht nur den einen Parameter R sondern viele Modellparameter, dann würden wir den Gradienten bilden.

$$\frac{\partial}{\partial R} \sum_{i=1}^J \left(\frac{\varepsilon_i}{\sigma_\varepsilon} \right)^2 = 0 \quad (1.12)$$

d.i. in unserem Beispiel

$$\frac{\partial}{\partial R} \sum_{i=1}^J \left(\frac{U_i - RI_i}{\sigma_\varepsilon} \right)^2 = 0 \quad (1.13)$$

Wir multiplizieren mit dem Faktor σ_ε

$$\frac{\partial}{\partial R} \sum_{i=1}^J (U_i - RI_i)^2 = 0 \quad (1.14)$$

und führen die Differentiation durch

$$\sum_{i=1}^J 2(U_i - RI_i) I_i = 0. \quad (1.15)$$

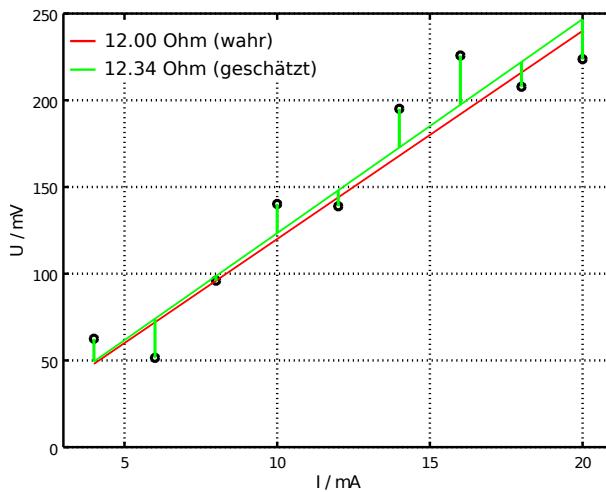


Abbildung 1.2: Regressionsbeispiel: Der geschätzte Wert des Modellparameters im Vergleich zum wahren.

Für viele Modellparameter gäbe es entsprechend viele solche Gleichungen wie es Modellparameter gibt mit den jeweiligen Ableitungen, so dass dann hier ein lineares Gleichungssystem stünde, das zu lösen wäre. Dazu mehr im nächsten Kapitel. Ein numerisches Verfahren zur Lösung linearer Gleichungssysteme ist das Gauß-Jordan-Eliminationsverfahren. Nun wir nur diese eine Gleichung haben, können wir einfach nach R auflösen.

$$\sum_{i=1}^J U_i I_i = \sum_{i=1}^J R I_i I_i \quad (1.16)$$

d.h.

$$R = \frac{\sum_{i=1}^N U_i I_i}{\sum_{i=1}^N I_i^2} \quad (1.17)$$

Der ermittelte Wert für R mit den Daten aus der Tabelle ist 12.34Ω , siehe Abb. 1.2

Zuvor haben wir die Likelihood ohne Normierungsfaktor aufgeschrieben. Eine Wahrscheinlichkeitsdichtevertteilung ist so definiert, dass die Fläche unter ihrer Kurve 1 ist, also die Wahrscheinlichkeit für das Beobachten jedes beliebigen Wertes immer voll eintritt.

$$p(\varepsilon) = \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} e^{-\frac{1}{2}\left(\frac{\varepsilon}{\sigma_\varepsilon}\right)^2} \quad (1.18)$$

Ein Maß für die Breite der Gaußglocke ist das σ_ε bei etwa 60 Prozent der Höhe ($e^{-\frac{1}{2}} \approx 0.6$). Ihr Quadrat wird **Varianz** genannt. Das zweite statistische Moment dieser Verteilung

$$\int_{-\infty}^{\infty} \varepsilon^2 p(\varepsilon) d\varepsilon \quad (1.19)$$

ist gleich dem Quadrat von σ_ε , also der Varianz der Gaußverteilung

$$\int_{-\infty}^{\infty} \varepsilon^2 \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} e^{-\frac{1}{2}\left(\frac{\varepsilon}{\sigma_\varepsilon}\right)^2} d\varepsilon = \sigma_\varepsilon^2. \quad (1.20)$$

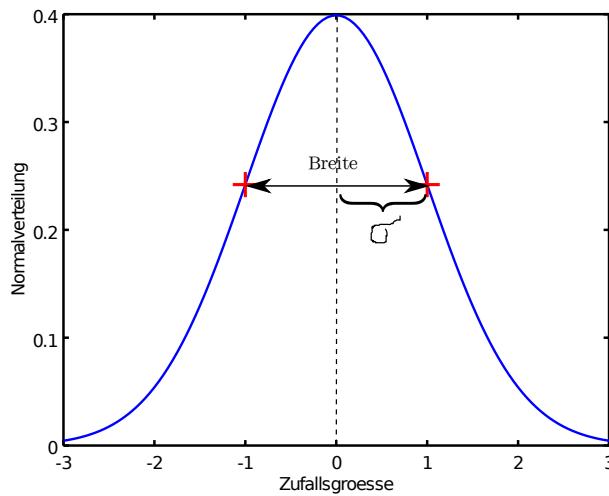


Abbildung 1.3: Breite der Normalverteilung.

Diskretisiert entspricht dies

$$\frac{1}{\sqrt{2\pi}\sigma_\varepsilon} e^{-\frac{1}{2}\left(\frac{\varepsilon}{\sigma_\varepsilon}\right)^2} d\varepsilon \quad (1.21)$$

einer relativen Häufigkeit $\frac{n_k}{J-1}$ in der k -ten Klasse der Breite $d\varepsilon$. Wenn wir nun kein Histogramm mit gleichgroßen Klassenbreiten bilden, sondern zu jedem ε_i eine Klasse mit der relativen Häufigkeit $\frac{1}{J-1}$ betrachten, dann haben wir so ungefähr für das σ_ε^2

$$\sum_{i=1}^J \varepsilon_i^2 \frac{1}{J-1} \approx \sigma_\varepsilon^2. \quad (1.22)$$

Wir nennen diese Breite wegen der Näherung nicht σ_ε^2 . Diese Näherung ist nur ein ungefährer Schätzwert und wird **empirische Varianz** genannt:

$$s_\varepsilon^2 = \frac{1}{J-1} \sum_{i=1}^J \varepsilon_i^2 \quad (1.23)$$

Wir unterscheiden zwischen der Varianz und der empirischen Varianz durch die Verwendung der unterschiedlichen Formelzeichen σ und s .

Uns interessiert aber nicht so sehr die Varianz der Residuen ε , die direkt der Varianz der Spannungsmessung entspricht, als viel mehr die Varianz des Modellparameters R . Betrachte die Modellgleichung $R = \frac{U}{I}$, wie empfindlich reagiert die Größe R auf Veränderung der Größe U . Wählt man anstelle eines Wertes U einen Wert $U + \Delta U$, so reagiert die Größe R über die Modellgleichung wie folgt

$$R + \Delta R = \frac{1}{I} (U + \Delta U) \quad (1.24)$$

d.i. mit $R = \frac{U}{I}$

$$\Delta R = \frac{1}{I} (\Delta U) \quad (1.25)$$

Der Differenzenquotient ist also

$$\frac{\Delta R}{\Delta U} = \frac{1}{I} \quad (1.26)$$

was bei diesem linearen Zusammenhang direkt hinkommt. Allgemein passt es für stetige Modelle auch bei nichtlinearem Zusammenhang für genügend kleine Änderungen. Deshalb wird die Empfindlichkeit, die **Sensitivität**, definiert als die lokale Steigung

$$c = \frac{\partial R}{\partial U} \quad (1.27)$$

Die Varianz ist eine quadratische Veränderung, so dass

$$s_R^2 = \left(\frac{\partial R}{\partial U} \right)^2 s_\varepsilon^2 = \frac{1}{I^2} s_\varepsilon^2 \quad (1.28)$$

Hier hätte ich auch direkt s_R aufschreiben können, aber die quadratische Schreibweise wird erforderlich, wenn es um mehrere Größen geht. In den folgenden Vorlesungen werden wir das Konzept der Fortpflanzung der Varianzen kennen lernen.

Die Varianzen sind ein Maß für die Unsicherheit von Messungen. Die **Messunsicherheit** ist zentrales Thema in der Messtechnik. Im Wörterbuch der Metrologie (VIM)

https://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf

wird sie wie folgt definiert:

Messunsicherheit; Unsicherheit (engl. *measurement uncertainty; uncertainty*): nicht-negativer Parameter, der die Streuung der Werte kennzeichnet, die der Messgröße auf der Grundlage der benutzten Information beigeordnet ist. [VIM2.26]

Sie drückt den Mangel einer genauen Kenntnis des Wertes der Messgröße aus. Selbst wenn systematische Effekte korrigiert werden, bleibt der **Messwert** lediglich ein **Schätzwert der Messgröße**, siehe auch GUM:2008, Abschnitt 3.3.1 .

Im Bereich der Metrologie wird zum Thema Messunsicherheitsberechnung eine Richtlinie verwendet, um die Auswerteverfahren im gesetzlichen Messwesen zu vereinheitlichen und die Resultate besser vergleichbar zu machen. Die Richtlinie wurde von der *Working Group 1* des *Joint Committee for Guides in Metrology* (JCGM/WG 1) entwickelt. Sie hat den Titel *Evaluation of measurement data - Guide to the expression of uncertainty in measurement* und wird mit GUM abgekürzt. Die derzeit gültige Version ist *JCGM 100:2008 GUM 1995 with minor corrections* und ist in englischer Fassung auf der Webseite des *Bureau international des poids et mesures* (BIPM) unter folgendem Link frei erhältlich:

https://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf

In vielen Fällen reicht es, die Varianzen der einzelnen direkten Messgrößen zu untersuchen sowie den Zusammenhang zwischen der Variation einer direkten Messgröße und die dadurch verursachte Veränderung der indirekten, also die Empfindlichkeit/Sensitivität. Dieser Zusammenhang ist zentrales Thema der Bestimmung von Messunsicherheiten und wird im Laufe dieser Vorlesung eingehend beleuchtet. Vielfach gibt es nicht nur einen Zusammenhang zwischen einer direkten mit einer indirekten Messgröße, sondern auch einen Zusammenhang zwischen den direkten Messgrößen untereinander.

In dieser Vorlesungsreihe liegt das Hauptaugenmerk auf linearen Zusammenhängen zwischen Messgrößen. Für viele Anwendungen greifen lineare Ansätze. Wenn die betrachteten Bereiche der Variation von Größen genügend klein sind, so lassen sich auch nichtlineare Modelle durch eine Linearisierung in einer gewissen Umgebung approximieren.

Das Maß für den linearen Zusammenhang wird **Korrelation** genannt. Bei der linearen Regression geht es um Modelle mit indirekten Messgrößen, die als Koeffizienten eines Polynoms auftreten. Um den Grad des Polynoms zu ermitteln, wird untersucht mit welchem Potenzgesetz ein Paar direkter Messgrößen zu verknüpfen ist (welchen Grad das Polynom hat), so dass die Korrelation groß ist.

1.3 Ko-Varianzen von Zufallsgrößen

Bei dem Beispiel mit dem Ohm'schen Gesetz hatten wir uns mit der Modellgleichung $U = RI$ befasst. Dabei waren U und I die beiden physikalischen Größen, die proportional unterschiedliche Werte annehmen. Sie sind direkt über einen linearen Zusammenhang verknüpft. Der lineare Zusammenhang wird durch eine Gerade, ein Polynom vom Grad 1, dargestellt.

Für die Charakterisierung linearer Zusammenhänge ist der **Korrelationskoeffizient** ρ ein Maß, mit dem quantifiziert wird, wie stark zwei Größen linear miteinander verknüpft sind. Der Korrelationskoeffizient ist so definiert, dass er Werte zwischen -1 und 1 annehmen kann. Gibt es einen direkten linearen Zusammenhang zwischen zwei Zufallsgrößen X_1 und X_2 , so wird man zu einer Beobachtung der einen Größe mit kleinem Wert auch einen kleinen Wert bei der anderen Größe beobachten und wenn die eine Größe einen großen Wert annimmt, so wird die andere auch einen großen Wert annehmen, und ρ wird in der Nähe von 1 liegen. Haben zwei Zufallsgrößen einen ebenso direkten Zusammenhang, aber in umgekehrter Weise, dass zu kleinen Werten von der ersten Größe große Werte zur zweiten Größe auftreten und umgekehrt, so liegt ρ in der Nähe von -1 . In beiden Fällen sagt man, dass die Größen **korreliert** seien.

Gibt es überhaupt gar keinen Zusammenhang zwischen dem, was an Werten zu der einen Größe beobachtet wird mit dem, was an Werten zur anderen Größe beobachtet wird, so ist $\rho = 0$ und man sagt: Die beiden Größen seien **unkorreliert**.

Die Erwartungswerte zu jeder der Größen X_1 und X_2 sind jeweils $E(X_1)$ und $E(X_2)$. Sie werden aus den Mittelwerten der jeweiligen Stichproben geschätzt

$$\bar{x}_1 = \frac{1}{J} \sum_{j=1}^J X_{1,j} \quad \bar{x}_2 = \frac{1}{J} \sum_{j=1}^J X_{2,j} \quad (1.29)$$

Dann ist

$$\bar{\mathbf{X}} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix} \quad (1.30)$$

der Schwerpunkt der Beobachtungen der Größen. Sind die Entfernungen der Beobachtungspärchen $(X_{1,j}, X_{2,j})$ vom Schwerpunkt so, dass die Entfernung der Größe X_1 in die gleiche Richtung weist wie die Entfernung der Größe X_2 , so liefern die Produkte $(X_{1,j} - \bar{x}_1)(X_{2,j} - \bar{x}_2)$ gleiche Vorzeichen. Streuen diese aber unzusammenhängend, so mitteln sich die unterschiedlichen Terme $(X_{1,j} - \bar{x}_1)(X_{2,j} - \bar{x}_2)$

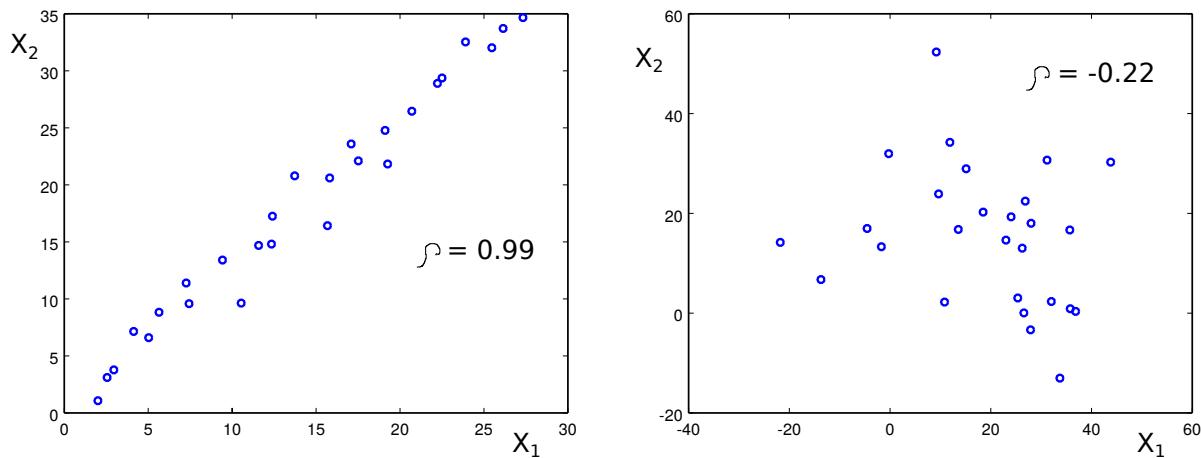


Abbildung 1.4: Wenn es einen linearen Zusammenhang zwischen zwei Zufallsgrößen X_1 und X_2 gibt, so sagt man, dass sie korreliert sind und der Korrelationskoeffizient ρ liegt bei Eins (links) oder bei minus Eins. Wenn es keinen Zusammenhang zwischen zwei Zufallsgrößen X_1 und X_2 gibt, so sagt man, dass sie unkorreliert sind und der Korrelationskoeffizient ρ liegt bei Null (rechts).

bei Summation weg.

Wir definieren die Größe

$$s_{1,2} := \frac{1}{J-1} \sum_{j=1}^J (X_{1,j} - \bar{x}_1)(X_{2,j} - \bar{x}_2) \quad (1.31)$$

und nennen sie die **empirische Kovarianz**. Allgemein ist die **Kovarianz** für Zufallsgrößen, deren gemeinsame Verteilung ihre Wahrscheinlichkeitsdichte $p(X_1, X_2)$ sei, wie folgt definiert

$$\text{Cov}(X_1, X_2) := \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (X'_1 - \text{E}(X_1))(X'_2 - \text{E}(X_2)) p(X'_1, X'_2) dX'_1 dX'_2. \quad (1.32)$$

Wir sehen hier, dass die Kovarianz einer Zufallsgröße mit sich selbst die **Varianz** ist

$$\text{Var}(X_1) := \text{Cov}(X_1, X_1) = \int_{-\infty}^{\infty} (X'_1 - \text{E}(X_1))^2 p(X'_1) dX'_1 \quad (1.33)$$

und für die empirische Varianz und Kovarianz gilt entsprechend

$$s_1^2 := s_{1,1} = \frac{1}{J-1} \sum_{j=1}^J (X_{1,j} - \bar{x}_1)^2. \quad (1.34)$$

Wir werden im nächsten Kapitel sehen, dass bei der Steigung der Regressionsgraden (Aufgabe 1) genau so ein Term $\sum_{j=1}^J (X_{1,j} - \bar{x}_1)(X_{2,j} - \bar{x}_2)$ im Zähler steht.

Für standardnormalverteilte Zufallsgrößen Z_i mit $Z \sim \mathcal{N}(0, 1)$ und $i = 1, 2$ in die sich die Größen X_i wie folgt umrechnen lassen

$$Z_i = \frac{X_i - \text{E}(X_i)}{\sqrt{\text{Var}(X_i)}} \quad (1.35)$$

ist die Kovarianz gleich dem Korrelationskoeffizienten

$$\rho_{i,k} = \text{Cov}(Z_i, Z_k) \quad (1.36)$$

was äquivalent ist zu

$$\rho_{i,k} = \frac{\text{Cov}(X_i, X_k)}{\sqrt{\text{Var}(X_i)} \sqrt{\text{Var}(X_k)}}. \quad (1.37)$$

Für das Verständnis der Gesamtzusammenhänge ist der Sprachgebrauch für Zufallsgrößen sowie der Sprachgebrauch des Erwartungswertes von Zufallgrößen von Bedeutung.

Verknüpfungen von Zufallsgrößen sind ihrerseits wieder Zufallsgrößen. Betrachten wir eine Größe X , die ganz abstrakt und allgemein eine Zufallsgröße ist, und $p(X)$ die Wahrscheinlichkeitsdichteverteilung dazu, so heißt

$$E(X) := \int_{-\infty}^{\infty} X' p(X') dX' \quad (1.38)$$

Erwartungswert der Zufallsgröße X .

Zur Berechnung der Varianz einer indirekten Messgröße, die linear abhängig ist von direkten Messgrößen, benötigen wir die algebraischen Grundregeln für Erwartungswerte: Addition, Multiplikation und Skalierung.

- Summe: $X = X_1 + X_2$

$$E(X_1 + X_2) := \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (X'_1 + X'_2) p(X'_1, X'_2) dX'_1 dX'_2 \quad (1.39)$$

d.h.

$$\begin{aligned} E(X_1 + X_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X'_1 p(X'_1, X'_2) dX'_1 dX'_2 \\ &\quad + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X'_2 p(X'_1, X'_2) dX'_1 dX'_2 \end{aligned} \quad (1.40)$$

und mit der Randverteilung (engl. *marginal distribution*)

$$p(X_2) = \int_{-\infty}^{\infty} p(X'_1, X_2) dX'_1 \quad (1.41)$$

erhalten wir

$$E(X_1 + X_2) = \int_{-\infty}^{\infty} X'_1 p(X'_1) dX'_1 + \int_{-\infty}^{\infty} X'_2 p(X'_2) dX'_2 \quad (1.42)$$

also

$$E(X_1 + X_2) = E(X_1) + E(X_2). \quad (1.43)$$

In Worten heißt dies: Der Erwartungswert einer Summe von Zufallsgrößen ist die Summe der Erwartungswerte der Zufallsgrößen.

- Produkt: $X = X_1 \cdot X_2$

$$\mathbb{E}(X_1 \cdot X_2) := \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X'_1 X'_2 p(X'_1, X'_2) dX'_2 dX'_1. \quad (1.44)$$

Die Kovarianz ist der Erwartungswert folgenden Produktes $(X_1 - \mathbb{E}(X_1)) \cdot (X_2 - \mathbb{E}(X_2))$

$$\mathbb{E}((X_1 - \mathbb{E}(X_1)) \cdot (X_2 - \mathbb{E}(X_2))) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (X'_1 - \mathbb{E}(X_1)) \cdot (X'_2 - \mathbb{E}(X_2)) p(X'_1, X'_2) dX'_1 dX'_2 \quad (1.45)$$

also

$$\text{Cov}(X_1, X_2) = \mathbb{E}((X_1 - \mathbb{E}(X_1)) \cdot (X_2 - \mathbb{E}(X_2))). \quad (1.46)$$

und auch

$$\begin{aligned} \text{Cov}(X_1, X_2) &= \mathbb{E}(X_1 X_2 - X_1 \mathbb{E}(X_2) - X_2 \mathbb{E}(X_1) + \mathbb{E}(X_1) \mathbb{E}(X_2)) \\ &= \mathbb{E}(X_1 X_2) - \mathbb{E}(X_2) \mathbb{E}(X_1) - \mathbb{E}(X_1) \mathbb{E}(X_2) + \mathbb{E}(X_1) \mathbb{E}(X_2) \\ &= \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1) \mathbb{E}(X_2). \end{aligned} \quad (1.47)$$

Per Definitionem gilt

$$\text{Var}(X) := \text{Cov}(X, X) \quad (1.48)$$

- Skalierung: Zufallsgröße X mit einem konstanten, festen reellen Faktor b multiplizieren

$$\begin{aligned} \text{Var}(bX) &= \mathbb{E}((bX)^2) = \int_{-\infty}^{\infty} (bX')^2 p(X') dX' \\ &= \int_{-\infty}^{\infty} b^2 (X')^2 p(X') dX' \\ &= b^2 \underbrace{\int_{-\infty}^{\infty} (X')^2 p(X') dX'}_{\text{Var}(X)} \end{aligned} \quad (1.49)$$

$$\text{Var}(bX) := b^2 \text{Var}(X) \quad (1.50)$$

Wenn wir die Varianz einer indirekten Messgröße, die linear von mehreren direkten Messgrößen abhängt, ermitteln wollen, so entspricht dies der Bestimmung der **Varianz einer Summe von Zufallsgrößen**. Sie ist gleich der Summe der paarweisen Kovarianzen aller Zufallsgrößen (hier der direkten Messgrößen)

$$\text{Var}\left(\sum_{i=1}^N c_i X_i\right) = \sum_{i,j=1}^N \text{Cov}(c_i X_i, c_j X_j) \quad (1.51)$$

Wir leiten jetzt her, weshalb Gl. (7.10) gilt: Die Varianz ist gemäß ihrer Definition

$$\text{Var}\left(\sum_{i=1}^N c_i X_i\right) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left[\left(\sum_{i=1}^N c_i X_i \right) - E\left(\sum_{i=1}^N c_i X_i\right) \right]^2 p(X_1, \dots, X_N) dX_1 \dots dX_N. \quad (1.52)$$

Da der Erwartungswert einer Summe von Zufallsgrößen gleich der Summe der Erwartungswerte jeder

einzelnen dieser Zufallsgrößen ist, siehe Gl. (1.43), gilt

$$\left(\sum_{i=1}^N X_i \right) - E\left(\sum_{i=1}^N X_i\right) = \left(\sum_{i=1}^N X_i \right) - \sum_{i=1}^N E(X_i)$$

so dass gilt

$$\text{Var}\left(\sum_{i=1}^N c_i X_i\right) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\sum_{i=1}^N (c_i X_i - E(c_i X_i)) \right]^2 p(X_1, \dots, X_N) dX_1 \dots dX_N \quad (1.53)$$

und mit $E(c_i X_i) = c_i E(X_i)$ also

$$\text{Var}\left(\sum_{i=1}^N c_i X_i\right) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\sum_{i=1}^N c_i (X_i - E(X_i)) \right]^2 p(X_1, \dots, X_N) dX_1 \dots dX_N. \quad (1.54)$$

Mit Anwendung des Assoziativgesetzes

$$\begin{aligned} \left[\sum_{i=1}^N c_i (X_i - E(X_i)) \right]^2 &= \left[\sum_{i=1}^N c_i (X_i - E(X_i)) \right] \left[\sum_{k=1}^N c_k (X_k - E(X_k)) \right] \\ &= \sum_{i=1}^N \sum_{k=1}^N c_i c_k (X_i - E(X_i)) (X_k - E(X_k)) \end{aligned}$$

gilt

$$\text{Var}\left(\sum_{i=1}^N c_i X_i\right) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \sum_{i=1}^N \sum_{k=1}^N c_i c_k (X_i - E(X_i)) (X_k - E(X_k)) p(X_1, \dots, X_N) dX_1 \dots dX_N. \quad (1.55)$$

Durch Berechnung der Marginalverteilungen und weil $\int p(X_j) dX_j = 1$ für alle j , die weder i noch k sind, erhalten wir

$$\text{Var}\left(\sum_{i=1}^N c_i X_i\right) = \sum_{i=1}^N \sum_{k=1}^N c_i c_k \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (X_i - E(X_i)) (X_k - E(X_k)) p(X_i, X_k) dX_i dX_k, \quad (1.56)$$

so dass

$$\begin{aligned} \text{Var}\left(\sum_{i=1}^N c_i X_i\right) &= \sum_{i=1}^N \sum_{k=1}^N \text{Cov}(c_i X_i, c_k X_k) \\ &= \sum_{i=1}^N c_i^2 \text{Var}(X_i) + \sum_{i,k=1, i \neq k}^N c_i c_k \text{Cov}(X_i, X_k) \\ &= \sum_{i=1}^N c_i^2 \text{Var}(X_i) + 2 \sum_{i=1}^{N-1} \sum_{k=i+1}^N c_i c_k \text{Cov}(X_i, X_k). \end{aligned} \quad (1.57)$$

Die ist das **Gesetz der Fortpflanzung von Messunsicherheiten**, mit dem wir uns im Laufe dieser Vorlesungsreihe eingehend befassen werden.

1.4 Beispiel zu Methode der kleinsten Quadrate zum Selbststudium

Zu Bestimmen ist ein Ohmscher Widerstand R sowie eine Offsetspannung U_0 bei gegebenen Werten einer Präzisionsstromquelle und eines Voltmeters:

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

Es wird angenommen, dass die Stromstärken ohne Streuung vorliegen (als Regressoren) und die Spannungen normalverteilt streuen (als Regressanden), so dass

$$\min_{R, U_0} \sum_{i=1}^J (U_i - R I_i - U_0)^2 \quad (1.58)$$

(a) Berechnen Sie den Korrelationskoeffizienten $\rho(I, U)$

(b) Lösen Sie das Gleichungssystem

$$\begin{pmatrix} \frac{\partial}{\partial R} \\ \frac{\partial}{\partial U_0} \end{pmatrix} \sum_{i=1}^J (U_i - R I_i - U_0)^2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (1.59)$$

(c) Berechnen Sie die aus der Lösung des Gleichungssystems aus (a) gewonnenen Schätzwerte für R und U_0 .

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

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

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

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

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

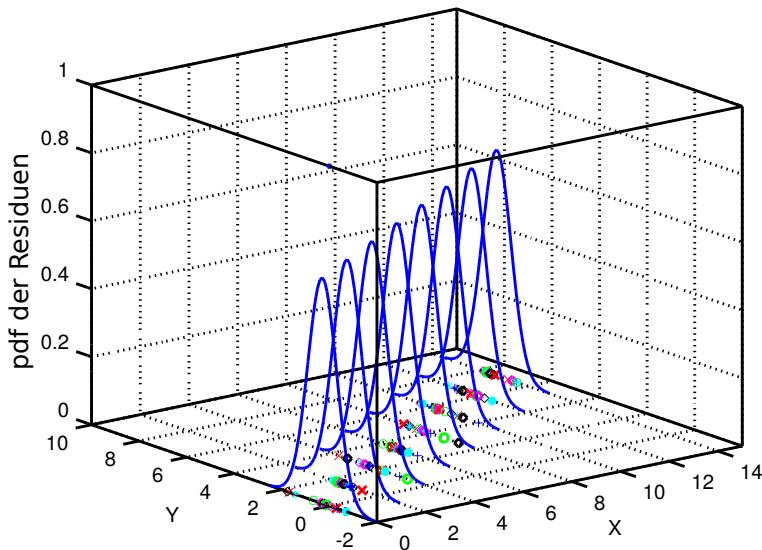


Abbildung 2.1: Lineare Regression für unabhängig und identisch verteilte Residuen.

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.

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

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

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

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)

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

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) $\hat{\theta}_k$:

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

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:

$$\begin{aligned} \left. \frac{\partial Q(\theta_0, \theta_1)}{\partial \theta_0} \right|_{\theta_0=\hat{\theta}_0, \theta_1=\hat{\theta}_1} &= \sum_{j=1}^J Y_j - J \hat{\theta}_0 - \hat{\theta}_1 \sum_{j=1}^J X_j = 0 \\ \left. \frac{\partial Q(\theta_0, \theta_1)}{\partial \theta_1} \right|_{\theta_0=\hat{\theta}_0, \theta_1=\hat{\theta}_1} &= \sum_{j=1}^J Y_j X_j - \hat{\theta}_0 \sum_{j=1}^J X_j - \hat{\theta}_1 \sum_{j=1}^J X_j^2 = 0 \end{aligned}$$

Mit den Mittelwerten \bar{X} , der sich zu $\sum_{j=1}^J X_j = J \bar{X}$ ergibt und dem Mittelwert \bar{Y} , der sich zu $\sum_{j=1}^J Y_j = J \bar{Y}$ ergibt, 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 $(\sum X_j^2) - J\bar{X}^2 = (\sum X_j^2) - 2J\bar{X}^2 + J\bar{X}^2 = \sum(X_j^2 - 2X_j\bar{X} + \bar{X}^2) = \sum(X_j - \bar{X})^2$

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

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 $\hat{\theta}_1$:

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

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)

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

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

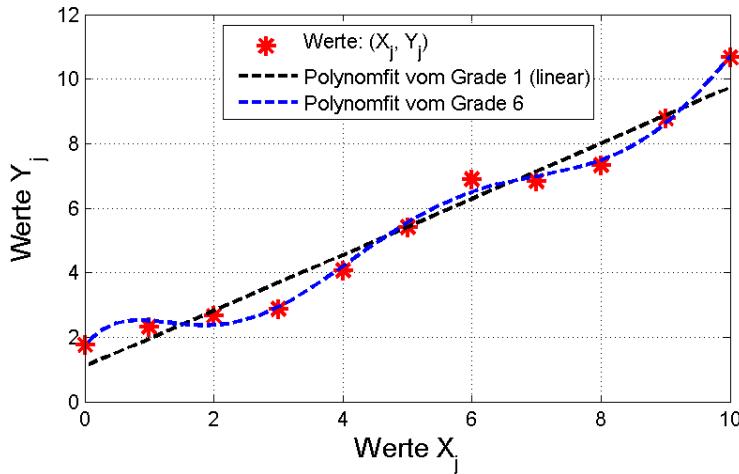


Abbildung 2.2: Beispiel für mögliche Fits, linearer Fit und Polynomfit vom Grade 6

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.

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

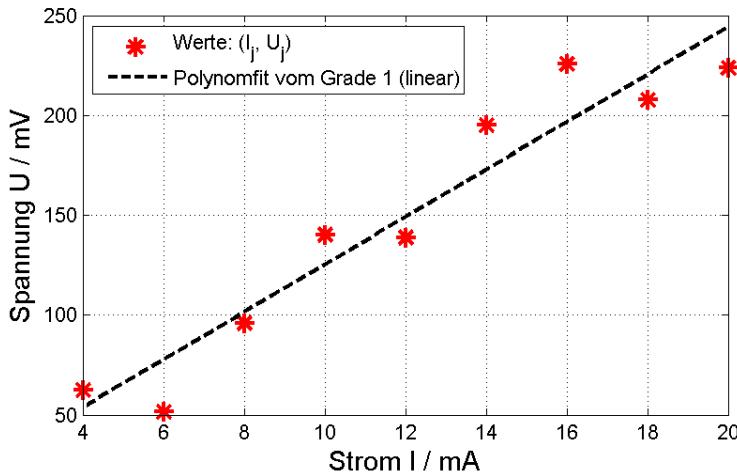


Abbildung 2.3: Beispiel: Bestimmung des Ohmschen Widerstand und einer Offset-Spannung durch lineare Regression

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

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

das heisst

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

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

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

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

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

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

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

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

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

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

$$\frac{\partial}{\partial \theta_l} \sum_{j=1}^J \varepsilon_j^2 = 2 \sum_{j=1}^J \varepsilon_j \frac{\partial}{\partial \theta_l} \varepsilon_j = 2 \sum_{j=1}^J \varepsilon_j (-1) \frac{\partial}{\partial \theta_l} (1 \theta_0 + X_{1,j} \theta_1) = 0 \quad (2.47)$$

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.

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

d.h.

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

das ist äquivalent zu

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

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

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

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:

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

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

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

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:

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

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

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

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.

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

das ist

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

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

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

mit

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

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

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

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

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

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

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

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

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

d.h.

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

d.h.

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

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 Paraboloide 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.

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

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

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

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

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

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

$$\hat{\Sigma}_{\boldsymbol{\theta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \hat{\sigma}_\varepsilon^2 \quad (2.82)$$

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.

$$\mathbf{e}_l = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (2.83)$$

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

$$\hat{\sigma}_l^2 = \mathbf{e}_l^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{e}_l \hat{\sigma}_\varepsilon^2. \quad (2.84)$$

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

$$\hat{\sigma}_l = \sqrt{\hat{\sigma}_l^2}. \quad (2.85)$$

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
$\rho_{XY} >$ bzw. < 0	gleich-/gegenläufige lineare Abhängigkeit

Für die Bestimmung des Bestimmtheitsmaßes oder des empirischen Korrelationskoeffizienten 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:

$$s_T^2 = \frac{1}{J-1} \sum_{j=1}^J (Y_j - \bar{Y})^2 \quad \text{mit } Y_j : \text{Messwert zu } X_j \quad (2.86)$$

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

$$s_M^2 = \frac{1}{J-1} \sum_{j=1}^n (\hat{\theta}_0 + \hat{\theta}_1 X_j - \bar{Y})^2 \quad (2.87)$$

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

$$s_R^2 = \frac{1}{J-1} \sum_{j=1}^J (Y_j - (\hat{\theta}_0 + \hat{\theta}_1 X_j))^2 \quad (2.88)$$

Man kann nun folgendes Bemerkenswerte zeigen:

$$s_T^2 = s_M^2 + s_R^2 \quad (2.89)$$

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:

$$\begin{aligned} \rho_{XY}^2 &= \frac{\left[\sum_{j=1}^J (X_j - \bar{X})(Y_j - \bar{Y}) \right]^2}{\sum_{j=1}^J (X_j - \bar{X})^2 \cdot \sum_{j=1}^J (Y_j - \bar{Y})^2} \\ &= \frac{s_{XY}^2}{s_X^2 \cdot s_Y^2} \quad \text{mit } 0 \leq \rho_{XY}^2 \leq 1 \end{aligned} \tag{2.90}$$

Den empirischen Korrelationskoeffizienten kann man somit wie folgt berechnen:

$$\rho_{XY} = \frac{s_{XY}}{s_X \cdot s_Y} \quad \text{mit } -1 \leq \rho_{XY} \leq 1 \tag{2.91}$$

2.7 Beispiele zur linearer Regression

2.7.1 Fit einer Geraden mittels linearer Regression

Gegeben sind 5 Messpunkte

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

- (a) 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.
- (b) 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?
- (c) 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. Die Regressormatrix \mathbf{X} lautet:

$$\mathbf{X} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ 1 & X_4 \\ 1 & X_5 \end{pmatrix} \quad (2.92)$$

Die Regressandenmatrix \mathbf{Y} lautet:

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{pmatrix} \quad (2.93)$$

Die gesuchten Modellparameter sind:

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}$$

Die optimalen Parameter $\hat{\boldsymbol{\theta}}$ ergeben sich durch Lösen des folgenden Gleichungssystems, siehe Gl.(2.79)

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

Der Matlab/Octave-Code lautet:

```

1 close all; clear all;
2 % Eingabe der Messwertepaare
3 input = [...
4 1, 0.4; ...
5 2, 0.55; ...
6 3, 0.7; ...
7 4, 0.75; ...
8 5, 0.8];
9 m = length(input); % Anzahl der Messpunkte
10 X = [ones(m,1), input(:,1)]; % Aufstellen der X-Matrix
11 Y = input(:,2); % Aufstellen der Y-Matrix
12 % degree of freedom
13 df = 5-2;
14 % Berechnung der Schätzwerte \thetaHat;
15 % y-Abschnitt = \thetaHat(1); Steigung = \thetaHat(2)
16 thetaHat = (X' * X) \ (X' * Y);
17 % Anzeige der Schätzwerte
18 disp(thetaHat);
19 % Plotten der Regressionsgeraden
20 xx = linspace(0,6,100);
21 yy = thetaHat(1) + thetaHat(2)*xx;
22 figure(10)
23 plot(xx, yy,'linewidth',1.5)
24 % Plotten der Datenwerte
25 hold on
26 plot(input(:,1), input(:,2), 'or','linewidth',2, 'MarkerSize',6)
27 hold off
28 grid on;
29 xlabel('Werte X_j','fontsize',14)

```

```

30 ylabel('Werte Y_j','fontsize',14)
31 set(gcf, 'PaperUnits', 'centimeters');
32 x_width=15 ;y_width= 9;
33 set(gcf, 'PaperPosition', [0 0 x_width y_width]);
34 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.

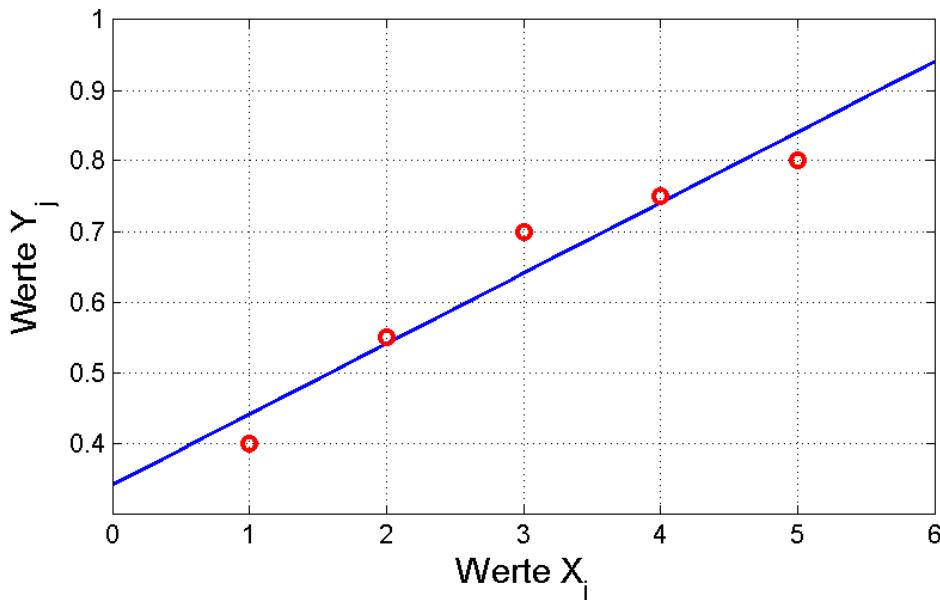


Abbildung 2.4: Lineare Regression. Als Schätzparameter erhält man $\hat{\theta} = [0.3400, 0.1000]^T$

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:

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

Die Varianz der Residuen ist wie folgt gegeben:

$$\text{Var}(\varepsilon) = \frac{1}{J-M} \cdot (\mathbf{Y} - \mathbf{X}\hat{\theta})^T (\mathbf{Y} - \mathbf{X}\hat{\theta}) \quad (2.94)$$

Die Kovarianzen der Schätzerparameter θ sind gegeben durch:

$$\text{Cov}(\theta) = (\mathbf{X}^T \mathbf{X})^{-1} \text{Var}(\varepsilon) \quad (2.95)$$

Matlab-Code zur Bestimmung des Qualitätsmaßes und der Varianz

```

1 % Qualitaetsmass Q
2 Q = (Y-X*thetaHat)' * (Y-X*thetaHat)
3 % Varianz der Residuen
4 Var_epsilon = 1/df * (Y-X*thetaHat)' * (Y-X*thetaHat)
5 % Wurzel aus Var_epsilon
6 RMSE = sqrt(Var_epsilon)
7 % Kovarianzen
8 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:

$$\begin{aligned}
\hat{\sigma}_{\theta_0}^2 &= s^2(\hat{\theta}_0, \hat{\theta}_1) \left[\frac{1}{J} + \frac{\bar{X}^2}{(J-1)s_x^2} \right] \\
&= 0.0023 \left[\frac{1}{5} + \frac{3^2}{(5-1) \cdot 1.5811^2} \right] \\
&= 0.00253
\end{aligned}$$

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

$$s_{\hat{\theta}_0} = \sqrt{\hat{\sigma}_{\theta_0}^2} = 0.0503$$

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

$$\begin{aligned}
\hat{\sigma}_{\theta_1}^2 &= \frac{\hat{s}^2(\hat{\theta}_0, \hat{\theta}_1)}{s_x^2 \cdot (J-1)} \\
&= \frac{0.0023}{1.5811^2 \cdot (5-1)} \\
&= 0.00023
\end{aligned} \quad (2.96)$$

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:

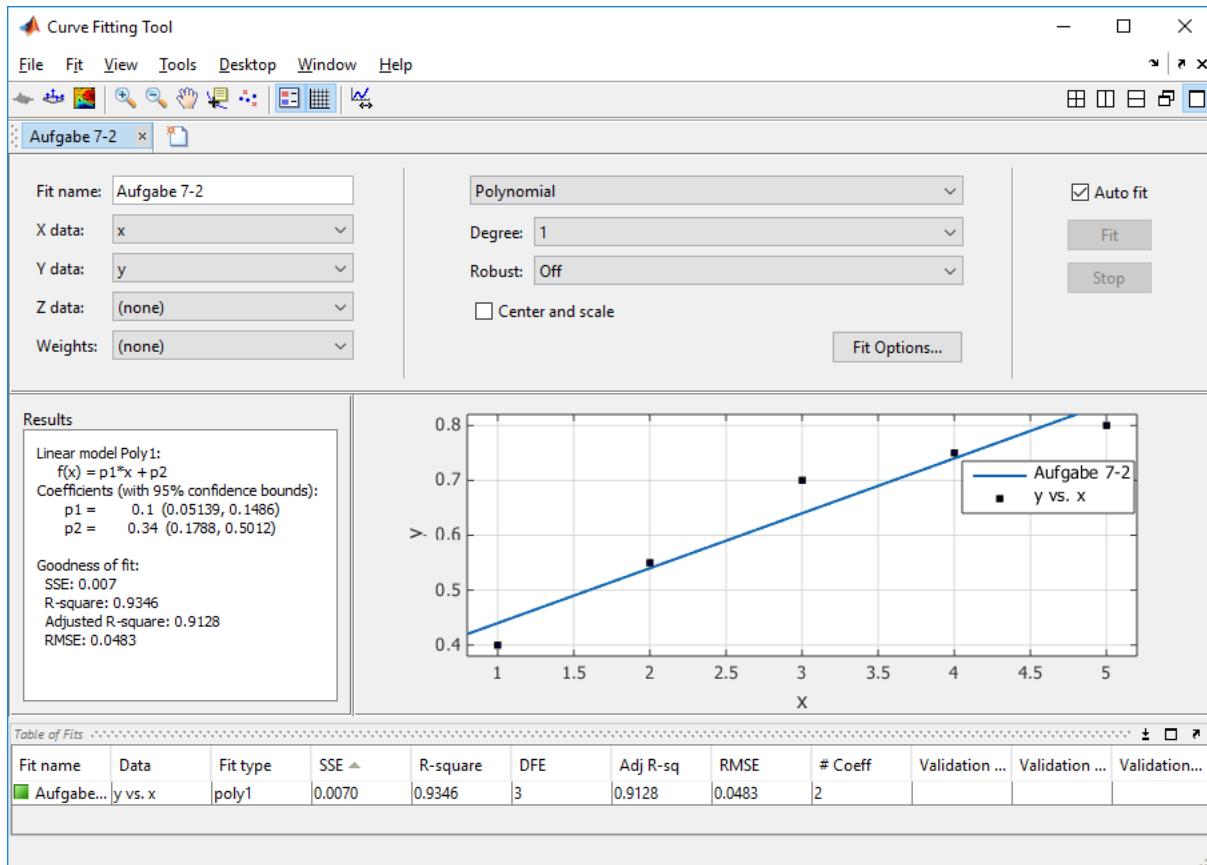


Abbildung 2.5: Lösung der Aufgabe mit dem polyfit Befehl von Matlab/Octave. Es wird hier u. a. der 95%ige Vertrauensbereich berechnet und angezeigt (siehe Mitte links in der Abbildung).

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:

$$\hat{\theta}_1 = 0.1000 \pm 0.0486 \quad \text{bzw.} \quad [0.0514; 0.1486] \quad (2.97)$$

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

$$\hat{\theta}_0 = 0.3400 \pm 0.1612 \quad \text{bzw.} \quad [0.1788; 0.5012] \quad (2.98)$$

2.7.2 Fit einer kubischen Funktion (Polynom 3. Ordnung) mittels linearer Regression

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

$$Y = \hat{\theta}_3 X^3 + \hat{\theta}_2 X^2 + \hat{\theta}_1 X^1 + \hat{\theta}_0 \quad (2.99)$$

Mit der Gl.(2.54) können wir das Beispiel mit dem Matrixformalismus auch lösen. Die Regressormatrix \mathbf{X} lautet:

$$\mathbf{X} = \begin{pmatrix} 1 & X_1 & X_1^2 & X_1^3 \\ 1 & X_2 & X_2^2 & X_2^3 \\ 1 & X_3 & X_3^2 & X_3^3 \\ 1 & X_4 & X_4^2 & X_4^3 \\ 1 & X_5 & X_5^2 & X_5^3 \end{pmatrix} \quad (2.100)$$

Die Regressandenmatrix \mathbf{Y} lautet:

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{pmatrix} \quad (2.101)$$

Die gesuchten Modellparameter sind:

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}$$

Die optimalen Parameter $\hat{\boldsymbol{\theta}}$ ergeben sich durch Lösen des folgenden Gleichungssystems, siehe Gl.(2.79)

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

Als Ergebnis für die Schätzparameter erhalten wir dann:

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

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 Kovarianzamtrix 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 (Klicke hier für interaktive Session):

```

1 close all; clear all;
2 % Eingabe der Messwertepaare
3 input = [...
4 1, 0.4; ...
5 2, 0.55; ...
6 3, 0.7; ...
7 4, 0.75; ...
8 5, 0.8];
9 m = length(input); % Anzahl der Messpunkte
10 x_Werte = input(:,1);
11 X = [ones(m,1), x_Werte, x_Werte.^2,x_Werte.^3];
12 Y = input(:,2);
13 % Aufstellen der X-Matrix % Aufstellen der Y-Matrix
14 % degree of freedom
15 df = 5-4;
16 % Berechnung der Schätzwerte \theta\hat{theta};
17 % y-Abschnitt = \theta\hat{theta}(1); Steigung = \theta\hat{theta}(2)
18 thetaHat = (X' * X) \ (X' * Y);
19 % Anzeige der Schätzwerte
20 disp(thetaHat);
21 % Plotten der Fitfunktion
22 xx = linspace(0,6,100);
23 yy = thetaHat(1) + thetaHat(2)*xx+ thetaHat(3)*xx.^2+thetaHat(4)*xx.^3;
24 figure(10)
25 plot(xx, yy,'linewidth',1.5)
26 % Plotten der Datenwerte
27 hold on
28 plot(input(:,1), input(:,2), 'or','linewidth',2, 'MarkerSize',6)
29 hold off
30 grid on;
31 xlabel('Werte X_j','fontsize',14)
32 ylabel('Werte Y_j','fontsize',14)
33 set(gcf, 'PaperUnits', 'centimeters');
34 x_width=15 ;y_width= 9;
```

```

35 set(gcf, 'PaperPosition', [0 0 x_width y_width]);
36 print '-dpng' kubischerFit.png
37 % Qualitaetsmaß Q
38 Q = (Y-X*thetaHat)' * (Y-X*thetaHat)
39 % Varianzen der Residuen
40 Var_epsilon = 1/df *(Y-X*thetaHat)' * (Y-X*thetaHat)
41 % Wurzel aus Var_epsilon
42 RMSE = sqrt(Var_epsilon)
43 % Kovarianzen
44 Cov = Var_epsilon * inv(X'*X)

```

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

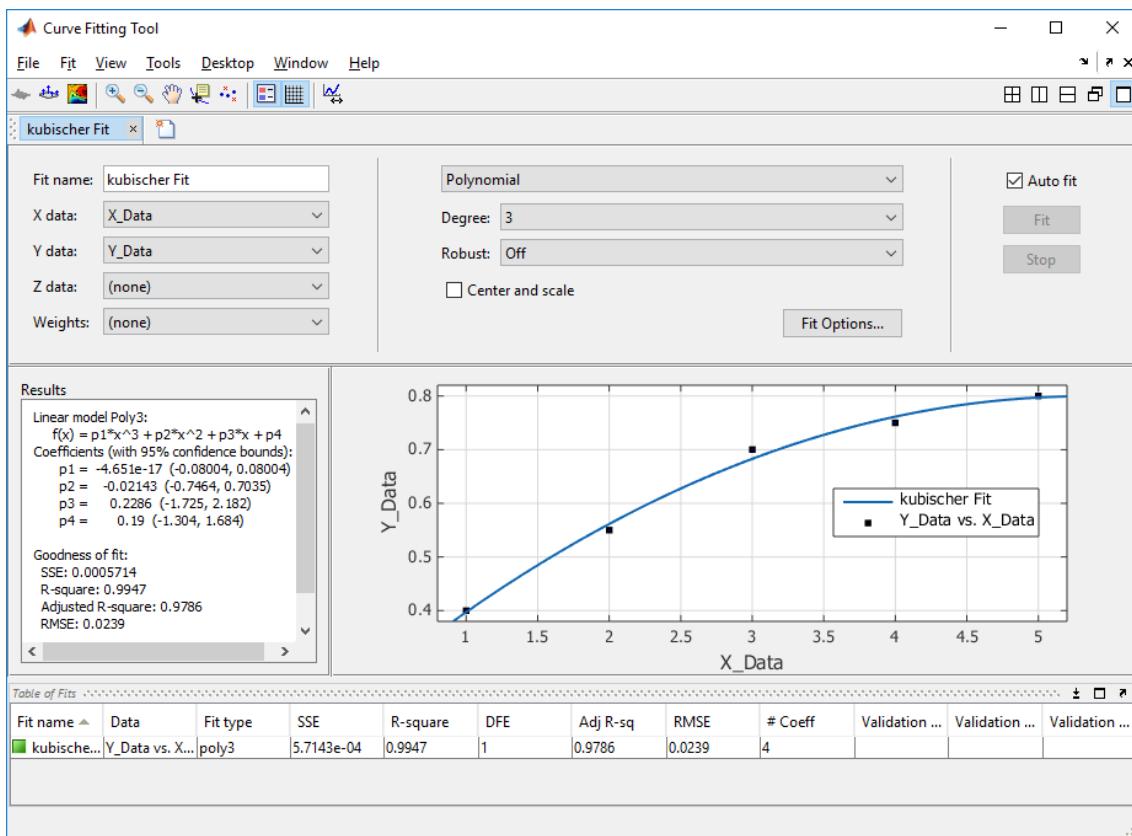
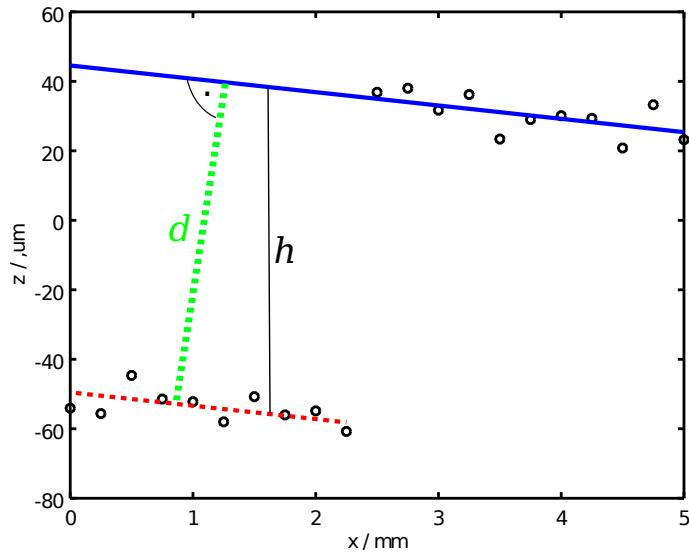


Abbildung 2.6: Kubischer Fit mit der Curve Fitting Toolbox von Matlab. Zum jedem geschätzten Parameter wird der 95%ige Vertrauensbereich angegeben. Zu sehen sind u.a. das Qualitätsmaß Q (engl. SSE: Sum squared error), das Bestimmtheitsmaß ρ_{XY}^2 (engl. R-square), Standardabweichung der Residuen (engl. 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 $100 \mu\text{m}$ 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

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

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 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.

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

Kapitel 3

Optimierung nichtlinearer Modelle

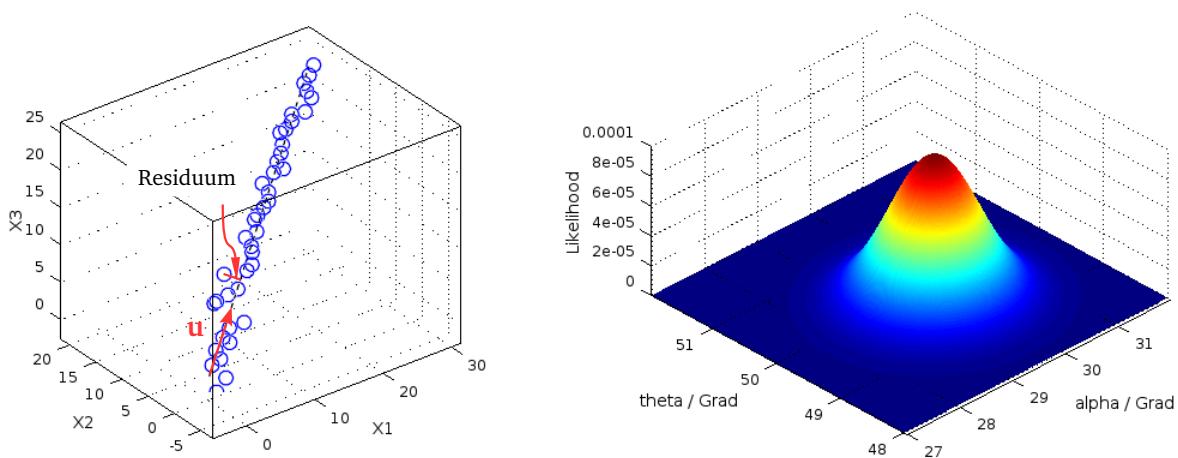


Abbildung 3.1: Modell mit 2 zu schätzenden Parametern (links) und der dazugehörigen Likelihood für normalverteilte Residuen (rechts).

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) $\mathbf{r}_{\text{Mess},j} = (X_{1,j}, X_{2,j}, X_{3,j})^T$ 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 $|$ 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 Minimumssuche.

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 genutzer 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 `fminsearch` in Gnu-Octave oder in Matlab 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:

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

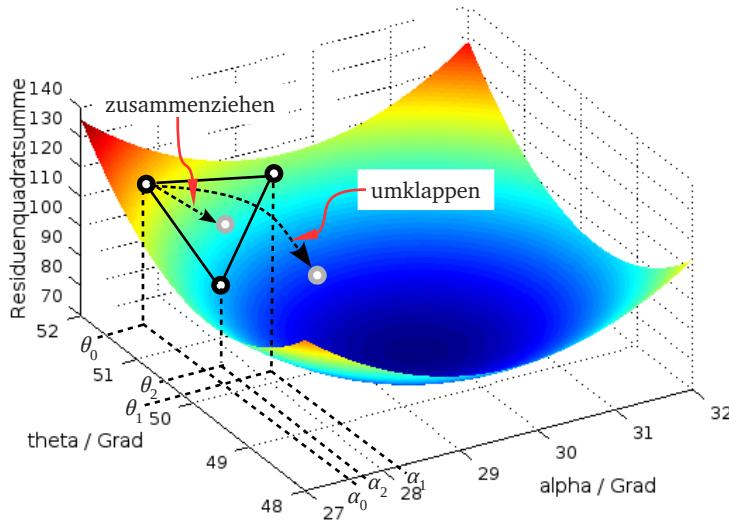


Abbildung 3.2: Residuenquadratsumme des Beispielmodells mit 2 zu schätzenden Parametern und Simplex für Optimierungsverfahren nach Nelder und Mead.

Die Strategie ist es, diesen Simplex durch Zusammenziehen oder durch Umlappen 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 Zielfunktionalen 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 waagerecht 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

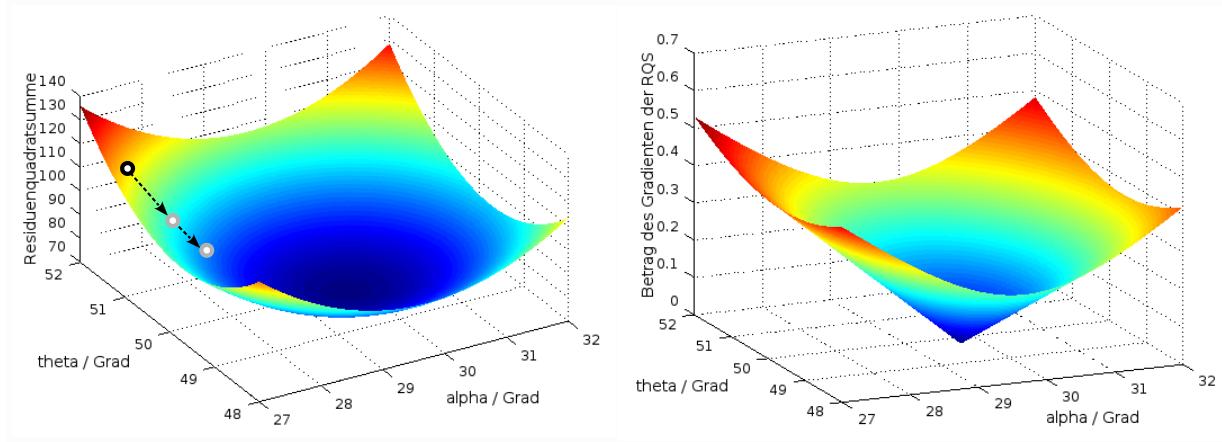


Abbildung 3.3: Residuenquadratsumme des Beispielmodells mit 2 zu schätzenden Parametern und Gradienten eingezeichnet als Pfeile (links) und Beträge der Gradienten als Funktion der 2 zu schätzenden Parameter (rechts).

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 schwerwiegenderen werden.

Für zu minimierende Zielfunktionale wie wir sie bei normalverteilten Residuen, also bei den runden, stetig differenzierbaren Kostenfunktionen als Residuenquadrate 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 Startwerttupels 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 Δp 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 Startwertevektors (α_0, θ_0) ist der Vektor der partiellen Ableitungen

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

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 RQS-Funktion befindet. 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 $\varepsilon(\mathbf{p}) = (\varepsilon_1(\mathbf{p}), \dots, \varepsilon_J(\mathbf{p}))^\top$, so dass die Kostenfunktion Q entsprechend Funktion des Parametervektors ist

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

für die wir das Minimum suchen mit

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

indem wir nach dem Nullstellenvektor der Gradientenfunktion

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

suchen, für den Q ein Minimum und nicht ein Maximum oder Sattelpunkt annimmt.

(Hinweis zu Gl. (3.13): $\frac{\partial}{\partial p} \varepsilon^2(p) = 2\varepsilon(p) \frac{\partial}{\partial p} \varepsilon(p)$)

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 Iterationschritten, 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 Zielfunktionalen 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:

$$\mathbf{J}^{(\kappa)} = \begin{pmatrix} \frac{\partial \varepsilon_1}{\partial P_1} \Big|_{\mathbf{p}_\kappa} & \cdots & \frac{\partial \varepsilon_1}{\partial P_M} \Big|_{\mathbf{p}_\kappa} \\ \ddots & & \\ \frac{\partial \varepsilon_J}{\partial P_1} \Big|_{\mathbf{p}_\kappa} & \cdots & \frac{\partial \varepsilon_J}{\partial P_M} \Big|_{\mathbf{p}_\kappa} \end{pmatrix}, \text{ bzw. } J_{jm}^{(\kappa)} = \frac{\partial \varepsilon_j}{\partial p_m} \Big|_{\mathbf{p}_\kappa} \text{ mit } j = 1 \dots J; m = 1 \dots M. \quad (3.17)$$

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 $\boldsymbol{\varepsilon}$ hier als Spaltenvektor der einzelnen Residuen.

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

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

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

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:

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

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

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

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

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

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:

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

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.

3.3 Beispiel zu Gradienten-Verfahren nach Gauß-Newton für den Fall gleicher Varianzen

Gegeben ist die Gl.(3.23) mit

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

zur Bestimmung der Modellparameter. Mit dieser Gleichung wird mit jeder Iteration die Änderung des Modellparameters $\Delta \mathbf{p}_\kappa$ bestimmt. Nach jeder Iteration κ erhält man einen neuen Modellparameter $\mathbf{p}_{\kappa+1}$ erhält.

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

Die Optimierung endet entweder wenn die Residuen kleiner sind als vorgegeben oder bei einer vorgegebenen Iterationsanzahl. Die Jakobimatrix J sind die partiellen Ableitungen (=Gradienten) nach den Modellparametern p :

$$(J)_{jl} = \frac{\partial \varepsilon_j(\boldsymbol{\theta}^{(\kappa)})}{\partial p_l}$$

Hierbei läuft der Index der Wertepaare von $j = 1, \dots, J$ und der Index der gesuchten Modellparameter von $l = 1, \dots, M$. (Weitere Infos zum Gauß-Newton Verfahren, siehe z.B. [Wiki02])

Die Gl. (3.23) kann in einen Programmiercode umgesetzt werden.

Algorithmus des Gauß-Newton-Verfahrens

Input: Startwert: \mathbf{p}_0 , Zielfunktional $F(p_\kappa) := \boldsymbol{\varepsilon}(\mathbf{p}^{(\kappa)})$, Partielle Ableitungen $F'(\mathbf{p}_\kappa) := \mathbf{J}$
 for $\kappa = 0, 1, \dots$

- i) Berechne $F(\theta_\kappa)$, $F'(\theta_\kappa)$

- ii) Bestimme den Korrekturvektor $s_\kappa := \Delta p^{(\kappa)}$ über $F'(x_k)^T F'(x_\kappa) s_\kappa = -F'(x_\kappa)^T F(x_\kappa)$
- iii) Setze $p_{\kappa+1} = p_\kappa + s_\kappa$

end

Eine einfache mögliche Realisierung in Octave-/Matlab ist nachfolgend dargestellt:

```
function theta = gauss_newton(F,DF,p0,maxit,tol)
% F: Zielfunktional,DF:Ableitung des Zielfunktionalen (Jakobimatrix)
% p0: Startwerte für die gesuchten Modellparameter
% maxit: Anzahl der maximalen Iterationen
% tol: Toleranz / Residuum
k=0;
p=p0;
s=-(DF(p0)'*DF(p0))\ (DF(p0)'*F(p0));
while norm(s)>tol && k<maxit
k=k+1;
p=p+s;
s=-(DF(p)'*DF(p))\ (DF(p)'*F(p));
end
end
```

Nichtlinearer Fit eines Kreises mit dem Gauß-Newton-Verfahren

Es soll im folgenden ein Kreisfit, d.h. nichtlinearer Fit mit dem Gauß-Newton-Verfahren durchgeführt werden. Gegeben ist die Punktwolke mit den Wertepaaren X_j, Y_j mit $j = 1, \dots, J$. Alle Punkte sollen näherungsweise auf einem Kreis liegen. Geht man von einem Kreis mit Mittelpunkt $M = (x_c, y_c)$ und einem Radius $r > 0$ aus, so müssten die Punkte die folgende Bedingung erfüllen:

$$\sqrt{(X_j - x_c)^2 + (Y_j - y_c)^2} \approx r; \quad j = 1, \dots, J \quad (3.25)$$

Die Aufgabe besteht also darin, den Mittelpunkt M und den Radius r so zu bestimmen, dass die Summe über die Quadrate der einzelnen Abstände minimiert wird. Die Residuen sind gegeben durch:

$$\epsilon = \sqrt{(X_j - x_c)^2 + (Y_j - y_c)^2} - r; \quad j = 1, \dots, J \quad (3.26)$$

Gesucht ist das Minimum der Summe der Residuenquadrate. Das Problem lässt sich dann wie folgt formulieren:

$$\|\boldsymbol{\varepsilon}\|^2 \rightarrow \min \quad \leftrightarrow \quad g(x_c, y_c, r) = Q := \sum_{j=1}^J \left(\sqrt{(X_j - x_c)^2 + (Y_j - y_c)^2} - r \right)^2 \rightarrow \min \quad (3.27)$$

Es besteht also **kein linearer** Zusammenhang zwischen den Parametern $\boldsymbol{\theta} = [x_c, y_c, r]$ und den Residuen. Zur Demonstration des Gauß-Newton Algorithmus nehmen wir zur Vereinfachung an, das wir nur den ersten Parameter, also den x-Wert des Mittelpunktes x_c schätzen. Die beiden anderen

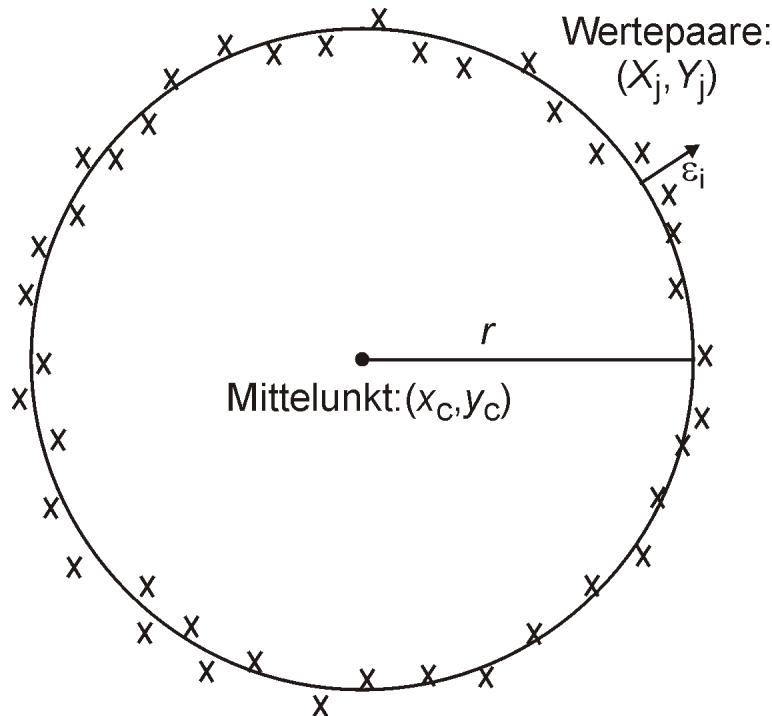


Abbildung 3.4: Skizze für den Fit eines Kreises

Parameter setzen wir zu $y_c = 0$ und $r = 1$. Dann besteht die Jakobi-Matrix nur aus einer partiellen Ableitung nach dem x-Wert des Mittelpunktes x_c . Wenn alle drei Parameter geschätzt werden sollen, müssten entsprechend die partiellen Ableitungen nach y_c und r mitverwendet werden.

Die Messpunkte erfüllen annähernd die Bedingung in Gl.(3.25):

$$\sqrt{(X_j - x_c)^2 + (Y_j - y_c)^2} \approx r; \quad j = 1, \dots, J$$

Die Zielfunktion Q lautet dann:

$$F(x_c) = Q(x_c) := \sum_{j=1}^J \left(\sqrt{(X_j - x_c)^2 + (Y_j - y_c)^2} - r \right)^2 \rightarrow \min \quad (3.28)$$

Die partielle Ableitung nach x_c lautet:

$$DF(x_c) = \frac{\partial Q(x_c)}{\partial x_c} = \sum_{j=1}^J \left[\frac{2(X_j - x_c) \cdot (\sqrt{(X_j - x_c)^2 + (Y_j - y_c)^2} - r)}{\sqrt{(X_j - x_c)^2 + (Y_j - y_c)^2}} \right] \quad (3.29)$$

Als Ergebnis erhält man den Halbkreisfit in Abb. 3.5 mit einem Schätzwert x_c in dem dargestellten Beispiel von -0.0067. Das Octave-/Matlab-Skript lautet:

```
% Anzahl der Messpunkte
M = 41;
Xj=linspace(-0.95,0.95,M);
r=1;
```

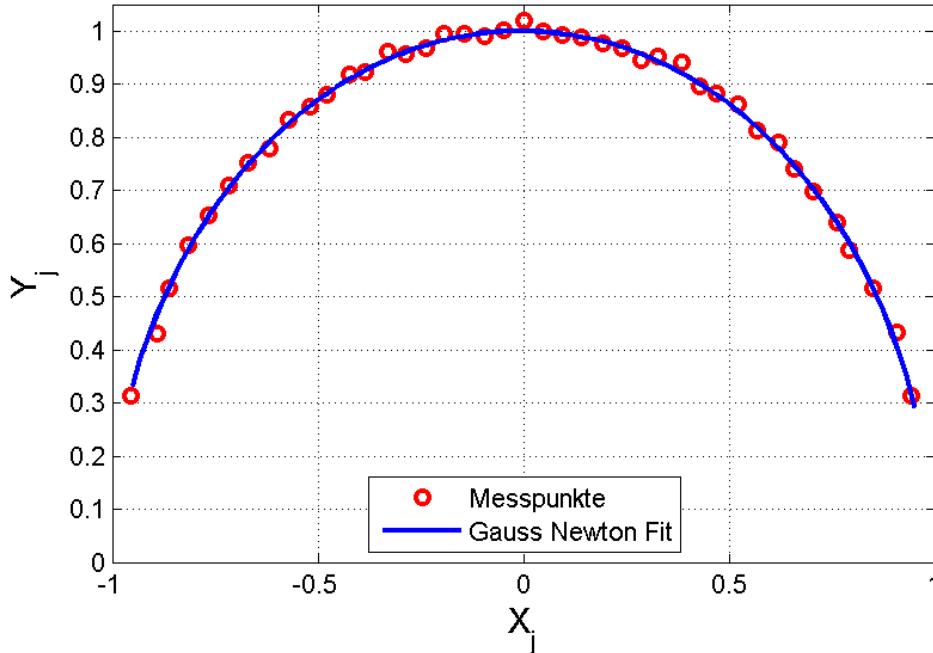


Abbildung 3.5: Ergebnis des Gauss Newton Fits

```

Yj = sqrt(r.^2-Xj.^2);

figure
plot(Xj,Yj,'*')

Xj_Noise = (1-0.01*randn(1,M)).*Xj;
Yj_Noise = (1-0.01*randn(1,M)).*Yj;

figure
plot(Xj_Noise,Yj_Noise,'ro','linewidth',2)
hold on
Xj=Xj_Noise;
Yj=Yj_Noise;
yc = 0;
r=1;
% Suche das Minimum der Funktion F(theta), die xc minimiert
F = @(theta) sum((sqrt((Xj-theta).^2+(Yj-yc).^2)-r).^2);

DF = @(theta) sum(2*(sqrt((Xj-theta).^2+(Yj-yc).^2)-r).*...
((1/2) * ((Xj-theta).^2 +(Yj-yc).^2).^(-0.5)) .*...
(2*(Xj-theta)*(-1)));
% Startparameter
theta0 = 0.1;
maxit = 1000; % Anzahl der Iterationen
tol = 1E-15; % Abruchbedingung

```

```
xc_estimate = gauss_newton(F,DF,theta0,maxit,tol);
disp(xc_estimate);

Xjfit=linspace(-0.95,0.95,101);
%
Yjfit = sqrt(r.^2-(Xjfit-xc_estimate).^2)+ yc;
plot(Xjfit,Yjfit,'-','linewidth',2)
grid on;
xlim([-1,1])
ylim([0,1.05])
legend('Messpunkte','Gauss Newton Fit','Location','South')
 xlabel('X_j','fontsize',14)
 ylabel('Y_j','fontsize',14)
set(gcf, 'PaperUnits', 'centimeters');
x_width=15 ;y_width= 10;
set(gcf, 'PaperPosition', [0 0 x_width y_width]);
print '-dpng' Kreisfit_Gauss_Newton.png
```

3.4 Levenberg-Marquardt-Verfahren

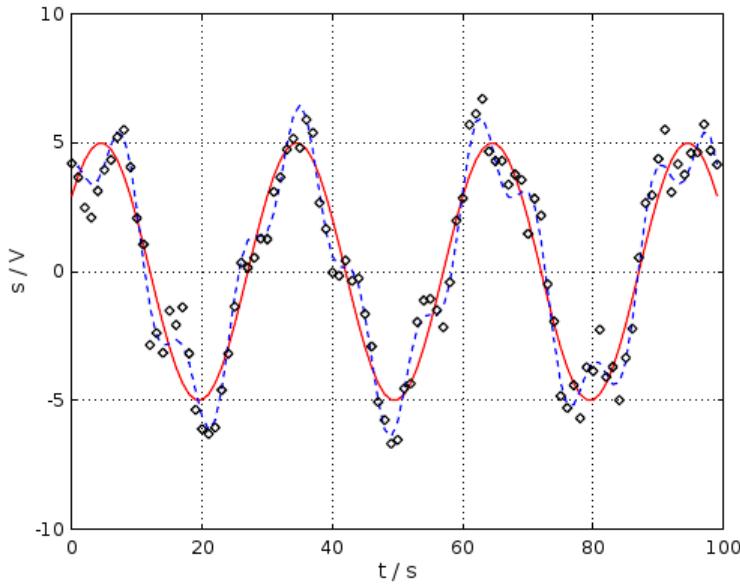


Abbildung 3.6: Beispiel eines sinusförmigen Signals.

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. 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 Beispieldesignals 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. 3.6 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. 3.7 zeigt die Summe der Quadrate der Residuen RQS als Kostenfunktion in drei verschiedenen Ansichten, die zeigen, dass es eine ausgeprägte Minimumsmulde 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 Maxima als auch die Minima die Nullstellen bilden.

Der Levenberg-Marquardt-Algorithmus ist ein Gradientenverfahren, bei dem das lineare Gleichungssys-

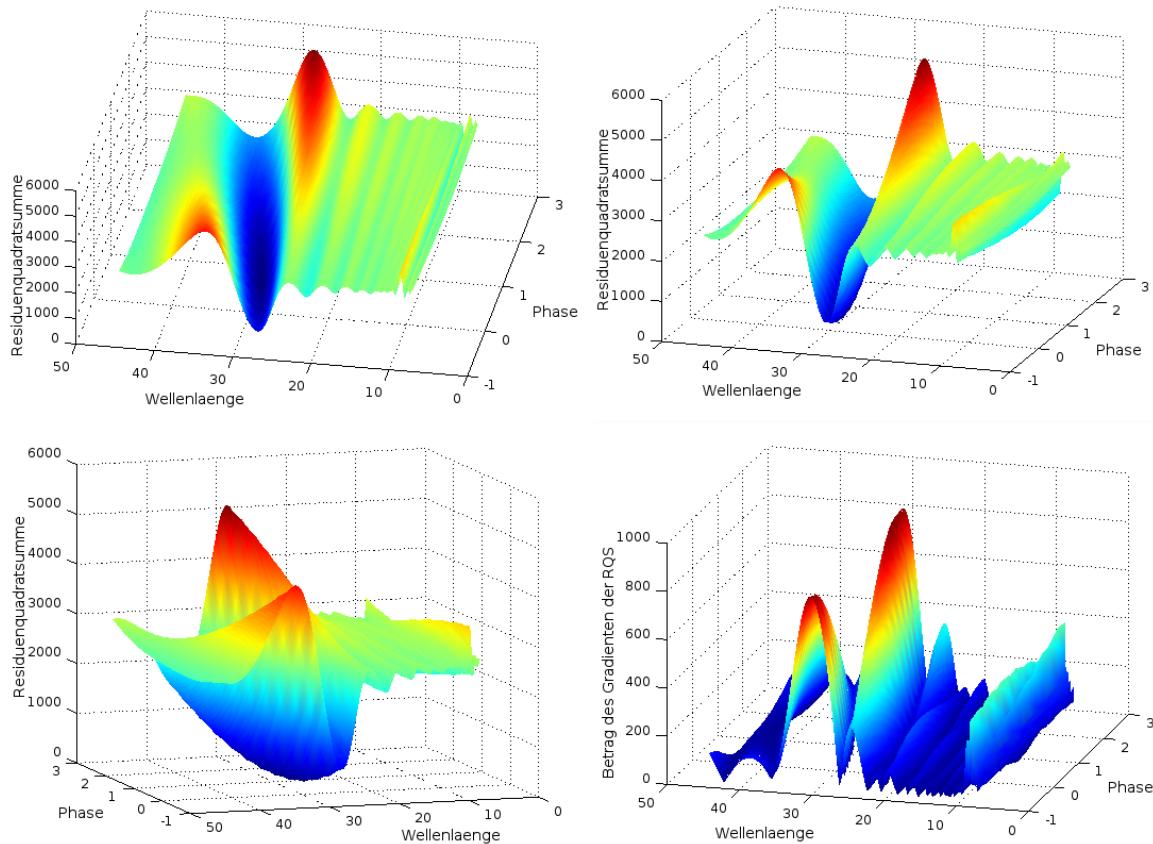


Abbildung 3.7: Residuenquadratsumme des Beispielmodells eines sinusförmigen Signals in drei verschiedenen Perspektiven (oben und unten links) und die Beträge der Gradienten als Funktion der 2 zu schätzenden Parameter (unten rechts).

tems (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:

$$Q(\mathbf{p}_\kappa + \alpha \cdot \Delta \mathbf{p}_\kappa) < Q(\mathbf{p}_\kappa) \quad (3.30)$$

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:

$$\mathbf{p}_{\kappa+1} = \mathbf{p}_\kappa + \alpha \cdot \Delta \mathbf{p}_\kappa \quad (3.31)$$

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 $\boldsymbol{\varepsilon}(\mathbf{p})$ (Gl. 3.19) auf

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

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

$$Q = \left(\boldsymbol{\varepsilon}(\mathbf{p}_\kappa) + \mathbf{J}^{(\kappa)} \Delta \mathbf{p} \right)^\top \left(\boldsymbol{\varepsilon}(\mathbf{p}_\kappa) + \mathbf{J}^{(\kappa)} \Delta \mathbf{p} \right). \quad (3.32)$$

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

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

Nun modifizieren wir das Gleichungssystem durch Hinzufügen eines Dämpfungsterms wie folgt

$$\left(\mathbf{J}^{(\kappa)\top} \mathbf{J}^{(\kappa)} + \mu \operatorname{diag} \left(\mathbf{J}^{(\kappa)\top} \mathbf{J}^{(\kappa)} \right) \right) \Delta \mathbf{p}_\kappa \stackrel{!}{=} -\mathbf{J}^{(\kappa)\top} \boldsymbol{\varepsilon}(\mathbf{p}_\kappa) \quad (3.33)$$

wobei $\operatorname{diag} \left(\mathbf{J}^{(\kappa)\top} \mathbf{J}^{(\kappa)} \right)$ 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

$$\rho_\mu := \frac{Q(\mathbf{p}_\kappa) - Q(\mathbf{p}_\kappa + \Delta\mathbf{p}_\kappa)}{Q(\mathbf{p}_\kappa) - (\boldsymbol{\varepsilon}(\mathbf{p}_\kappa) + \mathbf{J}^{(\kappa)}\Delta\mathbf{p}_\kappa)^\top (\boldsymbol{\varepsilon}(\mathbf{p}_\kappa) + \mathbf{J}^{(\kappa)}\Delta\mathbf{p}_\kappa)} =: \frac{\Delta R(\mathbf{p}_\kappa, \Delta\mathbf{p}_\kappa)}{\Delta \tilde{R}(\mathbf{p}_\kappa, \Delta\mathbf{p}_\kappa)}. \quad (3.34)$$

Die Prüfgröße ρ_μ ist das Verhältnis

- der Differenz ΔR des Zielfunktional 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 $\varepsilon(\mathbf{p}_\kappa)$, Partielle Ableitungen $J = \varepsilon'(\mathbf{p}_\kappa)$, Marquardt-Parameter μ , weitere Parameter $0 < \beta_0 < \beta_1 < 1$

for $k = 0, 1, \dots$

- i) Berechne $\varepsilon(\mathbf{p}_\kappa), F'(\mathbf{p}_\kappa)$
- 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 wird nun auf das zuvor beschriebene Beispiel des in Abb. 3.6 dargestellten sinusförmigen Signals mit einem Gnu-Octave/Matlab-Skript getestet. Ziel ist es, durch die gegebenen Spannungswerte (s_j, t_j) eine Sinusfunktion durchzulegen. Der Modellansatz für dieses Beispiel lautet:

$$s(t) = a \sin\left(2\pi \frac{t}{\lambda} + \varphi\right) + \varepsilon \quad \text{mit} \quad \varepsilon \sim \mathcal{N}(0, \sigma) \quad (3.35)$$

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

$$\varepsilon_j = Y_j - f(X_j, \mathbf{p}) \quad \text{mit} \quad j = 1, 2, \dots, J \quad (3.36)$$

vorliegen haben. \mathbf{p} sind die gesuchten Modellparameter. Die Kostenfunktion sieht damit wie folgt aus

$$Q(\mathbf{p}) = \sum_{j=1}^J (Y_j - f(X_j, \mathbf{p}))^2. \quad (3.37)$$

In dem vorliegenden Fall haben wir zwei Modellparameter $p(1) = \lambda$ und $p(2) = \varphi$

$$\mathbf{p} = \begin{pmatrix} p(1) \\ p(2) \end{pmatrix} = \begin{pmatrix} \lambda \\ \varphi \end{pmatrix} \quad (3.38)$$

Die Residuen sind wie folgt gegeben:

$$\varepsilon_j = s_j - a \cdot \sin\left[\frac{2\pi}{p(1)} t_j + p(2)\right] \quad (3.39)$$

Für die Jakobimatrix benötigen wir die beiden partiellen Ableitungen nach den Modellparametern $p(1)$ und $p(2)$:

$$\mathbf{J} = \begin{pmatrix} J_{11} \\ J_{21} \end{pmatrix} = \begin{pmatrix} \frac{\partial \varepsilon}{\partial p(1)} \\ \frac{\partial \varepsilon}{\partial p(2)} \end{pmatrix} \quad (3.40)$$

$$\begin{aligned} J_{11} = \frac{\partial \varphi(\mathbf{p})}{\partial p(1)} &= -a \cdot \cos \left[\frac{2\pi}{p(1)} \cdot t + p(2) \right] \cdot \left(\frac{-2\pi t}{p(1)^2} \right) \\ &= \frac{4a \cdot \pi}{p(1)^2} \cdot t \cdot \cos \left[\frac{2\pi}{p(1)} t + p(2) \right] \end{aligned} \quad (3.41)$$

$$J_{21} = \frac{\partial \varphi(\mathbf{p})}{\partial p(2)} = -a \cos \left[\frac{2\pi}{p(1)} t + p(2) \right] \quad (3.42)$$

Das Octave/Matlab-Skript sieht wir folgt aus:

```
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
%
```

siehe Gl.(3.41) und (3.42)

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

```

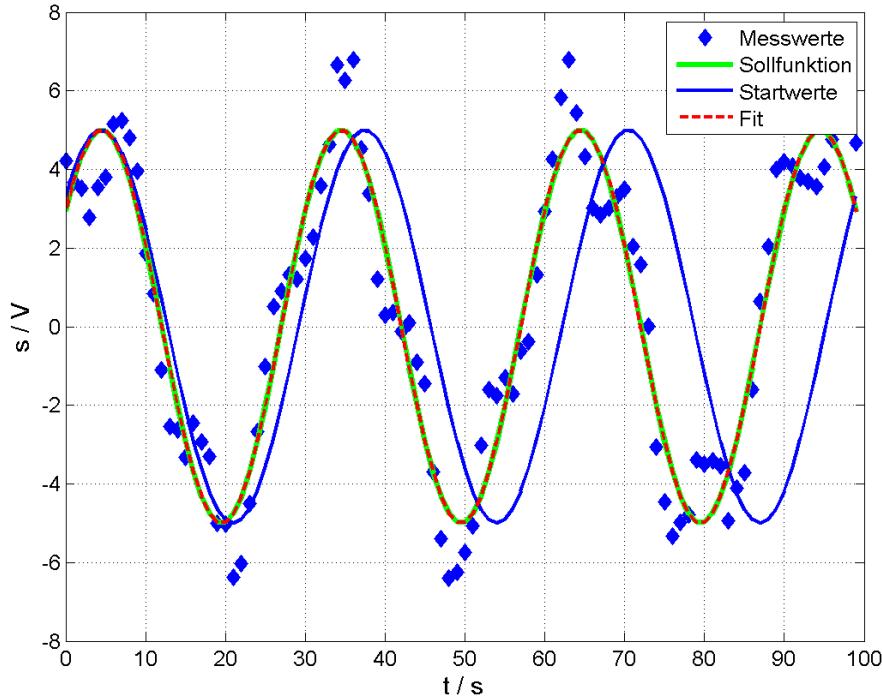


Abbildung 3.8: Beispiel eines sinusförmigen Signals mit Levenberg-Marquadt-Fit gemäß Gln. (3.33) und (3.34) zusammen mit den Entscheidungen aus Tabelle 1.

Dieses Gnu-Octave-/Matlab-Skript hat dazu das in Abb. 3.8 dargestellte Ergebnis mit den Werten

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

geliefert.

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
```

3.5 Matlabfunktionen zur Berechnung nichtlinearer Fits

Es gibt eine Vielzahl an Matlab-Bibliotheksfunktionen, die nichtlineare Optimierungsverfahren implementiert haben. Der Levenberg-Marquardt-Algorithmus ist in Matlab in der Funktion `lsqnonlin(fun,x0,lb,ub, options)` mit `options.Algorithm = 'levenberg-marquardt'` implementiert. Für unser Beispiel mit dem Sinusfit verwenden wir beispielhaft die Funktion `fminsearch(fun,x0,options)`, in der die Simplexmethode nach Nelder und Mead implementiert ist. Es muss ein Schätzwert für den Startwert x_0 angegeben werden. Als „OPTIONS“ können u.a. gewählt werden `TolX` (Toleranz, wann abgebrochen werden soll) oder `MaxIter` (Maximale Anzahl an Iterationen, die durchgeführt werden sollen). Die zu minimierende Kostenfunktion Q ist für unser Beispiel mit dem Sinusfit gegeben durch:

$$Q(\mathbf{p}) = \sum_{j=1}^J (a \cdot \sin(2 \cdot \pi \cdot t(j)/p(1) + p(2)) - s(j))^2 \rightarrow \text{Min.} \quad (3.43)$$

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
```

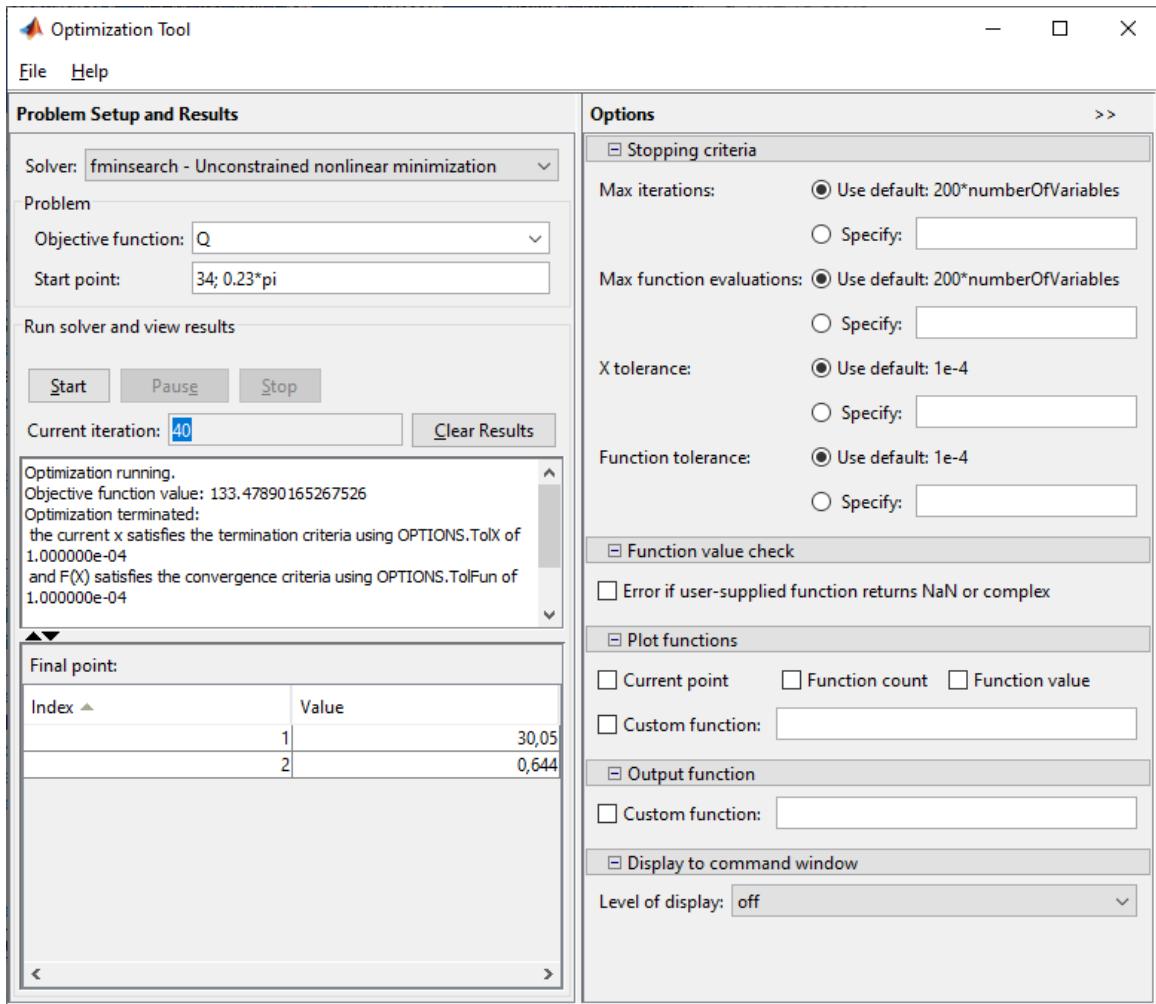


Abbildung 3.9: Optimtool von Matlab: Ergebnis $p(1) = 30,05$ und $p(2) = 0,644$

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. 3.9 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. Hinweis: Simplexmethode nach Nelder und Mead	fminunc oder fminsearch
Lineares Problem	Finde \mathbf{x} durch Minimierung von $\mathbf{f}^T \mathbf{x}$ mit $A\mathbf{x} \leq b$, $A_{eq}\mathbf{x} = \mathbf{b}_{eq}$, $\mathbf{l}_b \leq \mathbf{x} \leq \mathbf{u}_b$	linprog
Quadratisches Problem	Finde \mathbf{x} durch Minimierung von $\frac{1}{2}\mathbf{x}^T \mathbf{f}^T \mathbf{x}$ mit $A\mathbf{x} \leq \mathbf{b}$, $A_{eq}\mathbf{x} = \mathbf{b}_{eq}$, $\mathbf{l}_b \leq \mathbf{x} \leq \mathbf{u}_b$	quadprog
Minimierung von Funktionen von mehreren Variablen mit Nebenbedingungen	Finde \mathbf{x} durch Minimierung von $\frac{1}{2}\mathbf{x}^T \mathbf{f}^T \mathbf{x}$ mit $\mathbf{c}(\mathbf{x}) \leq \mathbf{0}$, $\mathbf{c}_{eq} = \mathbf{0}$ $A\mathbf{x} \leq \mathbf{b}$, $A_{eq}\mathbf{x} = \mathbf{b}_{eq}$, $\mathbf{l}_b \leq \mathbf{x} \leq \mathbf{u}_b$	fmincon
Minimax Problem	Finde das \mathbf{x} welches die Maxima von mehreren gegebenen Funktionen $F_i(\mathbf{x})$ minimiert, so dass $\mathbf{c}(\mathbf{x}) \leq \mathbf{0}$, $\mathbf{c}_{eq} = \mathbf{0}$ $A\mathbf{x} \leq \mathbf{b}$, $A_{eq}\mathbf{x} = \mathbf{b}_{eq}$, $\mathbf{l}_b \leq \mathbf{x} \leq \mathbf{u}_b$	fminimax

3.6 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 Summe der kleinsten Residuenquadrate weggekürzt. Sie lassen sich für den Fall ausklammern, bei dem die Streuung bezüglich des Modells für alle Messpunkte gleich ist, also

$$Q(\mathbf{p}) = \frac{\boldsymbol{\varepsilon}(\mathbf{p})^T \boldsymbol{\varepsilon}(\mathbf{p})}{\sigma^2(\mathbf{p})} \quad (3.44)$$

und

$$\nabla_{\mathbf{p}} Q(\mathbf{p}) = \frac{2}{\sigma^2(\mathbf{p})} \boldsymbol{\varepsilon}^T(\mathbf{p}) \mathbf{J} \stackrel{!}{=} \begin{pmatrix} 0 & \dots & 0 \end{pmatrix} \quad (3.45)$$

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

$$Q(\mathbf{p}) = \begin{pmatrix} \frac{\varepsilon_1(\mathbf{p})}{\sigma_1(\mathbf{p})} & \dots & \frac{\varepsilon_J(\mathbf{p})}{\sigma_J(\mathbf{p})} \end{pmatrix} \begin{pmatrix} \frac{\varepsilon_1(\mathbf{p})}{\sigma_1(\mathbf{p})} \\ \vdots \\ \frac{\varepsilon_J(\mathbf{p})}{\sigma_J(\mathbf{p})} \end{pmatrix} \quad (3.46)$$

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

$$Q(\mathbf{p}) = \boldsymbol{\varepsilon}^\top \begin{pmatrix} \frac{1}{\sigma_1^2} & \dots & 0 \\ & \ddots & \\ 0 & \dots & \frac{1}{\sigma_J^2} \end{pmatrix} \boldsymbol{\varepsilon} \quad (3.47)$$

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

$$Q(\mathbf{p}) = \boldsymbol{\varepsilon}^\top \mathbf{W} \boldsymbol{\varepsilon} \quad (3.48)$$

und mit

$$\frac{1}{2} \nabla_{\mathbf{p}} Q(\mathbf{p}) = \boldsymbol{\varepsilon}^\top(\mathbf{p}) \mathbf{W} \mathbf{J} \quad (3.49)$$

sieht unser Optimierungsproblem wie folgt aus

$$\lim_{\mathbf{p} \rightarrow \hat{\mathbf{p}}} \boldsymbol{\varepsilon}^\top(\mathbf{p}) \mathbf{W}(\mathbf{p}) \mathbf{J}(\mathbf{p}) = \begin{pmatrix} 0 & \dots & 0 \end{pmatrix}. \quad (3.50)$$

Die Gleichungen (3.23) und (3.33) werden damit zu

$$\mathbf{J}^{(\kappa)\top} \mathbf{W}(\mathbf{p}_\kappa) \mathbf{J}^{(\kappa)} \Delta \mathbf{p}_\kappa \stackrel{!}{=} -\mathbf{J}^{(\kappa)\top} \mathbf{W}(\mathbf{p}_\kappa) \boldsymbol{\varepsilon}(\mathbf{p}_\kappa). \quad (3.51)$$

und mit Dämpfungs- bzw. Marquardtparameter

$$\left(\mathbf{J}^{(\kappa)\top} \mathbf{W}(\mathbf{p}_\kappa) \mathbf{J}^{(\kappa)} + \mu \text{diag}(\mathbf{J}^{(\kappa)\top} \mathbf{W}(\mathbf{p}_\kappa) \mathbf{J}^{(\kappa)}) \right) \Delta \mathbf{p}_\kappa \stackrel{!}{=} -\mathbf{J}^{(\kappa)\top} \mathbf{W}(\mathbf{p}_\kappa) \boldsymbol{\varepsilon}(\mathbf{p}_\kappa). \quad (3.52)$$

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

$$\varepsilon_j = \begin{pmatrix} X_{1,j} \\ X_{2,j} \end{pmatrix} \cdot \begin{pmatrix} -\sin(\alpha) \\ \cos(\alpha) \end{pmatrix} - d. \quad (3.53)$$

Hier stehen die Residuen ε_j des j -ten Messpunktes $(X_{1,j}, X_{2,j})$ senkrecht auf der Modellgeraden, siehe Abb. 3.10, 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.

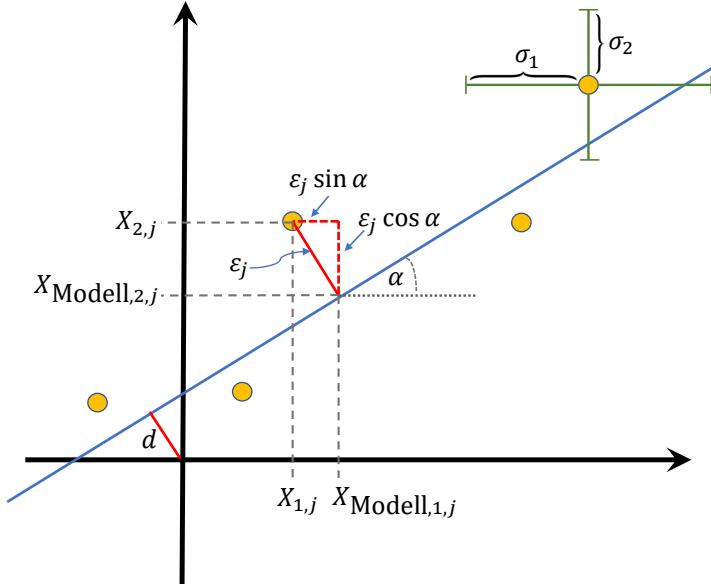


Abbildung 3.10: Veranschaulichung für Schätzung von Geradenparametern durch Minimierung der Summe der Quadrate der Residuen bei verschiedenen Varianzen σ_1^2 und σ_2^2 der beiden direkten Messgrößen X_1 und X_2

$$Q(\alpha, d) = \sum_{j=1}^J \left(\frac{\varepsilon_j(\alpha, d) \sin(\alpha)}{\sigma_1} \right)^2 + \sum_{j=1}^J \left(\frac{\varepsilon_j(\alpha, d) \cos(\alpha)}{\sigma_2} \right)^2. \quad (3.54)$$

Hier haben wir also die σ_i^2 als Varianzen der Abweichungskomponenten der Messgröße X_i (vektorwertige 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 so aus

$$Q(\mathbf{p}) = \sum_{j=1}^J \vec{\varepsilon}(X_{1,j}, \dots, X_{N,j}, \mathbf{p})^\top \Sigma(X_{1,j}, \dots, X_{N,j}, \mathbf{p})^{-1} \vec{\varepsilon}(X_{1,j}, \dots, X_{N,j}, \mathbf{p})_j. \quad (3.55)$$

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

3.7 Robuste Schätzverfahren

Wir haben nun eine Kostenfunktion Q mit Gewichtsfaktoren kennengelernt, Gl. (3.48), 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 Umgewichten so zu verändern, dass sie wieder die Gestalt einer Normalverteilung bekommt. Wir führen Gewichtsfaktoren $w = \delta$ 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

$$\tilde{\varepsilon} = \delta(\varepsilon) \varepsilon \quad \Rightarrow \quad \tilde{\varepsilon} \sim \mathcal{N}(0, \sigma). \quad (3.56)$$

Mit Hilfe der Gewichtsfunktion werden aus den Residuen die Gewichtsfaktoren $w_j = \delta_j = \delta(\varepsilon_j)$ 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

$$Q(\mathbf{p}) = \sum_{j=1}^J \delta_j \varepsilon_j(\mathbf{p})^2. \quad (3.57)$$

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

$$\min_{\mathbf{p}} Q(\mathbf{p}) \quad \text{mit} \quad Q(\mathbf{p}) = \sum_{j=1}^J \varepsilon_j(\mathbf{p})^2. \quad (3.58)$$

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. 3.11 dargestellt. Für das Histogramm werden $K = 80$ Klassen (engl. *bins*) gewählt, die Anzahl der Beobachtungen umfasst $J = 2200$. Die Klassenbreite ist dann

$$\Delta\xi = \frac{1}{K} (\max \{X_{1,j}\} - \min \{X_{1,j}\}) \quad (3.59)$$

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

$$\xi_{G,k} = k \Delta\xi + \min \{X_{1,j}\} \quad k = 0, \dots, K \quad (3.60)$$

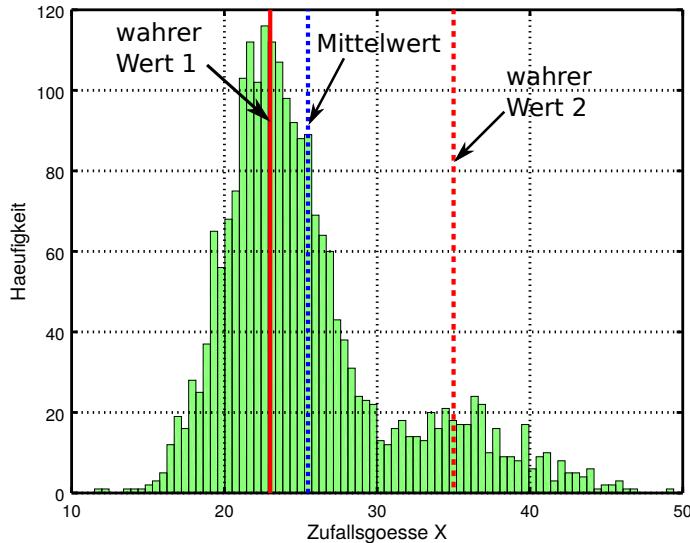


Abbildung 3.11: Beispiel für nicht erwartungsgemäße Beobachtungen

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

$$\xi_k = k \Delta\xi + \frac{1}{2}\Delta\xi + \min\{X_{1,j}\} \quad k = 1, \dots, K \quad (3.61)$$

wobei die Anzahl der Klassen K ist.

Die Anzahl der Einzelbeobachtungen, für die gilt

$$\xi_{G,k-1} \leq X_{1,j} < \xi_{G,k}, \quad (3.62)$$

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 Gewichtsfaktoren δ_j .

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

$$\delta_j = \begin{cases} \left(1 - \left(\frac{|X_{1,j} - y^{(\kappa)}|}{c^{(\kappa)}}\right)^2\right)^2 & \text{falls } |X_{1,j} - y^{(\kappa)}| \leq c^{(\kappa)} \\ 0 & \text{sonst} \end{cases} \quad (3.63)$$

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

$$c^{(\kappa)} = \text{median } |X_{1,j} - y^{(\kappa)}| \quad (3.64)$$

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

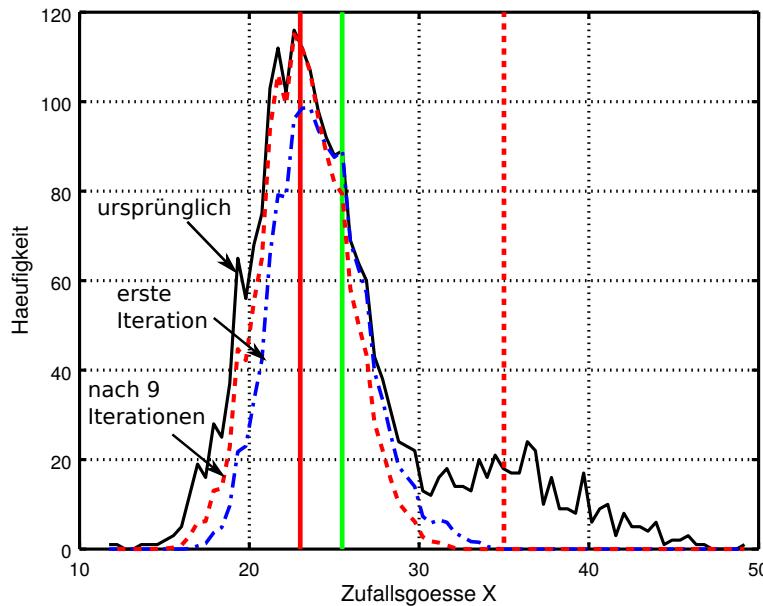


Abbildung 3.12: Umgewichten um die Verteilungsform der Normalverteilung anzupassen

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

$$Q(\mathbf{p}) = \sum_{j=1}^J \varepsilon_j^2 \left(1 - \left(\frac{\varepsilon_j(\mathbf{p})}{c(\mathbf{p})} \right)^2 \right)^2 H(|\varepsilon_j(\mathbf{p})| - c(\mathbf{p})) H(c(\mathbf{p}) - |\varepsilon_j(\mathbf{p})|), \quad (3.65)$$

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. 3.12 stellt die ursprüngliche Häufigkeitsverteilung aus Abb. 3.11 als durchgezogene, schwarze Linie dar. Nach dem ersten Iterationsschritt der Umgewichtung, hat sich die Verteilung so verändert, wie es durch die blaue gestrichelpunktete 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 1.2, das Beispiel „Beugung am Gitter“ betrachtet, um die Aufgabenstellung zu verstehen, dass man aus dem 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 1.2 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:

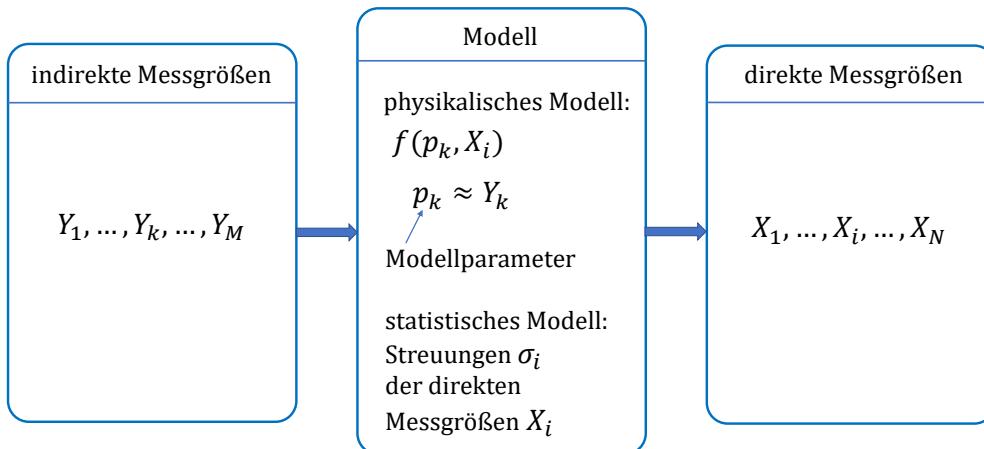


Abbildung 4.1: Modellbildung zur Gewinnung indirekter Messgrößen aus direkten Messgrößen

- 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 3.6 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. 3.10.

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

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 1.2 gesehen (Gln. (1.9), (1.10) und Gl. (1.11)), 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 sorum 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 taktiler 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

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

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

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

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 *a 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.

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

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

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

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

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 „*upgedated*“ werden.

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

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

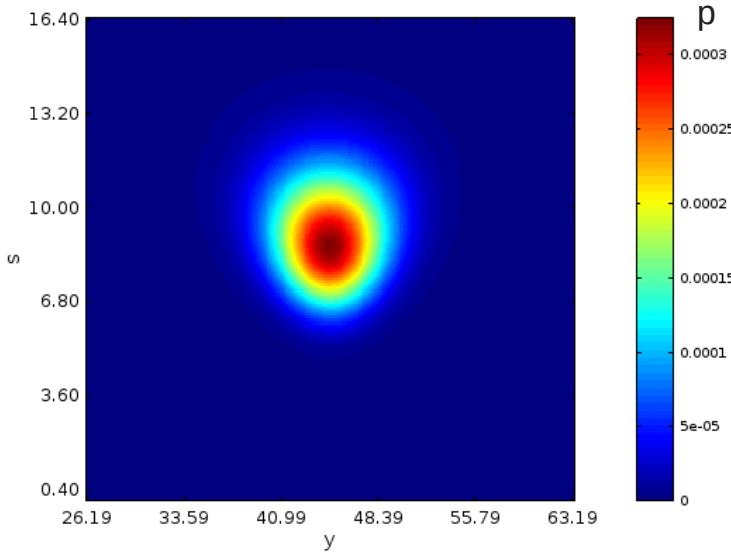


Abbildung 4.2: Posterior-Wahrscheinlichkeitsdichte als Funktion der beiden zu schätzenden Größen y und s .

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 anzudeuten, 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

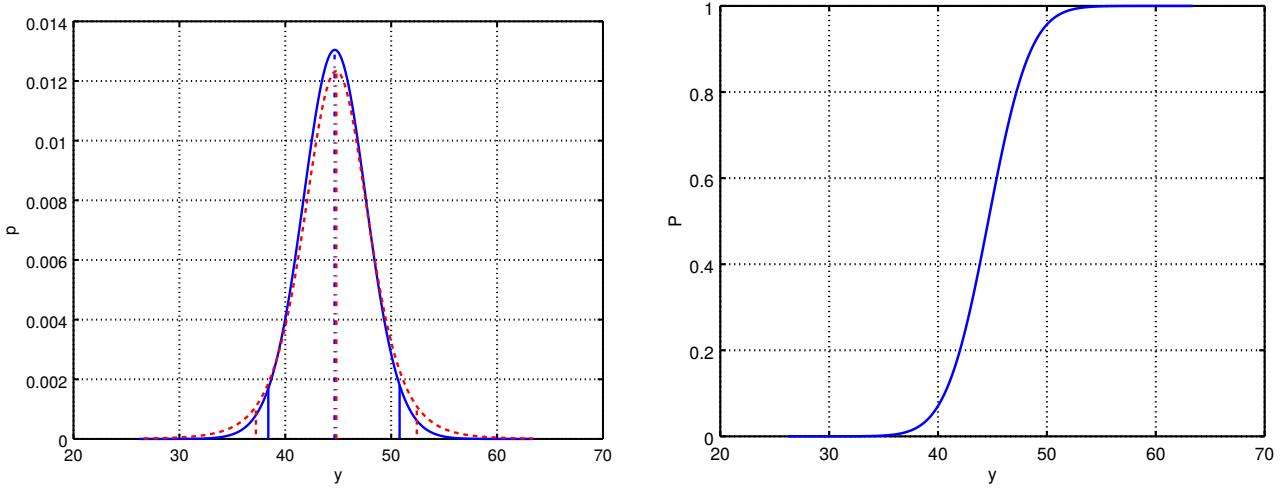


Abbildung 4.3: Links: Die blaue Kurve stellt die Randverteilung der Posterior-Wahrscheinlichkeitsdichte als Funktion der beiden zu schätzenden Größen Y mit *Credible Interval* dar und die rot gestrichelte Kurve die Student-t-Verteilung mit Vertrauensintervall. Rechts: Kumulierte Randverteilung der Posterior-Wahrscheinlichkeitsdichte,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

der *frequentistischen* Statistik *Signifikanzniveau* genannt wird.

Im folgenden werden wir sehen, wie eine *Messunsicherheit* im Rahmen der *bayesianischen* 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 Binomial-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

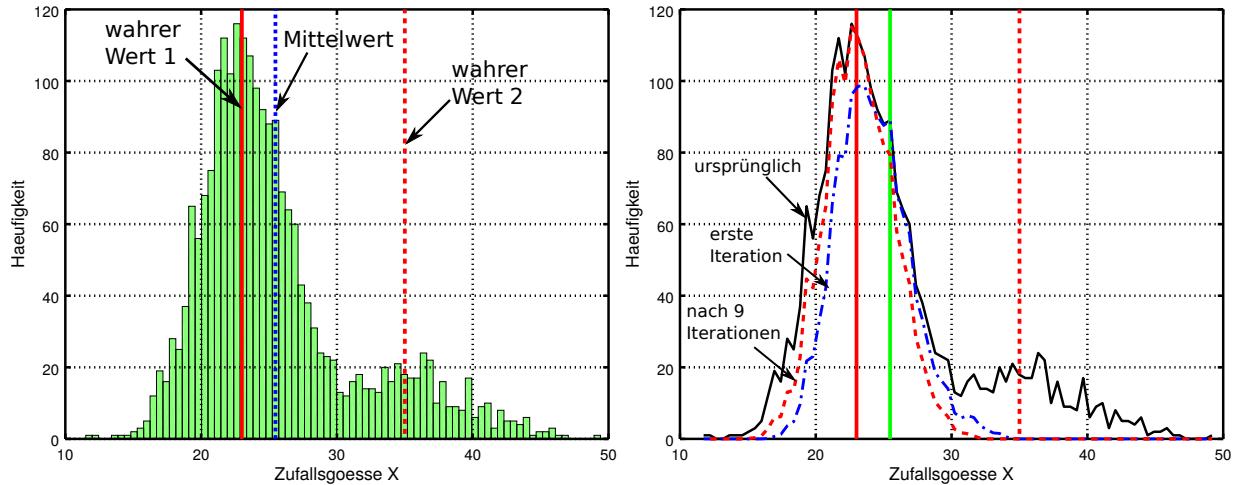


Abbildung 5.1: Beispiel für nicht erwartungsgemäße Beobachtungen (rechts) und Umgewichten um die Verteilungsform der Normalverteilung anzupassen links

(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 3.7 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 Beobachtungen 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 Beobach-

tungswerten werden geprüft. Diese Art der Prüfung wird *Hypothesentest* genannt. Es gibt unterschiedliche 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 3.7 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

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

Für die kumulierte Gaußverteilung entspricht im wesentlichen der Fehlerfunktion erf, englisch *Error Function*.

$$\text{erf}(X) := \frac{2}{\sqrt{\pi}} \int_0^X e^{-X'^2} dX'. \quad (5.4)$$

Die unterschiedlichen Definitionen für die kumulierte Gaußverteilung und die Fehlerfunktion führen zu unterschiedlichen Wertebereichen. Während die kumulierte Gaußverteilung einen Wertebereich von 0 bis 1 hat, ist die Fehlerfunktion so definiert, dass sie einen Wertebereich von -1 bis 1 hat. Sie lassen sich wie folgt in einander überführen

$$P(X) = \frac{1}{2} \left(1 + \text{erf}\left(\frac{X-y}{\sqrt{2}s}\right) \right). \quad (5.5)$$

Für standardnormalverteilte Zufallszahlen $Z = \frac{X-\mu}{\sigma}$ oder $Z = \frac{X-y}{s}$ wird für die kumulierte Gaußverteilung der Bezeichner Φ verwendet

$$\Phi(Z) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^Z e^{-\frac{1}{2}Z'^2} dZ' \Rightarrow \Phi(Z) = \frac{1}{2} \left(1 + \text{erf}\left(\frac{Z}{\sqrt{2}}\right) \right). \quad (5.6)$$

Die Kenntnis dieser unterschiedlichen Definitionen und wie sie ineinander umzurechnen sind ist wichtig,

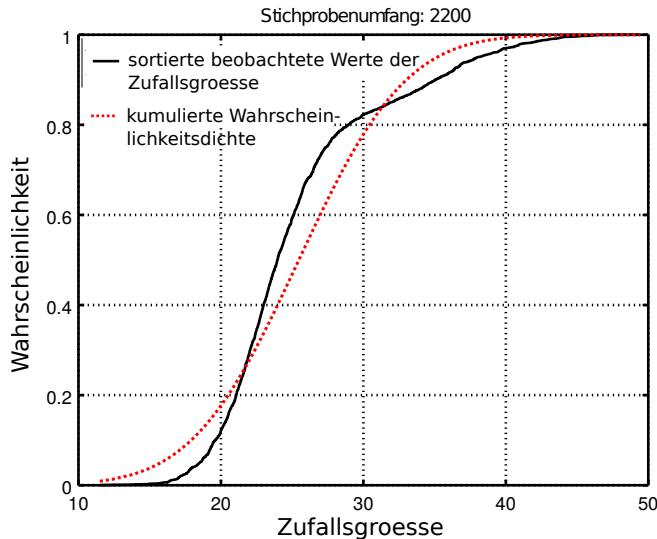


Abbildung 5.2: Vergleich der Wahrscheinlichkeiten (kumulierte Verteilungsdichten) aus der empirischen gewonnenen Häufigkeit der beobachteten Werte der Zufallsgröße (schwarze, durchgezogene Kurve) mit der kumulierten Normalverteilung, die sich auf die empirischen Erwartungswerte, den Mittelwert und die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

weil die statistischen Bibliotheksfunktionen von Matlab, Gnu-Octave, Python, etc. nicht die kumulierte Gaußverteilung bzw. die kumulierte Standardnormalverteilung, sondern die Fehlerfunktion zur Verfügung stellen.

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

$$X_{1,k_1} \leq X_{1,k_2} \leq \dots \leq X_{1,k_J} \quad (5.7)$$

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 mit den j kleinsten Beobachtungen

$$(X_{1,k_1}, X_{1,k_2}, \dots, X_{1,k_j}) \quad \text{die rel. kum. Häufigk.} \quad H(X_{1,k_j}) = \frac{j}{J} \quad (5.8)$$

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, durchgezogen gezeichnete Kurve in einem Diagramm dar.

Die Prüfgröße des *Kolmogorow-Smirnow*-Tests ist

$$\sup_X |H(X_{1,k_j}) - P(X)| = \max_{k_j} \left\{ \max \left\{ |H(X_{1,k_j}) - P(X_{1,k_j})|, \lim_{X \rightarrow X_{1,k_j}} |H(X) - P(X_{1,k_j} - 1)| \right\} \right\}. \quad (5.9)$$

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

$$\max_{k_j} \{ |H(X_{1,k_j}) - P(X_{1,k_j})| \}. \quad (5.10)$$

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

$$\sup_X |H(X_{1,k_j}) - P(X)| > K_{\alpha,J} \quad (5.11)$$

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:

$$K_{\alpha,J} = \sqrt{-\frac{1}{2J} \ln(\frac{\alpha}{2})} \quad (5.12)$$

Hypothesentests in Python (Klicke hier für interaktive Session):

```

1 import numpy as np
2 import matplotlib.pyplot as plt
3 import scipy.special
4 def cumulatives(x):
5     xbar = np.mean(x)
6     s = np.std(x, ddof=1)
7     J = len(x)
8     xsort = np.sort(x)
9     H = np.arange(1,J+1)/J;
10    z = (xsort - xbar)/s
11    # The cumulative of the unit normal distribution is given
12    # by Phi(z) = 1/2[1 + erf(z/sqrt(2))]
13    P = 0.5 * (1 + scipy.special.erf(z/np.sqrt(2)))

```

```
14     return P, H, xsort, xbar, s
15 #
16 # Hauptprogramm
17 #
18 # generiere synthetische Stichprobe
19 stichprobenumfang = 51
20 mu_1 = 42
21 sigma_1 = 5
22 stichprobe = np.random.normal(mu_1, sigma_1, stichprobenumfang)
23 #
24 # Aufgabenstellung Hypothesentest
25 alpha = 0.05
26 print('pruefe auf einem Signifikanzniveau von alpha = ',\
27       alpha*100, '% die Hypothese')
28 print('H0: Die Stichprobe ist normalverteilt')
29 #
30 # Durchfuehrung des Kolmogorow-Smirnow-Tests:
31 #
32 # Berechnung der kumulativen Normalverteilung P und der empirischen
33 # kumulativen Verteilung der Stichprobe H
34 P, H, xsort, xbar, s = cumulatives(stichprobe)
35 #
36 # Berechnung der Pruef- oder Testgroesse
37 Testgroesse = np.max(np.abs(H - P))
38 #
39 # Berechnung des Schwellwertes
40 K_alphaJ = np.sqrt(-np.log(alpha/2)/(2*stichprobenumfang))
41 #
42 # Antwortsatz
43 if (Testgroesse > K_alphaJ):
44     print("{:1.2f}".format(Testgroesse), ' > ', "{:1.2f}".format(K_alphaJ),\
45           ' ==> Die Nullhypothese H0 wird verworfen')
46 else:
47     print("{:1.2f}".format(Testgroesse), ' < oder = ', "{:1.2f}".format(K_alphaJ),\
48           ' ==> Die Nullhypothese H0 wird akzeptiert')
49 #
50 # graphische Darstellung
51 plt.figure(1)
52 plt.plot(xsort, P, 'r-')
53 plt.plot(xsort, H, 'kd')
54 plt.grid()
55 plt.xlabel('Zufallsgroesse')
56 plt.ylabel('Wahrscheinlichkeit')
57 plt.show()
```

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

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} X p(X) \mathrm{d}X =: \mu^{(1)}(X) \quad (5.13)$$

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

$$\int_{-\infty}^{\infty} X^2 p(X) \mathrm{d}X =: \mu^{(2)}(X) \quad (5.14)$$

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

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (X - \mathrm{E}(X))^2 p(X) \mathrm{d}X \quad (5.15)$$

deren Wurzel

$$\sigma = \sqrt{\mathrm{Var}(X)} \quad (5.16)$$

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

$$\int_{-\infty}^{\infty} X^3 p(X) \mathrm{d}X =: \mu^{(3)}(X) \quad (5.17)$$

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

$$\mathrm{Skew}(X) = \int_{-\infty}^{\infty} \left(\frac{X - \mathrm{E}(X)}{\sigma} \right)^3 p(X) \mathrm{d}X \quad (5.18)$$

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

$$\int_{-\infty}^{\infty} X^4 p(X) dX =: \mu^{(4)}(X) \quad (5.19)$$

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

$$\text{Kurt}(X) = \int_{-\infty}^{\infty} \left(\frac{X - E(X)}{\sigma} \right)^4 p(X) dX \quad (5.20)$$

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

$$\int_{x_1}^{x_2} p(X) dX = 1 - \alpha. \quad (5.21)$$

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

$$\int_{-\infty}^{\infty} p(X) dX = 1 \quad (5.22)$$

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

$$E(X) = \int_{-\infty}^{\infty} X p(X) dX \quad (5.23)$$

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

$$[P^{-1}\left(\frac{\alpha}{2}\right), P^{-1}\left(1 - \frac{\alpha}{2}\right)] \quad (5.24)$$

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 Modellparametern (kein *Prior*) und keine weiteren beein-

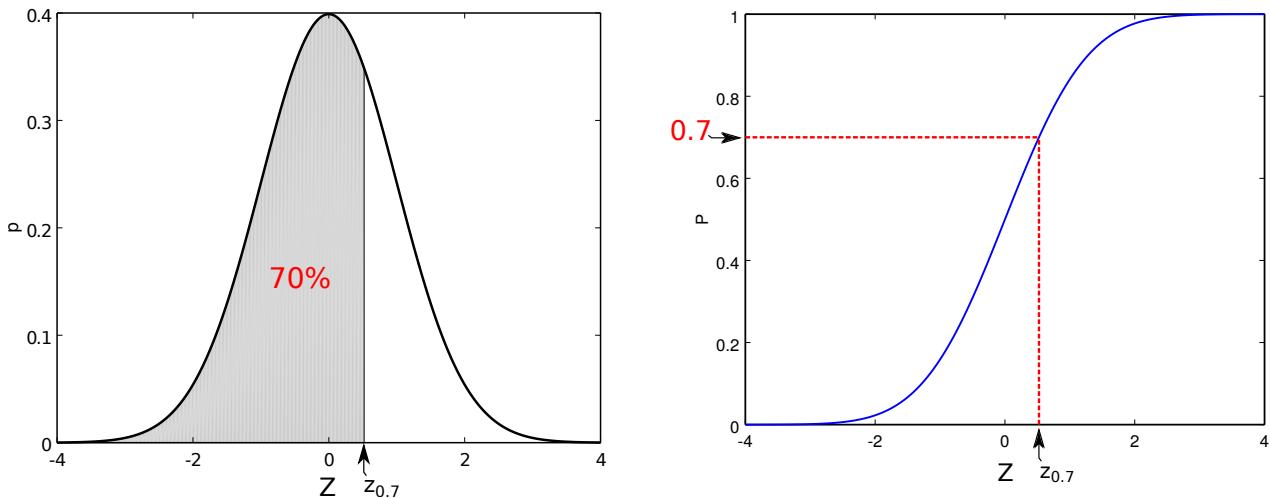


Abbildung 5.3: Quantil $z_\alpha = 0.5244$ und Wahrscheinlichkeit $\alpha = 0.7$ der Standardnormalverteilung.

flussenden 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*

$$Z \sim \mathcal{N}(0, 1). \quad (5.25)$$

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

$$p(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2} \quad \text{mit} \quad Z = \frac{X - \mu}{\sigma}. \quad (5.26)$$

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

$$\int_{-\infty}^{z_\alpha} p(Z)dZ = \alpha. \quad (5.27)$$

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

$$z_{0.7} = 0.5244$$

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 $1 - 0.05 = 0.95$

$$\int_{-1.96}^{1.96} p(Z) dZ = 0.95 \quad (5.28)$$

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

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

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 Vertrauensintervalls zurückgerechnet werden

$$Z = \frac{X - \mu}{\sigma} \quad \Leftrightarrow \quad X = \mu + Z\sigma \quad (5.29)$$

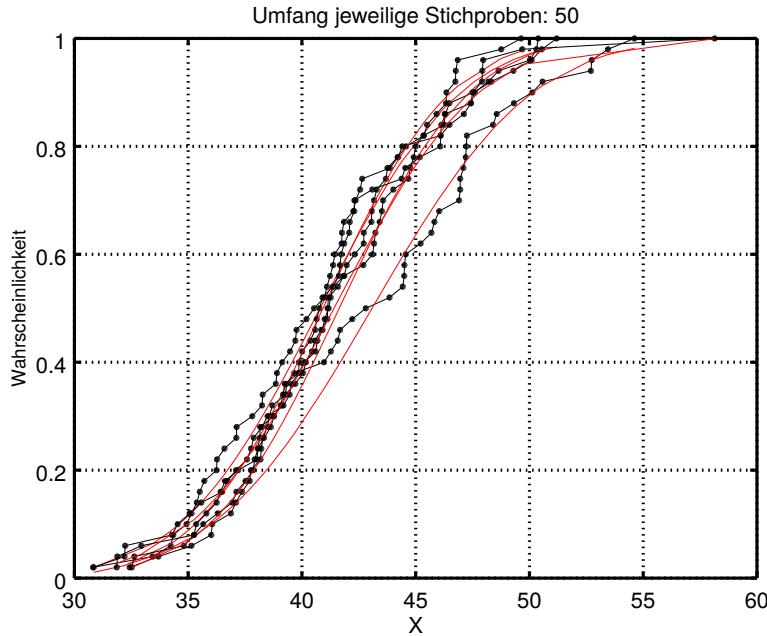


Abbildung 5.4: Kumulative relative Wahrscheinlichkeitsfunktionen und relative Summenhäufigkeiten von verschiedenen Stichproben mit jeweils $J = 50$ Werten, entnommen aus einer Grundgesamtheit mit $\mu = 42$ und $\sigma = 5$.

also

$$[z_{\frac{1}{2}\alpha}, z_{1-\frac{1}{2}\alpha}] = [-z_{1-\frac{1}{2}\alpha}, z_{1-\frac{1}{2}\alpha}] \quad \Leftrightarrow \quad [\mu - z_{1-\frac{1}{2}\alpha} \sigma, \mu + z_{1-\frac{1}{2}\alpha} \sigma] \quad (5.30)$$

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

Die Likelihood wird maximal für

$$y = \frac{1}{J} \sum_{j=1}^J X_{1,j}, \quad (5.31)$$

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

$$s^2 = \frac{1}{J-1} \sum_{j=1}^J (X_{1,j} - y)^2. \quad (5.32)$$

Nachdem wir gesehen haben, wie wir für eine normalverteilte Zufallsgröße das Vertrauensintervall ermitteln können, wenden wir uns jetzt der Thematik zu, was passiert, 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:

```
1 function example_t()
2 J1 = 50;
```

```

3 mu1 = 42;
4 sigma1 = 5;
5 samples = [];
6 for kappa = 1:5
7     samples = [samples (mu1 + sigma1*randn(J1,1))];
8 end
9 figure(3000); hold on;
10 [J1, nS] = size(samples);
11 for kappa = 1:5
12     [P, H, xsort, xbar, s] = cumulatives(samples(:,kappa));
13     plot(xsort, H, 'k.-');
14     plot(xsort, P, 'r-');
15 end
16 xlabel('X', 'fontsize', 14);
17 ylabel('Wahrscheinlichkeit', 'fontsize', 12);
18 title(['Umfang jeweilige Stichproben: ' num2str(J1)], 'fontsize', 14);
19 grid on;
20 set(gca, 'fontsize', 14, 'linewidth', 2);
21 hold off;
22 end
23 function [P, H, xsort, xbar, s] = cumulatives(x)
24 xbar = mean(x);
25 s = std(x);
26 J1 = length(x);
27 [xsort, isort] = sort(x, 'ascend');
28 H = [1:J1]/J1;
29 P = normcdf(xsort, xbar, s);
30 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

$$y = \frac{1}{J} \sum_{j=1}^J X_{1,j}. \quad (5.33)$$

Die empirische Standardabweichung wird berechnet mit

$$s = \sqrt{\frac{1}{J-1} \sum_{j=1}^J (X_{1,j} - y)^2}. \quad (5.34)$$

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

$$\text{skew} = \frac{1}{J} \sum_{j=1}^J \left(\frac{X_{1,j} - y}{s} \right)^3. \quad (5.35)$$

Die empirische Kurtosis wird berechnet mit

$$\text{kurt} = \frac{1}{J} \sum_{j=1}^J \left(\frac{X_{1,j} - y}{s} \right)^4. \quad (5.36)$$

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

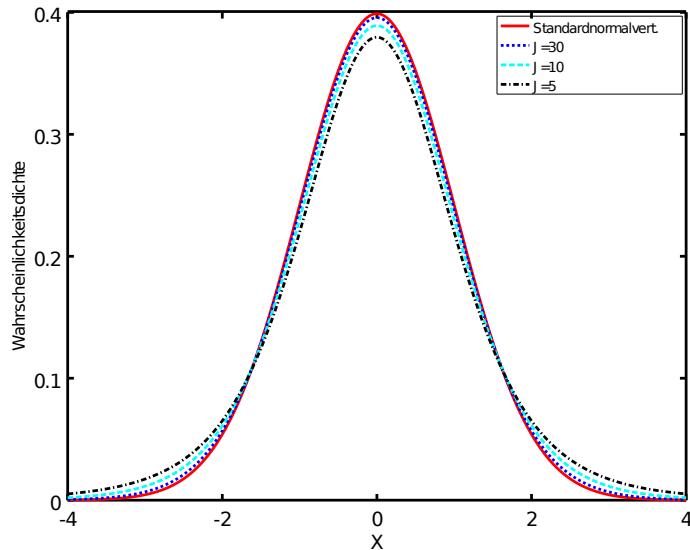


Abbildung 5.5: Student-t Verteilungen für unterschiedliche Stichprobenumfänge.

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

$$p_t(Z, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} \left(1 + \frac{Z^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad (5.37)$$

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:

$$\Gamma\left(\frac{\nu}{2}\right) = \sqrt{\pi} \frac{(\nu - 2)!!}{2^{\frac{\nu-1}{2}}} \quad \nu \in \mathbb{N} \quad (5.38)$$

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 beigefügt.

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:

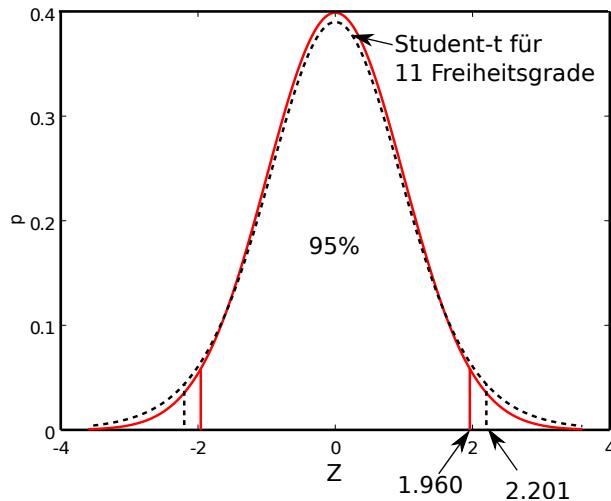


Abbildung 5.6: Vertrauensintervalle in Vergleich für die Standardnormalverteilung (rote Kurve) und die Student-t-Verteilung für $\nu = 11$ Freiheitsgrade (schwarze gestrichelte Kurve).

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.

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 Freiheitgrade beeinflusst, je kleiner die Anzahl der Freiheitsgrade ist, desto breiter wird das Intervall. Die beiden Abb. 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 Ver-

trauensintervall. Eine gebräuchliche Schreibweise ist

$$Y = y \pm t_{1-\alpha/2, \nu} s. \quad (5.39)$$

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 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

$$l(Y, \sigma | \{X_{1,1}, \dots, X_{1,J}\}) = \prod_{j=1}^J \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{1}{2} \left(\frac{(X_{1,j}-Y)}{\sigma} \right)^2} \quad (5.40)$$

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

$$l(Y, \sigma | \{X_{1,1}, \dots, X_{1,J}\}) = \prod_{j=1}^J \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{1}{2} \left(\frac{(X_{1,j}-y)-(Y-y)}{\sigma} \right)^2} \quad (5.41)$$

d.h.

$$l(Y, \sigma | \{X_{1,1}, \dots, X_{1,J}\}) = \frac{1}{(\sqrt{2\pi} \sigma)^J} e^{-\frac{1}{2} \sum_{j=1}^J \left(\frac{(X_{1,j}-y)-(Y-y)}{\sigma} \right)^2}. \quad (5.42)$$

mit folgender Nebenrechnung für die Summe im Exponenten

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

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

$$l(Y, \sigma | \{X_{1,1}, \dots, X_{1,J}\}) \propto e^{-\frac{1}{2} J \left(\frac{(Y-y)}{\sigma} \right)^2} \quad (5.43)$$

und wir definieren folgende Größe

$$\bar{\sigma}^2 := \frac{\sigma^2}{J} \quad (5.44)$$

als *Varianz des Mittelwertes*.

Damit lässt sich Gl. (5.43) umschreiben in

$$l(Y, \sigma | \{X_{1,1}, \dots, X_{1,J}\}) \propto e^{-\frac{1}{2} \left(\frac{(Y-y)}{\bar{\sigma}} \right)^2}. \quad (5.45)$$

Nun setzen wir in die Definitionsgleichung der Varianz des Mittelwertes Gl. (5.44) bzw. die Wurzel daraus die empirisch ermittelten Werte unseres Beispiels ein:

$$\bar{s} = \frac{s}{\sqrt{J}} \quad (5.46)$$

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*

$$Y = y \pm t_{1-\alpha/2, \nu} \bar{s}. \quad (5.47)$$

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 denselben Erwartungswert haben, was soviel bedeutet wie $\mu_1 = \mu_2$. Die Alternativhypothese H_a lautet, dass sie unterschiedliche Erwartungswerte haben, was soviel bedeutet wie $\mu_1 \neq \mu_2$, und damit zu unterschiedlichen Grundgesamtheiten gehören.

Umgekehrt, bedeutet aber nicht, dass wenn sie denselben Erwartungswert haben, sie zwingend zu derselben Grundgesamtheit gehören. Das zu prüfen erfordert die in den nachfolgenden Abschnitten behandelten Tests der Varianzen, den χ^2 -Test bei Test der empirischen Varianz einer Stichprobe mit einer bekannten Varianz, den F-Test bei Test der Varianzen zweier Stichproben.

Der Test auf Erwartungswerte, der t-Test wird wie folgt formuliert:

H_0	$\mu_1 = \mu_2$	beide Stichproben den gleichen Erwartungswert
H_a	$\mu_1 \neq \mu_2$	beide Stichproben haben unterschiedliche Erwartungswerte und gehören somit unterschiedlichen Grundgesamtheiten an

Die Prüfgröße ist die Differenz der empirisch berechneten Mittelwerte normiert auf die Standardabweichungen der Mittelwerte. Mit Gl. (5.34) 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. (5.46) darüber abgeschätzt, dass sie durch den Stichprobenumfang geteilt wird:

$$\bar{s}_i^2 = \frac{1}{J_i(J_i - 1)} \sum_{j=1}^{J_i} (X_{i,j} - y_i)^2 \quad (5.48)$$

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

$$T = \frac{y_1 - y_2}{\sqrt{\bar{s}_1^2 + \bar{s}_2^2}} \quad (5.49)$$

mit

$$y_i = \frac{1}{J_i} \sum_{j=1}^{J_i} X_{i,j} \quad \text{mit} \quad i = 1, 2 \quad (5.50)$$

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

$$|T| > t_{1-\frac{1}{2}\alpha, \nu} \quad \Rightarrow \quad H_0 \text{ ablehnen} \quad (5.51)$$

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

$$\nu = J_1 + J_2 - 2 \quad (5.52)$$

und für NICHT *gepoolte*

$$\nu = \frac{\left(\frac{s_1^2}{J_1} + \frac{s_2^2}{J_2}\right)^2}{\frac{\left(\frac{s_1^2}{J_1}\right)^2}{J_1-1} + \frac{\left(\frac{s_2^2}{J_2}\right)^2}{J_2-1}}. \quad (5.53)$$

Die Herleitung zur Berechnung der Anzahl der Freiheitsgrade für nicht gepoolte Stichproben wird in Kapitel 7 dran kommen. Gl. (5.53) wird auch Satterthwaite'sche Gleichung genannt.

Bei $\nu = J_1 + J_2 - 2$ Freiheitsgraden, bei dem hier betrachteten Beispiel ist dies $\nu = 98$ und $t_{1-\frac{1}{2}\alpha, \nu} = t_{0.975, 98} = 1.985$. 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 Test einer Varianz

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.

$$X \sim \mathcal{N}(\mu, \sigma). \quad (5.54)$$

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

$$y_i = \frac{1}{J_i} \sum_{j=1}^{J_i} X_{i,j} \quad s_i^2 = \frac{1}{J_i - 1} \sum_{j=1}^{J_i} (X_{i,j} - y_i)^2 \quad (5.55)$$

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

$$Q = \sum_{j=1}^J Z_j^2 \quad (5.56)$$

gemäß folgender Wahrscheinlichkeitsdichtefunktion verteilt

$$p_{\chi^2}(Q, \nu) = \frac{Q^{\frac{\nu}{2}-1} e^{-\frac{Q}{2}}}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})} \quad Q > 0 \quad (5.57)$$

mit ν für die Anzahl der Freiheitsgrade und Γ für die im Abschnitt zuvor als Gl. (5.38) angegebenen Gammafunktion.

Eine Schreibweise für die Aussage

Q ist Chi2-verteilt

ist

$$Q \sim \chi^2(\nu). \quad (5.58)$$

Die Definition der χ^2 -Verteilungsdichtefunktion braucht nicht auswendig gelernt zu werden. Der Umgang mit den Quantitabellen 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

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

$$s_\mu^2 = \frac{1}{J} Q \sigma^2 \quad \Leftrightarrow \quad Q = J \left(\frac{s_\mu}{\sigma} \right)^2 \quad (5.60)$$

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.

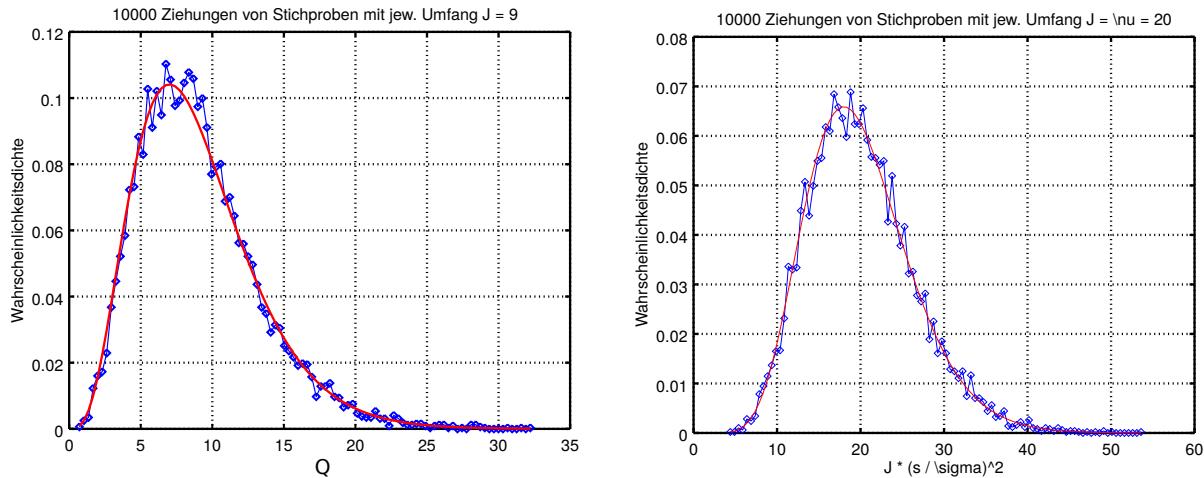


Abbildung 5.7: χ^2 -Verteilung der normierten Zufallsgröße $Q = J \left(\frac{s_\mu}{\sigma} \right)^2$ für $s_\mu^2 = \frac{1}{J} \sum_{j=1}^J (X_{..j} - \mu)^2$ für $J = 9$ (links) und $J = 20$ (rechts).

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

```

1 function understandchi2()
2 Jz = 9;
3 nbin = 100;
4 n = 10000;
5 d = zeros(n,1);
6 for k=1:n
7     x = randn(Jz,1);
8     d(k) = sum(x.^2);
9 end
10 [haeuf, bin] = hist(d,nbin);
11 Deltabin = bin(2)-bin(1);
12 figure(1000);
13 plot(bin,haeuf/(n*Deltabin), 'bd-', bin, chi2pdf(bin,Jz), 'r-');
14 xlabel('J * (s / \sigma)^2', 'fontsize', 14);
15 ylabel('Wahrscheinlichkeitsdichte', 'fontsize', 14);
16 title([num2str(n) 'Stichprobenumfang J = \nu = ' num2str(Jz)], 'fontsize', 14);
17 grid on;
18 set(gca, 'fontsize', 14, 'linewidth', 2);
19 print(1000,['understandchi2_df' num2str(Jz) '_solid.svg'], '-dsvg');
20 end

```

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

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 die Größe Q Werte beobachtet werden,

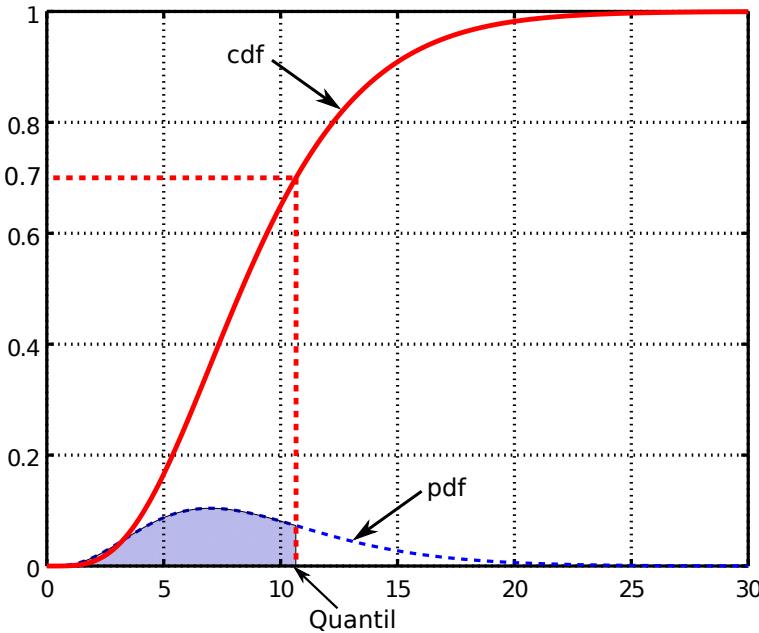


Abbildung 5.8: Wahrscheinlichkeitsverteilungsdichte (pdf) und kumulierte Wahrscheinlichkeitsfunktion (cdf) der normierten Varianzen, das Quantil wird χ^2 genannt. Hier für $\nu = 9$ Freiheitgrade und $1 - \alpha = 0.7$, somit $\chi_{\nu,1-\alpha}^2 = 10.66$ dargestellt.

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.

$$P(Q) = \int_0^{\chi_{1-\alpha,\nu}^2} p(Q', \nu) \, dQ' = 1 - \alpha \quad (5.61)$$

Bei dem χ^2 -Test wird die empirische Varianz einer Stichprobe

$$s_{\mu,i}^2 = \frac{1}{J_i} \sum_{j=1}^{J_i} (X_{i,j} - \mu)^2 \approx \frac{1}{J_i - 1} \sum_{j=1}^{J_i} (X_{i,j} - y_i)^2 = s_i^2 \quad (5.62)$$

mit der bekannten Varianz einer Grundgesamtheit verglichen.

```

1 for k=1:n
2   x = randn(Jz,1);
3   y = mean(x);
4   d(k) = sum((x-y).^2);
5 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

$$s^2 = \frac{1}{\nu} Q \sigma^2 \quad \Leftrightarrow \quad Q = \nu \left(\frac{s}{\sigma} \right)^2, \quad (5.63)$$

so dass die χ^2 -Verteilungsdichte, in Gnu-Octave mit dem Funktionsaufruf `chi2pdf` für $\nu = J - 1$ zu

verwenden ist:

```

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

```

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

H_0	$\sigma^2 = \sigma_0^2$	die Stichprobe gehört zu einer Grundgesamtheit mit der Varianz σ_0^2
H_a	$\sigma^2 \neq \sigma_0^2$	die Stichprobe gehört nicht zu einer Grundgesamtheit mit der Varianz σ_0^2

Die Testgröße ist

$$T = \nu \left(\frac{s}{\sigma_0} \right)^2. \quad (5.64)$$

Die Nullhypothese wird mit einem Signifikanzniveau von α verworfen, falls die Testgröße außerhalb des Intervalls liegt, das durch die in Gl. (5.61) definierten Quantile begrenzt wird, d.h.

$$T < \chi_{\alpha/2, \nu}^2 \quad \text{oder} \quad T > \chi_{1-\alpha/2, \nu}^2. \quad (5.65)$$

Es ist zu beachten, dass beim zweiseitigen Test für jede Seite die Hälfte des Signifikanzniveaus verwendet wird, also $\alpha/2$.

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

5.7 FisherVerteilung und Vergleich zweier empirischer Varianzen

Der Fall, bei dem zu überprüfen ist, ob die Varianzen zweier Stichproben zur gleichen Grundgesamtheit gehören müssten, bzw. ob diese im wesentlichen übereinstimmen, wird als Hypothesentest wie folgt formuliert:

H_0	$\sigma_1^2 = \sigma_2^2$	die beiden Stichproben haben die gleichen Varianzen
H_a	$\sigma_1^2 \neq \sigma_2^2$	die Stichproben haben unterschiedliche Varianzen und gehören somit zu unterschiedlichen Grundgesamtheiten

Die Testgröße ist

$$T = \left(\frac{s_1}{s_2} \right)^2. \quad (5.66)$$

Sie wird verglichen mit den beiden Quantilen, dem linksseitigen und dem rechtsseitigen, der FisherVerteilung. Der Test wird in der Literatur Fisher-Test oder kurz F-Test genannt.

Die FisherVerteilung ist die Marginalverteilung des Produkts der Wahrscheinlichkeitsdichtefunktionen

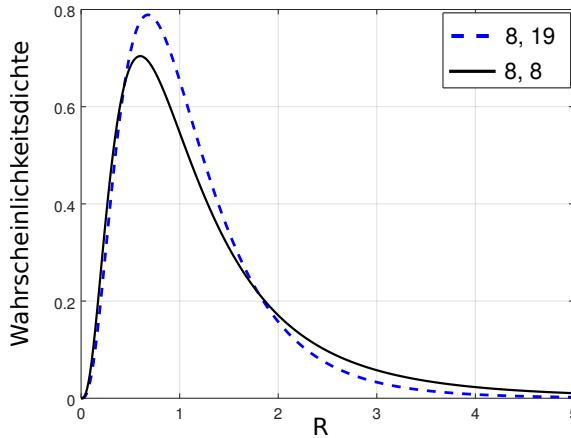


Abbildung 5.9: Fisherverteilung des Quotienten R zweier χ^2 -verteilter Zufallsgrößen Q_1 und Q_2 , also $R = \frac{\frac{1}{\nu_1}Q_1}{\frac{1}{\nu_2}Q_2} = \frac{s_1^2}{s_2^2}$ mit jeweils der Anzahl der Freiheitsgrade ν_1 und ν_2 . Die beiden Kurven zeigen zwei Beispiele, die blau gestrichelte für $\nu_1 = 8$ und $\nu_2 = 19$ und die schwarze durchgezogene für $\nu_1 = \nu_2 = 8$ Freiheitsgrade.

zweier χ^2 -verteilter Zufallsgrößen Q_1 und Q_2

$$p_F(R, \nu_1, \nu_2) = \int_0^\infty p_{\chi^2}(R, \nu_1) p_{\chi^2}(Q_2, \nu_2) \frac{\nu_1}{\nu_2} Q_2 dQ_2 \quad (5.67)$$

mit

$$R = \frac{\frac{1}{\nu_1}Q_1}{\frac{1}{\nu_2}Q_2}. \quad (5.68)$$

Dabei ist der Faktor $\frac{\nu_1}{\nu_2}Q_2$ die Funktionaldeterminante aus der Transformation $(Q_1, Q_2) \rightarrow (R, Q_2)$, um aus der Wahrscheinlichkeitsdichtefunktion der beiden χ^2 -verteilten Zufallszahlen

$$p_{\chi^2}(Q_1, \nu_1) p_{\chi^2}(Q_2, \nu_2)$$

die Fisherverteilung zu berechnen. Nach Einsetzen der Gleichung für die χ^2 -Verteilung, Gl. 5.57, und Ausführen der Integration in Gl. 5.67, sieht die Fisherverteilung wie folgt aus

$$p_F(R, \nu_1, \nu_2) = \nu_1^{\frac{\nu_1}{2}} \nu_2^{\frac{\nu_2}{2}} \frac{\Gamma(\frac{\nu_1+\nu_1}{2})}{\Gamma(\frac{\nu_1}{2}) \Gamma(\frac{\nu_2}{2})} \frac{R^{\frac{\nu_1}{2}-1}}{(\nu_1 R + \nu_2)^{\frac{\nu_1+\nu_2}{2}}}. \quad (5.69)$$

Abb. 5.9 zeigt beispielhaft zwei Fisherverteilungen, die mit blau gestrichelter Kurve dargestellte für $\nu_1 = 8$ und $\nu_2 = 19$ und die mit schwarz durchgezogener Kurve für $\nu_1 = \nu_2 = 8$ Freiheitsgrade.

Die Nullhypothese wird mit einem Signifikanzniveau von α verworfen, falls die Testgröße außerhalb des Intervalls liegt, das durch die Quantile der Fisherverteilung begrenzt wird, d.h.

$$T < F_{\alpha/2, \nu_2, \nu_1} \quad \text{oder} \quad T > F_{1-\alpha/2, \nu_2, \nu_1}. \quad (5.70)$$

Dabei ist ν_1 die Anzahl der Freiheitsgrade zu s_1 , die im Zähler stehende Standardabweichung aber in den Tabellen Nennerfreiheitsgrad genannt wird, und ν_2 die Anzahl der Freiheitsgrade zu s_2 , die im Nenner stehende Standardabweichung deren Freiheitsgrad aber in den Tabellen Nennerfreiheitsgrad

genannt wird. Das kommt daher, dass in Gl. 5.68 die Freiheitsgrade ν_1 zur Zufallsgröße Q_1 im Nenner stehen.

Beispiel: Der *General Social Survey (GSS)* hatte in den USA im Jahr 2000 anhand einer zufallsgenerierten Teilstichprobe mit $J = 652$ Erwachsenen Menschen untersucht, wie sich die Nutzungsdauer t in Stunden/Woche der Internetnutzung nach Geschlecht unterscheidet.

Geschlecht	J_i	t_i	s_i
männlich	305	5.48	7.80
weiblich	347	3.84	5.86

mit J_i für die Stichprobenumfänge und s_i für die Standardabweichung mit $i = 1, 2$. Die Standardabweichung zur Nutzungsdauer t beträgt bei den Männern $s_1 = 7.80$ Stunden/Woche und bei den Frauen $s_2 = 5.86$ Stunden/Woche. Der F-Test überprüft, ob der Unterschied der Varianzen s_1^2 und s_2^2 signifikant ist. Im vorliegenden Beispiel beträgt der Nennerfreiheitsgrad $\nu_1 = J_1 - 1 = 305 - 1 = 304$ und der Zählerfreiheitsgrad $\nu_2 = J_2 - 1 = 347 - 1 = 346$.

```

1 import numpy as np
2 import scipy.stats
3 alpha = 0.05
4 s1 = 7.80
5 s2 = 5.86
6 J1 = 305
7 J2 = 347
8 nu1 = J1-1
9 nu2 = J2-1
10 Ttruef = (s1/s2)**2
11 nuZaehl = nu2
12 nuNenn = nu1
13 Flinks = scipy.stats.f.ppf(alpha/2, nuZaehl, nuNenn)
14 Frechts = scipy.stats.f.ppf(1-alpha/2, nuZaehl, nuNenn)
15 print(Flinks, Ttruef, Frechts)

```

$T = 1.772$ und $F_{\alpha/2, \nu_2, \nu_1} = 0.805$ und $F_{1-\alpha/2, \nu_2, \nu_1} = 1.245$. Mit $T = 1.772 > F_{1-\alpha/2, \nu_2, \nu_1} = 1.245$ liegt die Prüfgröße signifikant höher als das rechte Quantil. Es kann also davon ausgegangen werden, dass sich die Varianzen der beiden Geschlechter signifikant voneinander unterscheiden.

5.8 Anwendung von Hypothesentests

Anwendung finden die Hypothesentests im Bereich der Qualitätssicherung in der Produktion (siehe beispielsweise zur Prüfung des Stout in der Guinessbrauerei schon vor 100 Jahren), oder im Bereich der Naturwissenschaften (Biologie, Chemie, Pharmazie, Medizien) zum Nachweis von Substanzen. Es geht darum, dass zu bewerten ist, mit welcher Wahrscheinlichkeit Beobachtungen zu welcher Verteilung

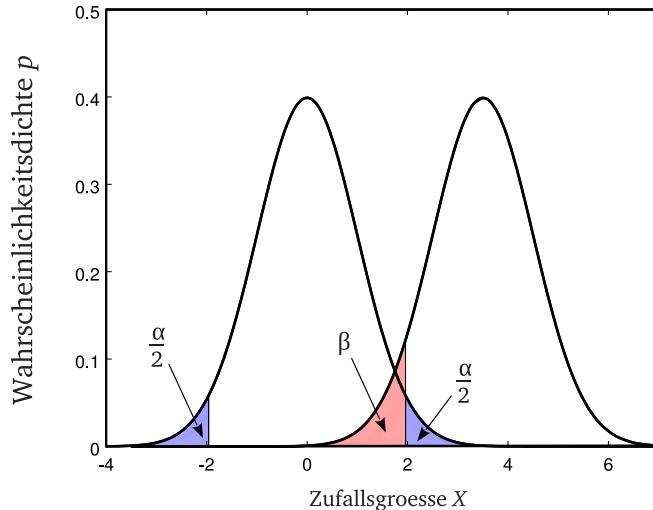


Abbildung 5.10: Die Ereignisse (beobachtete Werte) aus der linken Wahrscheinlichkeitsdichteverteilung, die in deren *Tails* liegen, werden mit einem Signifikanzniveau von α verworfen. Die Ereignisse (beobachtete Werte) aus der rechten Wahrscheinlichkeitsdichteverteilung, die in dessen linkem *Tail* liegen, werden mit Wahrscheinlichkeit β als der linken Wahrscheinlichkeit zugehörig erachtet.

gehören. Einer Entscheidung liegen Regeln zugrunde. Mit dem t-Test haben wir die Regeln kennengelernt, über die wir entscheiden, ob zwei Datensätze den gleichen Erwartungswert haben.

Als nächstes wollen wir untersuchen, welche Fehler mit welcher Wahrscheinlichkeit dabei entstehen können. Anschließend wollen wir uns ansehen, wie eine Stichprobe mit einer Toleranzvorgabe verglichen wird. Die Toleranzvorgabe liefert ein Intervall, dessen Lage mit der Lage der Wahrscheinlichkeitsverteilung der zu untersuchenden Stichprobe verglichen wird.

5.8.1 Entscheidung bei Vergleich zweier Stichproben

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

1. eine Nullhypothese H_0 (evl. 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. 5.10 soll illustrieren, dass Beobachtungswerte, die zu einer Grundgesamtheit mit einer Wahrscheinlichkeitsdichtevertteilung 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

```

1 function plot_entscheidungsfehler()
2 dz = 0.02;
3 lim = 7;
4 z = [-lim:dz:lim];
5 p1 = normpdf(z,0,1);
6 t = 1.96;
7 zleft = [-lim:dz:-t];
8 pleft = normpdf(zleft,0,1);
9 zright = [t:dz:lim];
10 pright = normpdf(zright,0,1);
11 mu2 = 3.5;
12 p2 = normpdf(z+mu2,mu2,1);
13 zbeta = [mu2-lim:dz:t];
14 pbeta = normpdf(zbeta,mu2,1);
15 figure(200); hold on;
16 area( zleft, pleft, 'Facecolor', [0.5 0.5 1]);
17 area( zright, pright, 'Facecolor', [0.5 0.5 1]);
18 area( zbeta, pbeta, 'Facecolor', [1 0.5 0.5]);
19 plot( z, p1, 'k-', 'linewidth', 2);
20 plot( z+mu2, p2, 'k—', 'linewidth', 2);
21 xlabel('X', 'fontsize', 14);
22 ylabel('p', 'fontsize', 14);
23 axis([-4 7 0 0.5]);
24 set(gca, 'fontsize', 14);
25 hold off;
26 print(200, 'entscheidungsfehler.svg', '-dsvg');
27 end

```

5.8.2 Konformitätsprüfung

Nachdem wir gesehen haben wie wir zwei Hypothesen über die Lage zweier Wahrscheinlichkeitsdichtevereiblungen verglichen haben, befassen wir uns jetzt mit dem Vergleich von Grenzen, den Toleranzgrenzen, mit einer Wahrscheinlichkeitsdichtevereiblung. Für die Fertigung werden in den Konstruktionszeichnungen zu den Merkmalen von Bauteilen die Toleranzgrenzen 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 Wahrscheinlichkeitsdichtevereiblung 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 (*Supplement*) zum „Guide of Uncertainty“ gibt es das Dokument *Evaluation of measurement data - The role of measurement uncertainty in conformity assessment*, das darlegt, wie Konformitätsbewertungen vorzunehem 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

$$p_c = P(Y \in C|X_1, \dots, X_J) = \int_C p(y|X_1, \dots, X_J) dy. \quad (5.71)$$

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 $C = [T_L, T_U]$ im Falle der zweiseitigen Grenzen, $C = [-\infty, T_U]$ für den Fall, dass es eine Obergrenze gibt, $C = [T_L, \infty]$, dass es eine Untergrenze gibt. Dabei sollen der Index L für „lower limit“ und der Index U für „upper limit“ stehen. Der Bezeichner T ist in der Normung das üblicherweise für Toleranzgrenzen verwendete Symbol und bedeutet an dieser Stelle nicht Testgröße! Er unterscheidet sich von den Testgrößen durch seine Indizes.

Die Wahrscheinlichkeit p_c , dass die Beobachtungen X_1, \dots, X_J innerhalb des Toleranzintervalls $C =$

$[T_L, T_U]$ liegen, ist

$$p_c = \int_{T_L}^{T_U} p(y|X_1, \dots, X_J) dy. \quad (5.72)$$

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 ($\bar{y} = \frac{1}{J} \sum_{j=1}^J X_j$) und die Varianz als empirische Varianz s aus $s = \frac{1}{J-1} \sum_{j=1}^J (X_j - \bar{y})^2$ berechnet und für die Wahrscheinlichkeitsdichte p die Gaußverteilung verwendet

$$p_c = \frac{1}{\sqrt{2\pi}s} \int_{T_L}^{T_U} e^{-\frac{1}{2}\left(\frac{y-\bar{y}}{s}\right)^2} dy. \quad (5.73)$$

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 für die kumulative standardnormalverteilte Gaußfunktion oder die Bibliotheksfunktion für die gauß'sche Fehlerfunktion (error function) wie folgt verwendet werden kann:

$$p_c = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_U} e^{-\frac{1}{2}z^2} dz - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_L} e^{-\frac{1}{2}z^2} dz \quad (5.74)$$

d.h.

$$p_c = \Phi(z_U) - \Phi(z_L) \quad (5.75)$$

mit $\Phi(Z) = \frac{1}{2} \left(1 + \operatorname{erf} \left(\frac{Z}{\sqrt{2}} \right) \right)$.

5.9 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:

```
1 J1 = 2000; % Stichprobenumfang
2 mu1 = 23; % Erwartungswert der Zufallsgröesse
3 s1 = 3;   % Wurzel aus dem Erwartungswert fuer die Varianz
4 Xarray = s1*randn(J1,1) + mu1;
```

Das Sortieren mit Matlab/Gnu-Octave können Sie mit

```
1 [xsort, isort] = sort(Xarray, 'ascend');
```

und die Wahrscheinlichkeiten mit

1	<code>h = [1:J1]'/J1;</code>
---	------------------------------

berechnen.

1. Verwenden Sie nicht die à priori eingesetzten Erwartungswerte, sondern Mittelwert $y = \text{mean}(X\text{array})$ und empirische Standardabweichung $s = \text{std}(X\text{array})$, um die Funktion $P(X)$ aus Gl. (5.3) zu realisieren. Berechnen Sie die relativen Häufigkeiten H gemäß Gl. (5.8), sowie den in Gl. (5.12) 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 $\text{mu2} = 35$ und $\text{s2} = 5$. 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
----	----	----	----	----	----	----	----	----

- Geben Sie zu jeder der beiden Stichproben die Mittelwerte und Standardabweichungen an.
- 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?
- 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.
- Führen Sie denselben Test wie in (c) durch, aber dieses Mal für $\sigma_0 = 9$.

Verwenden Sie die Quantiltabellen aus dem Anhang Kapitel 12.

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 $Var(X) = \sigma^2$ 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:

- $E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}(E(X) - \mu) = 0$
- $Var(Z) = Var\left(\frac{X-\mu}{\sigma}\right) = Var\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}Var(X) = 1$

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

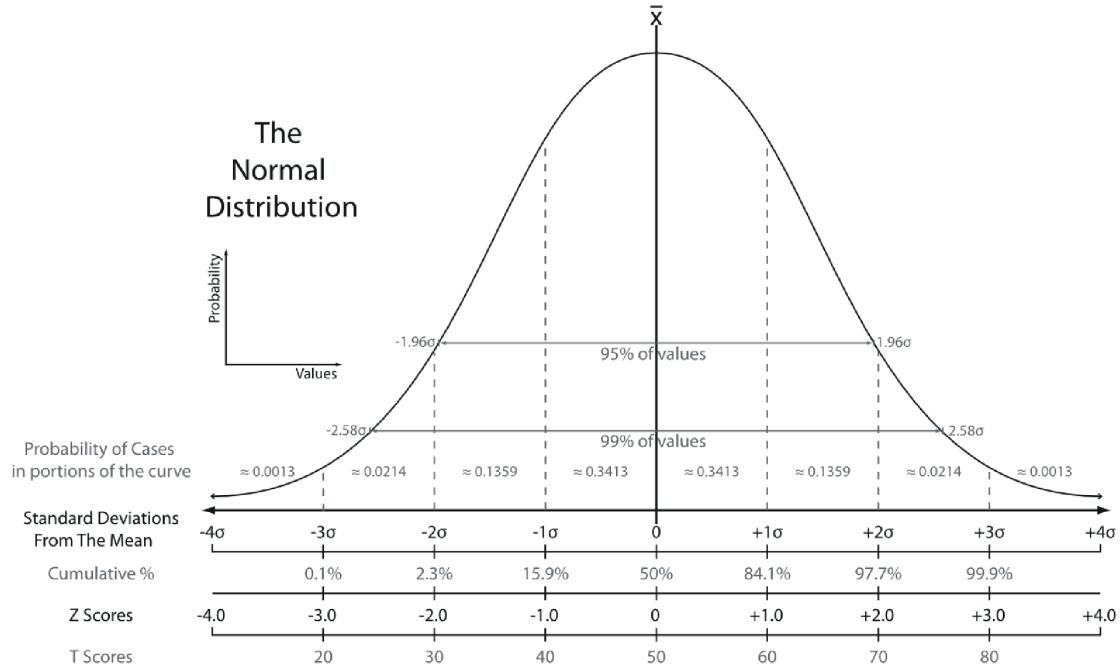


Abbildung 6.1: Die Standard-Normalverteilung $\mathcal{N}(0, \sigma^2)$ mit den entsprechenden Z-Werten.
Quelle: https://en.wikipedia.org/wiki/Standard_score

weitere 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 $[2u; 2u]$ zum 95,45% 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 statistisch 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):

$$T := \chi^2_{\text{obs}} = \frac{(x_1 - \tilde{x})^2}{u^2(x_1)} + \dots + \frac{(x_N - \tilde{x})^2}{u^2(x_N)} \quad (6.6)$$

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) ergibt sich wie folgt:

$$x_{\text{ref}} = \sum_{i=1}^N \tilde{w}_i x_i \quad (6.10)$$

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: $u^2(y) = \left(\frac{\partial f}{\partial x_1}\right) u^2(x_1) + \left(\frac{\partial f}{\partial x_2}\right) u^2(x_2)$)

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 angeben 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
$u(x_i)/g$	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` und evtl. den Pfad hinzufügen):

```

1 % addpath('C:\Octave\Octave-4.2.1\share\octave\packages');
2 % clear all;
3 x = [99.82, 99.05, 99.17 , 99.20 ,99.38];
4 u_x = [0.80 0.63, 0.86, 0.82, 0.98];
5 U_x = 2*u_x;
6

```

```

7 % Anzahl der Freiheitsgrade
8 nu = 4;
9
10 % Schritt 1:
11 x_ref = sum(x./(u_x).^2) ./ sum(1./(u_x).^2)
12
13 % Schritt 2:
14 u_ref = sqrt(1./sum(1./(u_x).^2))
15
16 % Schritt 3:
17 T = sum((x-x_ref).^2./(u_x.^2))
18
19 % Bestimme die Wahrscheinlichkeit Pr der Chi^2-Verteilung fuer Chi^2(\nu) > T:
20 Pr = 1 - chi2pdf(T,nu)
21
22 % Schritt 4: Grad der Uebereinstimmung
23 d = x-x_ref;
24 u_d = sqrt(u_x.^2 - u_ref.^2)
25 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). 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. 5.64, 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

$$R_B = \sqrt{\frac{1}{N-1} \sum_{i=1}^N \left(\frac{x_i - x_{\text{ref}}}{u(x_i)} \right)^2}. \quad (6.21)$$

Vergleicht man das Birge-Verhältnis mit der Testgröße $T = \nu \left(\frac{s}{\sigma_0} \right)^2$ in Gl.(6.6) so ergibt sich der Zusammenhang:

$$R_B^2 = \frac{1}{N-1} T \quad (6.22)$$

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:

$$R_B^2 = \sum_{i=1}^N \frac{w_i(x_i - x_{\text{ref}})^2}{N - 1} \quad (6.23)$$

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 Gewichtsfaktoren

Bei Ringvergleichen kommt es auch mal vor, dass die Messmethoden nicht immer genau die gleichen sind und der Einfluss der verschiedenen Methoden des Vergleichs unterschätzt wird. Wenn sich herausstellt, dass sich Gruppen von Ergebnissen ausbilden, so kann eine Zwischengruppenvarianz ermittelt werden. Dies kann beispielsweise mit der Paule-Mandel Methode erfolgen.

Es kann bei Ringvergleichen vorkommen, dass jeder am Vergleich beteiligte Partner eine Unsicherheit U angibt, die kleiner ist als der Abstand des Messwertes zu anderen Partnern, so dass sich ein En-Wert bzw. des Birge-Verhältnisses unter Verwendung der Gewichte gemäß Gl. (6.8) größer als Eins ergibt. Durch eine Anpassung der Gewichtsfaktoren 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 Gewichtsfaktoren 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 ist:

$$\bar{x}_I = \frac{1}{N} \sum_{i=1}^N x_i = \frac{1}{6} (1.1 + 1.9 + 1.5 + 16 + 25 + 3) \approx 8.1 \quad (6.24)$$

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 gleichgewichteten Mittelwert über alle Werte erhalten wir:

$$\bar{x}_{II} = \frac{1}{N} \sum_{i=1}^N x_i = \frac{1}{8} (2.0 + 1.0 + 1.5 + 1.8 + 1.2 + 1.7 + 16.3 + 16.8) = 5.2875 \quad (6.25)$$

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

$$\bar{x}_{II,A} = \frac{1}{6} (2.0 + 1.0 + 1.5 + 1.8 + 1.2 + 1.7) = 1.533, \quad \bar{x}_{II,B} = \frac{1}{2} (16.3 + 16.8) = 16.550 \quad (6.26)$$

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 Reziproker Wert der Varianzen der Einzelwerte berechnet werden. In 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

$$x_{ref} = \frac{\sum_{i=1}^N w_i x_i}{\sum_{i=1}^N w_i}.$$

Wenn die Varianzen der Methode A und B ähnlich sind, so kann eine gemeinsame (gepoolte) Varianz der beiden Methoden berechnet werden (siehe 5. Vorlesung):

$$\text{Var}(x_{ref}) = \frac{\sum_{i=1}^6 (x_{A,i} - \bar{x}_A)^2 + \sum_{i=1}^2 (x_{B,i} - \bar{x}_B)^2}{(6-1) + (2-1)} = 0.1397 \quad (6.27)$$

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 $u_i = \text{Var}(\bar{x}_i)$. Sie ist definiert als

$$\text{Var}(\bar{x}_i) = \frac{\text{Var}(x_i)}{N_i}$$

mit N_i den Anzahl der Wiederholungen für den Mittelwert \bar{x}_i .

Für Beispiel II erhalten wir dann folgende Werte:

$$\text{Var}(\bar{x}_{II,A}) = \frac{1}{6(6-1)} \sum_{k=1}^6 (x_{II,k} - \bar{x}_{II,A})^2 = 0.023778 \quad \text{d.h.} \quad \bar{s}_A = \sqrt{\text{Var}(\bar{x}_{II,A})} = 0.154$$

und

$$\text{Var}(\bar{x}_{II,B}) = \frac{1}{2(2-1)} \sum_{k=7}^8 (x_{II,k} - \bar{x}_{II,B})^2 = 0.062500 \quad \text{d.h.} \quad \bar{s}_B = \sqrt{\text{Var}(\bar{x}_{II,B})} = 0.250$$

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 aufzuschreiben 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 $\bar{s}_{i,c}^2 = \text{Var}(\bar{x}_{S_i,c})$

$$\bar{s}_{i,c}^2 = \bar{s}_j^2 + s_b^2 \quad (6.28)$$

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

$$w_{i,c} = \frac{1}{\bar{s}_{i,c}^2} = \frac{1}{\bar{s}_j^2 + s_b^2}. \quad (6.29)$$

Der Konsenzreferenzwert aus den Werten aller N Partner ist dann

$$x_{\text{ref},c} = \frac{\sum_{i=1}^N w_{i,c} x_i}{\sum_{i=1}^N w_{i,c}}. \quad (6.30)$$

Das Birge-Verhältnis für diese Konsensuswerte ist somit

$$R_B = \sqrt{\frac{1}{N-1} \sum_{i=1}^N \frac{1}{\bar{s}_i^2 + s_b^2} (\bar{x}_i - x_{\text{ref},c})^2}. \quad (6.31)$$

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 Kosten-

funktion zu haben, die sich einfacher linearisieren lässt. Dies geht ganz gut, weil $1^2 = 1$ ist, also $R_B^2 \rightarrow 1$, d.h.

$$\frac{1}{N-1} \sum_{i=1}^N \frac{1}{\bar{s}_i^2 + s_b^2} (\bar{x}_i - x_{ref,c})^2 \rightarrow 1$$

oder genauer ausgedrückt

$$\min_{s_b^2} \left\{ \frac{1}{N-1} \left(\sum_{i=1}^N \frac{1}{\bar{s}_i^2 + s_b^2} (\bar{x}_i - x_{ref,c})^2 \right) - 1 \right\}. \quad (6.32)$$

Diese Optimierungsaufgabe hat eine Kostenfunktion $Q(s_b^2) = \frac{1}{N-1} \left(\sum_{i=1}^N \frac{1}{\bar{s}_i^2 + s_b^2} (\bar{x}_i - x_{ref,c})^2 \right) - 1$ 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

$$Q(v_0 + dv_0) \approx Q_0 + \left. \left(\frac{\partial Q}{\partial v} \right) \right|_{v=v_0} dv_0 \quad (6.33)$$

Mit $Q(v) \rightarrow 0$ setzen wir für den ersten Iterationsschritt

$$Q_0 + \left. \left(\frac{\partial Q}{\partial v} \right) \right|_{v=v_0} dv_0 = 0 \quad (6.34)$$

und lösen dies nach dv_0 auf zu

$$dv_0 = - \left. \left(\frac{Q_0}{\frac{\partial Q}{\partial v}} \right) \right|_{v=v_0} \quad (6.35)$$

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

$$v_{k+1} = v_k + dv_k \quad (6.36)$$

bis Q zu einem Minimum möglichst nahe bei Null konvergiert.

Die Ableitung der Kostenfunktion Q nach der Zwischengruppenvarianz ist

$$\frac{\partial Q}{\partial v} = \frac{1}{N-1} \left(\sum_{i=1}^N \frac{\partial}{\partial v} \frac{1}{\bar{s}_i^2 + v} (\bar{x}_i - x_{ref,c})^2 \right) - 0$$

d.h.

$$\frac{\partial Q}{\partial v} = \frac{1}{N-1} \sum_{j=1}^N (\bar{x}_j - x_{ref,c})^2 \frac{\partial}{\partial v} \frac{1}{\bar{s}_i^2 + v}$$

d.h.

$$\frac{\partial Q}{\partial v} = \frac{1}{N-1} \sum_{i=1}^N (\bar{x}_i - x_{ref,c})^2 \frac{-1}{(\bar{s}_i^2 + v)^2} \quad (6.37)$$

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.

$$v = 112.707 \quad \text{d.h.} \quad s_b = 10.6$$

Die Konsensusvarianz für den Referenzwert können wir dann wie folgt berechnen:

$$\text{Var}(x_{\text{ref}}) = \left(\sum_{i=1}^N \frac{1}{\bar{s}_i^2 + s_b^2} \right)^{-1} = 56.375; \quad s_{\text{ref}} = \sqrt{\text{Var}(x_{\text{ref}})} = 7.5$$

Als Ergebnis ergibt sich damit:

$$x_{\text{ref}} = 9.040 \quad \text{mit} \quad s_{\text{ref}} = 7.5 \quad (6.38)$$

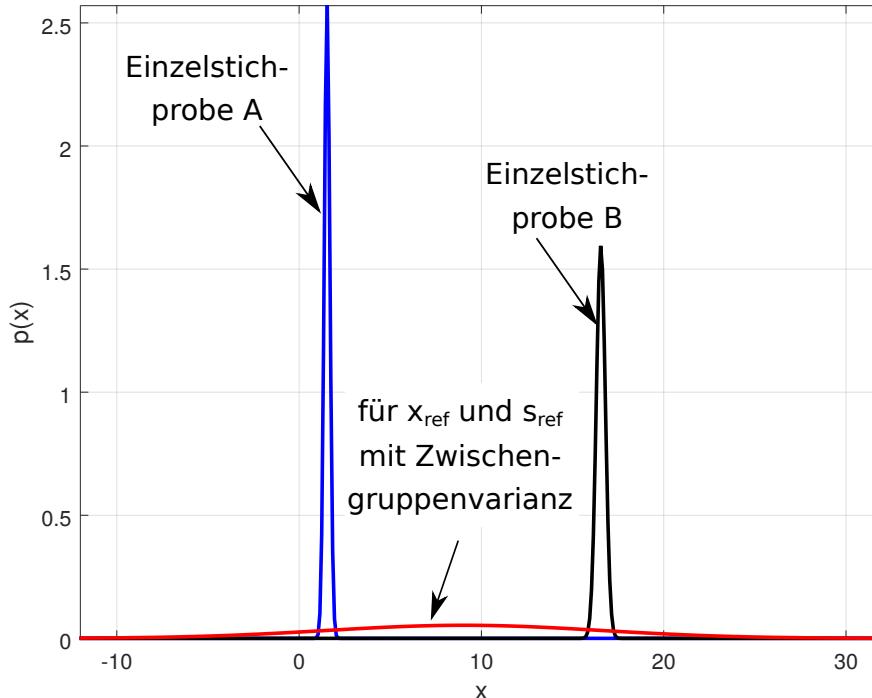


Abbildung 6.2: Wahrscheinlichkeitsdichteverteilungen als Normalverteilungen zu Beispiel II mit blauer Kurve für die Einzelstichprobe A mit $\bar{x}_{II,A} = 1.533$ und $\bar{s}_A = 0.1542$, schwarzer Kurve für die Einzelstichprobe B mit $\bar{x}_{II,B} = 16.550$ und $\bar{s}_B = 0.2500$, sowie die gemeinsame Verteilung nach Hinzufügen der Zwischengruppenvarianz mit $x_{\text{ref}} = 9.040$ und $\sqrt{\text{Var}(x_{\text{ref}})} = 7.5$ als sehr viel breitere und flachere Gaußverteilung dargestellt als rote Kurve.

Der dazu gehörige Quellcode für Gnu-Octave/Matlab sieht wie folgt aus:

```

1 function consensus_fit()
2 % Programm laeuft unter Matlab und Gnu-Octave
3 x_A = [2.0,1.0,1.5,1.8,1.2,1.7];
4 N_A = length(x_A);
5 x_B = [16.3,16.8];

```

```

6 N_B = length(x_B);
7
8 bar_x_A = mean(x_A);
9 bar_x_B = mean(x_B);
10 fprintf('Mittelwerte: %f und %f \n', bar_x_A, bar_x_B);
11
12 % Varianzen der Mittelwerte
13 Var_A_mittel = var(x_A)/N_A;
14 Var_B_mittel = var(x_B)/N_B;
15 fprintf('Varianzen der Mittelwerte innerhalb jeder Stichprobe: \n %f und %f \n',...
16 , Var_A_mittel, Var_B_mittel);
17
18 % Anzahl der Methoden/ Partner M
19 M = 2;
20 % Startwert fuer Zwischengruppenvarianz
21 v_0 = 0;
22
23 % Berechnung des Birge ratio fuer den Startwert
24 w_A = 1/Var_A_mittel;
25 w_B = 1/Var_B_mittel;
26 x_ref = (w_A*bar_x_A + w_B*bar_x_B) / (w_A + w_B);
27 RB = sqrt( (1/(M-1))*(w_A * (bar_x_A - x_ref)^2 + ...
28 w_B * (bar_x_B - x_ref)^2));
29 fprintf('Birge-Verhaeltnis fuer den Startwert %f\n\n',RB)
30 %
31 max_num_iterations = 50;
32 v = v_0;
33 k = 0;
34 while (RB > 1) && (k < max_num_iterations)
35 w_A = 1/(Var_A_mittel + v);
36 w_B = 1/(Var_B_mittel + v);
37
38 x_ref = (w_A*bar_x_A + w_B*bar_x_B) / (w_A + w_B);
39 RBsq = (1/(M-1))*(w_A * (bar_x_A - x_ref)^2 + ...
40 w_B * (bar_x_B - x_ref)^2);
41 RB = sqrt(RBsq);
42 Q = RBsq - 1;
43 dQ_dv = (-1/(M-1))*(w_A^2 * (bar_x_A - x_ref)^2 + ...
44 w_B^2 * (bar_x_B - x_ref)^2);
45 dv = -Q / dQ_dv;
46 v = v + dv;
47 k = k + 1;
48 fprintf('x_ref: %f, RB: %.3f, dv: %e --> v: %.3f\n', x_ref, sqrt(RBsq), dv,v);
49 end

```

```

50 fprintf('\n s_b = %f bei Anzahl Iterationen %d\n', ...
51 sqrt(v-dv), k);
52 end

```

mit dem Output:

```

1 >> consensus_fit
2 Mittelwerte: 1.533333 und 16.550000
3 Varianzen der Mittelwerte innerhalb jeder Stichprobe:
4 0.023778 und 0.062500
5 Birge-Verhaeltnis fuer den Startwert 51.123910
6
7 x_ref: 5.671861, RB: 51.124, dv: 4.312238e-02 --> v: 0.043
8 x_ref: 7.356441, RB: 36.154, dv: 8.619528e-02 --> v: 0.129
9 x_ref: 8.198732, RB: 25.569, dv: 1.721928e-01 --> v: 0.302
10 x_ref: 8.619877, RB: 18.087, dv: 3.435958e-01 --> v: 0.645
11 x_ref: 8.830449, RB: 12.799, dv: 6.840440e-01 --> v: 1.329
12 x_ref: 8.935734, RB: 9.064, dv: 1.355587e+00 --> v: 2.685
13 x_ref: 8.988376, RB: 6.429, dv: 2.661878e+00 --> v: 5.347
14 x_ref: 9.014695, RB: 4.574, dv: 5.132109e+00 --> v: 10.479
15 x_ref: 9.027851, RB: 3.273, dv: 9.539961e+00 --> v: 20.019
16 x_ref: 9.034421, RB: 2.371, dv: 1.649219e+01 --> v: 36.511
17 x_ref: 9.037690, RB: 1.756, dv: 2.470307e+01 --> v: 61.214
18 x_ref: 9.039294, RB: 1.357, dv: 2.797615e+01 --> v: 89.190
19 x_ref: 9.040038, RB: 1.124, dv: 1.861186e+01 --> v: 107.802
20 x_ref: 9.040319, RB: 1.022, dv: 4.691663e+00 --> v: 112.494
21 x_ref: 9.040375, RB: 1.001, dv: 2.129841e-01 --> v: 112.707
22 x_ref: 9.040377, RB: 1.000, dv: 4.038510e-04 --> v: 112.707
23 x_ref: 9.040377, RB: 1.000, dv: 1.446530e-09 --> v: 112.707
24 x_ref: 9.040377, RB: 1.000, dv: 0.000000e+00 --> v: 112.707
25
26 s_b = 10.616355 bei Anzahl Iterationen 18

```


Literaturverzeichnis

zu Kapitel 6:

- [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
- [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/utils/common/pdf/CC/CCPR/CCPR-G2.pdf>
- [Pau82] R.C. Paule, J. Mandel: Consensus Values and Weighting Factors, J. Res. NBS, Vol.87, No.5,377-385 (1982)

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 nichtdeterministische Verhalten eines Systems wird in vielen Fällen von nicht vorhersehbaren, kleinen Störeinflüssen und Auflösungsbegrenzungen beim Beobachtungsvorgang verursacht. Es gibt aber auch Prozesse, bei denen der physikalische Vorgang, der zu untersuchen ist, selber stochastischen Charakter hat, wie beispielsweise bei radioaktiven Zerfällen, Streuprozessen 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 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 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

$$Y = f(\mathbf{X}) = f(\mathbf{X})|_{\bar{\mathbf{x}}} + \sum_{i=1}^N \underbrace{\frac{\partial f}{\partial X_i}}_{c_i} \Big|_{\bar{\mathbf{x}}} \Delta X_i \quad (7.6)$$

- 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. (1.57) aus Abschnitt 1.3 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 physikalische 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*)

$$\mathbf{x}_1 = (x_{1,1}, \dots, x_{N,1})^\top, \dots, \mathbf{x}_J = (x_{1,J}, \dots, x_{N,J})^\top$$

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.

→ Dieses Verfahren wird Kapitel 10 dargelegt.

- 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 Kapitel 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 bayesischer Methoden

→ Kapitel 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 der in JCGM 101 spezifizierten Monte-Carlo-Methode ermöglicht, dass man nicht auf die Bestimmung der Unsicherheit expliziter, univarter, linearisierbarer Messgrößen begrenzt ist. Es bietet die Möglichkeit auch die Unsicherheit für implizite, multivariate Messgrößen, deren direkte Eingangsgrößen beliebig verteilt sein können zu ermitteln. Die Verteilung der Größen können eine von Null verschiedene Skewness aufweisen, sie können die Form einer U-Verteilung haben. Die Größen kön-

nen irgendwelchen 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 Abb. 3.2 und 3.3 mit dem der Abb. 3.2 in Kapitel 3.7 vergleichen, erkennen wir, dass die Modellparameter bei Abb. 3.2 und 3.3 in etwa die gleiche Skalierung aufweisen, während die in Abb. 3.2 stark unterschiedlich sind. In Abb. 3.2 und 3.3 sehen wir eine schöne, runde Kuhle und in Abb. 3.3 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, 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 Kapitel 2 gesehen, in der ein Vergleich zwischen einer Regressiongeraden und einem Polynom 6. Grades gemacht wurde, siehe dort das in Abb. 2.2 dargestellte Beispiel.

7.2 M.U.F. für univariate, explizite indirekte Größe, linear von direkten Größen abhängig

In der ersten Vorlesung, in Abschnitt 1.3, 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. (1.57) 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 1.3 Gl. (1.57) 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 Korrelationkoeffizienten, definiert in Abschnitt 1.3 Gl. (1.37) $\rho_{i,k} = \frac{\text{Cov}(X_i, X_k)}{\sqrt{\text{Var}(X_i) \text{Var}(X_k)}}$, angegeben werden und die Standardabweichungen $\sigma_i = \sqrt{\text{Var}(X_i)}$, so dass das Unsicherheitsfortpflanzungsgesetz auch in dieser Form aufgeschrieben wird

$$\text{Var}(Y) = \sigma_Y^2 = \sum_{i=1}^N c_i^2 \sigma_i^2 + 2 \sum_{i=1}^{N-1} \sum_{k=i+1}^N c_i c_k \rho_{i,k} \sigma_i \sigma_k \quad (7.14)$$

Die aus dem Fortpflanzungsgesetz resultierende Messunsicherheit σ_Y wird auch *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 Standardabweichungen σ als Kürzel für *uncertainty* oder *Unsicherheit*, und heißt in diesem Zusammenhang *Standardunsicherheit* (engl. *standard uncertainty*).

$$\begin{aligned} u^2(Y) &= \sum_{i=1}^N c_i^2 \operatorname{Var}(X_i) + 2 \sum_{i=1}^{N-1} \sum_{k=i+1}^N c_i c_k \operatorname{Cov}(X_i, X_k) \\ &= \underbrace{\sum_{i=1}^N c_i^2 u^2(X_i)}_{\text{unkorrelierter Fall}} + 2 \underbrace{\sum_{i=1}^{N-1} \sum_{k=i+1}^N c_i c_k \rho_{i,k} u(X_i) u(X_k)}_{\text{Mischterme: korrelierter Fall}}. \end{aligned} \quad (7.15)$$

mit

$$\operatorname{Cov}(X_i, X_i) \equiv \operatorname{Var}(X_i) \equiv \sigma_i^2 \equiv u^2(X_i) \quad (7.16)$$

und

$$\operatorname{Cov}(X_i, X_k) \equiv \sigma_{i,k} \equiv u(X_i, X_k) \quad (7.17)$$

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:

$$\operatorname{Cov}(X_i, X_k) = \frac{1}{J-1} \sum_{j=1}^J (X_{i,j} - \bar{x}_i)(X_{k,j} - \bar{x}_k) \quad \text{und mit} \quad \bar{x}_i = \frac{1}{J} \sum_{j=1}^J X_{i,j}$$

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 Modelfunktion, 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

$$c_1 := \frac{\Delta Y}{\Delta X_1} = X_2 \quad (7.23)$$

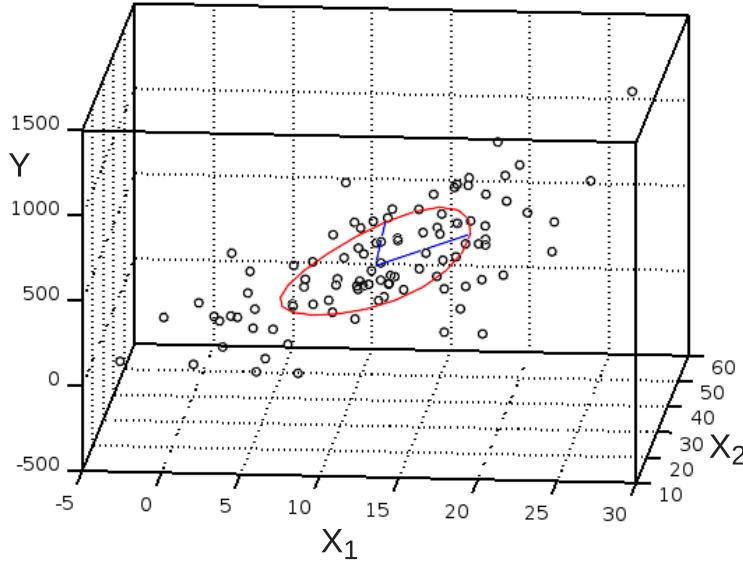


Abbildung 7.1: Eine anschauliche Deutung für die pythagoräische Addition der Standardabweichungen liefert die Betrachtung der mit roter Kurve dargestellten Ellipse. Die blauen Geraden repräsentieren die Abstände vom Mittelpunkt als Halbmesser der Ellipse, die aus den Standardabweichungen jeweils von X_1 und X_2 berechnet wurden.

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 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 $c_i = \frac{\partial f}{\partial X_i} \Big|_{\bar{x}}$ dar. 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 sensistiv das Modell an der Position \bar{x} ist.

Gegeben sei ein univariates, expliziten Modell $Y = f(\mathbf{X})$, das linearisierbar ist also in Taylorreihe entwickelbar bis zum linearen Term mit dem Erwartungswertvektor $\bar{\mathbf{x}}$ als Entwicklungspunkt

$$Y = f(\mathbf{X}) \approx f(\mathbf{X})|_{\bar{\mathbf{x}}} + \sum_{i=1}^N \frac{\partial f}{\partial X_i} \Big|_{\bar{\mathbf{x}}} \Delta X_i \quad \text{und} \quad \mathbf{X} = \bar{\mathbf{x}} + \Delta \mathbf{X} \quad (7.25)$$

Die direkten Messgrößen X_i sind Zufallsgrößen. Bei der Linearisierung nehmen die Abweichungen

ΔX_i der direkten Größen von ihren Erwartungswerten $E(X_i) = \bar{x}_i$ oder als Vektor mit den Erwartungswerten aller direkten Größen $\bar{\mathbf{x}}$ die Rolle der Zufallsgrößen ein. Ihr Erwartungswert ist dann Null $E(\Delta X_i) = 0$.

Der Erwartungswert der indirekten Messgröße Y wird durch Einsetzen der Erwartungswerte der direkten Messgrößen in die Modellgleichung $Y = f(\mathbf{X})$ gewonnen

$$E(Y) = f(\mathbf{X})|_{\bar{\mathbf{x}}} \quad (7.26)$$

Zu der indirekten Messgröße soll die Varianz ermittelt werden

$$\text{Var}(Y) = \int (Y - E(Y))^2 p(Y) dY = \int (Y^2 - 2YE(Y) + E(Y)^2) p(Y) dY \quad (7.27)$$

mit $\int YE(Y)p(Y)dY = E(Y) \int Yp(Y)dY = E(Y)E(Y)$ wird dies zu

$$\text{Var}(Y) = \int (Y^2 - E(Y)^2) p(Y) dY = \left(\int Y^2 p(Y) dY \right) - E(Y)^2,$$

so dass gilt

$$\text{Var}(Y) = \int (Y^2 - E(Y)^2) p(Y) dY = \int Y^2 p(Y) dY - \int E(Y)^2 p(Y) dY. \quad (7.28)$$

Hier wird der Ansatz Gl. (7.25) eingesetzt und mit $p(Y)$ aus $p(X_1, \dots, X_N) = p(\mathbf{X})$ und mit $d\mathbf{X} = dX_1 \dots dX_N$, sowie mit $\int E(Y)^2 p(Y) dY = E(Y)^2$

$$\begin{aligned} \text{Var}(Y) &= \int \left(f(\mathbf{X})|_{\bar{\mathbf{x}}} + \sum_{i=1}^N \frac{\partial f}{\partial X_i} \Big|_{\bar{\mathbf{x}}} \Delta X_i \right)^2 p(\mathbf{X}) d\mathbf{X} - E(Y)^2 \\ &= \int \left[f(\mathbf{X})|_{\bar{\mathbf{x}}}^2 + 2 f(\mathbf{X})|_{\bar{\mathbf{x}}} \sum_{i=1}^N \frac{\partial f}{\partial X_i} \Big|_{\bar{\mathbf{x}}} \Delta X_i + \left(\sum_{i=1}^N \frac{\partial f}{\partial X_i} \Big|_{\bar{\mathbf{x}}} \Delta X_i \right)^2 \right] p(\mathbf{X}) d\mathbf{X} - E(Y)^2 \\ &= \int \left[2 f(\mathbf{X})|_{\bar{\mathbf{x}}} \sum_{i=1}^N \frac{\partial f}{\partial X_i} \Big|_{\bar{\mathbf{x}}} \Delta X_i + \left(\sum_{i=1}^N \frac{\partial f}{\partial X_i} \Big|_{\bar{\mathbf{x}}} \Delta X_i \right)^2 \right] p(\mathbf{X}) d\mathbf{X} \\ \text{Var}(Y) &= \int \left[2 f(\mathbf{X})|_{\bar{\mathbf{x}}} \sum_{i=1}^N \frac{\partial f}{\partial X_i} \Big|_{\bar{\mathbf{x}}} \Delta X_i + \left(\sum_{i=1}^N \frac{\partial f}{\partial X_i} \Big|_{\bar{\mathbf{x}}} \Delta X_i \right)^2 \right] p(\mathbf{X}) d\mathbf{X} \\ &= 2 \int f(\mathbf{X})|_{\bar{\mathbf{x}}} \sum_{i=1}^N \frac{\partial f}{\partial X_i} \Big|_{\bar{\mathbf{x}}} \Delta X_i p(\mathbf{X}) d\mathbf{X} + \int \left(\sum_{i=1}^N \frac{\partial f}{\partial X_i} \Big|_{\bar{\mathbf{x}}} \Delta X_i \right)^2 p(\mathbf{X}) d\mathbf{X} \\ &= 2 f(\mathbf{X})|_{\bar{\mathbf{x}}} \sum_{i=1}^N \frac{\partial f}{\partial X_i} \Big|_{\bar{\mathbf{x}}} \underbrace{\int \Delta X_i p(\mathbf{X}) d\mathbf{X}}_{=0} + \int \left(\sum_{i=1}^N \frac{\partial f}{\partial X_i} \Big|_{\bar{\mathbf{x}}} \Delta X_i \right)^2 p(\mathbf{X}) d\mathbf{X} \end{aligned}$$

$$\text{Var}(Y) = \int \left(\sum_{i=1}^N \frac{\partial f}{\partial X_i} \Big|_{\bar{\mathbf{x}}} \Delta X_i \right)^2 p(\mathbf{X}) d\mathbf{X} = \text{Var} \left(\sum_{i=1}^N \frac{\partial f}{\partial X_i} \Big|_{\bar{\mathbf{x}}} \Delta X_i \right) \quad (7.30)$$

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 $\frac{\partial f}{\partial X_i}$ an der Stelle der geschätzten Erwartungswerte sind, d.h. $c_i = \frac{\partial f}{\partial X_i} \Big|_{\bar{\mathbf{x}}}$

$$\text{Var}(Y) = \text{Var} \left(\sum_{i=1}^N c_i \Delta X_i \right) \quad (7.31)$$

$$\text{Var}(Y) = \sum_{i=1}^N c_i^2 \text{Var}(\Delta X_i) + 2 \sum_{i=1}^{N-1} \sum_{k=i+1}^N c_i c_k \text{Cov}(\Delta X_i, \Delta X_k) \quad (7.32)$$

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 $\frac{\partial f}{\partial X_i} \Big|_{\bar{\mathbf{x}}}$ oder $\frac{\partial f}{\partial X_i} \Big|_{\bar{x}_1, \dots, \bar{x}_N}$. Im GUM wird angemerkt, dass die saloppe Notation $\frac{\partial f}{\partial x_i}$ für $\frac{\partial f}{\partial X_i}$ unter Verwendung der Schätzwerte x_i zu den X_i gewählt wird. 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

Am Ende von Kapitel 4 haben wir das Überdeckungsintervall eingeführt und nach Einführung der Quantile der t-Verteilung in Kapitel 5 genauer behandelt, als *Credible interval* und als *Vertrauensintervall*. Die Intervallgrenzen des Vertrauensintervalls haben wir auch als Produkt aus einem Quantil und einer Standardabweichung repräsentiert, im Fall, dass eine Normalverteilung zugrunde gelegt wird in der Form

$$[\bar{x} - z_{1-\frac{1}{2}\alpha} \sigma_X, \bar{x} + z_{1-\frac{1}{2}\alpha} \sigma_X] \quad (7.33)$$

und im Fall, dass eine *t*-Verteilung zugrunde gelegt wird

$$[\bar{x} - t_{1-\frac{1}{2}\alpha, \nu} \sigma_X, \bar{x} + t_{1-\frac{1}{2}\alpha, \nu} \sigma_X]. \quad (7.34)$$

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

$$p_{\text{L},i}(\{X_{i,1}, \dots, X_{i,J}\}|X_i, s_i) = \prod_{j=1}^J \frac{1}{\sqrt{2\pi} s_i} e^{-\frac{1}{2} \left(\frac{X_{i,j}-X_i}{s_i} \right)^2} = l(X_i, s_i | \{X_{i,1}, \dots, X_{i,J}\}) \quad (7.35)$$

was zwei Verteilungen liefert, und als dritte Verteilung noch

$$p_{\text{L,y}}(X_1, X_2 | Y, s) = \frac{1}{\sqrt{2\pi} s} e^{-\frac{1}{2} \left(\frac{Y-X_1 X_2}{s} \right)^2} \quad (7.36)$$

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 kolmogoroff'sche 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

$$u^2(Y_l) = \sum_{i=1}^N \left(\frac{\partial f_l(\mathbf{X})}{\partial X_i} \Big|_{\bar{\mathbf{x}}} \right)^2 u^2(X_i) + 2 \sum_{i=1}^{N-1} \sum_{k=i+1}^N \frac{\partial f_l(\mathbf{X})}{\partial X_i} \Big|_{\bar{\mathbf{x}}} \frac{\partial f_l(\mathbf{X})}{\partial X_k} \Big|_{\bar{\mathbf{x}}} \rho_{i,k} u(X_i) u(X_k) \quad (7.37)$$

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, die in Abschnitt 2.4 dargelegt wird, siehe Gl. (2.72) bzw. Gl. (2.82) mit Gl. (2.84). 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

$$\text{Cov}(\boldsymbol{\theta}) = (\mathbf{X}^\top \mathbf{X})^{-1} \text{Var}(\boldsymbol{\varepsilon}) \quad (7.38)$$

mit J für den Stichprobenumfang und M für die Anzahl der Modellparameter und mit

$$\text{Var}(\boldsymbol{\varepsilon}) = \frac{1}{J-M} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\theta}})^\top (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\theta}}). \quad (7.39)$$

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

$$\text{Var}(s_y^2) = \text{Var}(c_1^2 s_1^2) + \text{Var}(c_2^2 s_2^2) \quad (7.40)$$

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

$$Q_i = \nu_i \left(\frac{s_i}{\sigma_i} \right)^2 \sim \chi^2(\nu_i). \quad (7.41)$$

Der Erwartungswert für Q_i ist

$$\text{E}(Q_i) = \int_0^\infty Q p_{\chi^2}(Q) \, dQ = \nu_i. \quad (7.42)$$

Die Varianz $\text{Var}(Q_i)$ ist

$$\text{Var}(Q_i) = \int_0^\infty (Q - \nu_i)^2 p_{\chi^2}(Q) \, dQ = 2\nu_i. \quad (7.43)$$

Die Herleitungen zu den beiden Gln. (7.42) und (7.43) befinden sich in Anhang 7.7.

Ferner gilt allgemein für die Varianz einer Zufallszahl R multipliziert mit einem konstanten Faktor a

$$\begin{aligned} \text{Var}(aR) &= \int_0^\infty (aR - \text{E}(aR))^2 p(R) \, dR \\ &= \int_0^\infty a^2 (R - \text{E}(R))^2 p(R) \, dR \\ &= a^2 \int_0^\infty (R - \text{E}(R))^2 p(R) \, dR \\ &= a^2 \text{Var}(R) \end{aligned} \quad (7.44)$$

so dass

$$\text{Var}(Q_i) = \text{Var}\left(\nu_i \left(\frac{s_i}{\sigma_i} \right)^2\right) = \left(\frac{\nu_i}{\sigma_i^2} \right)^2 \text{Var}(s_i^2) \quad (7.45)$$

mit Gl. (7.43) wird dies zu

$$2\nu_i = \left(\frac{\nu_i}{\sigma_i^2} \right)^2 \text{Var}(s_i^2). \quad (7.46)$$

Das heißt

$$2\nu_i = \frac{\nu_i^2}{\sigma_i^4} \text{Var}(s_i^2) \quad (7.47)$$

also

$$\text{Var}(s_i^2) = \frac{2\sigma_i^4}{\nu_i}. \quad (7.48)$$

so dass nach Satterthwaite aus Gl. (7.40) 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

$$\frac{\sigma_y^4}{\nu_y} = \frac{c_1^4\sigma_1^4}{\nu_1} + \frac{c_2^4\sigma_2^4}{\nu_2} \quad (7.49)$$

so dass

$$\nu_y = \frac{\sigma_y^4}{\frac{(c_1\sigma_1)^4}{\nu_1} + \frac{(c_2\sigma_2)^4}{\nu_2}}. \quad (7.50)$$

Wenn mit Gleichung (7.42) 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

$$E(s^2) = E\left(\frac{\sigma^2}{\nu}Q\right) \quad (7.51)$$

und mit $\frac{\sigma^2}{\nu}$ als konstanter Faktor bezüglich der Wahrscheinlichkeitsdichte und mit Gl. (7.42) erhalten wir

$$E(s^2) = \frac{\sigma^2}{\nu} E(Q) = \frac{\sigma^2}{\nu} \nu = \sigma^2 \quad (7.52)$$

und wir setzen in Gl. (7.50) 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

$$Y = y \pm t_{1-\alpha/2, \nu_y} s_y. \quad (7.53)$$

Verallgemeinert für N direkte Größen ist dies

$$\nu_{\text{eff}} = \frac{\sigma_y^4}{\sum_{i=1}^N \frac{(c_i\sigma_i)^4}{\nu_i}}. \quad (7.54)$$

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

$$\bar{Y} = \int_{-\infty}^{\infty} Y p(Y| \dots) dY \quad (7.55)$$

und deren Varianz

$$\text{Var}(Y) = \int_{-\infty}^{\infty} (Y - \bar{Y})^2 p(Y| \dots) dY \quad (7.56)$$

ist, so ist der *Erweiterungsfaktor* k für symmetrische Verteilungen mit $z_{\max} - \bar{Y} = \bar{Y} - z_{\min}$

$$k = \frac{z_{\max} - \bar{Y}}{\sqrt{\text{Var}(Y)}} \quad (7.57)$$

was für klassische Berechnungsverfahren dem t-Quantil $t_{1-\alpha/2, \nu_{\text{eff}}}$ entspricht

$$k = t_{1-\alpha/2, \nu_{\text{eff}}} \cdot \quad (7.58)$$

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 erweiterten Unsicherheiten sind das Produkt aus dem Erweiterungsfaktor k , der in der klassischen Statistik das t -Quantil ist, und der Unsicherheit u .

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, deren Konzept wir in Abschnitt 4.2 vorgestellt haben, 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 Temperatur 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 Beispiele 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

$$Y = f(X_K, X_M) = X_K X_M \quad (7.59)$$

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 Messignal 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} wesentlich kleiner ist als die Streuung der Messvorgänge beim Erfassen der Rohdaten und der Sensorkalibrierung. Die beruht auf der Annahme das sich der Sensor linear verhält, also dass Linearitätsabweichungen vernachlässigbar klein sind. Der Ansatz ist

$$p(Y, X_K, X_M | (X_{M,1}, \dots, X_{M,J}), K_0, s_K) \propto \underbrace{e^{-\frac{1}{2} \left(\frac{Y - X_K X_M}{\sigma_{\text{winzig}}} \right)^2}}_{\text{Modellprior}} \underbrace{e^{-\frac{1}{2} \left(\frac{X_K - K_0}{s_K} \right)^2}}_{\text{Prior}} \underbrace{\prod_{j=1}^J e^{-\frac{1}{2} \left(\frac{X_M - X_{M,j}}{s_M} \right)^2}}_{\text{Likelihood}} \quad (7.60)$$

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 Unsicher-

heitsfortpflanzung gemäß (7.15) und zum anderen aus der Marginalverteilung zu Gl. (7.60):

$$\begin{aligned} p(Y|(X_{M,1}, \dots, X_{M,J}), K_0, s_K) &= \\ \frac{\int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{Y-X_K X_M}{\sigma_{\text{winzig}}}\right)^2} e^{-\frac{1}{2}\left(\frac{X_K-K_0}{s_K}\right)^2} \prod_{j=1}^J e^{-\frac{1}{2}\left(\frac{X_M-X_{M,j}}{s_M}\right)^2} dX_M dX_K}{\int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{Y-X_K X_M}{\sigma_{\text{winzig}}}\right)^2} e^{-\frac{1}{2}\left(\frac{X_K-K_0}{s_K}\right)^2} \prod_{j=1}^J e^{-\frac{1}{2}\left(\frac{X_M-X_{M,j}}{s_M}\right)^2} dX_M dX_K dY} \end{aligned} \quad (7.61)$$

Wir erhalten den Schätzwert \bar{y} aus

$$\bar{y} = \int_{-\infty}^{\infty} Y p(Y|(X_{M,1}, \dots, X_{M,J}), K_0, s_K) dY \quad (7.62)$$

und für die Bestimmung der Intervallgrenzen y_1, y_2 für das Überdeckungsintervall berechnen wir die kumulierte Verteilung

$$P(Y) = \int_{-\infty}^Y p(Y'|(X_{M,1}, \dots, X_{M,J}), K_0, s_K) dY' \quad (7.63)$$

und dann für 95% Wahrscheinlichkeit

$$P(y_1) = 0.025 \Leftrightarrow y_1 = P^{-1}(0.025) \quad \text{und} \quad P(y_2) = 0.975 \Leftrightarrow y_2 = P^{-1}(0.975). \quad (7.64)$$

Durch numerische Integration der Gl. (7.62) erhalten wir

$$\bar{y} = 46.278 \text{ E}$$

und für das *Credible Interval* durch numerische Integration von Gl. (7.63)

$$[36.855 \text{ E}, 56.185 \text{ E}] = [\bar{y} - 9.423, \bar{y} + 9.907] \text{ E}$$

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.

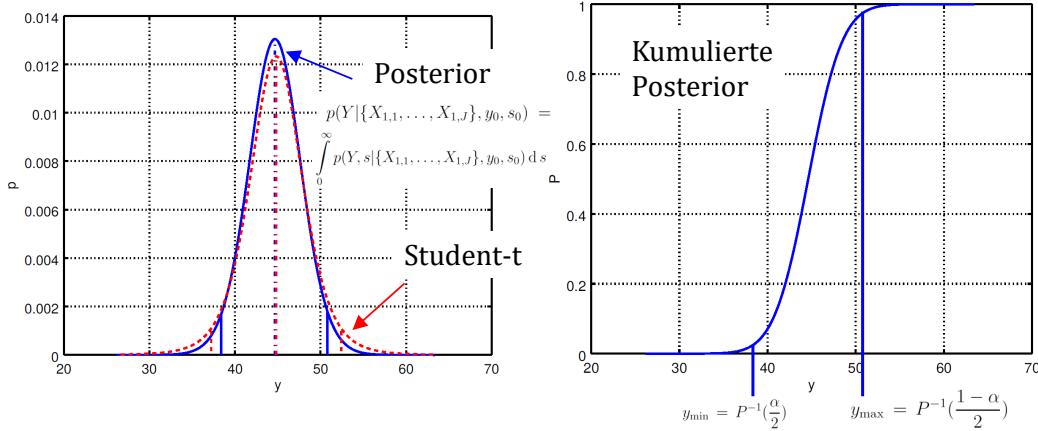


Abbildung 7.2: Wahrscheinlichkeitsdichtefunktionen und kumulierte Posterior für die Bestimmung des Überdeckungsintervalls einer indirekten Messgröße

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 (7.54), 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 $K_0 = 0.0925 \frac{\text{E}}{\text{V}}$, $s_K = 0.009 \frac{\text{E}}{\text{V}}$, $\bar{x}_M = 500.83 \text{ V}$, $s_M = 17.89 \text{ V}$

$$\nu_y = u(Y)^4 \left(\frac{K_0^4 s_M^4}{\nu_M} + \frac{\bar{x}_M^4 s_K^4}{\nu_K} \right)^{-1} = 52.58 \approx 52$$

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.

```

1 >> tinv(0.975,52)
2 ans = 2.0066
3 >> tinv(0.975,52.58)
4 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 Freiheits-

graden also auch wie folgt geschrieben werden

$$\nu_y = \lfloor u(Y)^4 \left(\sum_{i=1}^N \frac{(c_i s_i)^4}{\nu_i} \right)^{-1} \rfloor \quad (7.65)$$

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 $U = 9.635 \text{ E}$ 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.

$$\lim_{\nu_i \rightarrow \infty} \left\{ \frac{(c_i s_i)^4}{\nu_i} \right\} = 0 \quad (7.66)$$

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

$$\nu \sim \frac{\sigma_i^2}{\text{Var}(s_i)} \quad (7.67)$$

betrachtet, siehe 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 (7.67) die Anzahl der Freiheitsgrade abgeschätzt mit

$$\nu \approx \frac{1}{2\alpha^2}. \quad (7.68)$$

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 Wahrscheinlichkeitsdichteverteilun-

gen 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 Rechnen mit Wahrscheinlichkeitsdichteveerteilungen 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'sche 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 Integrationen 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 *kolmogoroff'schen 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:

$$d = (4.997 \pm 0.011) \mu\text{m} \quad \text{mit} \quad k = 2.1, \nu = 26 \quad (7.69)$$

Mit dem induktiven Wegaufnehmer wurden auf dem Bezugsnormal folgende Stufenhöhen in der Di-

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

$$d_W = \frac{d}{U_B} U_W \quad (7.70)$$

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.

$$c_B = \left. \frac{\partial}{\partial U_B} d_W \right|_{\bar{U}_B, \bar{d}, \bar{U}_W} \quad (7.71)$$

Die Sensitivitäten c_d und c_w von d_W bezüglich Änderungen von d und U_W sind aufgrund des linearen Zusammenhangs genau

$$c_d = \left. \frac{U_W}{U_B} \right|_{\bar{U}_B, \bar{U}_W} \quad c_w = \left. \frac{d}{U_B} \right|_{\bar{U}_B, \bar{d}} \quad (7.72)$$

definiert. Die kombinierte Varianz unter der Voraussetzung, dass die gemessenen Größen und die Angabe aus dem Kalibrierschein unkorreliert sind, ist

$$s_{dW}^2 = (c_d s_d)^2 + (c_B s_B)^2 + (c_w s_w)^2 \quad (7.73)$$

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.

- Ermitteln Sie aus den Angaben des Kalibrierscheins (hier Gl. (13.6)) 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.

g) Bestimmen Sie das vollständige Messergebnis.

7.7 Anhang: Erwartungswerte zur χ^2 -Verteilung

Die χ^2 -Verteilung ist definiert durch

$$p(Q, \nu) := \frac{Q^{\frac{\nu}{2}-1} e^{-\frac{Q}{2}}}{\Gamma(\frac{\nu}{2}) 2^{\frac{\nu}{2}}} \quad \nu \in \mathbb{N} \quad (7.74)$$

mit

$$\Gamma\left(\frac{\nu}{2}\right) = \sqrt{\pi} \frac{(\nu-2)!!}{2^{\frac{\nu-1}{2}}} \quad (7.75)$$

und

$$\nu!! = \nu(\nu-2)(\nu-4)\dots 4 \cdot 2. \quad (7.76)$$

Der Erwartungswert für die Zufallsgröße Q ist

$$E(Q) = \int_0^\infty Q' p(Q', \nu) dQ' \quad (7.77)$$

und nach Integration mit Computeralgebrassoftware (beispielsweise Maxima)

$$E(Q) = \frac{2\Gamma\left(\frac{\nu}{2} + 1\right)}{\Gamma\left(\frac{\nu}{2}\right)} \quad (7.78)$$

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

$$\begin{aligned} 2^{\frac{\Gamma(\nu/2+1)}{\Gamma(\nu/2)}} &= 2^{\frac{(\nu)!!}{(\nu-2)!!}} \frac{2^{(\nu-1)/2}}{2^{(\nu-1)/2+1}} \\ &= 2^{\frac{(\nu)!!}{(\nu-2)!!}} \frac{1}{2} \\ &= \nu \end{aligned} \quad (7.79)$$

Der Erwartungswert für die Varianz $\text{Var}(Q)$ von Q ist

$$\text{Var}(Q) = \int_0^\infty (Q' - \nu)^2 p(Q', \nu) dQ' \quad (7.80)$$

und nach Integration mit dem Computeralgebraprogramm Maxima

$$\text{Var}(Q) = \nu^2 - 2\nu \frac{\Gamma(\nu/2+1)}{\Gamma(\nu/2)} + 2 \frac{\Gamma(\nu/2+2)}{\Gamma(\nu/2)} \quad (7.81)$$

d.h.

$$\text{Var}(Q) = \nu^2 - 2\nu^2 + \nu(\nu+2) = 2\nu \quad (7.82)$$

7.8 Anhang: Beispiel indirekte Größe aus Produkt zweier direkter Größen

Gnu-Octave Code zu dem Beispiel aus Abschnitt 7.5

```

1 function bayes_indirect_quantity_product()
2 % Kalibrierfaktor
3 K_faktor_0 = 0.0925;
4 sigma_K = 0.0090;
5 %
6 %
7 JM = 9;
8 sig = 12;
9 mue = 500;
10 data_M = [479.58; 526.47; 516.77; 522.01; 506.61; 497.99; 481.71; 484.90; 491.41];
11 %
12 %
13 xK = [-4*sigma_K:0.005:4*sigma_K] + K_faktor_0;
14 nK = length(xK)
15 xM = [-4*sig:0.05:4*sig] + mue;
16 nM = length(xM)
17 y = [25:0.005:65];
18 nY = length(y)
19 %
20 std_M = std(data_M);
21 std_KM = K_faktor_0 * std_M;
22 printf('s_M = %1.4f, s_KM = %1.4f\n', std_M, std_KM);
23 %
24 % Messungen in Volt
25 XM = data_M * ones(1,nM) - ones(JM,1) * xM;
26 %
27 % Das Modell: y = f(xK, xM)
28 % fuer jedes xK und jedes xM kombiniert
29 % y_{KM,i,j} = xK_i * xM_j fuer alle i=1,...,nK und j=1,...,nM
30 y_KM_matrix = xK' * xM;
31 % in einen langen Spaltenvektor gebracht
32 y_KM = y_KM_matrix(:);
33 nKM = length(y_KM);
34 delta_y = y_KM * ones(1,nY) - ones(nKM, 1)*y;
35 %
36 % direkte Groessen
37 % Kalibrierfaktor
38 p_K = exp(-0.5 * ( (xK - K_faktor_0)/sigma_K ).^2 );
39 p_K = p_K / sum(p_K);
40 % Messungen: Likelihood

```

```

41 p_M_matrix = exp(-0.5 * ( XM/std_M ).^2 );
42 % Summe ueber alle Messungen
43 p_M_2 = sum( p_M_matrix );
44 p_M_2 = p_M_2 / sum(p_M_2);
45 %
46 p_KM_matrix = p_K' * p_M_2;
47 % in einen langen Spaltenvektor gebracht
48 p_KM_2 = p_KM_matrix(:);
49 %
50 % Modellprior
51 p_y2 = exp(-0.5 * ( delta_y/(std_KM*0.003) ).^2 );
52 % Posterior
53 post_matrix = p_y2 .* (p_KM_2 * ones(1,nY));
54 % Normierung des Posteriors
55 sumpost = sum(post_matrix(:));
56 post_matrix = post_matrix/sumpost;
57 %
58 % Marginalverteilung durch Summation ueber xM, xK
59 posterior3 = sum(post_matrix);
60 %
61 % kumulative Marginale-PDF also die CDF
62 cdf3 = cumsum(posterior3);
63 %
64 figure(200);
65 plot( y, cdf3, 'b-', 'linewidth', 2);
66 xlabel('indirekte Groesse / Einheit', 'fontsize', 14);
67 ylabel('cdf', 'fontsize', 14);
68 set(gca, 'fontsize', 12);
69 %
70 % Ergebnis
71 y_bar = sum( y.*posterior3 );
72 [dmy, imin] = min( abs(cdf3-0.025) )
73 y1 = y(imin);
74 [dmy, imin] = min( abs(cdf3-0.975) )
75 y2 = y(imin);
76 printf('Credible interval = [%1.4f, %1.4f]\n', y1, y2);
77 printf('y_bar = %1.4f + (%1.4f) + (%1.4f)\n', y_bar, y1-y_bar, y2-y_bar);
78 % —
79 % Vergleich mit Fortpflanzungsgesetz
80 % —
81 xMbar = mean(data_M);
82 printf('mean(X_M): %1.4f\n', xMbar);
83 printf('s_M = %1.4f, s_KM = %1.4f\n', std_M, std_KM);
84 u = sqrt((K_faktor_0*std_M)^2 + (xMbar*sigma_K)^2);

```

```
85 printf('u(Y) = %1.4f\n',u);
86 hlp = (K_faktor_0*std_M)^4/8 + (xMbar*sigma_K)^4/45;
87 nu_eff = u^4/hlp
88 t = tinv(0.975,floor(nu_eff))
89 printf('U = %1.5f \n', t*u);
90 printf('ymean = %1.3f, [%1.3f,%1.3f]\n', ...
91 K_faktor_0*xMbar, K_faktor_0*xMbar-t*u,K_faktor_0*xMbar+t*u);
92 end
```

Kapitel 8

Numerische Werkzeuge für Messunsicherheitsberechnungen

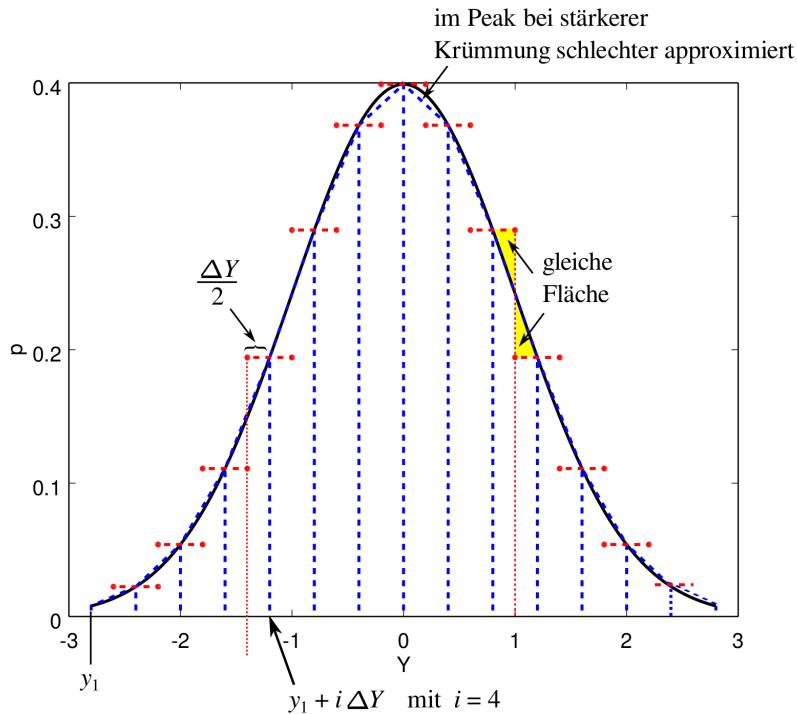


Abbildung 8.1: Trapezapproximation für die Berechnung der Fläche unter einer Kurve zur numerischen Integration.

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 Stücken 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 Wikipedia-Seite 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 Bespielen 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 samplen. 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, bei denen Zufallszahlen verwendet werden, 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.

Um rechentechnisch mit Zufallszahlen zu arbeiten, müssen diese synthetisch generiert werden. Synthetisch gerechnete Zufallszahlen beruhen auf definierten Algorithmen und sind deshalb nicht wirklich zufällig, sondern deterministisch. Sie werden deshalb *Pseudozufallszahlen* genannt. Ein deterministischer Algorithmus berechnet eine Zahlenfolge, die nur 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 Programmertools 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. In Python sind diese in der zu `numpy` gehörenden Bibliothek `random` zu finden. Mit `randn` werden standardnormalverteilte Zufallszahlen erzeugt:

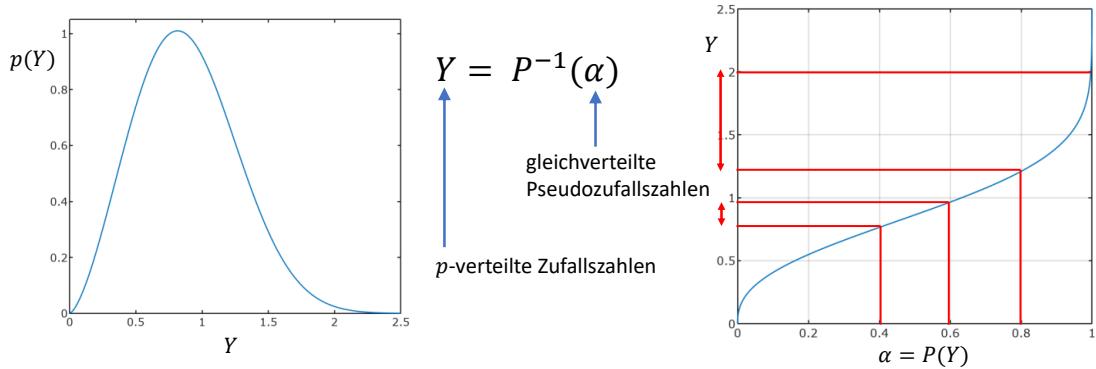


Abbildung 8.2: Eine gleichverteilte Größe wie beispielsweise eine Wahrscheinlichkeit α wird über die inverse Funktion P^{-1} der kummulierten Wahrscheinlichkeitsverteilung P , also der aufintegrierten Verteilung der Wahrscheinlichkeitsdichte p , abgebildet auf die Größe Y , die dann p -verteilt ist.

```

1 import numpy as np
2 nrow = 5
3 ncol = 20
4 a = np.random.randn(nrow, ncol)

```

Mit `rand` werden gleichverteilte Zufallszahlen $R \in [0, 1]$ erzeugt.

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 $R = \{R_1, \dots, R_J\}$ 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. 8.3.

Als Beispiel wollen wir in Python 1000 Zufallszahlen generieren, die verteilt sind gemäß einer t -Verteilung für 45 Freiheitsgrade:

```

1 import numpy as np
2 import scipy
3 from scipy import stats
4 Jz = 1000
5 nu = 45
6 R = np.random.rand(1, Jz)
7 Z = scipy.stats.t.ppf(R[0], nu)

```

8.3 Monte-Carlo-Verfahren gemäß GUM-supplement 1 JCGM 101

In Kapitel 10 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. (7.60), 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 Modelfunktionen und für normalverteilte Größen X_i sieht die Posterior mit *scharfem Modellprior* dann wie folgt aus

$$p(Y, X_1, \dots, X_N | (X_{1,1}, \dots, X_{1,J_1}), \dots, (X_{K,1}, \dots, X_{K,J_1}), \bar{x}_{K+1}, s_{K+1}, \dots, \bar{x}_N, s_N) \propto \delta(Y - f(X_1, \dots, X_N)) \left(\prod_{i=1}^K \prod_{j=1}^{J_i} e^{-\frac{1}{2} \left(\frac{X_{i,j} - \bar{x}_{i,j}}{s_i} \right)^2} \right) \left(\prod_{i=K+1}^N e^{-\frac{1}{2} \left(\frac{X_i - \bar{x}_i}{s_i} \right)^2} \right). \quad (8.11)$$

Wir hatten 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

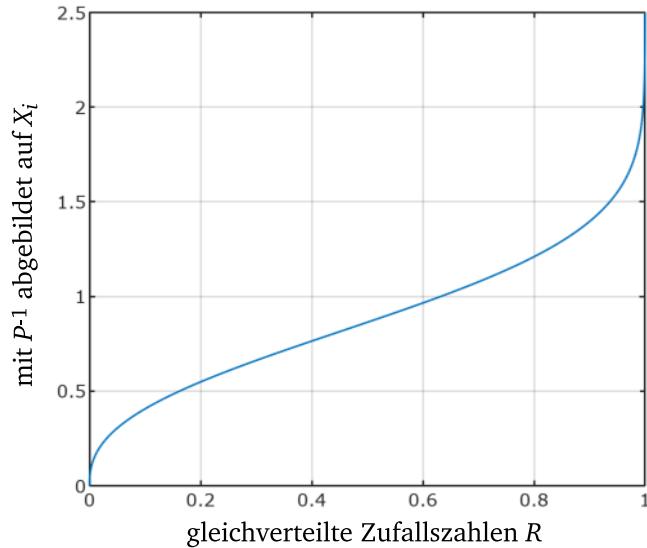


Abbildung 8.3: Ein Sample gleichverteilter Zufallszahlen $\mathbf{R} = \{R_1, \dots, R_L\}$ wird über die inverse Funktion P^{-1} der kumulierten Wahrscheinlichkeitsverteilung P abgebildet auf ein Sample $\mathbf{X} = \{X_1, \dots, X_L\}$, deren Werte dann p -verteilt sind.

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 Vibratoren 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

$$p(Y, X_1, \dots, X_N | (X_{1,1}, \dots, X_{1,J_1}), \dots, (X_{K,1}, \dots, X_{K,J_1}), \theta_{K+1,1}, \dots, \theta_{N,M_K}) \propto \underbrace{\delta(Y - f(X_1, \dots, X_N))}_{\text{scharfer Modellprior}} \underbrace{\left(\prod_{i=1}^K \prod_{j=1}^{J_i} p_i(X_{i,j} | X_i) \right)}_{\text{Likelihood}} \underbrace{\left(\prod_{i=K+1}^N p_i(X_i | \theta_{i,1}, \dots, \theta_{i,M_i}) \right)}_{\text{Prior}}. \quad (8.12)$$

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. $p_i(X_i | \theta_{i,1}, \dots, \theta_{i,M_i})$ 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, das 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

Messunsicherheitsberechnung mit bayesischer Statistik

9.1 Bayes-Theorem

9.1.1 Einleitung: Klassische und Bayessche Statistik

Grob gesagt verwendet die klassische Statistik zum Schätzen von Parametern und zum Testen von Hypothesen nur die Stichprobe; die bayessche stellt zusätzlich in Rechnung, was man sonst noch über das Problem weiß oder annimmt. Dies hängt mit den unterschiedlichen Meinungen darüber zusammen, was Wahrscheinlichkeit bedeutet: relative Häufigkeit in Zufallsexperimenten (die klassisches Sicht) oder einen Ausdruck des Wissens (die bayessche Sicht). Siehe hierzu auch die 2. Vorlesung.

9.1.2 Das Bayes-Theorem

(Hinweis: Unter einer *bedingten Wahrscheinlichkeit* $P(A|B)$ versteht man die Wahrscheinlichkeit, dass ein Ereignis A unter der Bedingung eintritt, dass das Eintreten eines Ereignisses B bereits bekannt ist, d.h. Ereignis B bereits eingetreten ist. Im Gegensatz dazu gibt es die *totale Wahrscheinlichkeit* $p(A)$.)

In der Statistik wird der zu schätzende Parameter häufig mit θ bezeichnet. Deshalb nehmen wir hier auch für den Modellparameter θ anstatt Y . Gegeben sind Messdaten / Beobachtungen einer Messgröße X mit X_j , $j = 1 \dots J$. Die dazugehörige bedingte Wahrscheinlichkeitsdichte in Abhängigkeit von dem Parameter θ ist $p(X|\theta)$. Unter der Annahme, dass θ eine Wahrscheinlichkeitsdichte $p(\theta)$ besitzt, gibt es nach Bayes folgenden Zusammenhang

$$p(X|\theta) \cdot p(\theta) = p(\theta|X) \cdot p(X) \quad (9.1)$$

Die gesuchte bedingte Wahrscheinlichkeit $p(\theta|X)$ ist dann gegeben durch:

$$p(\theta|X) = \frac{p(X|\theta) \cdot p(\theta)}{p(X)} \quad (9.2)$$

Die totale Wahrscheinlichkeitsdichte erhalten wir durch:

$$p(X) = E(p(X|\theta)) = c^{-1} = \begin{cases} \int p(X|\theta)p(\theta)d\theta & \text{mit } \theta \text{ kontinuierlich} \\ \sum p(X|\theta)p(\theta) & \text{mit } \theta \text{ diskret} \end{cases} \quad (9.3)$$

Das Integral oder die Summe geht über den erlaubten Bereich von θ . $E()$ bezeichnet den Erwartungswert.

Damit können wir die Gl. (9.2) umschreiben:

$$p(\theta|X) = c \cdot p(X|\theta) \cdot p(\theta) \quad (9.4)$$

Gl.(9.2) bzw. Gl.(9.4) wird in der Regel als **Bayes-Theorem** bezeichnet. Wir bezeichnen:

- $p(\theta|X)$: Gesuchte Posterior-Wahrscheinlichkeitsdichte, kurz: **Posterior**
- $p(X|\theta)$: Bedingte Wahrscheinlichkeitsdichte der Messdaten X

- $p(\theta)$: Prior-Wahrscheinlichkeitsdichte von θ , kurz: **Prior**
- c : Normierungskonstante, die sicherstellt, dass die Summe bzw. das Integral über die Posterior $p(\theta|X)$ gleich Eins ist.

Da c „nur“ eine Normierungskonstante ist wird das Bayes-Theorem auch oft als Proportionalitätsgleichung geschrieben:

$$p(\theta|X) \propto p(X|\theta) \cdot p(\theta) \quad (9.5)$$

Zusammenhang zwischen Bayes-Theorem und Likelihood-Funktion:

Die bedingte Wahrscheinlichkeitsdichte $p(X|\theta)$ kann

(a) als Funktion von X gesehen werden

oder

(b) als Funktion von θ , d.h. welches θ passt am besten zu den gemessenen Daten X ; dieses bezeichnet man als **Likelihood-Funktion**: $l(\theta|X)$

Somit können wir die Gl.(9.5) auch schreiben als:

$$p(\theta|X) \propto l(\theta|X) \cdot p(\theta) \quad (9.6)$$

Das Bayes-Theorem gibt uns die Wahrscheinlichkeitsfunktion für den Posterior $p(\theta|X)$ durch das Produkt aus der Likelihoodfunktion und der Prior-Funktion $p(\theta)$

Posteriorverteilung \propto Likelihoodfunktion \times Prior – Verteilung

Die Likelihood-Fkt. spielt eine entscheidende Rolle bei Bayes, da sie durch Aufnehmen von Messdaten X meine Priorkenntnis über θ updated.

Eine Konstante bei der Likelihood-Funktion hat keinen Einfluss auf das erweiterte Wissen, da schlussendlich ja die Posterior-Verteilung normiert bleiben muss, d.h. $\int p(\theta|X)d\theta = 1$. Deshalb sind nur die relativen Werte der Likelihood-Funktion von Relevanz.

Sequentielle Natur von Bayes

Ein wichtige Eigenschaft des Bayes-Theorems ist, dass Vorwissen mit neuem Wissen kombiniert werden kann. Es gilt bei Aufnahme von zwei Messungen (zwei Beobachtungen) X_1 und X_2

$$\begin{aligned} p(\theta|X_2, X_1) &\propto p(\theta) \cdot l(\theta|X_1) \cdot l(\theta|X_2) \\ &\propto p(\theta|X_1) \cdot l(\theta|X_2) \end{aligned}$$

D.h. die Posterior $p(\theta|X_1)$ der 1. Beobachtung wird zum Prior für die 2. Beobachtung. Dieser Prozess kann beliebig oft wiederholt werden, d.h. wir erhalten:

$$p(\theta|X_1, \dots, X_J) \propto p(\theta|X_1, \dots, X_{J-1}) \cdot l(\theta|X_J) \quad (9.7)$$

Diese Gleichung beschreibt den **Prozess des Lernens durch kontinuierliche Updates**. Dies ist auch für Ringvergleiche interessant. Hier messen mehrere Institute denselben Prüfling (siehe 6. Vor-

lesung „Auswertung von Mess- und Ringvergleichen“). Darüber lässt sich auch herleiten, wie bei einem Ringvergleich (siehe 6. Vorl.) die Erwartungswerte und Varianzen der teilnehmenden Labore zu kombinieren sind.

9.2 Parameterschätzung mit Bayesscher Statistik

9.2.1 Einführung

Wie in der klassischen Statistik (frequentistische Statistik), so kann man auch in der bayesschen Statistik eine Parameterschätzung durchführen. Eine Schätzung hat zum Ziel einen Wert anzugeben, der dem wahren Wert möglichst nahe kommt. Man gibt zum Schätzwert noch ein Intervall an, indem vermutlich der wahre Wert liegt. Sowohl der Schätzwert $\hat{\theta}$ als auch das dazugehörige Intervall kann aus der Posterior-Verteilungsdichte $p(\theta|X)$ bestimmt werden.

9.2.2 Beispiel: Schätzung einer physikalischen Konstante

Dieses etwas fiktive Beispiel soll zeigen, wie sich unterschiedliches Vorwissen oder unterschiedliche Beobachtungen auf die Schätzung einer physikalischen Konstante auswirkt. Wir nehmen an, dass Beobachter A sehr vertraut auf diesem Gebiet ist und großes Vorwissen hat und somit einen guten Schätzwert für den Prior $p_A(\theta)$ angeben kann. Wir nehmen an, dass das Prior-Wissen des Beobachters A durch eine Normalverteilung $\mathcal{N}(\mu = 900, \sigma^2 = 20^2)$ approximiert werden kann. D.h.

$$p_A(\theta) = \frac{1}{\sqrt{(2\pi)20}} \exp \left[-\frac{1}{2} \left(\frac{\theta - 900}{20} \right)^2 \right]$$

Im Gegensatz dazu nehmen wir an, dass Beobachter B weniger Vorwissen hat, hier beispielweise nur eine Varianz von $\sigma^2 = 80^2$ hat. Wir nehmen für B als Prior eine Normalverteilung $\mathcal{N}(800, 80^2)$ an. D.h.

$$p_B(\theta) = \frac{1}{\sqrt{(2\pi)80}} \exp \left[-\frac{1}{2} \left(\frac{\theta - 800}{80} \right)^2 \right]$$

Abb.9.1 zeigt die beiden Prior-Verteilungen: Wir nehmen zunächst an, dass die physik. Konstante einmal beobachtet wird mit dem Mittelwert $\theta = 850$ und der Standardabweichung 40, d.h. als Likelihood der Messdaten haben wir eine Normalverteilung $l(\theta|X) \sim \mathcal{N}(850, 40)$. Abb.9.2 zeigt die Wahrscheinlichkeitsdichtefunktion (engl. *probability density function*, kurz pdf) der Likelihood und Abb.9.3 die Posteriorverteilung $p_A(\theta|X)$ für den Beobachter A und die Posterior-Verteilung $p_B(\theta|X)$ für den Beobachter B. Abb. 9.3 zeigt, dass Beobachter A nicht viel gelernt hat, während Beobachter B viel gelernt hat.

Wenn sowohl der Prior $\theta \sim N(\theta_0, \sigma_0^2)$ als auch die Likelihood normalverteilt ist, dann ist auch der Posterior $p(\theta|X)$ normalverteilt mit $\mathcal{N}(\bar{\theta}, \bar{\sigma}^2)$, wobei

$$\bar{\theta} = \frac{1}{w_0 + w_1}(w_0\theta_0 + w_1x) \quad , \quad \frac{1}{\bar{\sigma}^2} = w_0 + w_1 \quad (9.8)$$

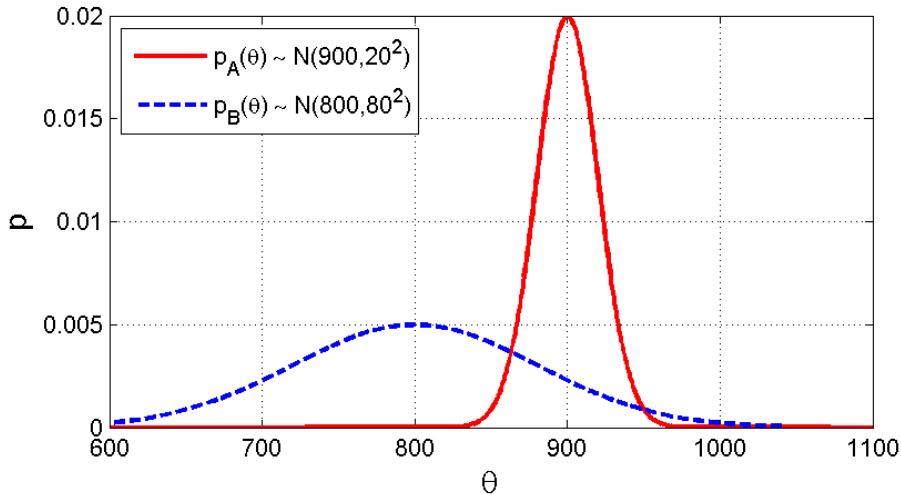
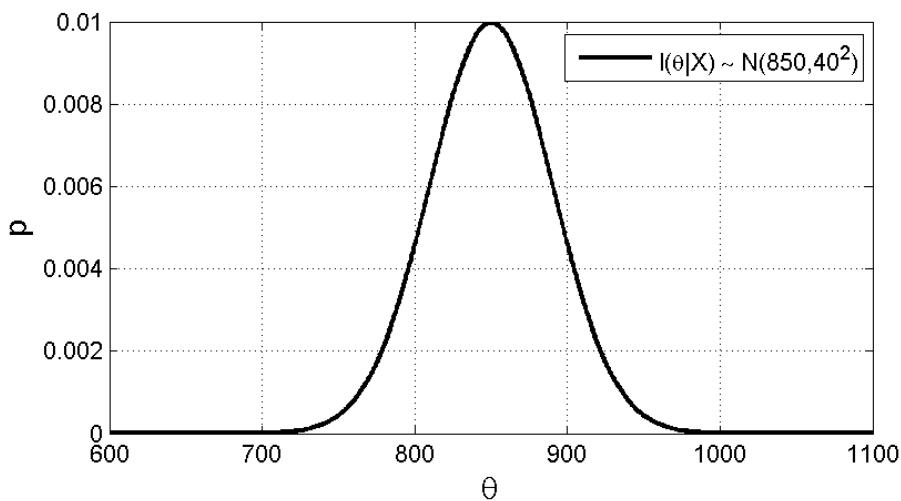


Abbildung 9.1: Prior-Verteilungen für A und B

Abbildung 9.2: Standardisierte Likelihood für 1 Beobachtung mit $x = 850$

$$\text{mit } w_0 = \frac{1}{\sigma_0^2} \quad \text{und} \quad w_1 = \frac{1}{\sigma_1^2}$$

Beweis dazu siehe weiter unten.

Mit $\bar{\sigma}^2$ bezeichnen wir hier die Varianz der Posterior und nicht die Varianz des Mittelwertes. D.h. durch die Beobachtung wird meine Varianz upgedatet. Bei J Beobachtungen wäre dann die Varianz $\bar{\sigma}^2$ des Posteriors die Kombination aus der Varianz des Priors σ_0^2 und den Messungen σ_j^2 :

$$\frac{1}{\bar{\sigma}^2} = \left(\sum_{j=1}^J \frac{1}{\sigma_j^2} \right) + \frac{1}{\sigma_0^2}$$

Der Posterior Mittelwert $\bar{\theta}$ ist das gewichtete Mittel des Prior-Mittelwertes und des Mittelwertes x der Beobachtung X .

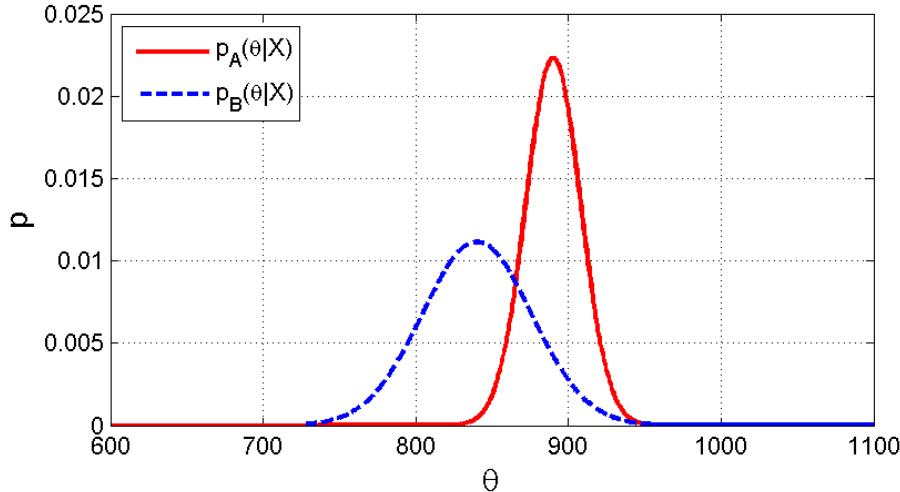


Abbildung 9.3: Posterior Verteilung für A und B nach einer Beobachtung mit $\sigma_A = 17.9$ und $\sigma_B = 35.7$. D.h. A hat wenig gelernt, B hat viel gelernt.

Für den Beobachter A ergibt sich der Mittelwert des Posteriors zu:

$$\bar{\theta}_A = \frac{1}{1/20^2 + 1/40^2} \left(\frac{1}{20^2} \cdot 900 + \frac{1}{40^2} \cdot 850 \right) = 890$$

Die Standardabweichung ergibt sich zu:

$$\frac{1}{\bar{\sigma}_A^2} = \frac{1}{20^2} + \frac{1}{40^2} \Rightarrow \bar{\sigma}_A = 17.9$$

Die Ergebnisse sind in Tab.9.1 zusammengefasst.

Tabelle 9.1: Zusammenfassung der Ergebnisse für Beobachter A und B. A lernt wenig. B lernt viel.

Prior Verteilung	Likelihood von den Daten	Posterior Verteilung
A $\theta \sim N(900, 20^2)$	A $N(850, 40^2)$	A $\theta \sim N(890, 17.9^2)$
B $\theta \sim N(800, 80^2)$	B $N(850, 40^2)$	B $\theta \sim N(840, 35.7^2)$

Nehmen wir jetzt an, dass 99 weitere unabhängige Messungen zur Bestimmung der phys. Konstante gemacht werden mit einem Mittelwert $\bar{x} = 1/100 \cdot \sum_{j=1}^{100}$ für 100 Beobachtungen mit Mittelwert 870. Die Likelihood-Funktion von θ bei J unabhängigen Beobachtungen, die normalverteilt $N(\theta, \sigma^2)$ sind, ist gegeben durch:

$$l(\theta|X) \propto \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^J \exp \left[-\frac{1}{2\sigma^2} \sum_{j=1}^J (X_j - \theta)^2 \right]$$

Da

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

und die gegebenen Daten $\sum(X_j - \bar{Y})^2$ eine feste Konstante ist, so ist die Likelihood:

$$l(\theta|X) \propto \exp \left[-\frac{1}{2} \left(\frac{\theta - \bar{X}}{\sigma/\sqrt{J}} \right)^2 \right] \quad (9.9)$$

D.h. dies ist eine Normalverteilung zentriert bei \bar{X} mit Standardabweichung des Mittelwertes σ/\sqrt{J} . Im gegebenem Beispiel ist die Likelihood die Normalverteilung mit dem Mittelwert $\bar{X} = 870$ und der Standardabweichung des Mittelwertes $\sigma/\sqrt{J} = 40/\sqrt{100} = 4$. Dies ist in Abb. 9.4 gezeigt.

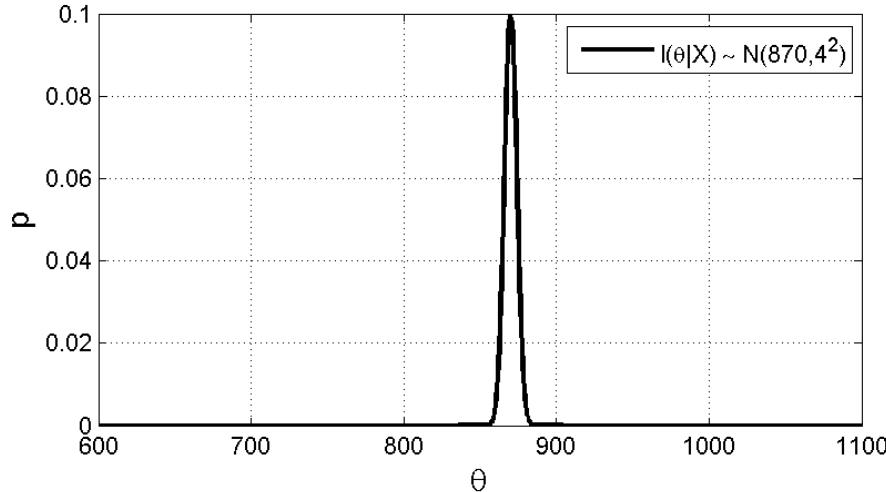


Abbildung 9.4: Standardisierte Likelihood für 100 Beobachtungen mit $\bar{X} = 870$

Mit der Formel in Gl.(9.8) errechnen sich die Posterior-Verteilungen von A und B zu $N(871.2, 3.9^2)$ und $N(869.8, 3.9995^2)$. Diese sind in Abb.9.5 dargestellt.

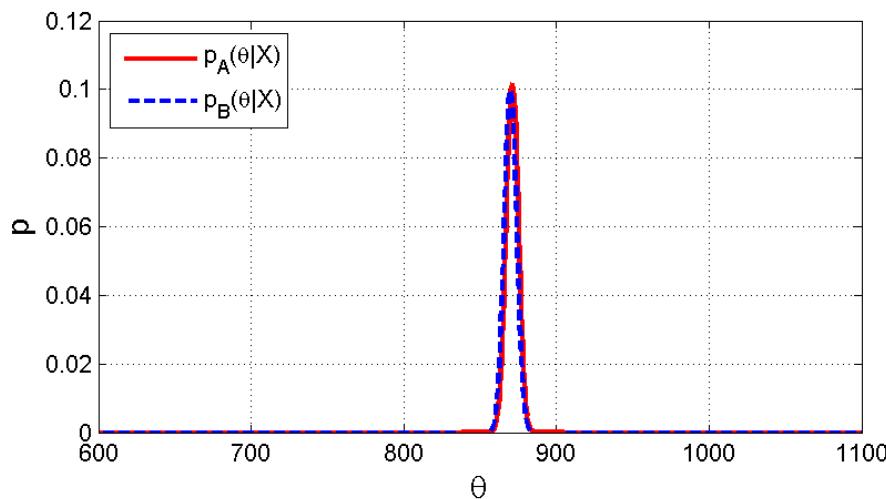


Abbildung 9.5: Posterior-Verteilungen für A und B nach 100 Beobachtungen

Nach 100 Beobachtungen sind die Beobachter A und B fast in Übereinstimmung. Das Prior-Wissen von A und B spielt fast keine Rolle mehr. In Tabelle 9.2 sind die Ergebnisse dargestellt.

Anhang (a): Herleitung von Gl.(9.8)

Tabelle 9.2: Zusammenfassung der Ergebnisse für Beobachter A und B bei 100 Beobachtungen. Vorwissen von Beobachter A und B spielt fast keine Rolle mehr

Prior Verteilung	Likelihood von den Daten	Posterior Verteilung
A $\theta \sim N(900, 20^2)$	A $N(870, 4^2)$	A $\theta \sim N(871.2, 3.9^2)$
B $\theta \sim N(800, 80^2)$	B $N(870, 4^2)$	B $\theta \sim N(869.8, 3.995^2)$

Im Folgenden wollen wir noch die Gl.(9.8) herleiten, mit der wir die Posterior berechnen können, wenn die Prior-Verteilung und die Likelihood normalverteilt sind. Wir nehmen an, dass die Prior normalverteilt ist mit:

$$p(\theta) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp \left[-\frac{1}{2} \left(\frac{\theta - \theta_0}{\sigma_0} \right)^2 \right], \quad -\infty < \theta < \infty \quad (9.10)$$

Die Likelihood-Funktion von θ ist ebenfalls normalverteilt:

$$l(\theta|X) \propto \exp \left[-\frac{1}{2} \left(\frac{\theta - X}{\sigma_1} \right)^2 \right] \quad (9.11)$$

Die Posterior-Verteilung erhalten wir dann:

$$p(\theta|X) = \frac{p(\theta) l(\theta|X)}{\int_{-\infty}^{\infty} p(\theta) l(\theta|X) d\theta} = \frac{f(\theta|X)}{\int_{-\infty}^{\infty} f(\theta|X) d\theta} \quad (9.12)$$

mit

$$f(\theta|X) = \exp \left\{ -\frac{1}{2} \left[\left(\frac{\theta - \theta_0}{\sigma_0} \right)^2 + \left(\frac{\theta - X}{\sigma_1} \right)^2 \right] \right\} \quad (9.13)$$

Unter der Verwendung der Identität

$$A(z - a)^2 + B(z - b)^2 = (A + B)(z - c)^2 + \frac{AB}{A + B}(a - b)^2 \quad (9.14)$$

mit

$$c = \frac{1}{A + B}(Aa + Bb) \quad (9.15)$$

können wir den folgenden Ausdruck wie folgt schreiben:

$$\left(\frac{\theta - \theta_0}{\sigma_0} \right)^2 + \left(\frac{\theta - X}{\sigma_1} \right)^2 = (\sigma_0^{-2} + \sigma_1^{-2}) (\theta - \bar{\theta})^2 + d \quad (9.16)$$

wobei

$$\bar{\theta} = \frac{1}{\sigma_0^{-2} + \sigma_1^{-2}} (\sigma_0^{-2} \theta_0 + \sigma_1^{-2} X)$$

und d ist konstant, d.h. unabhängig von θ . Somit

$$f(\theta|X) = \exp \left(-\frac{d}{2} \right) \exp[-1/2 \cdot (\sigma_0^{-2} + \sigma_1^{-2})(\theta - \bar{\theta})^2] \quad (9.17)$$

so dass

$$\begin{aligned} \int_{-\infty}^{\infty} f(\theta|X)d\theta &= \exp\left(-\frac{d}{2}\right) \cdot \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(\sigma_0^{-2} + \sigma_1^{-2})(\theta - \bar{\theta})^2\right] d\theta \\ &= \sqrt{2\pi}(\sigma_0^{-2} + \sigma_1^{-2})^{-1/2} \exp(-d/2) \end{aligned} \quad (9.18)$$

Es folgt daraus, dass

$$p(\theta|X) = \frac{(\sigma_0^{-2} + \sigma_1^{-2})^{1/2}}{\sqrt{2\pi}} \exp\left[-1/2 \cdot (\sigma_0^{-2} + \sigma_1^{-2})(\theta - \bar{\theta})^2\right] \quad (9.19)$$

mit der Normalverteilung

$$N\left[\bar{\theta}, (\sigma_0^{-2} + \sigma_1^{-2})^{-1}\right] \quad (9.20)$$

Anhang (b): Octave/Matlab-Skript zum Beispiel

```

1 % by Ehret, date: 2019-12-10
2 clear all; close all;
3 % Beispiel: Bayes for a physical constants
4
5 % Definition der X-Werte
6 X = linspace(600,1100,1000);
7
8 % Beispiel
9 % Prior-Verteilung des Beobachters A
10 % Normalverteilt mit mu_A = 900 und sigma_a = 20
11 mu_A = 900;
12 sigma_A = 20;
13 DeltaX = X(2)-X(1);
14 p_A = normpdf(X,mu_A, sigma_A);
15
16 % Prior-Verteilung des Beobachters B
17 % Normalverteilt mit mu_B = 800 und sigma_a = 80
18 mu_B = 800;
19 sigma_B = 80;
20 p_B = normpdf(X,mu_B, sigma_B);
21
22 % Plotten der Prior-Verteilungen von A und B
23 figure(1)
24 plot(X,p_A,'r-','linewidth',2)
25 hold on;
26 plot(X,p_B,'b--','linewidth',2)
27 grid on;
28 legend('p_A(\theta) ~ N(900,20^2)', 'p_B(\theta) ~ N(800,80^2)', ...
29 'Location','NorthWest')
30 xlabel('\theta','fontsize',14)
31 ylabel('p','fontsize',14)
32
33 set(gcf, 'PaperUnits', 'centimeters');
34 x_width=15 ;y_width= 8;
35 set(gcf, 'PaperPosition', [0 0 x_width y_width]);
36 print '-dpng' prior_A_Prior_B.png
37
38 % Annahme der Likelihood (Messung) als Normalverteilung mit l ~ N(850,40^2 )
39 mu_likeli = 850;
40 sigma_likeli = 40;
41 likelihood = normpdf(X,mu_likeli, sigma_likeli);
42
43 % Plotten der Likelihood

```

```

44 figure(2)
45 plot(X,likelihood,'k-','linewidth',2)
46 grid on;
47 legend('l(\theta|X) \sim N(850,40^2)')
48 xlabel('\theta','fontsize',14)
49 ylabel('p','fontsize',14)
50
51 set(gcf, 'PaperUnits', 'centimeters');
52 x_width=15 ;y_width= 8;
53 set(gcf, 'PaperPosition', [0 0 x_width y_width]);
54 print '-dpng' likelihood_1_Beobachtung.png
55
56
57 % Berechnung der Posterior fuer eine Beobachtung fuer A und B
58 % Posterior ~ Prior * Likelihood
59 p_post_A_1 = p_A .* likelihood;
60 p_post_B_1 = p_B .* likelihood;
61
62 % Normierung des posterior mit Normierungskonstante
63 % Nicht ganz korrekt, da eigentlich
64 % von -\inf bis +\inf integriert bzw. summiert werden muessste
65
66 const_post_A1 = 1/sum(p_post_A_1*DeltaX);
67 p_post_A_1 = const_post_A1 * p_post_A_1 ;
68
69 const_post_B1 = 1/sum(p_post_B_1*DeltaX);
70 p_post_B_1 = const_post_B1 * p_post_B_1 ;
71
72 % Plotten der Posterior
73 figure(3)
74 plot(X,p_post_A_1,'r-','linewidth',2)
75 hold on;
76 plot(X,p_post_B_1,'b-','linewidth',2)
77 grid on;
78 legend('p_A(\theta|X)', 'p_B(\theta|X) ', ...
79 'Location','NorthWest')
80 xlabel('\theta','fontsize',14)
81 ylabel('p','fontsize',14)
82
83 set(gcf, 'PaperUnits', 'centimeters');
84 x_width=15 ;y_width= 8;
85 set(gcf, 'PaperPosition', [0 0 x_width y_width]);
86 print '-dpng' posterior_A_posterior_B.png
87

```

```

88 % Jetzt besteht die Likelihood aus 100 Stichproben
89 % Bei 100 Stichproben kann die Verteilung schmäler werden
90 % Wir nehmen eine Normalverteilung  $N(870, \sigma / \sqrt{N})$  an
91 N = 100;
92 sigma_likeli_100 = sigma_likeli / sqrt(N);
93 mu_likeli_100 = 870;
94 % Likelihood Funktion fuer 100 Stichproben
95 likelihood_100 = normpdf(X,mu_likeli_100, sigma_likeli_100);
96
97 figure(4)
98 plot(X,likelihood_100, 'k-', 'linewidth',2)
99 grid on;
100 legend('l(\theta|X) ~ N(870,4^2)')
101 xlabel('\theta', 'fontsize',14)
102 ylabel('p', 'fontsize',14)
103
104 % Berechnung der Posterior fuer 100 Beobachtungen fuer A und B
105 p_post_A_100 = p_A .* likelihood_100;
106 p_post_B_100 = p_B .* likelihood_100;
107
108 % Normierung des posterior mit Normierungskonstante
109 % Nicht ganz korrekt, da eigentlich
110 % von -\inf bis +\inf integriert bzw. summiert werden müsste
111 const_post_A_100 = 1/sum(p_post_A_100*DeltaX);
112 p_post_A_100 = const_post_A_100 * p_post_A_100 ;
113
114 const_post_B_100 = 1/sum(p_post_B_100*DeltaX);
115 p_post_B_100 = const_post_B_100 * p_post_B_100 ;
116
117 figure(5)
118 plot(X,p_post_A_100, 'r-', 'linewidth',2)
119 hold on;
120 plot(X,p_post_B_100, 'b--', 'linewidth',2)
121 grid on;
122 legend('p_A(\theta|X)', 'p_B(\theta|X)', ...
123 'Location','NorthWest')
124 xlabel('\theta', 'fontsize',14)
125 ylabel('p', 'fontsize',14)

```

9.2.3 Priorverteilungen

Das Bayes-Theorem verwendet eine Prior-Verteilung, also ein Vorwissen über die zu schätzenden Modellparameter. Für Situationen, in denen es im wesentlichen kein Vorwissen gibt, wird eine einfache Rechteckverteilung als Prior angesetzt, was bedeutet, dass die Wahrscheinlichkeit innerhalb eines vorgegebenen Intervalls $[\theta_{\text{li}}, \theta_{\text{re}}]$ konstant ist:

$$p(\theta) = \begin{cases} \text{const.} & \text{falls } \theta \in [\theta_{\text{li}}, \theta_{\text{re}}] \\ 0 & \text{sonst.} \end{cases}$$

Dieser Ansatz ist mit Vorsicht zu verwenden: Die Posterior kann mehrere lokale Maxima, die ähnlich groß sind, aufweisen. Sie kann auch ihrerseits nahezu rechteckverteilt sein. Je nach Modell kann dieser Ansatz mit ungeschickt gewähltem Intervall $[\theta_{\text{li}}, \theta_{\text{re}}]$ zu numerischen Problemen führen. Besonders problematisch ist die Situation, wenn mangels Vorwissen für die Intervallgrenzen oder für eine der beiden Intervallgrenzen der Grenzübergang ins Unendliche vorgenommen wird. Das kann dann dazu führen, dass kein physikalisch oder statistisch sinnvoller Erwartungswert bestimmt werden kann.

Ein Beispiel hierfür ist die Bestimmung der Fläche eines Rechteckes $F = a \cdot b$. Wenn wir kein Wissen über das Verhältnis der Seitenlängen a und b haben, dann würden wir zunächst sagen, dass alle Verhältnisse $R = a/(a+b)$ gleichverteilt sind. Wir können dies auch als $R = 1/(1+b/a)$ umschreiben. Für $b/a \rightarrow 0$ ergibt sich für das Verhältnis $R = 1$, für $b/a \rightarrow \infty$ ergibt sich für das Verhältnis $R = 0$. Wir können somit annehmen, dass die Verhältnisse zwischen 0 und 1 gleichmäßig verteilt sind, wenn wir keine Vorkenntnisse haben. Umgekehrt, wenn wir nichts über die Fläche wissen, würden wir annehmen, dass alle Flächen gleichverteilt sind. Beide Priors sind in Abb. 9.6 dargestellt.

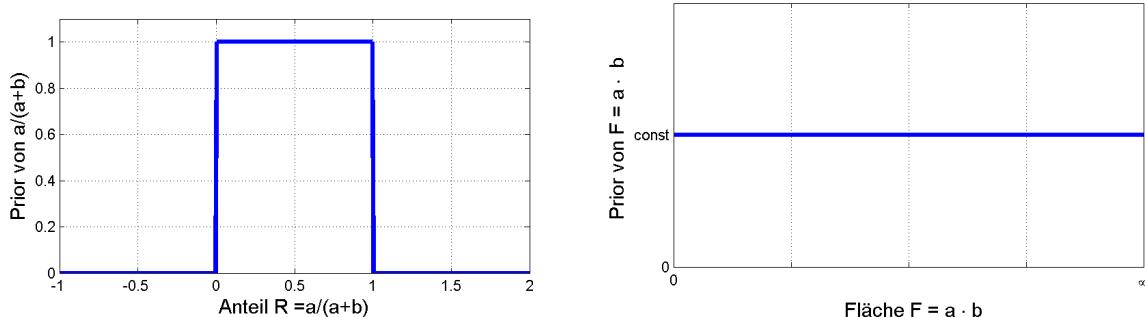


Abbildung 9.6: Plausible Priors für den Anteil $a/(a+b)$ und die Fläche $F = a \cdot b$. Beide Priors sind jedoch nicht konsistent zueinander.

Wenn nun der Anteil $R = 0$ ist, so muss auch die Fläche $F = 0$ sein. Wenn der Anteil $R = 1$ ist, so muss die Fläche $F = \infty$ sein. Bei $a = b$ ergibt sich ein Anteil $R = 0.5$. Wenn wir die Priorverteilung des Anteiles anschauen, so liegt der Anteil zu 50% zwischen 0 und 0.5. Dies würde auch bedeuten, dass 50% der Fläche zwischen $F = 0$ und $F = a^2 = b^2$ liegen müssten. Dies steht jedoch im Widerspruch zu der uniformen Annahme in Abb. 9.6. Hier sind nicht 50% der Fläche zwischen $[0, a^2 = b^2]$ und 50% zwischen $[a^2 = b^2, \infty]$. Somit haben wir hier einen Widerspruch. Deshalb hat sich hier Jeffrey [Hel08] etwas einfallen lassen, er schlägt für den Prior des Anteiles, eine Beta-Verteilung mit $\alpha = \beta = 0.5$ vor, wie sie Abb. 9.7 dargestellt ist. Nun haben wir in Abb. 9.7 zwei in sich konsistente Prior-Verteilungen. Unwissenheit kann also nicht grundsätzlich durch eine uniforme Verteilung dargestellt werden [Tsc14].

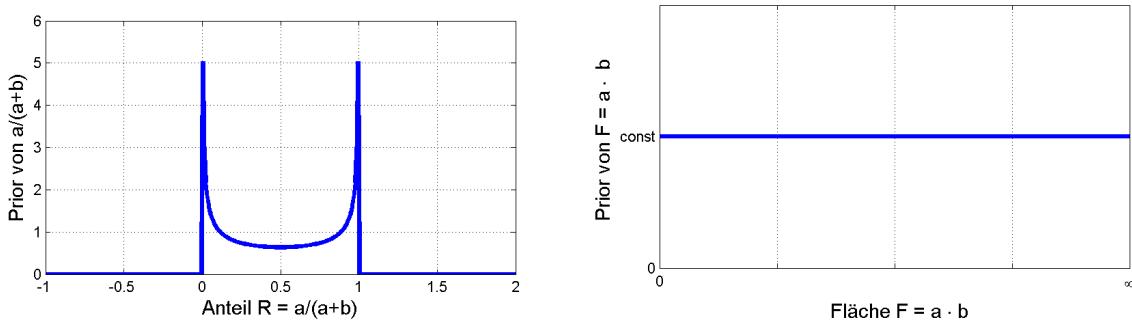


Abbildung 9.7: links: Jeffreys Prior für den Anteil $R = a/(a+b)$, eine BetaVerteilung $(0,5;0,5)$, rechts: Jeffreys Prior für den Erwartungswert einer normalverteilten Größe, hier ist dies die Fläche $F = a \cdot b$. Beide Priors sind jetzt konsistent zueinander.

Die BetaVerteilung ist eine stetige Wahrscheinlichkeitsverteilung über dem Intervall $(0, 1)$. Die BetaVerteilungsdichte ist definiert durch (siehe z.B. Wikipedia):

$$p(X) = \frac{1}{B(\alpha, \beta)} X^{\alpha-1} (1-X)^{\beta-1}.$$

Außerhalb des Intervalls $(0, 1)$ wird sie durch $p(X) = 0$ fortgesetzt. Für $\alpha, \beta \geq 1$ lässt sich $(0, 1)$ durch $[0, 1]$ ersetzen. Die BetaVerteilung besitzt die reellen Parameter α und β . Um ihre Normierbarkeit zu garantieren, wird $\alpha, \beta > 0$ gefordert.

Der Vorfaktor $1/B(\alpha, \beta)$ dient der korrekten Normierung. Der Ausdruck

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} = \int_0^1 u^{\alpha-1} (1-u)^{\beta-1} du$$

steht für die Betafunktion, nach der die Verteilung benannt ist. Dabei bezeichnet Γ die Gammafunktion.

Einige nützliche Priors sind beispielhaft in Tab.9.3 dargestellt.

Tabelle 9.3: Häufig verwendete Priors.

Vorhandenes Vorwissen	PDF und Illustration	x und $u(x)$
Schätzwert x und Standardunsicherheit $u(x)$	Normal: $\mathcal{N}(x, u(x)^2)$ 	$x,$ $u(x)$
Es sind schon vorhandene (historische) Werte bekannt. Der Erwartungswert und die Varianz ist jedoch unbekannt.	Skalierte und verschobene t-Verteilung $t_{n-1}(\bar{\xi}, s^2/n)$ 	$x = \bar{\xi},$ $u(x) = \left(\frac{n-1}{n-3}\right) \frac{s}{\sqrt{n}}$
Kenntnisse über die Varianzen sind bekannt	χ^2 	$x,$ $u(x)$
Obere und untere Grenzen a, b sind bekannt	Rechteck: $\mathcal{R}(a, b)$ 	$x = \frac{a+b}{2},$ $u(x) = \frac{b-a}{\sqrt{12}}$

Hinweise zur skalierten und verschobenen t-Verteilung als Prior

(Wir bezeichnen hier die Anzahl der Beobachtungen mit n und nicht mit J um mit den Guides zur Bestimmung der Messunsicherheit konsistent zu sein, also statt $\xi_{i,1}, \dots, \xi_{i,J}$ verwenden wir hier $\xi_{i,1}, \dots, \xi_{i,n}$) Wenn angenommen wird, dass wir aus n Beobachtungen $\xi_{i,1}, \dots, \xi_{i,n}$, die aus einer Gauß- bzw. Normalverteilung (Erwartungswert und Varianz unbekannt) gezogen wurden und diese den Mittelwert $\bar{\xi}$ und die empirische Standardabweichung s besitzt, so können wir eine mit der Varianz des Mittelwertes s_i^2/n skalierte und um den Mittelwert $\bar{\xi}$ verschobene t-Verteilung mit $n-1$ Freiheitsgraden annehmen: $t_{n-1}(\bar{\xi}, s_i^2/n)$.

Der beste Schätzwert für X und die dazugehörige Varianz, die die beste Schätzung für die Unsicherheit darstellt, sind

$$x = \bar{\xi}$$

$$u(x_i) = \left(\frac{n-1}{n-3} \right)^{1/2} \frac{s_i}{\sqrt{n}}$$

mit

$$s_i = \left[\frac{1}{n-1} \sum_{r=1}^n (\xi_{i,r} - \bar{\xi}_i)^2 \right]^{1/2}$$

Eine Herleitung des Vorfaktors $\sqrt{\frac{n-1}{n-3}}$ ist z. B. in [Kac03] zu finden.

9.2.4 Posteriorverteilungen von Modellparametern

Wir geben hier ein paar Beispiele für Posterior-Verteilungen an:

(a) Priorverteilung ist konstant und die Varianz σ der Messdaten X ist bekannt

$$\text{Priorverteilung : } p(\theta) \sim U(-\infty, +\infty)$$

$$\text{Likelihood (siehe auch Gl.(9.9)) : } l(\theta|X) \sim N(\bar{X}, \sigma^2/J) \propto \exp \left[-\frac{1}{2} \left(\frac{\theta - \bar{X}}{\sigma/\sqrt{J}} \right)^2 \right]$$

$$\text{Posteriorverteilung : } p(\theta|X) \sim N(\bar{X}, \sigma^2/J)$$

(b) Priorverteilung ist normalverteilt und die Varianz σ der Messdaten X ist bekannt

$$\text{Priorverteilung : } p(\theta) \sim N(\theta_0, \sigma_0^2)$$

$$\text{Likelihood (siehe auch Gl.9.9) : } l(\theta|X) \sim N(\bar{X}, \sigma^2/J) \propto \exp \left[-\frac{1}{2} \left(\frac{\theta - \bar{X}}{\sigma/\sqrt{J}} \right)^2 \right]$$

$$\text{Posteriorverteilung : } p(\theta|X) \sim N(\bar{\theta}, \bar{\sigma}^2)$$

mit dem Erwartungswert

$$\bar{\theta} = \frac{\sigma_0^2 \cdot \bar{X} + \frac{\sigma^2}{J} \cdot \theta_0}{\sigma_0^2 + \frac{\sigma^2}{J}}$$

und der Standardabweichung

$$\bar{\sigma}^2 = \frac{\sigma_0^2 \cdot \frac{\sigma^2}{J}}{\sigma_0^2 + \frac{\sigma^2}{J}}$$

(a) Priorverteilung ist normalverteilt und die Varianz σ^2 der Messdaten X ist unbekannt

Ist die Varianz der Messgröße unbekannt, so muss die Varianz noch geschätzt werden, z. B. über die χ^2 -Verteilung.

9.3 MU-Bestimmung mittels Posterior aus Bayesansatz

9.3.1 Bayessche Intervallschätzung

Wie eine klassische Punktschätzung z.B. mit MLE oder LS, so sagt auch eine bayessche nicht aus, wie nahe der Schätzwert dem wahren Wert kommt. Doch auch hier kann man ein Intervall angeben, die den wahren Wert mit einer gewissen Wahrscheinlichkeit enthalten. Der üblichen Sprachregelung folgend bezeichnen wir die Intervalle der Bayes-Statistik nicht als *Vertrauens-* bzw. *Konfidenz-Intervalle* sondern als *Kredibilitätsintervalle*. (credibility: Glaubwürdigkeit, confidence: Vertrauen). Nicht weil das Wort credibility treffender wäre als das Wort confidence. Beide entstammen der Umgangssprache. Wir wählen den anderen Ausdruck, damit jederzeit klar wird, ob wir ein klassisches oder ein bayessches Intervall meinen. Der ISO-Guide verwendet den Begriff *Überdeckungsintervall*, engl. *coverage interval*, das das Intervall angibt, das die Menge der wahren Werte einer Messgröße mit einer angegebenen Wahrscheinlichkeit enthält, auf der Grundlage der verfügbaren Information [VIM08]. (siehe zu den Begrifflichkeiten: Vertrauens-, Credible- und Überdeckungintervall auch die Ausführungen der 4. Vorlesung) Den Wahrscheinlichkeitswert, den ein credible interval einschließt, bezeichnen wir als Kredibilitätsniveau γ .

Beispiel: Symmetrische/ nichtsymmetrische Verteilungen

Bei symmetrischen Posterior-Verteilungen wird in der Regel das symmetrische Intervall genommen. So ist in Abb. 9.8 das 95%ige Intervall einer Normalverteilung $N(10, 4^2)$ dargestellt. Die untere und obere Grenze kann hier noch per Hand ausgerechnet werden. Mit $Z = (\theta - \mu)/\sigma$ (Standardisieren, siehe 3. Vorlesung). In der 3. Vorlesung haben die kummulierten Wahrscheinlichkeitsfunktionen mit $P(X) = \int_{-\infty}^X p(X')dX'$ bezeichnet. Im Falle der Standardnormalverteilung $\mathcal{N}(0, 1)$ bezeichnen wir diese hier mit Φ . Die Quantile der rechtsseitigen Verteilungsfunktion sind im Anhang in der Tabelle dargestellt. Die gesuchte obere Grenze θ_{oG} des Intervalls erhalten wir durch:

$$\Phi\left(\frac{\theta_{oG} - \mu}{\sigma}\right) = 0.5 + 0.95/2 = 0.975$$

(Hinweis: beidseitiges Quantil: 95%, einseitiges Quantil 97.5%:)

In einer Quantittabelle der Normalverteilung (Anhang Kapitel 12) kann der Wert 1.96 abgelesen werden. Wir erhalten damit:

$$\frac{\theta_{oG} - 10}{4} = 1.96 \rightarrow \theta_{oG} = 17.84 \rightarrow \theta_{uG} = 17.84 - 2 \cdot 7.84 = 2.16$$

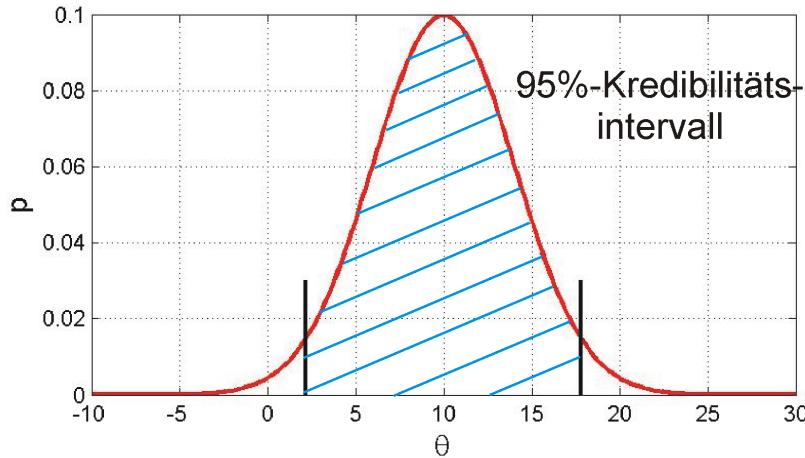


Abbildung 9.8: Bestimmung des Kredibilitätsintervalls der Posterior-Verteilung $N(10, 4^2)$ für das Niveau $\gamma = 0.95$, untere Grenze: $\theta_{uG} = 2.16$, obere Grenze $\theta_{oG} = 17.84$

Bei nichtsymmetrischen Intervallen lassen sich unterschiedliche obere und untere Grenzen finden, die zum gleichen Kredibilitätsintervall gehören. Dies ist in Abb. 9.9 schematisch in dem Beispiel mit der χ^2 -Verteilung mit $N = 10$ Freiheitsgraden dargestellt.

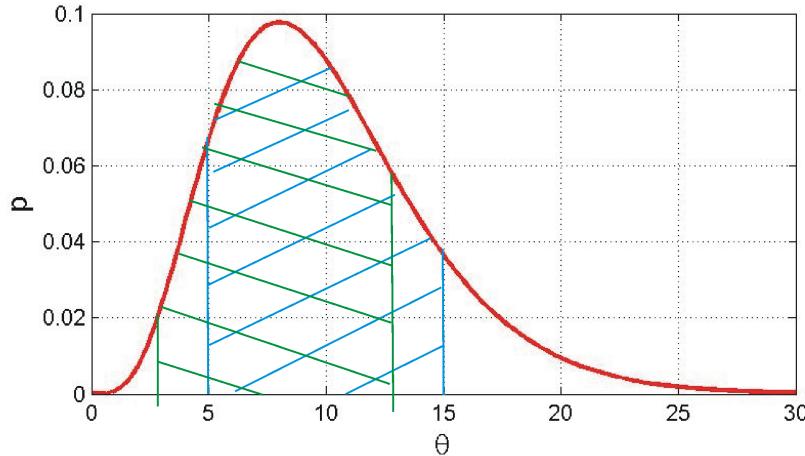


Abbildung 9.9: Bestimmung des Credible-Intervalls einer χ^2 -Verteilung für $N = 10$ Freiheitsgrade. Der grün und blau schraffierte Bereich hat dasselbe Kredibilitätsniveau.

9.3.2 MU-Bestimmung mit der Bayesschen Statistik

In der 9. Vorlesung werden wir den ISO-Guide 1995 / 2008 [GUM95] noch näher kennenlernen. Er basiert im Wesentlichen auf der klassischen Statistik und hat folgende Einschränkungen:

- Der ISO Guide verwendet keine Informationen über Wahrscheinlichkeitsverteilungen, die eventuell ja vorhanden sind. Es werden nur die Unsicherheiten fortgepflanzt jedoch keine Wahrscheinlichkeitsverteilungen.
- Das Messergebnis gibt keine vollständige Information wieder, da keine Wahrscheinlichkeitsverteilung

angegeben wird.

- Die Fehlerfortpflanzung erfordert Sensitivitäten.

Der ISO-Guide ist dennoch heutzutage der Leitfaden zur Angabe der Messunsicherheit z.B. für Angabe der Messunsicherheit auf Kalibrierscheinen. Wenn die pdf der Messgröße Y normalverteilt ist, hat das Überdeckungsintervall $[y \pm 2u(y)]$ eine Überdeckungswahrscheinlichkeit von 95.45% (siehe Tabelle G.2 im ISO-Guide [GUM95]). Bei anderen Verteilungen kann die Überdeckungswahrscheinlichkeit p des Intervalls $[y \pm 2u(y)]$ kleiner oder größer sein. Der ISO-Guide (Abschnitt 2.3.5 und 6.3.2) schreibt, dass ohne Kenntnis der Verteilung auch keine Überdeckungswahrscheinlichkeiten bestimmt werden können. Schlussendlich schlägt der ISO-Guide jedoch dann vor, dass für die Messgröße Y eine Normalverteilung oder eine skalierte und geschiftete Normalverteilung angenommen werden soll. Es wird dann vorgeschlagen, dass die erweiterte Messunsicherheit $U_p = k_p u_c(y)$ (u_c : Kombinierte Standardunsicherheit) für einen Grad des Vertrauens p (siehe Guide G 6.4) mit Hilfe der Welch-Satterthwaite Formel berechnet werden soll. Diese Formel bietet eine Näherung bei gemischten Verteilungen der Messgrößen X_i (siehe 7. Vorlesung).

In der Bayesschen Statistik wird die Standard-Unsicherheit vom Typ-A-Fehler nicht mehr als ein Schätzwert sondern als ein Parameter des Kenntnisstandes der pdf betrachtet. Es wird auch das Konzept der effektiven Freiheitsgrade nicht mehr benötigt. Die Typ-A-Unsicherheit wird anders ausgewertet. Es ergibt sich für die Standardunsicherheit $u(x_i) = \sqrt{(n-1)/(n-3)} \cdot s_i / \sqrt{n}$ [Kac03].

Den aktuellen Leitfaden ISO-Guide komplett auf die Bayes-Statistik umzuschreiben und für alle (z.B. auch Kalibrierlaboratorien) verpflichtend einzuführen, fand keine Mehrheit. Dennoch ist es in vielen Fällen sinnvoller und richtiger die Messunsicherheitsbestimmung mit der Bayes-Statistik durchzuführen. Ziel der Bayesschen Statistik ist es die PDF (probability density function) $p(Y|X)$ der Messgröße Y zu bestimmen. Bestimmt man davon den Erwartungswert y , erhält man den besten Schätzwert für die Messgröße. Die Standardabweichung von Y ergibt die zugeordnete Standardunsicherheit $u(y)$. Das Überdeckungsintervall (engl. *coverage integral*) wird dann aus der Posterior $p(Y|X)$ bestimmt. Der Erwartungswert der Posterior ist der beste Schätzer für die Messgröße Y :

$$E(Y) = y = \int_{-\infty}^{\infty} Y p(Y|X) dY$$

Die Varianz ist ein Maß für die Messunsicherheit:

$$u^2(y) =: \text{Var}(Y) = E((y - E(Y))^2) \quad (9.21)$$

$$= \int_{-\infty}^{\infty} (y - Y)^2 \cdot p(Y|X) dY \quad (9.22)$$

Zur Bestimmung der Posterior kann wiederum das Bayes-Theorem angewendet werden.

$$p(Y|X) \propto l(Y|X) \cdot p(Y) \quad (9.23)$$

Anstelle der Likelihood $l(Y|X)$ wird häufig die bedingte Wahrscheinlichkeitsdichte $p(X|Y)$ geschrieben.

$$p(Y|X) \propto p(X|Y) \cdot p(Y)$$

Wir bezeichnen $p(Y)$ als die Prior-Wahrscheinlichkeit, $l(Y|X)$ Likelihood der Messdaten bzw. $p(X|Y)$ die Wahrscheinlichkeitsdichte der Messdaten und $p(Y|X)$ die gesuchte Posterior-Wahrscheinlichkeitsdichte. Oftmals ist es nicht möglich die Integrale analytisch zu lösen und es kommen numerische Integrationsmethoden zum Einsatz. Bei komplexen Modellen mit vielen Parametern sind numerische Verfahren mit gleichförmigem Raster zu kostspielig /langwierig. Für solche Fälle werden statistische Integrationsmethoden eingesetzt, bei denen die Positionen, an denen die zu integrierende Funktion auszurechnen ist, gewürfelt werden. Solche Methoden nennt man wegen des Auswürfels, also des Verwendens von Pseudozufallszahlen, in Erinnerung an Spielcasinos Monte-Carlo-Methoden. Es gibt dazu unterschiedliche Strategien wie beispielsweise das Markov-Ketten-Monte-Carlo.

9.4 Regression mit Vorkenntnis (Prior)

Wir hatten klassisch die Regressionsrechnung nach der Methode der kleinsten Quadrate (*Least Squares*) berechnet, also

$$\min_{\theta} \left\{ (\mathbf{Y}^T - \boldsymbol{\theta}^T \mathbf{X}^T) (\mathbf{Y} - \mathbf{X}\boldsymbol{\theta}) \right\}$$

was äquivalent dem Maximieren der Likelihood ist

$$\max_{\theta} \left\{ e^{-\frac{1}{2\sigma_\varepsilon^2} (\mathbf{Y}^T - \boldsymbol{\theta}^T \mathbf{X}^T) (\mathbf{Y} - \mathbf{X}\boldsymbol{\theta})} \right\}. \quad (9.24)$$

Wir stellen uns vor, dass von einer vorherigen Messkampagne $\kappa - 1$ zu der gleichen Messaufgabe für alle Regressionparameter $\hat{\theta}_l^{(\kappa-1)}$ mit $l = 1, \dots, M$ schon ein vollständiges Messergebnis mit

$$\begin{aligned} \hat{\theta}_l^{(\kappa-1)} &= \mathbf{e}_l^T \hat{\boldsymbol{\theta}}^{(\kappa-1)} = \mathbf{e}_l^T \left((\mathbf{X}^{(\kappa-1)})^T \mathbf{X}^{(\kappa-1)} \right)^{-1} (\mathbf{X}^{(\kappa-1)}) \mathbf{Y}^{(\kappa-1)} \\ &\pm t_{1-\alpha/2, \nu^{(\kappa-1)}} \mathbf{e}_l^T \left((\mathbf{X}^{(\kappa-1)})^T \mathbf{X}^{(\kappa-1)} \right)^{-1} \mathbf{e}_l \left(\hat{\sigma}_\varepsilon^{(\kappa-1)} \right)^2 \end{aligned} \quad (9.25)$$

und

$$\rho_{l,m} = \frac{\left(\hat{\sigma}_\varepsilon^{(\kappa-1)} \right)^2}{\hat{\sigma}_l^{(\kappa-1)} \hat{\sigma}_m^{(\kappa-1)}} \mathbf{e}_l^T \left((\mathbf{X}^{(\kappa-1)})^T \mathbf{X}^{(\kappa-1)} \right)^{-1} \mathbf{e}_m \quad (9.26)$$

vorliegt, das wir mit neuen Informationen aus unserer neuen Messkampagne κ updaten wollen.

Die Likelihood mit den Beobachtungen unserer aktuellen Kampagne ist

$$l(\boldsymbol{\theta}, \sigma_\varepsilon | \mathbf{Y}^{(\kappa)}) \propto e^{-\frac{1}{2\sigma_\varepsilon^2} (\mathbf{Y}^{(\kappa)} - \mathbf{X}^{(\kappa)}\boldsymbol{\theta})^T (\mathbf{Y}^{(\kappa)} - \mathbf{X}^{(\kappa)}\boldsymbol{\theta})} \quad (9.27)$$

wobei die Regressanden $\mathbf{Y}^{(\kappa)}$ mit σ_ε streuen, nicht aber die Regressoren $\mathbf{X}^{(\kappa)}$ und wobei die Modellparameter $\boldsymbol{\theta}$ und σ_ε die zu schätzenden Größen sind.

Die Wahrscheinlichkeitsdichtevertteilung der vorherigen Messkampagne $\kappa - 1$ ist

$$p_{\kappa-1}(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(\kappa-1)}, \hat{\Sigma}_{\boldsymbol{\theta}}^{(\kappa-1)}) \propto e^{-\frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^{(\kappa-1)})^T (\hat{\Sigma}_{\boldsymbol{\theta}}^{(\kappa-1)})^{-1} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^{(\kappa-1)})} \quad (9.28)$$

mit

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

Die Likelihood wird im folgenden formal so aufgeschrieben, dass zu lesen ist, dass es die Wahrscheinlichkeitsdichteverteilung des Regressanden ist, gegeben die Parameter θ und σ_{ε}

$$p_{L,\kappa}(\mathbf{Y}^{(\kappa)} | \theta, \sigma_{\varepsilon}) \propto e^{-\frac{1}{2\sigma_{\varepsilon}^2} (\mathbf{Y}^{(\kappa)} - \mathbf{X}^{(\kappa)}\theta)^T (\mathbf{Y}^{(\kappa)} - \mathbf{X}^{(\kappa)}\theta)}. \quad (9.29)$$

Wir sehen, dass dies an der Verteilung selber hier nichts verändert, sondern nur der Interpretation dient, ob man auf die Verteilung des Regressanden (der direkten Messgröße) guckt oder auf die Verteilung der Modellparameter (der indirekten Messgrößen).

Zur Schätzung der Modellparameter θ und σ_{ε} berechnen wir die Wahrscheinlichkeitsdichteverteilung als Funktion von θ und σ_{ε} .

Bei der Methode der kleinsten Quadrate fiel das σ_{ε} zunächst raus, so dass der Schätzvorgang nur für das θ durchzuführen war und nach Erhalt der Schätzer $\hat{\theta}$ für die Modellparameter θ die entsprechende empirische Standardabweichung $\hat{\sigma}_{\varepsilon}$ ausgerechnet wurde.

Dass für die Likelihood und für die Verteilungsdichte der Parameter der vorherigen Messkampagne $\kappa-1$ Normalverteilungen angesetzt wurden, bedeutet folgendes: Als à priori-Information wurde in das Modell die Voraussetzung hineingesteckt, dass

$$Y_j \sim \mathcal{N}(\theta^T \begin{pmatrix} 1 \\ X_{1,j} \\ \vdots \\ X_{M-1,j} \end{pmatrix}, \sigma_{\varepsilon}) \quad \Leftrightarrow \quad \varepsilon_j \sim \mathcal{N}(0, \sigma_{\varepsilon}) \quad (9.30)$$

und

$$\theta^{(\kappa-1)} \sim \mathcal{N}(\hat{\theta}^{(\kappa-1)}, \hat{\Sigma}_{\theta}^{(\kappa-1)}). \quad (9.31)$$

Für das σ_{ε} setzen wir voraus, dass es χ^2 -verteilt sei

$$\nu \left(\frac{\sigma_{\varepsilon}}{\hat{\sigma}_{\varepsilon}^{(\kappa,R)}} \right)^2 \sim \chi_{\nu}^2 \quad (9.32)$$

wobei wir für $\hat{\sigma}_{\varepsilon}^{(\kappa,R)}$ beispielsweise das Ergebnis für die lineare Regression ohne Berücksichtigung der vorherigen Messkampagne, also gemäß der im vorigen Kapitel beschriebenen Methode, verwenden.

Als Prior-Verteilungen haben wir nun zum einen die Verteilung aus der Erfahrung der vorherigen Messkampagne(n) $\kappa-1$, Gl. (9.28) und zum anderen die Verteilung für σ_{ε}

$$p_{\chi^2}(\sigma_{\varepsilon} | \hat{\sigma}_{\varepsilon}^{(\kappa,R)}, \nu). \quad (9.33)$$

Die Posterior-Verteilung ist damit

$$\begin{aligned} p_\kappa(\boldsymbol{\theta}, \Sigma_{\boldsymbol{\theta}} | \mathbf{Y}^{(\kappa)}, \hat{\sigma}_\varepsilon^{(\kappa, R)}, \hat{\boldsymbol{\theta}}^{(\kappa-1)}, \hat{\Sigma}_{\boldsymbol{\theta}}^{(\kappa-1)}) \\ \propto p_{L,\kappa}(\mathbf{Y}^{(\kappa)} | \boldsymbol{\theta}, \sigma_\varepsilon) p_{\chi^2}(\sigma_\varepsilon | \hat{\sigma}_\varepsilon^{(\kappa, R)}, \nu) p_{\kappa-1}(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(\kappa-1)}, \hat{\Sigma}_{\boldsymbol{\theta}}^{(\kappa-1)}). \end{aligned} \quad (9.34)$$

Da in den $\mathbf{Y}^{(\kappa)}$ nur die Regressoren und die σ_ε stecken

$$\Sigma_{\boldsymbol{\theta}} = \left((\mathbf{X}^{(\kappa)})^\top \mathbf{X}^{(\kappa)} \right)^{-1} \sigma_\varepsilon^2$$

brauchen wir nicht das gesamte $\Sigma_{\boldsymbol{\theta}}$ als Funktionsvariable, sondern lediglich σ_ε , so dass die Posterior nur noch Funktion von $\boldsymbol{\theta}$ und σ_ε ist

$$\begin{aligned} p_\kappa(\boldsymbol{\theta}, \sigma_\varepsilon | \mathbf{Y}^{(\kappa)}, \hat{\sigma}_\varepsilon^{(\kappa, R)}, \hat{\boldsymbol{\theta}}^{(\kappa-1)}, \hat{\Sigma}_{\boldsymbol{\theta}}^{(\kappa-1)}) \\ \propto p_{L,\kappa}(\mathbf{Y}^{(\kappa)} | \boldsymbol{\theta}, \sigma_\varepsilon) p_{\chi^2}(\sigma_\varepsilon | \hat{\sigma}_\varepsilon^{(\kappa, R)}, \nu) p_{\kappa-1}(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(\kappa-1)}, \hat{\Sigma}_{\boldsymbol{\theta}}^{(\kappa-1)}). \end{aligned} \quad (9.35)$$

Für das vollständige Messergebnis berechnen wir die Marginalverteilung als Funktion von $\boldsymbol{\theta}$, indem wir über σ_ε integrieren

$$p_\kappa(\boldsymbol{\theta} | \mathbf{Y}^{(\kappa)}, \hat{\sigma}_\varepsilon^{(\kappa, R)}, \hat{\boldsymbol{\theta}}^{(\kappa-1)}, \hat{\Sigma}_{\boldsymbol{\theta}}^{(\kappa-1)}) = \int_0^\infty p_\kappa(\boldsymbol{\theta}, \sigma_\varepsilon | \mathbf{Y}^{(\kappa)}, \hat{\sigma}_\varepsilon^{(\kappa, R)}, \hat{\boldsymbol{\theta}}^{(\kappa-1)}, \hat{\Sigma}_{\boldsymbol{\theta}}^{(\kappa-1)}) d\sigma_\varepsilon. \quad (9.36)$$

Dies ist eine Funktion von M Variablen $\theta_1, \dots, \theta_M$. Für die Wahrscheinlichkeitsdichten als Marginalverteilung je eines der Regressionsparameter wird über die anderen (numerisch) integriert

$$\begin{aligned} p_\kappa(\theta_m | \mathbf{Y}^{(\kappa)}, \hat{\sigma}_\varepsilon^{(\kappa, R)}, \hat{\boldsymbol{\theta}}^{(\kappa-1)}, \hat{\Sigma}_{\boldsymbol{\theta}}^{(\kappa-1)}) &= \\ \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty p_\kappa(\boldsymbol{\theta} | \mathbf{Y}^{(\kappa)}, \hat{\sigma}_\varepsilon^{(\kappa, R)}, \hat{\boldsymbol{\theta}}^{(\kappa-1)}, \hat{\Sigma}_{\boldsymbol{\theta}}^{(\kappa-1)}) d\theta_1 \dots d\theta_{m-1} d\theta_{m+1} \dots d\theta_M. \end{aligned} \quad (9.37)$$

Zur Berechnung der Erwartungswertes wird (numerisch) integriert

$$E(\theta_m) = \int_{-\infty}^\infty \theta_m p_\kappa(\theta_m | \mathbf{Y}^{(\kappa)}, \hat{\sigma}_\varepsilon^{(\kappa, R)}, \hat{\boldsymbol{\theta}}^{(\kappa-1)}, \hat{\Sigma}_{\boldsymbol{\theta}}^{(\kappa-1)}) d\theta_m. \quad (9.38)$$

Das *Credible Interval* wird ebenso mittels numerischer Verfahren berechnet

$$\alpha/2 = \int_{-\infty}^{\theta_{m,\min}} p_\kappa(\theta_m | \mathbf{Y}^{(\kappa)}, \hat{\sigma}_\varepsilon^{(\kappa, R)}, \hat{\boldsymbol{\theta}}^{(\kappa-1)}, \hat{\Sigma}_{\boldsymbol{\theta}}^{(\kappa-1)}) d\theta_m \quad (9.39)$$

und

$$1 - \alpha/2 = \int_{-\infty}^{\theta_{m,\max}} p_\kappa(\theta_m | \mathbf{Y}^{(\kappa)}, \hat{\sigma}_\varepsilon^{(\kappa, R)}, \hat{\boldsymbol{\theta}}^{(\kappa-1)}, \hat{\Sigma}_{\boldsymbol{\theta}}^{(\kappa-1)}) d\theta_m. \quad (9.40)$$

Für die Regressionsanalyse kann auch eine Möglichkeit gefunden werden, eine analytische Lösung zu erhalten, wenn anstelle der χ^2 -Verteilung für die σ_ε eine Verteilung gewählt wird, so dass sich das Produkt aus Likelihood und einer alternativen Verteilung $p_a(\sigma_\varepsilon)$ vereinfachen lässt, wie es in Klauenberg *et al.* dargelegt wird [Kla15].

Da $p_{\kappa-1}$ nicht von σ_ε abhängt, gilt

$$\begin{aligned} p_\kappa(\boldsymbol{\theta} | \mathbf{Y}^{(\kappa)}, \hat{\sigma}_\varepsilon^{(\kappa,R)}, \hat{\boldsymbol{\theta}}^{(\kappa-1)}, \hat{\Sigma}_{\boldsymbol{\theta}}^{(\kappa-1)}) \\ \propto p_{\kappa-1}(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(\kappa-1)}, \hat{\Sigma}_{\boldsymbol{\theta}}^{(\kappa-1)}) \int_0^\infty p_{L,\kappa}(\mathbf{Y}^{(\kappa)} | \boldsymbol{\theta}, \sigma_\varepsilon) p_a(\sigma_\varepsilon | \hat{\sigma}_\varepsilon^{(\kappa,R)}, \nu) d\sigma_\varepsilon \end{aligned} \quad (9.41)$$

und mit einer vereinfachten Verteilung $p_a(\sigma_\varepsilon)$

$$\int_0^\infty p_{L,\kappa}(\mathbf{Y}^{(\kappa)} | \boldsymbol{\theta}, \sigma_\varepsilon) p_a(\sigma_\varepsilon | \hat{\sigma}_\varepsilon^{(\kappa,R)}, \nu) d\sigma_\varepsilon \propto e^{-\frac{1}{2} \left(\frac{f(\nu)}{\hat{\sigma}_\varepsilon^{(\kappa,R)}} \right)^2 (\mathbf{Y}^{(\kappa)} - \mathbf{X}^{(\kappa)} \boldsymbol{\theta})^\top (\mathbf{Y}^{(\kappa)} - \mathbf{X}^{(\kappa)} \boldsymbol{\theta})} \quad (9.42)$$

mit $f(\nu)$ eine sich aus p_a ergebende Funktion der Anzahl der Freiheitsgrade.

Mit Gl. (9.28) bekommen wir

$$\begin{aligned} p_\kappa(\boldsymbol{\theta} | \mathbf{Y}^{(\kappa)}, \hat{\sigma}_\varepsilon^{(\kappa,R)}, \hat{\boldsymbol{\theta}}^{(\kappa-1)}, \hat{\Sigma}_{\boldsymbol{\theta}}^{(\kappa-1)}) &\propto \\ e^{-\frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^{(\kappa-1)})^\top (\hat{\Sigma}_{\boldsymbol{\theta}}^{(\kappa-1)})^{-1} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^{(\kappa-1)})} e^{-\frac{1}{2} \left(\frac{f(\nu)}{\hat{\sigma}_\varepsilon^{(\kappa,R)}} \right)^2 (\mathbf{Y}^{(\kappa)} - \mathbf{X}^{(\kappa)} \boldsymbol{\theta})^\top (\mathbf{Y}^{(\kappa)} - \mathbf{X}^{(\kappa)} \boldsymbol{\theta})} \end{aligned} \quad (9.43)$$

so dass die Schätzer für $\boldsymbol{\theta}$ und $\Sigma_{\boldsymbol{\theta}}$ wie folgt aussehen

$$\begin{aligned} \hat{\boldsymbol{\theta}}^{(\kappa)} &= \\ \left(\left(\frac{\hat{\sigma}_\varepsilon^{(\kappa,R)}}{f(\nu)} \right)^2 (\hat{\Sigma}_{\boldsymbol{\theta}}^{(\kappa-1)})^{-1} + (\mathbf{X}^{(\kappa)})^\top \mathbf{X}^{(\kappa)} \right)^{-1} \left(\left(\frac{\hat{\sigma}_\varepsilon^{(\kappa,R)}}{f(\nu)} \right)^2 (\hat{\Sigma}_{\boldsymbol{\theta}}^{(\kappa-1)})^{-1} \boldsymbol{\theta}^{(\kappa-1)} + (\mathbf{X}^{(\kappa)})^\top \mathbf{Y}^{(\kappa)} \right) \end{aligned} \quad (9.44)$$

und

$$\hat{\Sigma}_{\boldsymbol{\theta}}^{(\kappa)} = \left((\hat{\Sigma}_{\boldsymbol{\theta}}^{(\kappa-1)})^{-1} + (\mathbf{X}^{(\kappa)})^\top \mathbf{X}^{(\kappa)} \left(\frac{f(\nu)}{\hat{\sigma}_\varepsilon^{(\kappa,R)}} \right)^2 \right)^{-1} \quad (9.45)$$

siehe auch das von K. Klauenberg mitverfasste Tutorial [Els15], das ausführlicher in der Beschreibung ist als ihre Publikation [Kla15] und auch beispielhaft Matlab-Skripte enthält.

9.5 Median einer Verteilung

Wir haben bisher den Erwartungswert einer Verteilung (1. statistisches Moment) und höhere Momente von Verteilungen (Varianz, Skewness, Kurtosis) kennengelernt, siehe 5. Vorlesung. Unter den symmetrischen Verteilungen unterscheiden wir:

- die Normalverteilung,
- die t-Verteilung, deren Ausläufer (*Tails*) um so ausgeprägter sind je kleiner die Anzahl der Freiheitsgrade ist,
- die Laplace-Verteilungen, die sehr lange Ausläufer hat. Sie setzt sich aus zwei aneinandergefügten Exponentialverteilungen zusammen, siehe 4. Vorlesung Gl.(6).

Für die Schätzung eines Modellparameters θ haben wir in der 2. Vorlesung die Methode der Maximum Likelihood kennengelernt („Finde den Parameter θ , sodass die Likelihood maximal wird“). Diese Methode führt bei normalverteilten Größen zum Least Square (Minimierung der Residuen) und wird als *L2-Norm* bezeichnet.

Zuvor hatten wir gesehen wie die Maximierung der Likelihood $l(\theta) \propto e^{-\frac{1}{2} \sum_i \left(\frac{\varepsilon_i(\theta)}{\sigma} \right)^2}$ in die Minimierung der Summe der Quadrate der Residuen übergeht

$$\max_{\theta} \left\{ e^{-\frac{1}{2} \sum_i \left(\frac{\varepsilon_i(\theta)}{\sigma} \right)^2} \right\} \Rightarrow \min_{\theta} \left\{ -\log \left(e^{-\frac{1}{2} \sum_i \left(\frac{\varepsilon_i(\theta)}{\sigma} \right)^2} \right) \right\},$$

d.h.

$$\min_{\theta} \left\{ \sum_i \left(\frac{\varepsilon_i(\theta)}{\sigma} \right)^2 \right\}.$$

Mit $\text{Var}(\varepsilon) = \sum_i \varepsilon_i(\theta)^2$ bedeutet dies, dass die Varianz der Residuen minimiert wird

$$\min_{\theta} \left\{ \int_{-\infty}^{\infty} \text{Var}(\varepsilon) l(\theta) d\theta \right\}. \quad (9.46)$$

Analog lässt sich auch die Varianz bezüglich der Posterior $p(\theta|X)$ für die bayes'sche Statistik minimieren:

$$\min_{\theta} \left\{ \int_{-\infty}^{\infty} \text{Var}(\varepsilon) p(\theta|X) d\theta \right\}. \quad (9.47)$$

Betrachte einen skalaren Modellparameter θ mit Schätzwert $\hat{\theta}$ und $\varepsilon = \theta - \hat{\theta}$, so dass

$$\min_{\theta} \left\{ \int_{-\infty}^{\infty} \varepsilon^2 p(\theta|X) d\theta \right\}. \quad (9.48)$$

Dabei ist $\int_{-\infty}^{\infty} \varepsilon^2 p(\theta|X) d\theta$ gleich dem Erwartungswert der Residuenquadrate, mit $\varepsilon = \theta - \hat{\theta}$

$$\int_{-\infty}^{\infty} \varepsilon^2 p(\theta|X) d\theta = E((\theta - \hat{\theta})^2)$$

Die erste Ableitung wird zu Null gesetzt:

$$\frac{d \epsilon}{d \hat{\theta}} = \frac{d}{d \hat{\theta}} \int_{-\infty}^{\infty} (\hat{\theta} - \theta)^2 p(\theta|X) d\theta = 0 \quad (9.49)$$

$$2 \int_{-\infty}^{\infty} (\hat{\theta} - \theta) p(\theta|X) d\theta = 0$$

Daraus folgt:

$$\hat{\theta} \int_{-\infty}^{\infty} p(\theta|X) d\theta = \int_{-\infty}^{\infty} \theta p(\theta|X) d\theta$$

Das linke Integral ist auf Grund der Normierungsbedingung $\int_{-\infty}^{\infty} p(\theta|X) d\theta = 1$. Es folgt somit:

$$\hat{\theta} = \int_{-\infty}^{\infty} \theta p(\theta|X) d\theta = E(p(\theta|X)) \quad (9.50)$$

Der Erwartungswert $E(p(\theta|X))$ der Posterior-Verteilung $p(\theta|X)$ ist derjenige Schätzwert $\hat{\theta}$, der die Residuen minimiert.

Zur Behandlung von Beobachtungen, die nicht rein stochastischer Natur sind, sondern durch andere, nicht weiter modellierbare Einflüsse gegeben sind, wie z.B. mit einer größeren Wahrscheinlichkeit - als bei der Normalverteilung in den Ausläufern, wird beispielsweise auf die Laplaceverteilung zurückgegriffen werden. Die Maximum-Likelihood-Methode für die Laplace-Transformation führt auf die Minimierung der Absolutbeträge der Residuen (siehe 2. Vorlesung), d.h. $\sum |\epsilon| \rightarrow \min$. Man bezeichnet dies als die *L1-Norm*. Sie liefert den *Median*. Er ist definiert als das 0.5-Quantil $Q_{0.5}$. Ein Quantil Q_α ist ein Lagemaß in der Statistik (siehe 5. Vorlesung). So liegen beim 0.5-Quantil 50% der Verteilung unterhalb und 50% der Verteilung oberhalb des Quantils.

Wir schauen uns die Minimierung der Erwartungswert der absoluten Abweichung zwischen $\hat{\theta}$ und θ , d.h.

$$\epsilon(\hat{\theta}) := E(|\hat{\theta} - \theta|) \rightarrow \min.$$

mit

$$E(|\hat{\theta} - \theta|) = \int_{-\infty}^{\infty} |\hat{\theta} - \theta| p(\theta|X) d\theta$$

Das Minimum erhalten wir durch Differentiation $d\epsilon/d\hat{\theta} = 0$

$$\begin{aligned} \frac{d\epsilon}{d\hat{\theta}} &= \frac{d}{d\hat{\theta}} \int_{-\infty}^{\infty} |\hat{\theta} - \theta| p(\theta|X) d\theta \\ &= \frac{d}{d\hat{\theta}} \left[\int_{-\infty}^{\hat{\theta}} (\hat{\theta} - \theta) p(\theta|X) d\theta + \int_{\hat{\theta}}^{\infty} (\theta - \hat{\theta}) p(\theta|X) d\theta \right] = 0 \end{aligned}$$

Mit der Leibnizschen Regel für Parameterintegrale leiten wir die Gleichung ab und erhalten:

$$\int_{-\infty}^{\hat{\theta}} p(\theta|X) d\theta - \int_{\hat{\theta}}^{\infty} p(\theta|X) d\theta = 0$$

Die Summe der beiden Integrale ist als Dichte-Integral gleich 1. Daraus folgt:

$$\int_{-\infty}^{\hat{\theta}} p(\theta|X) d\theta = \int_{\hat{\theta}}^{\infty} p(\theta|X) d\theta = 0.5$$

Somit ist $\hat{\theta}$ der Median bzw. das 50-Prozent-Quantil $Q_{0.5}$.

Der Median ist robust gegenüber langen „tails“ und kann für jede beliebige Posterior-Verteilung bestimmt werden.

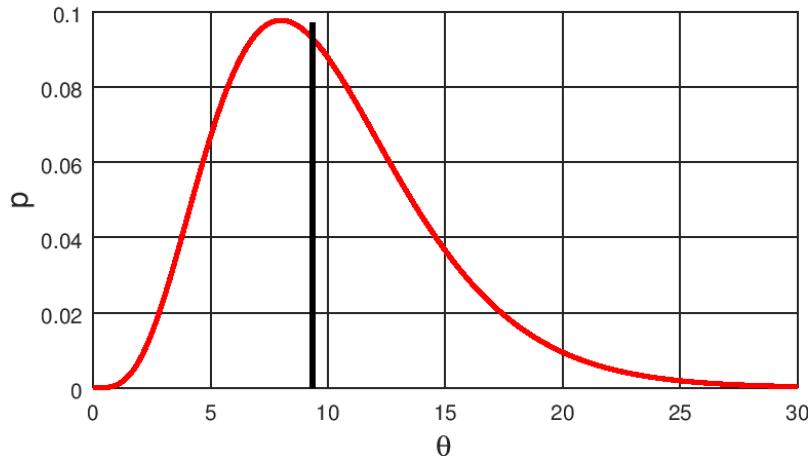


Abbildung 9.10: Beispiel für den Median, er entspricht hier nicht dem Maximum. Der Median ist definiert, so dass 50 % der Fläche unterhalb und 50 % der Fläche oberhalb des Medians liegen

Wesentliche Unterschiede zwischen der L1- und der L2-Norm sind in Tab. 9.4 aufgezählt.

Tabelle 9.4: Unterschiede zwischen L2- und L1-Norm

L2-Norm: Minimierung der quadr. Abweichungen	L1-Norm: Minimierung der Beträge
Nicht sehr robust	robust
stabile Lösung	instabile Lösung möglich
immer nur eine Lösung möglich	mehrere Lösungen möglich

Literaturverzeichnis

zu Kapitel 8:

- [VIM08] VIM International vocabulary of metrology – Basic and general concepts and associated terms (VIM) 3rd edition (2008)
- [GUM95] JCGM 100:2008 BIPM, IEC, IFCC, ISO, IUPAC, IUPAP and OIML 1995 Guide to the Expression of Uncertainty in Measurement (GUM) (Geneva, Switzerland: International Organization for Standardization) ISBN 92-67-10188-9
- [Kac03] R. Kacker, A. Jones: On use of Bayesian statistics to make the Guide to the Expression of Uncertainty in Measurement consistent, Metrologia, Volume 40, Issue 5, pp. 235-248 (2003).
- [Hel08] L. Held: Methoden der statistischen Inferenz. Likelihood und Bayes. Spektrum, Heidelberg (2008)
- [Tsc14] W. Tscheirik: Statistik: Klassisch oder Bayes. Zwei Wege im Vergleich, Springer Lehrbuch, (2014).
- [Kla15] K. Klauenberg et al. A tutorial on bayesian normal linear regression. Metrologia, 52, 2017.
- [Els15] C. Elster, K. Klauenberg, M. Walzel, G. Wübbeler, P. Harris, M. Cox, C. Matthews, I. Smith, L. Wright, A. Allard, N. Fischer, S. Cowen, S. Ellison, P. Wilson, F. Pennecchi, G. Kok, A. van der Veen, L. Pendrill, A Guide to Bayesian Inference for Regression Problems, Deliverable of EMRP project NEW04 “Novel mathematical and statistical approaches to uncertainty evaluation”, 2015
www.ptb.de/emrp/fileadmin/documents/nmasatue/NEW04/Papers/BPGWP1.pdf

9.6 Übungsaufgaben

Aufgabe 9-1: Likelihood des Erwartungswertes bei bekannter Varianz

- (a) Bestimmen Sie die Likelihood einer normalverteilten Zufallsgröße X , die sich aus J unabhängigen Beobachtungen X_1, \dots, X_J ergibt. Die Varianz σ^2 der Zufallsgröße X sei bekannt.

Lösung zu 9-1(a)

Die Likelihoodfunktion ist gegeben durch:

$$l(\theta|X) = \prod_{j=1}^J l(\theta|X_j)$$

Die Beobachtungen sind normalverteilt mit Erwartungswert $\hat{\theta}$ und Varianz σ^2 . Es ergibt sich:

$$l(\theta|X) = \prod_{j=1}^J \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}\left(\frac{X_j - \hat{\theta}}{\sigma}\right)^2}$$

Somit erhalten wir:

$$\begin{aligned} l(\theta|X) &\propto \prod_{j=1}^J e^{-\frac{(X_j - \hat{\theta})^2}{2\sigma^2}} \\ &= e^{-\frac{\sum_{j=1}^J (X_j - \hat{\theta})^2}{2\sigma^2}} \\ &= e^{-\frac{\sum_{j=1}^J (X_j^2 - 2X_j\hat{\theta} + \hat{\theta}^2)}{2\sigma^2}} \\ &= e^{-\frac{\sum_{j=1}^J X_j^2 - 2(\sum_{j=1}^J X_j)\hat{\theta} + J\hat{\theta}^2}{2\sigma^2}} \\ &= e^{-\frac{\bar{X}^2 - 2\bar{X}\hat{\theta} + \hat{\theta}^2}{2\sigma^2/J}} \end{aligned}$$

Die empirische Varianz s^2 ist gegeben durch:

$$\begin{aligned} s^2 &= \frac{1}{J} \sum_{j=1}^J (X_j - \bar{X})^2 \\ &= \frac{1}{J} (\sum_j X_j^2 - 2 \sum_j X_j \bar{X} + \sum_j \bar{X}^2) \\ &= \bar{X}^2 - (\bar{X})^2 \end{aligned}$$

Damit erhalten wir für die Likelihood:

$$\begin{aligned} l(\theta|X) &\propto e^{-\frac{\bar{X}^2 + s^2 - 2\bar{X}\hat{\theta} + \hat{\theta}^2}{2\sigma^2/J}} \\ &= e^{-\frac{\bar{X}^2 - 2\bar{X}\hat{\theta} + \hat{\theta}^2}{2\sigma^2/J}} \cdot e^{-\frac{s^2}{2\sigma^2/J}} \\ &= e^{-\frac{(\bar{X} - \hat{\theta})^2}{2\sigma^2/J}} \cdot e^{-\frac{s^2}{2\sigma^2/J}} \end{aligned}$$

Da der Term

$$e^{-\frac{s^2}{2\sigma^2/J}} = \text{const.}$$

ist, geht dieser Term nur in die Normierung ein. Es folgt für die Likelihood der Stichproben:

$$l(\theta|X) = l(\theta|\bar{X}) \propto e^{-\frac{(\bar{X} - \hat{\theta})^2}{2\sigma^2/J}}$$

Der Stichprobenmittelwert \bar{X} ist also normalverteilt mit dem Erwartungswert $\hat{\theta}$ und der Varianz σ^2/J . Hinweis / Ergänzung: Ist die Varianz von X unbekannt, dann schätzen wir diese durch die empirische Varianz s^2 (klassische Statistik). Für annähernd große Stichproben geht die t-Verteilung in die Normalverteilung über und der Stichprobenmittelwert ist normalverteilt mit der Varianz s^2/J . Dann erhalten wir analog für die Likelihood des Mittelwertes:

$$l(\theta|\bar{X}) \propto e^{-\frac{(\bar{X} - \hat{\theta})^2}{2s^2/J}}$$

Aufgabe 9-2: Bestimmung des Credible-Intervalls einer Posterior-Verteilung

Sie wollen die Temperatur T eines Laborraumes bestimmen. Da hier T der zu bestimmende Parameter ist, bezeichnen wir diesen im Folgenden mit θ . Aus Erfahrung wissen wir (wenn die Klimaanlage gerade nicht ausgefallen ist :-)), dass die Temperatur normalverteilt bei 20°C mit einer Steigung von 2°C ist. D.h. als Prior nehmen wir an $\theta \sim N(\mu = 20, \sigma^2 = 2^2)$. Zur genauen Bestimmung der Labortemperatur verwenden Sie nun einen Temperaturfühler, der einer Normalverteilung mit $N(20.5, 0.5^2)$ folgt.

- (a) Bestimmen sie die Posteriorverteilung, wenn eine Stichprobe von 4 Messungen durchgeführt wird.
- b) Bestimmen Sie die obere und untere Grenze des 95%igen Credible-Intervall der Posterior-Verteilung.

Lösung zu 9-2(a) In der Vorlesung wurde gezeigt, dass wenn Priorverteilung und die Likelihood normalverteilt sind, auch die Posteriorverteilung normalverteilt ist. Die Likelihood bei 16 Stichproben ist normalverteilt mit $\sigma/\sqrt(J) = 0.5/\sqrt{4} = 0.25$

$$w_0 = \frac{1}{\sigma_0^2} = 1/2^2 = 0.25; w_1 = \frac{1}{\sigma_0^2/J} = 1/0.25^2 = 16;$$

$$\bar{\sigma}^2 = 1/(w_0 + w_1) = 0.062$$

$$\bar{\theta} = \frac{1}{w_0 + w_1}(w_0\theta_0 + w_1x) = \frac{1}{0.25 + 16}(0.25 \cdot 20 + 16 \cdot 20.5) = 20.49$$

Die Posterior ist also eine Normalverteilung mit $N(20.49, 0.25^2)$. Das Vorwissen hat hier fast keinen Einfluss mehr auf die Posterior.

Lösung zu 9-2(b)

$N(20.49, 0.25^2)$ ist gegeben.

Mit $Z = (\theta - \mu)/\sigma$ (Standardisieren) und mit der rechtseitigen Verteilungsfunktion Φ erhält man die obere Grenze θ_{oG} durch:

$$\Phi\left(\frac{\theta_{oG} - \mu}{\sigma}\right) = 0.5 + 0.95/2 = 0.975$$

In einer Quantiltabelle der Normalverteilung (siehe Anhang Kapitel 12) kann der Wert 1.96 abgelesen werden. Wir erhalten damit:

$$\frac{\theta_{oG} - 20.49}{0.25} = 1.96 \rightarrow \theta_{oG} = 20.98 \rightarrow \theta_{uG} = 20.98 - 2 \cdot (20.98 - 20.49) = 20$$

Kapitel 10

Monte-Carlo Methoden für Messunsicherheitsberechnungen

10.1 Einbeziehen von a-Priori-Wissen, „scharfer“ Model-Prior

10.1.1 Kurze Wiederholung: Bayessche Statistik zur MU-Bestimmung

In der 8. Vorlesung haben wir das Bayes-Theorem kennengelernt. Mit diesem ist es möglich a-Priori-Wissen $p(Y)$ über die zu bestimmende indirekte Messgröße Y zu berücksichtigen. Wir haben folgenden Zusammenhang kennengelernt:

$$p(Y|X_1, \dots, X_N) \propto l(Y|X_1, \dots, X_N) \cdot p(Y) \quad (10.1)$$

bzw.

$$p(Y|X_1, \dots, X_N) \propto p(X_1, \dots, X_N|Y) \cdot p(Y)$$

mit $p(Y)$ dem Prior der indirekten Messgröße Y (auch oft als Ausgangsgröße bezeichnet), $l(Y|X_1, \dots, X_N)$ Likelihood der Messdaten (Eingangsgrößen) X_1, \dots, X_N .

Die Likelihood kann auch – wie in der 8. Vorlesung dargestellt – als bedingte Wahrscheinlichkeitsdichte $p(X_1, \dots, X_N|Y)$ aufgefasst werden.

Der *Erwartungswert* $y := E(Y)$ ist der *beste Schätzwert* für die Messgröße Y und die *Varianz* ist ein Maß für die Unsicherheit $u^2(y) := \text{Var}(Y)$.

Beispiel: Bestimmung einer physikalischen Größe

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 Temperatur 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 Beispiele denkbar. Sensoren sind oft so konzipiert, dass sie ein physikalisches Phänomen nutzen und zur elektronischen Erfassung der zu messenden Größe als direkte Größe ein Messignal 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 (Rohdaten), das als Spannung U bzw. direkte Größe X_M vorliegt, ist über einen Kalibrierfaktor K bzw. eine direkte Größe X_K in die indirekte physikalische Größe Y umzurechnen. Der Index M steht hier für Messung und der Index K für Kalibrierfaktor.

In dem betrachteten Beispiel ergebe das Produkt

$$Y = f(X_K, X_M) = X_K X_M \quad (10.2)$$

der beiden direkten Größen X_K und X_M die indirekte Größe Y .

Um den Unterschied zwischen der bayesischen Statistik im eigentlichen Sinne und einem Ansatz, der dem GUM Supplement 1 [GUMS1] zugrunde gelegt wird, zu veranschaulichen, werden wir uns klar machen, dass es einerseits die Sichtweise gibt,

- dass nicht nur die beiden direkten Messgrößen X_K und X_M Zufallsgrößen sind, sondern auch die indirekte Messgröße eine Zufallsgröße Y ist (**bayesische Statistik**), die ihrerseits zusätzlich zu der Streuung der beiden direkten Größen einen Beitrag zur Streuung liefert,

und andererseits,

- dass die Streuung der indirekten Größe Y von der Streuung der beiden direkten Messgrößen X_K und X_M herröhrt, dass aber Y selber von sich aus nicht streut, sondern dass es einen „**scharfen**“ **Model-Prior** gibt.

Mit dem Begriff „scharf“ ist hier gemeint, dass $Y = f(X_K, X_M)$ keine Streuung haben würde, wenn die beiden direkten Größen X_K und X_M keine Streuung hätten.

Der Einfachheit halber nehmen wir an, dass die Streuung der beiden direkten Größen normalverteilt sei, so dass für deren Wahrscheinlichkeitsdichteverteilungen, *Probability Density Functions* - kurz **pdfs**, folgendes gelte: Der Kalibrierfaktor sei nach Herstellung des Sensors bestimmt worden mit einem Wert K_0 und einer Unsicherheit U_K , zu der der Erweiterungsfaktor mit $k = 2$ angegeben sei, so dass wir die Standardabweichung $s_K = \frac{1}{2}U_K$.

$$p_K \propto e^{-\frac{1}{2} \left(\frac{X_K - K_0}{s_K} \right)^2} \quad (10.3)$$

Diese Verteilung betrachten wir als Prior, der in Abb. 10.1 dargestellt ist. Ferner sagen wir, dass wir

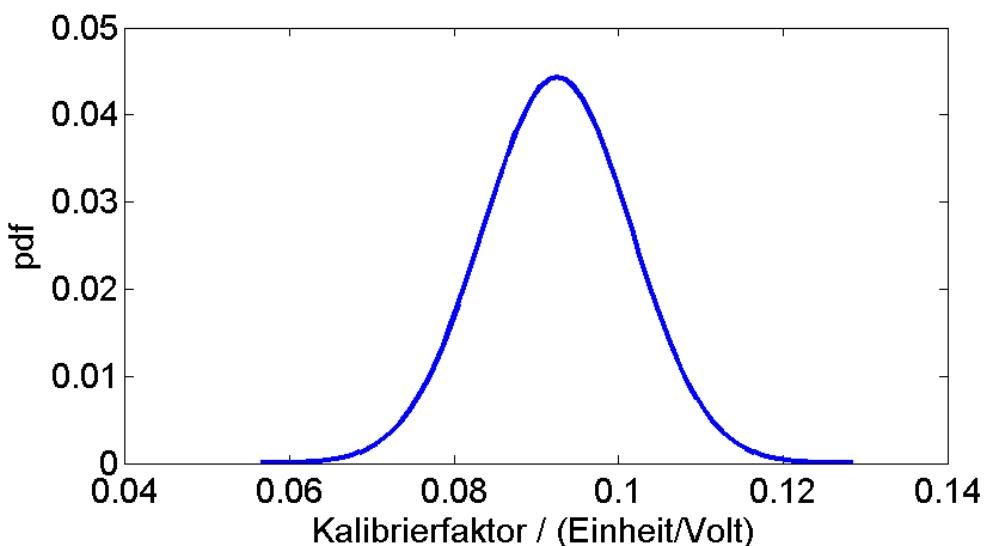


Abbildung 10.1: Wahrscheinlichkeitsdichteverteilung, Prior, des Kalibrierfaktors zur Umrechnung der Sensordaten, die die physikalische Dimension einer Spannung in der Einheit Volt hat, in die Dimension der indirekten Messgröße Y in ihrer Einheit.

in unserem Beispiel eine Stichprobe mit $J = 11$ Messwerten mit dem Sensor aufgenommen haben, $(X_{M,1}, \dots, X_{M,J})$. Der zu messende Gegenstand mit Messgröße Y , deren physikalische Einheit wir hier ganz allgemein einfach *Einheit* nennen, solle so streuen, dass die gemessenen Spannungswerte $(X_{M,1}, \dots, X_{M,J})$ mit $\sigma = 12$ Volt streuen. Wir machen uns dazu klar, dass die Streuung von $\sigma = 12$ V nicht die Eigenschaft des Sensors ist, sondern aus der Eigenschaft des Messgegenstands kommt!

Als à-priori Wissen zu dem Sensor kennen wir für einen Fall (A.) die Unsicherheit des Sensors unabhängig von dem aktuell vorliegenden Gegenstand der Messung. Die dazu proportionale Standardabweichung s_S des Sensors mit Index S für Sensor sei signifikant kleiner als die Streuung der Daten der Stichprobe.

Ferner untersuchen wir Auswerteverfahren ohne Kenntnis der Unsicherheit des Sensors, für die die gemeinsame Unsicherheit von Sensor und der Größe Y des Gegenstands der Messung verwendet wird, um die Likelihood zu berechnen, also eine Standardabweichung s_M , die sich aus der empirischen Varianz der Stichprobe ergibt. Dazu verwenden wir die Wurzel aus der empirischen Varianz. Dies nennen wir Fall (B.).

Wir betrachten nun folgende vier Ansätze zur Bestimmung einer Posterior:

- (A.) Präziser Sensor, der die Rohdaten in der physikalischen Einheit Volt ausgibt und sehr wenig streut (z. B. $s_S = 0.3$ V). Die sensorintrinsische Standardabweichung s_S wird für die Likelihood der direkten Messgröße X_M verwendet.

$$p_L \propto \prod_{j=1}^J e^{-\frac{1}{2} \left(\frac{X_M - X_{M,j}}{s_S} \right)^2} \quad (10.4)$$

- (B.) Das intrinsische Streuverhalten des Sensors sei unbekannt und damit nicht trennbar von der Streuung der Größe Y . Eine gemeinsame Streuung s_M , die signifikant größer ist als die intrinsische Sensorstreuung s_S wird für die Likelihood der Sensorspannung X_M verwendet.

$$p_L \propto \prod_{j=1}^J e^{-\frac{1}{2} \left(\frac{X_M - X_{M,j}}{s_M} \right)^2} \quad (10.5)$$

In Abb. 10.2 ist die Likelihood gemäß Gl. (10.4) mit schwarzer, durchgezogener Kurve dargestellt. Sie setzt sich zusammen aus einzelnen schmalen Peaks, Gaußverteilungen, weil die Breite der Normalverteilungen mit $s_S = 0.3$ V wesentlich schmäler ist als der Abstand zwischen den einzelnen Stichprobenelementen. Gl. (10.5) mit der gemeinsamen, breiten Streuung, die von der Streuung des Messgegenstandes dominiert wird, ist als rot gestrichelte Kurve eingezeichnet. Hier sehen wir eine gemeinsame, breite Normalverteilung.

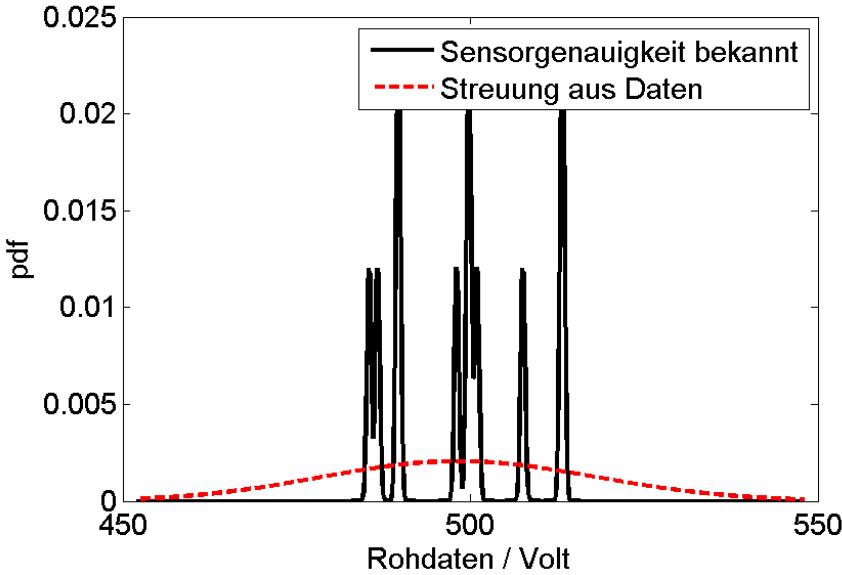


Abbildung 10.2: Likelihood (Wahrscheinlichkeitsdichtefunktion pdf) der direkt gemessenen Größe X_M , der Rohdaten: Schwarze, durchgezogene Linie für die pdf mit à-priori-Wissen über die Genauigkeit des Sensors/Messgerätes („Geraet genau“); rote gestrichelte Kurve für die pdf für unbekannte Streuung vom Gerät, mit der aus der empirischen Standardabweichung der Rohdaten bestimmten Streuung, so dass das Gerät als ungenau erscheint erscheint („Geraet ungenau“)

1. Im Fall 1 verwenden wir die Likelihood mit bekannter Streuung $s_S = 0.3$ V des Präzisionssensors (A.), Gl. (10.4), zusammen mit der Verteilung des Kalibrierfaktors, dem Prior Gl. (10.3), und einer Normalverteilung für die indirekte Messgröße Y mit einer zu schätzenden Standardabweichung σ_Y :

$$p_{\text{Modell}}(X_K, X_M | Y, \sigma_Y) \propto e^{-\frac{1}{2} \left(\frac{Y - X_K X_M}{\sigma_Y} \right)^2} \quad \text{mit} \quad (J-1) \left(\frac{\sigma_Y}{s_{KM}} \right)^2 \sim \chi^2 \quad (10.6)$$

Dabei ist s_{KM} eine zur Streuung der Rohdaten proportionale Größe $s_{KM} = K_0 s_M$. Der Faktor K_0 ist der à-priori bekannte Schätzer des Kalibrierfaktors. Die Posterior ist dann

$$\begin{aligned} p(Y, X_K, X_M, \sigma_Y | (X_{M,1}, \dots, X_{M,J}), K_0, s_K, s_S) &\propto \\ e^{-\frac{1}{2} \left(\frac{Y - X_K X_M}{\sigma_Y} \right)^2} \prod_{j=1}^J e^{-\frac{1}{2} \left(\frac{X_M - X_{M,j}}{s_S} \right)^2} e^{-\frac{1}{2} \left(\frac{X_K - K_0}{s_K} \right)^2} p_{\chi^2}((J-1) \left(\frac{\sigma_Y}{s_{KM}} \right)^2). \end{aligned} \quad (10.7)$$

2. Im Fall 2 betrachten wir für die Likelihood (B.) der direkten Messgröße X_M die Standardabweichung s_M , Gl. (10.5), und für Verteilung B („Gerät ungenau“) in Kombination mit der Verteilung des Kalibrierfaktors und für die Modellverteilung nehmen wir wieder eine Standardabweichung s_{KM} an.

$$\begin{aligned} p(Y, X_K, X_M, \sigma_Y | (X_{M,1}, \dots, X_{M,J}), K_0, s_K) &\propto \\ e^{-\frac{1}{2} \left(\frac{Y - X_K X_M}{\sigma_Y} \right)^2} \prod_{j=1}^J e^{-\frac{1}{2} \left(\frac{X_M - X_{M,j}}{s_M} \right)^2} e^{-\frac{1}{2} \left(\frac{X_K - K_0}{s_K} \right)^2} p_{\chi^2}((J-1) \left(\frac{\sigma_Y}{s_{KM}} \right)^2) \end{aligned} \quad (10.8)$$

mit s_M für die empirische Standardabweichung berechnet aus der Stichprobe $(X_{M,1}, \dots, X_{M,J})$.

3. Im Fall 3 nehmen wir genauso wie in Fall 2 wieder (B.) („Gerät ungenau“) dieselbe Likelihood und verwenden wie gehabt die Verteilung des Kalibrierfaktors. Wir wollen jetzt den Übergang zu dem Ansatz mit scharfem *Model-Prior* bekommen, indem wir für σ_Y einen sehr, sehr kleinen Wert einsetzen, so dass die Verteilungsdichte für das Modell Gl. (10.6) sehr schmal wird, weil die Dirac-Funktion als Normalverteilung für $\sigma \rightarrow 0$ interpretiert werden kann.

$$p(Y, X_K, X_M | (X_{M,1}, \dots, X_{M,J}), K_0, s_K) \propto e^{-\frac{1}{2} \left(\frac{Y - X_K X_M}{\sigma_{\text{winzig}}} \right)^2} \prod_{j=1}^J e^{-\frac{1}{2} \left(\frac{X_{M,j} - X_{M,j}}{s_M} \right)^2} e^{-\frac{1}{2} \left(\frac{X_K - K_0}{s_K} \right)^2} \quad (10.9)$$

mit $\sigma_{\text{winzig}} = 0.003 \cdot s_{KM}$.

4. Im Fall 4 wird für die pdf des Modells der Grenzübergang $\sigma_Y \rightarrow 0$ vollzogen, siehe Paper von Wuebbler et al. [Wue08], also

$$p_{\text{Modell}}(X_K, X_M | Y) \propto \lim_{\sigma_Y \rightarrow 0} \frac{1}{\sigma_Y} e^{-\frac{1}{2} \left(\frac{Y - X_K X_M}{\sigma_Y} \right)^2} = \delta(Y - X_K X_M) \quad (10.10)$$

oder allgemein für jedes Modell der Gestalt $Y = f(X_1, \dots, X_N)$

$$p_{\text{Modell}}(X_K, X_M | Y) \propto \lim_{\sigma_Y \rightarrow 0} \frac{1}{\sigma_Y} e^{-\frac{1}{2} \left(\frac{Y - f(X_1, \dots, X_N)}{\sigma_Y} \right)^2} = \delta(Y - f(X_1, \dots, X_N)). \quad (10.11)$$

Damit sieht der Posterior wie folgt aus

$$p(Y, X_K, X_M | (X_{M,1}, \dots, X_{M,J}), K_0, s_K) \propto \delta(Y - X_K X_M) e^{-\frac{1}{2} \left(\frac{X_M - X_{M,j}}{s_M} \right)^2} e^{-\frac{1}{2} \left(\frac{X_K - K_0}{s_K} \right)^2}. \quad (10.12)$$

Wir berechnen für die jeweiligen Posteriors wieder die jeweilige Marginalverteilung, bei der die Posterior als Funktion von Y gewonnen wird, indem über die Größen X_K , X_M und σ_Y integriert wird. Bei dem im Anhang abgedruckten Matlab/Gnu-Octaveskript wurde eine Vereinfachung vorgenommen, indem darauf verzichtet wurde, für σ_Y eine Verteilungsdichtefunktion $p_{\chi^2}((J-1) \left(\frac{\sigma_Y}{s_{KM}} \right)^2)$ zu verwenden und dann über σ_Y zu integrieren. Hier wurde einfach nur für σ_Y die zur Streuung der Rohdaten proportionale Streuung s_{KM} eingesetzt, also $\sigma_Y = s_{KM}$. Um von der Überdeckung her sicherer zu sein, haben wir diese aufgrund des sehr kleinen Stichprobenumfangs noch mit einem heuristischen Faktor $c > 1$ multipliziert

$$s_{KM} = c K_0 s_M,$$

so dass für die Posteriors der Fälle 1 und 2 folgende Integrationen durchgeführt wurden

$$p(Y|(X_{M,1}, \dots, X_{M,J}), K_0, s_K, s_S) \propto \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2} \left(\frac{Y - X_K X_M}{\sigma_Y} \right)^2} \prod_{j=1}^J e^{-\frac{1}{2} \left(\frac{X_M - X_{M,j}}{s_M} \right)^2} e^{-\frac{1}{2} \left(\frac{X_K - K_0}{s_K} \right)^2} dX_M dX_K \quad (10.13)$$

für Fall 1, bei dem s_S die aus vorheriger Charakterisierung des Sensors, d.h. aus à-priori Wissen, bekannte Standardabweichung der Sensorspannung ist und

$$p(Y|(X_{M,1}, \dots, X_{M,J}), K_0, s_K) \propto \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2} \left(\frac{Y - X_K X_M}{\sigma_Y} \right)^2} \prod_{j=1}^J e^{-\frac{1}{2} \left(\frac{X_M - X_{M,j}}{s_M} \right)^2} e^{-\frac{1}{2} \left(\frac{X_K - K_0}{s_K} \right)^2} dX_M dX_K \quad (10.14)$$

für Fall 2, bei dem s_M die aus der Stichprobe $(X_{M,1}, \dots, X_{M,J})$ berechnete empirische Standardabweichung ist.

Die Marginalverteilung zu Posterior Gl. (10.12) für Fall 4, also die pdf als Funktion von Y , sieht wie folgt aus

$$p(Y|(X_{M,1}, \dots, X_{M,J}), K_0, s_K) \propto \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(Y - X_K X_M) \prod_{j=1}^J e^{-\frac{1}{2} \left(\frac{X_M - X_{M,j}}{s_M} \right)^2} e^{-\frac{1}{2} \left(\frac{X_K - K_0}{s_K} \right)^2} dX_M dX_K. \quad (10.15)$$

Damit berechnen wir für alle möglichen Paare X_K, X_M direkt je ein Y mit $Y = X_K X_M$. Zu jedem der Paare X_K, X_M berechnen wir die Wahrscheinlichkeit

$$p(X_K X_M) \propto \prod_{j=1}^J e^{-\frac{1}{2} \left(\frac{X_M - X_{M,j}}{s_M} \right)^2} e^{-\frac{1}{2} \left(\frac{X_K - K_0}{s_K} \right)^2},$$

die dann die pdf als Funktion von Y ist.

In dem im Anhang gezeigten Matlab/Gnu-octaveskript haben wir für die Größen X_K und X_M ein äquidistant gerastertes Gitter verwendet und alle Möglichkeiten durchgerechnet. Das lässt sich bei diesem kleinen Beispiel und der Leistung der heutigen Rechentechnik leicht durchführen. Bei umfangreicheren Problemstellungen jedoch ist diese „brute force“ nicht sinnvoll. Für solche Fälle werden solche Samples für die direkten Messgrößen gewählt, die bereits berücksichtigen, wie ihre Wahrscheinlichkeitsdichteverteilungen aussehen. Dies können nichtäquidistante, deterministische Samples sein, deren Intervalle enger sind je größer die Wahrscheinlichkeit ist. Dies können aber auch stochastisch gewählte Samples sein, die sich ebenfalls an der Wahrscheinlichkeitsdichte der direkten Größen orientieren. Für die stochastischen Samples werden Pseudozufallszahlen eingesetzt, also Zahlen, die zufällig aussehen, die aber mit deterministischen, numerischen Algorithmen berechnet werden. Wegen des zufälligen Charakters der Zahlen, den auch Zahlen, die bei Spielen mit Würfeln und Roulettescheiben gefunden werden, haben, werden diese Methoden **Monte-Carlo-Methoden** genannt.

Abb. 10.3 zeigt alle vier Posteriomarginalverteilungen, die pdf für Fall 1 gemäß Gl. (10.13) als

schwarze, durchgezogene Linie mit der Bezeichnung in der Legende „Geraet genau, Messobjekt schlecht“, die pdf für Fall 2 gemäß Gl. (10.14) als rote, gestrichelte Linie mit der Bezeichnung „beides ungenau“, die Marginalverteilung zu Gl. (10.9) als blaue, durchgezogene Kurve und die Wahrscheinlichkeiten zu den paarweise berechneten Werten für Y gemäß der Marginalverteilung mit dem scharfen Model-Prior Gl. (10.15).

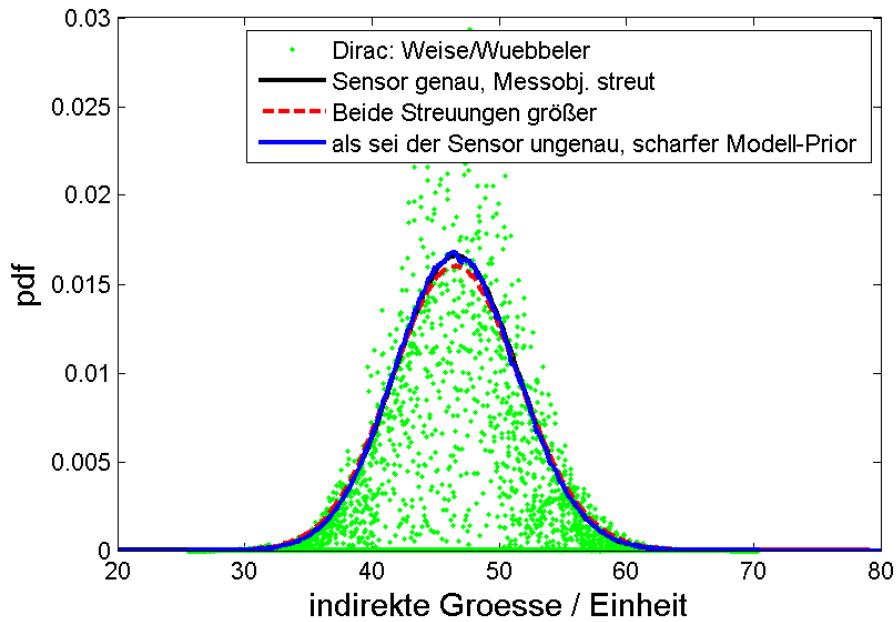


Abbildung 10.3: Posteriorverteilungen für die Bestimmung der indirekten Größe für die 4 Fälle

10.1.2 Modellgleichung als à-Priori-Wissen → klass. Statistik

Wenn eine Modellgleichung gegeben ist, so kann diese als à-Priori-Wissen miteinbezogen werden. Gegeben sei die Modellgleichung zwischen der gesuchten Messgröße Y und den direkten Messgrößen X_1, \dots, X_N , die wir auch in einem Vektor $\mathbf{X} := (X_1 \dots X_N)$ schreiben

$$Y = f(X_1, \dots, X_N) = f(\mathbf{X}) \quad (10.16)$$

Werden alle Größen, sowohl die direkten $\mathbf{X} := (X_1 \dots X_N)$ als auch die indirekte Größe Y als Zufallsgrößen angesehen, so gibt es zu jeder der Größen eine stochastische Komponente, die wir oftmals als normalverteilt erachten können. Wir können aber auch für bestimmte physikalische Gegebenheiten, wie beispielsweise Störungen durch Vibrationen, als verteilt nach einem Arcussinus-Verhalten berechnen. Zu jeder der Größen gibt es also Abweichungen, so dass wir schreiben

$$Y + \varepsilon_{Y-\text{all}} = f(X_1 + \varepsilon_1, \dots, X_N + \varepsilon_N) + \varepsilon_{Y-\text{intrinsisch}} \quad (10.17)$$

Wir wollen im weiteren Fall 4 des oben dargelegten Beispiels, nämlich den Fall eines **scharfen Model-Priors**, genauer beleuchten. Mit dem Begriff des scharfen Model-Priors ist gemeint, dass das Modell selber scharf ist, also Gl. (10.16) genau so gilt und nur die direkten Größen als Zufallsgrößen erachtet werden

$$Y + \varepsilon_Y = f(X_1 + \varepsilon_1, \dots, X_N + \varepsilon_N) \quad (10.18)$$

sodass die Abweichungen ε_Y der indirekten Größe Y allein von den Abweichungen ε_i der direkten Größen abhängt und keine der Größe Y eigene (intrinsische) Abweichung zugeordnet wird.

$$\varepsilon_Y \approx \sum_{i=1}^N \frac{\partial}{\partial X_i} f(X_1 + \varepsilon_1, \dots, X_N + \varepsilon_N) \varepsilon_i + \mathcal{O}(\varepsilon_i^2) \quad (10.19)$$

Das heißt es gelte der in Gl. (10.11) vorgenommene Grenzübergang, der auch dem entspricht, dass wir sagen

$$\varepsilon_{Y-\text{intrinsisch}} \rightarrow 0$$

und es wird für die indirekte Messgröße keine Verteilung mehr angenommen. Anders ausgedrückt heißt es, dass bei einem gemessenen Satz an Eingangsgrößen, sich die Ausgangsgröße nach der Modellgleichung Gl.(10.16) berechnet, so wird die gesuchte Messgröße Y festgelegt. Man kann sich das auch so vorstellen, dass die angenommene Modellverteilung immer schmäler wird und somit nur noch ein Y_j Wert bei jeder Beobachtung $\mathbf{X}_j = (X_{1,j}, \dots, X_{N,j})$ erlaubt ist. Diesen Übergang stellt auch Abb. 10.4 dar.

Wir verwenden gemäß Gl. (10.11) den **Model-Prior** als Dirac'sche delta-Funktion:

$$\delta(Y - f(\mathbf{X})) \quad (10.20)$$

Dadurch, dass wir hier Y festlegen sind wir nicht mehr in der Bayes-Welt (Bayes-Statistik), sondern wieder in der klassischen Statistik. Durch Messungen oder andere Informationen seien **Wahrscheinlichkeiten**

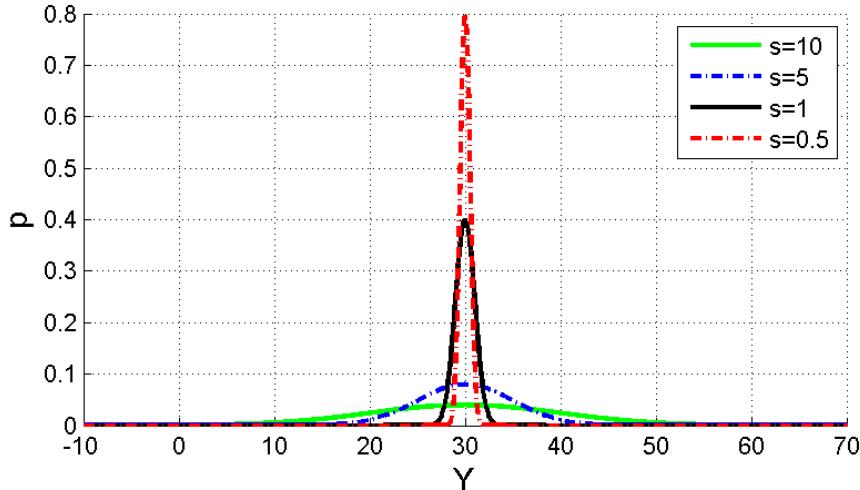


Abbildung 10.4: Übergang von Bayes-Statistik zur klassischen Statistik mit immer schmaler werdenden angenommenen Verteilungen für Y

lichkeitsdichteverteilungen der Messdaten \mathbf{X} gegeben, d.h.

$$p(\mathbf{X}) \quad (10.21)$$

Darüberhinaus können wir zusätzliche Informationen über die Größe Y in die statistische Analyse einarbeiten, beispielsweise Ergebnisse aus anderen Instituten oder aus vergangenen Messkampagnen. Die können wir ganz genau so behandeln wie den Prior in der bayes'schen Statistik. Da wir von den anderen Instituten oder Messkampagnen keine direkten Größen und somit auch keine Likelihood vorliegen haben, verwenden wird eine Wahrscheinlichkeitsdichteverteilung für Y gemäß den Annahmen über die Streuung von Y , die aus dem vollständigen Messergebnis des anderen Instituts oder der anderen Messkampagne hervorgeht. Dies kann eine Normalverteilung, eine Student-t-Verteilung oder dergleichen sein. Die Wahrscheinlichkeitsdichteverteilung (PDF) des Priors zu Y kennzeichnen wir hier mit einem Index „Null“, also $p_0(Y)$ (**Prior der Messgröße Y**)

$$p_0(Y) \quad (10.22)$$

Wir können nun eine gemeinsame PDF für $\mathbf{Z} = (\mathbf{X}, Y)$ aufstellen, die wie folgt gegeben ist [Els07]:

$$p(\mathbf{Z}) = p(\mathbf{X}, Y) \propto p(\mathbf{X}) \cdot p_0(Y) \cdot \delta(Y - f(\mathbf{X})) \quad (10.23)$$

Die gesuchte PDF für die Messgröße Y erhalten wir durch Marginalisierung der $p(\mathbf{X}, Y)$ über X :

$$p(Y) = \int_{-\infty}^{\infty} p(\mathbf{X}', Y) d\mathbf{X}' \propto p_0(Y) \int_{-\infty}^{\infty} p(\mathbf{X}') \cdot \delta(Y - f(\mathbf{X}')) d\mathbf{X}' \quad (10.24)$$

Den besten Schätzwert für Y erhalten wir wieder über den Erwartungswert $y = E(Y)$:

$$\begin{aligned}
 y &= \int Y' p(Y') dY' \\
 &= \int Y' \frac{p_0(Y) \int p(\mathbf{X}') \cdot \delta(Y' - f(\mathbf{X}')) d\mathbf{X}'}{\int p_0(Y'') \int p(\mathbf{X}') \cdot \delta(Y'' - f(\mathbf{X}')) d\mathbf{X}' dY''} dY' \\
 &= \frac{\int (\int Y' \cdot p_0(Y) \cdot p(\mathbf{X}') \cdot \delta(Y' - f(\mathbf{X}')) dY') d\mathbf{X}'}{\int (\int p_0(Y'') p(\mathbf{X}') \cdot \delta(Y'' - f(\mathbf{X}')) dY'') d\mathbf{X}'} \\
 &= \frac{\int f(\mathbf{X}') \cdot p_0(f(\mathbf{X}')) \cdot p(\mathbf{X}') d\mathbf{X}'}{\int p_0(f(\mathbf{X}')) \cdot p(\mathbf{X}') d\mathbf{X}'}
 \end{aligned} \tag{10.25}$$

Die Unsicherheit ergibt sich aus der Varianz von Y

$$\begin{aligned}
 u^2(y) &= \int (y - Y')^2 p(Y') dY' \\
 &= \frac{\int (y - f(\mathbf{X}'))^2 \cdot p_0(f(\mathbf{X}')) \cdot p(\mathbf{X}') d\mathbf{X}'}{p_0(f(\mathbf{X}')) \cdot p(\mathbf{X}') d\mathbf{X}'}
 \end{aligned} \tag{10.26}$$

Die Nenner in den beiden Gleichungen (10.25) und (10.26) sind für die Normierung da.

Für den Fall, dass wir keine Vorkenntnisse über die Messgröße Y haben und als Prior-Verteilung eine 1 verwenden (siehe auch die Hinweise der 8. Vorlesung zu non-informativen Priors), d.h. $p_0(Y) \propto 1$ ergibt sich für die Verteilungsdichte von Y aus der Gl.(10.24) die sogenannte **Markov-Formel** [Cox06]:

$$p(Y) = \int_{-\infty}^{\infty} p(\mathbf{X}') \cdot \delta(Y - f(\mathbf{X}')) d\mathbf{X}' \tag{10.27}$$

In der Regel benötigen die Formeln (10.27) und (10.24) numerische Lösungsverfahren. In einer späteren Vorlesung Monte-Carlo-Methoden als Lösungsverfahren vorgestellt werden.

Hier zeigen wir für den Fall, dass wir eine lineares Modell und Normalverteilungen haben, eine analytische Lösung. Dieser Ansatz geht natürlich auch, wenn die Modellgleichung (10.16) linearisiert werden kann.

10.1.3 Analyt. Lösung für lineares Modell und Normalverteilungsdichten

Die Herleitung findet sich im Paper [Els07]. Gegeben ist die lineare Modellgleichung:

$$Y = \sum_{i=1}^N c_i \cdot X_i = \mathbf{c}^T \mathbf{X} \quad (10.28)$$

Darüber hinaus seien die Kovarianzen der direkten Messgrößen \mathbf{X} gegeben, d.h. die Kovarianzmatrix $\mathbf{V}(\mathbf{X})$ bzw. äquivalent \mathbf{X} als Index geschrieben \mathbf{V}_X :

$$\begin{aligned} \mathbf{V}(\mathbf{X}) &= (\text{Cov}(X_i, X_j))_{i,j=1,\dots,N} \\ &= \begin{pmatrix} \text{E}((X_1 - \mu_1)(X_1 - \mu_1)) & \text{E}((X_1 - \mu_1)(X_2 - \mu_2)) & \cdots & \text{E}((X_1 - \mu_1)(X_N - \mu_N)) \\ \text{E}((X_2 - \mu_2)(X_1 - \mu_1)) & \text{E}((X_2 - \mu_2)(X_2 - \mu_2)) & \cdots & \text{E}((X_2 - \mu_2)(X_N - \mu_N)) \\ \vdots & \vdots & \ddots & \vdots \\ \text{E}((X_N - \mu_N)(X_1 - \mu_1)) & \text{E}((X_N - \mu_N)(X_2 - \mu_2)) & \cdots & \text{E}((X_N - \mu_N)(X_N - \mu_N)) \end{pmatrix} \\ &= \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_N) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \cdots & \text{Cov}(X_2, X_N) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_N, X_1) & \text{Cov}(X_N, X_2) & \cdots & \text{Var}(X_N) \end{pmatrix} \end{aligned} \quad (10.29)$$

Die Unsicherheiten $u(\mathbf{X})$ der direkten Messgrößen \mathbf{X} stehen auf der Hauptdiagonalen der Kovarianzmatrix \mathbf{V} , d. h. $u^2(X_i) = \text{Var}(X_i)$. (Hinweis: Zu den Begriffen *Varianz* und *Kovarianz*, siehe auch 5. Vorlesung)

Für die pdfs des linearen Modells nehmen wir eine Normalverteilung an, d.h.

$$p(\mathbf{X}|\mathbf{x}) = \frac{\exp\left[-\frac{1}{2}(\mathbf{X} - \mathbf{x})^T \mathbf{V}_X^{-1} (\mathbf{X} - \mathbf{x})\right]}{\sqrt{(2\pi)^N \det(\mathbf{V}_X)}} \quad (10.30)$$

wobei \mathbf{x} der Vektor mit beobachteten oder geschätzten Werten zu den Größen \mathbf{X} ist.

Die Priorinformationen über die zu bestimmende Messgröße Y seien gegeben durch die Standardabweichung σ_0 und den Schätzwert y_0 zur Größe Y :

$$p_0(Y|y_0, \sigma_0) = \frac{\exp\left[-\frac{1}{2}(Y - y_0)^2 / \sigma_0^2\right]}{\sqrt{(2\pi)} \sigma_0} \quad (10.31)$$

Wenn die Priorinformationen über Y nicht berücksichtigt werden, so erhalten wir aus den Schätzwerten \mathbf{x} , die wir aus den Beobachtungen der direkten Größen gewonnen haben, für die indirekte Größe Y

aufgrund des linearen Zusammenhangs, Gl. (10.28), folgendes Ergebnis für den Schätzwert y und die Standardabweichung $u(y)$ bzw. Varianz $u^2(y)$:

$$y = \mathbf{c}^T \mathbf{x}, \quad u^2(y) = \mathbf{c}^T \mathbf{V}_X \mathbf{c}. \quad (10.32)$$

Wenn die Information aus den Messungen, also aus \mathbf{x} , nicht berücksichtigt wird, sondern nur die Priorinformation σ_0 und y_0 vorliegt, dann gilt, dass das Ergebnis für Y gleich dem seiner Priorinformationen ist:

$$y = y_0, \quad u^2(y) = \sigma_0^2. \quad (10.33)$$

Wenn wir beide Informationen, die à-priori mit y_0 und σ_0 und die aktuelle Messung mit dem Schätzervektor \mathbf{x} , verknüpfen wollen, bilden wir wie gehabt das Produkt der entsprechenden Wahrscheinlichkeitsdichteveerteilungen (PDFs)

$$p(Y|\mathbf{x}, y_0, \sigma_0) \propto p(\mathbf{X}|\mathbf{x}) p_0(Y|y_0, \sigma_0) \quad (10.34)$$

d.h.

$$p(Y|\mathbf{x}, y_0, \sigma_0) \propto \frac{\exp\left[-\frac{1}{2}(\mathbf{X} - \mathbf{x})^T \mathbf{V}_X^{-1} (\mathbf{X} - \mathbf{x})\right]}{\sqrt{(2\pi)^N \det(\mathbf{V}_X)}} \frac{\exp\left[-\frac{1}{2}(Y - y_0)^2 / \sigma_0^2\right]}{\sqrt{(2\pi) \sigma_0}} \quad (10.35)$$

Hier haben wir zunächst „proportional“ geschrieben, denn das Produkt der beiden Verteilungen Gl. (10.30) und Gl. (10.31) muss noch derart normiert werden, dass gilt

$$\int_{-\infty}^{\infty} p(Y|\mathbf{x}, y_0, \sigma_0) dY = 1.$$

Die pdf $p(Y|\mathbf{x}, y_0, \sigma_0)$ für Y ist also gegeben durch das normierte Produkt der beiden Verteilungen Gl.(10.30) und Gl.(10.31) bzw. deren Marginalverteilung Gl.(10.24). Diese aus dem Produkt resultierende Verteilung $p(Y|\mathbf{x}, y_0, \sigma_0)$ ist wegen des Potenzgesetzes mit $e^a e^b = e^{a+b}$ ebenso normalverteilt. Also schreiben wir folgenden Ansatz auf:

$$p(Y|y) = \frac{\exp\left[-\frac{1}{2}(Y - y)^2 / u^2(y)\right]}{\sqrt{(2\pi)u^2(y)}} \quad (10.36)$$

Dabei müssen also der Schätzer y und dessen Varianz $u^2(y)$ eine Verknüpfung aus den Schätzern y_0 und \mathbf{x} und deren (Ko-)Varianzen σ_0^2 und \mathbf{V}_X darstellen.

Der Erwartungswert y ist gegeben durch:

$$y = u^2(y) \left[\frac{\mathbf{c}^T \cdot \mathbf{x}}{\mathbf{c}^T \cdot \mathbf{V}_X \cdot \mathbf{c}} + \frac{y_0}{\sigma_0^2} \right] \quad (10.37)$$

Die Varianz bzw. Unsicherheit $u^2(y)$ ist gegeben durch:

$$u^2(y) = \left[\frac{1}{\mathbf{c}^T \cdot \mathbf{V}_X \cdot \mathbf{c}} + \frac{1}{\sigma_0^2} \right]^{-1} \quad (10.38)$$

Der Erwartungswert in Gl.(10.37) ist der gewichtete Mittelwert y aus dem Schätzwert aus den der Messungen $\mathbf{c}^T \mathbf{x}$ gemäß Gl. (10.32) und der Priorinformation y_0 . Dabei sind die Gewichtsfaktoren, wie wir es für den Fall einer einzelnen Messgröße kennengelernt haben, die reziproken Varianzen bzw.

reziproken Kovarianzmatrizen. Als Beispiel siehe Übungsaufgabe 10-1.

Für den Fall, dass keine Priorinformationen vorhanden sind, d.h. $\sigma_0 \rightarrow \infty$ ist der Erwartungswert und die Unsicherheit durch die Messung gegeben. In der Gl. (10.37) und (10.38) verschwindet dann der 2. Term.

10.2 MU-Berechnung mit rechteck- und U-förmigen Messgrößen

10.2.1 Messunsicherheitsfortpflanzung linearisierbarer Modelle

Kurze Wiederholung: Das klassische Gesetz zur Fortpflanzung der Messunsicherheit mit der Gleichung

$$u^2(Y) = \sum_{i=1}^n c_i^2 u(X_i)^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n c_i c_j \rho_{i,j} u(X_i) u(X_j) \quad (10.39)$$

gilt für

1. eine indirekte Messgröße Y (univariat), die explizit abhängig ist von den direkten Messgrößen X_i

$$Y = f(X_1, \dots, X_n) \quad (10.40)$$

2. für lineare Abhängigkeiten der indirekten von den direkten Messgrößen, oder wenigstens für linearisierbare Abhängigkeiten, also

$$Y \approx f(X_1, \dots, X_n)|_{\bar{x}_1, \dots, \bar{x}_n} + \underbrace{\sum_{i=1}^n \frac{\partial f}{\partial X_i} \Big|_{\bar{x}_1, \dots, \bar{x}_n}}_{c_i} \Delta X_i \quad (10.41)$$

3. für glockenförmige Streuung der direkten Messgrößen X_i :

- Die Wahrscheinlichkeitsdichtevertteilung ist eine glockenförmige Kurve mit genau einem Maximum, d.h. sie ist unimodal. Die Position des Maximums bildet im wesentlichen den Erwartungswert.
- Die Wurzel der Varianz entspricht der Breite σ des Kerns der Kurve. Im Kern liegt der Großteil der Fläche unter der Kurve.

10.2.2 Wahrscheinlichkeitsdichtevertteilungen für Messgrößen

Es gibt zwei Grenzfälle, die in der Metrologie vorkommen können und notbehelfsmäßig in das Fortpflanzungsgesetz eingebastelt werden: Die U-Verteilung und die Rechteckverteilung.

Diese erfüllen nicht die oben genannte dritte Voraussetzung, dass diese glockenförmig sind, d.h. ein Maximum relativ mittig besitzen und nach außen mit abnehmender Wahrscheinlichkeit verlaufend sind.

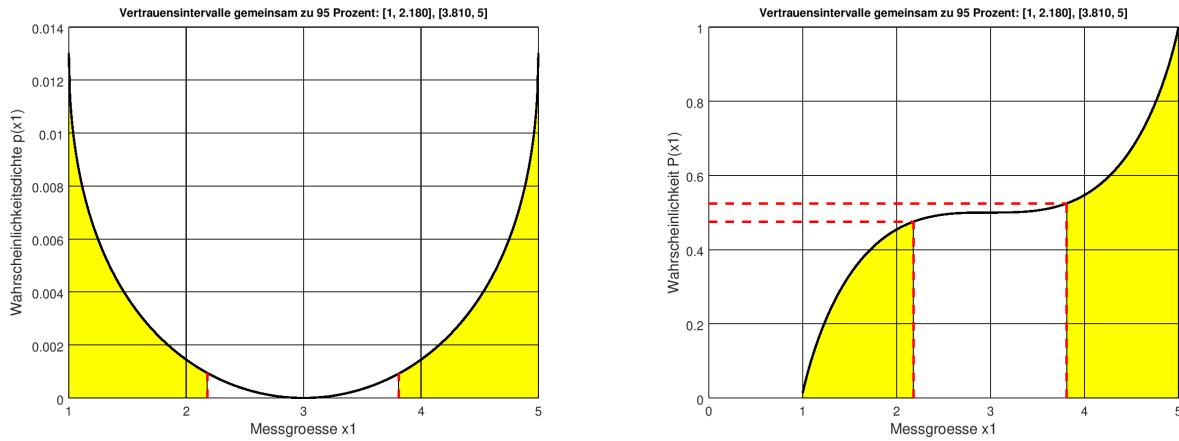


Abbildung 10.5: Für das Vertrauensniveau möchte man das häufigste Vorkommen angeben, bei bimodaler Verteilung als von jedem der beiden Peaks ausgehen: Bei 95% Vertrauensniveau also jeweils 47,5% für ein Intervall ganz rechts und 47,5% für ein Intervall ganz links, hier als gelbe Flächen markiert.

Die Bestimmung von Vertrauensintervallen zu bestimmten Vertrauensniveaus anhand der Quantile von der Normalverteilung und der t-Verteilung setzt voraus, dass die folgenden beiden Kriterien für die Verteilungen gegeben sind: univariat und Streubreite sind gemäß zweitem statistischen Moment gegeben. Auch die Bestimmung von Vertrauensintervallen zu bestimmten Vertrauensniveaus, wenn man zu einer Größe genau ein Intervall wählt mit zwei Grenzen, setzt voraus, dass die Verteilung unimodal ist, einen Häufungspunkt, den Erwartungswert, hat.

Je nachdem wie stark z.B. der Beitrag einer direkten Messgröße ist, deren Wahrscheinlichkeitsdichte einer U-Verteilung folgt, kann die Wahrscheinlichkeitsdichte der indirekten Größe eine bimodale Verteilung werden. U-Verteilungen liegen vor, wenn die Streuung einer Größe durch Schwingungseinflüsse hervorgerufen wird.

Die U-Verteilung hat 2 Peaks an ihren beiden äußersten Rändern. Wir betrachten folgende U-Verteilung

$$p(X) = \begin{cases} \arcsin^2\left(\frac{X-\bar{x}}{s}\right) & \text{für } \bar{x}-s \leq X \leq \bar{x}+s \\ 0 & \text{für } (X < \bar{x}-s) \vee (\bar{x}+s < X) \end{cases} \quad (10.42)$$

für eine Schwingung mit der Amplitude s als Störung, die die Streuung der Größe X um den Wert \bar{x} verursacht. In der Mitte, nämlich bei \bar{x} ist die Wahrscheinlichkeit am geringsten. Das Signifikanzniveau α liegt folglich in der Mitte und nicht am Rand. Vom 95% Vertrauensniveau wird jeweils die Hälfte auf jeden der beiden Peaks am Rand aufgeteilt, also 47.5% der Fläche der Wahrscheinlichkeitsdichtekurve auf den einen Peak und die anderen 47.5% der Fläche auf den anderen Peak. Dann müssen wir zwei Intervalle angeben und sagen: Mit 95% Wahrscheinlichkeit sind Werte entweder in dem einen oder dem anderen Intervall zu erwarten, siehe Abb. 10.5.

Die Rechteckverteilung liefert für jeden Wert der Größe X dieselbe Wahrscheinlichkeit für das liegen innerhalb einer Spanne und die Wahrscheinlichkeit Null außerhalb. Wenn s die halbe Spanne ist, dann

ist die rechteckverteilte Wahrscheinlichkeitsdichteverteilung

$$p(X) = \begin{cases} \frac{1}{2s} & \text{für } \bar{x} - s \leq X \leq \bar{x} + s \\ 0 & \text{für } (X < \bar{x} - s) \vee (\bar{x} + s < X) \end{cases}. \quad (10.43)$$

Hier kann man nicht entscheiden, wo der Bereich des Signifikanzniveaus und wo der des Vertrauensniveaus hinzulegen ist.

10.2.3 Wahrscheinlichkeitsdichten im Fortpflanzungsgesetz

Wir betrachten ein einfaches Beispiel eines linearen Modells, bei dem eine indirekte Messgröße Y von zwei direkten Messgrößen X_1 und X_2 linear abhängt, also

$$Y = aX_1 + bX_2 \quad (10.44)$$

wobei a und b Konstanten sind. Diese Konstanten sind für das Fortpflanzungsgesetz dann auch direkt die Sensitivitäten: $c_1 = a$ und $c_2 = b$.

Bei diesem Beispiel soll X_1 eine Größe darstellen, die nicht der Voraussetzung genügt, gauß- oder t-verteilt zu sein. Wir untersuchen zwei Fälle, einerseits den Fall, dass sie U-verteilt mit der Amplitude s_1 sei, also $p(X_1)$ gemäß Gl. (10.42), andererseits den Fall, dass sie rechteckverteilt sei, $p(X_1)$ gemäß Gl. (10.43), mit der halben Spanne s_1 .

Die Größe X_2 sei gaußverteilt.

Konkret verwenden wir hier beispielhaft folgende Zahlenwerte:

$c_1 = a$	$c_2 = b$	\bar{x}_1	s_1	\bar{x}_2	s_2
1	1	3	2	3.5	0.5

Die Wahrscheinlichkeitsdichtevertteilung als Funktion der beiden direkten Größen X_1 und X_2 ist in Abb. 10.6 dargestellt, für die U-Verteilung von X_1 links und für die Rechteckverteilung von X_1 rechts:

$$p(X_1, X_2) = p(X_1) \underbrace{\frac{1}{\sqrt{2\pi}s_2} e^{-\frac{1}{2}\left(\frac{X_2 - \bar{x}_2}{s_2}\right)^2}}_{p(X_2)}. \quad (10.45)$$

Die Wahrscheinlichkeitsdichtevertteilung als Funktion der indirekten Größe Y ist die Marginalverteilung $p(Y)$ der Verteilung $p(Y, X_1, X_2)$, die sich aus folgendem Produkt ergibt:

$$p(Y, X_1, X_2) = C_{\text{norm}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{Y - f(X_1, X_2)}{\sigma}\right)^2} p(X_1, X_2) \quad (10.46)$$

mit $f(X_1, X_2) = aX_1 + bX_2$ und σ für die Unsicherheit des Modells, sowie des Systems, das die Messvorgänge der direkten Größen verknüpft. Letztendlich steckt dahinter die Vorstellung, dass die indirekte Größe Y ihrerseits auch eine Zufallsgröße ist, nicht nur die direkten. Die Konstante C_{norm}

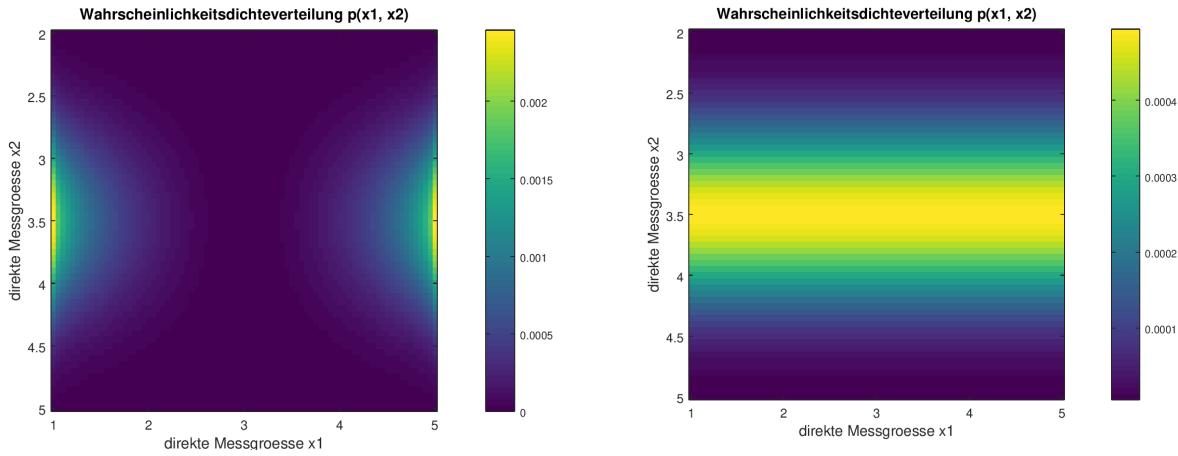


Abbildung 10.6: Die Wahrscheinlichkeitsdichteverteilung als Funktion der beiden direkten Größen X_1 und X_2 : Links: U-verteilte Größe X_1 , rechts: rechteckverteilte Größe X_1 .

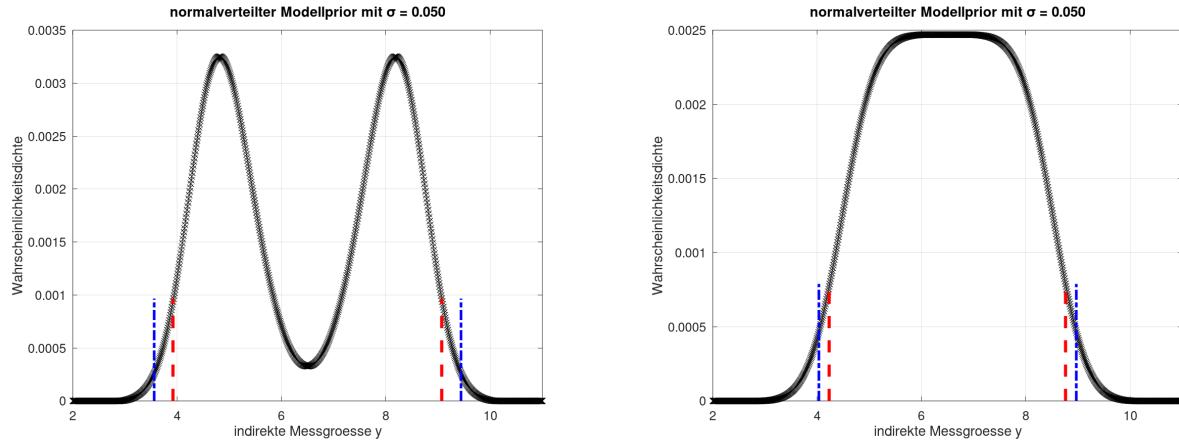


Abbildung 10.7: Die Wahrscheinlichkeitsdichteverteilung $p(Y)$ gemäß Gl. (10.49) als Funktion der indirekten Größe Y mit Unsicherheit $\sigma = 0.05$ für die Größe Y : Links: U-verteilte Größe X_1 , rechts: rechteckverteilte Größe X_1 .

ist so definiert, dass $\iiint p(Y, X_1, X_2) dX_1 dX_2 dY = 1$ (Normierungsbedingung) erfüllt wird, also

$$C_{\text{norm}} = \frac{1}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(Y, X_1, X_2) dX_1 dX_2 dY}. \quad (10.47)$$

Für die Vorstellung von einer indirekten Messgröße, die einen exakten wahren Wert haben muss und die für sich genommen keine Zufallsgröße ist, deren Streuung einzig durch die Streuung der direkten Größen verursacht wird, wird in der Messtechnik (nirgends sonst in der Statistik) ein Modellprior verwendet, der durch einen unendlich scharfen Peak repräsentiert wird. Wir bezeichnen den (siehe oben) als *scharfen Modellprior*.

Für den scharfen Modellprior für das Verfahren in GUM-Supplement 1 machen wir den Grenzübergang

$$\lim_{\sigma \rightarrow 0} \left\{ \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2} \left(\frac{Y-f(X_1, X_2)}{\sigma} \right)^2} \right\} = \delta(Y - f(X_1, X_2)) \quad (10.48)$$

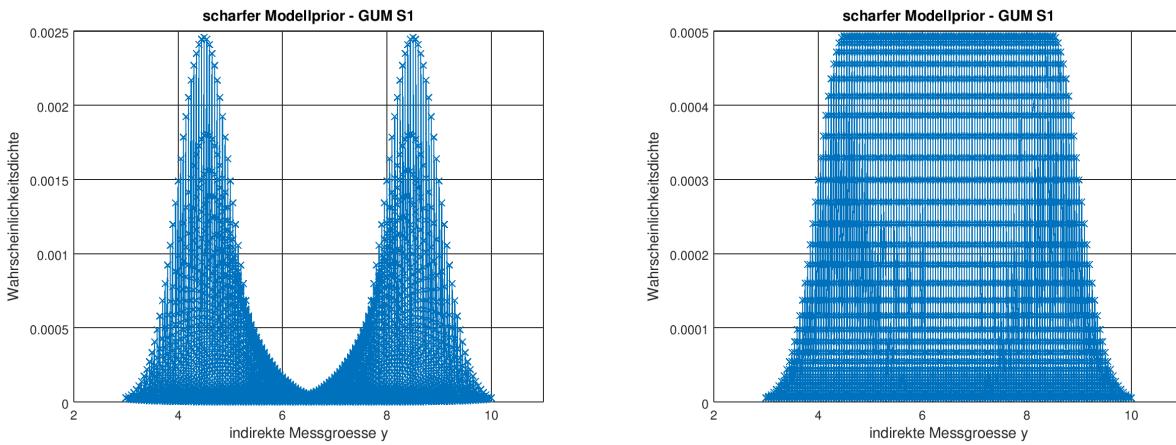


Abbildung 10.8: Die Wahrscheinlichkeitsdichteverteilung $p(Y)$ gemäß Gl. (10.49) als Funktion der indirekten Größe Y mit scharfem Modellprior: Links: U-verteilte Größe X_1 , rechts: rechteckverteilte Größe X_1 .

wobei δ die dirac'sche Deltadistribution ist.

Die Marginalverteilung $p(Y)$, die die Wahrscheinlichkeitsdichte der indirekten Größe darstellt, ist

$$p(Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(Y, X_1, X_2) dX_1 dX_2. \quad (10.49)$$

Sie ist in Abb. 10.7 für einen gaußglockenförmigen Modellprior mit $\sigma = 0.05$ und in Abb. 10.8 für einen scharfen Modellprior dargestellt.

In den beiden Abbn. 10.7 und 10.8 auf der linken Seite gezeigten Diagrammen ist die Marginalverteilung für Y für den Fall dargestellt, dass die direkte Größe X_1 U-verteilt ist. Sie weist zwei Peaks auf, d.h. sie ist bimodal. Hier müsste man sich also genau anschauen, wo das Signifikanzniveau anzusiedeln ist, also, ob ein Teil davon in den mittleren Bereich gehört. Der Fall mit rechteckverteilter Größe X_1 ist in dieser Hinsicht gutmütig, so dass wir hier wie gewohnt von der Umkehrfunktion P^{-1} der kumulativen Verteilung $P(Y)$ die entsprechenden Werte $P^{-1}(\alpha/2)$ und $P^{-1}(1 - \alpha/2)$ für das Vertrauens- oder Überdeckungsintervall nehmen können, siehe Abb. 10.9. Diese Vorgehensweise übernehmen wir hier einfach auch für das Beispiel mit der U-Verteilung. Die in den Diagrammen in Abb. 10.7 gestrichelt rot eingezeichneten Linien kennzeichnen die Vertrauensintervalle.

Zum Vergleich sind in diesen Abb. 10.7 die Überdeckungsintervalle, die man aus dem Fortpflanzungsgesetz des GUM von 1995/2008 erhalten würde mit blauer Strichpunktlinie eingezeichnet. Gemäß dem GUM soll für U-verteilte Größen der Ansatz für die Unsicherheit $u = \text{Spanne}/\sqrt{8}$ und für rechteckverteilte $u = \text{Spanne}/\sqrt{12}$ verwendet werden. Mit s_1 als halbe Spanne ist dies also

$$u^2(X_1) = \frac{(2s_1)^2}{8} = \frac{s_1^2}{2} \quad \text{bzw.} \quad u^2(X_1) = \frac{(2s_1)^2}{12} = \frac{s_1^2}{3} \quad (10.50)$$

so dass

$$u^2(Y) = \begin{cases} a^2 \frac{s_1^2}{2} + b^2 s_2^2 & \text{für U-verteilte } X_1 \\ a^2 \frac{s_1^2}{3} + b^2 s_2^2 & \text{für rechteckverteilte } X_1 \end{cases}. \quad (10.51)$$

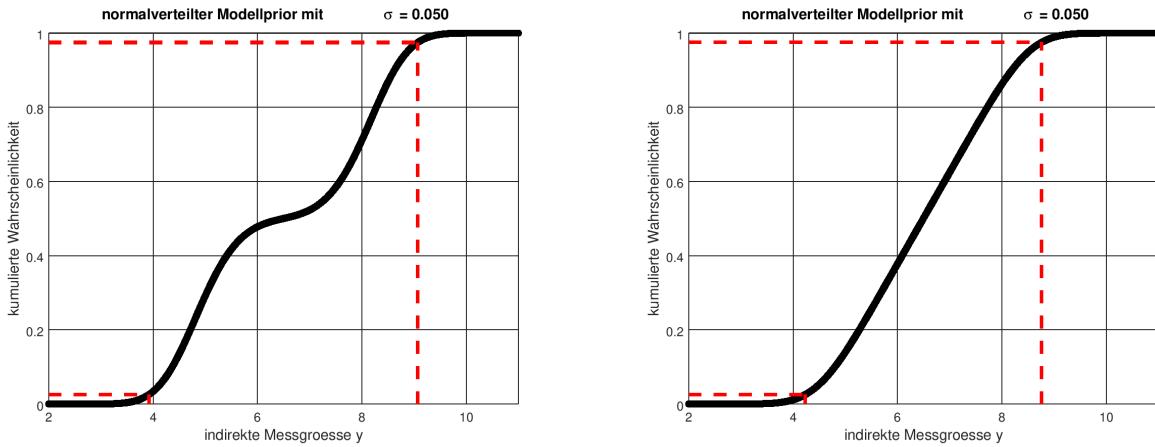


Abbildung 10.9: Die zur kumulierten Verteilung $P(Y)$ der Wahrscheinlichkeitsdichteverteilung $p(Y)$ erworbene Umkehrfunktion liefert mit $P^{-1}(\alpha/2)$ und $P^{-1}(1 - \alpha/2)$ zum Signifikanzniveau α das entsprechende Vertrauensintervall: Links: U-verteilte Größe X_1 , rechts: rechteckverteilte Größe X_1 .

Das Vertrauensintervall, durch die blaue Strichpunktlinie gekennzeichnet, ergeben sich mit den Quantilen der Gaußverteilung $z_{\alpha/2} = -1.96$ und $z_{1-\alpha/2} = 1.96$ zu:

$$[\bar{y} - 1.96 u(Y), \bar{y} + 1.96 u(Y)] = \begin{cases} [3.56, 9.44] & \text{für U-verteilte } X_1 \\ [4.03, 8.97] & \text{für rechteckverteilte } X_1 \end{cases} \quad (10.52)$$

mit $\bar{y} = a\bar{x}_1 + b\bar{x}_2$.

Die Werte aus der Marginalverteilung $p(Y)$ im Vergleich dazu sind

$$[P^{-1}(\alpha/2), P^{-1}(1 - \alpha/2)] = \begin{cases} [3.92, 9.07] & \text{für U-verteilte } X_1 \\ [4.23, 8.76] & \text{für rechteckverteilte } X_1 \end{cases} \quad (10.53)$$

für die Breite der Verteilung des Modellprior $\sigma = 0.05$. Sie bilden das Überdeckungsintervall, das mit roten gestrichelten Linien eingezeichnet ist. Es ist kleiner als das gemäß GUM, Gl (10.51), ermittelte und zeigt sehr schön, dass U-verteilte und rechteckverteilte Messgrößen bei der Behandlung mit dem klassischen Fehlerfortpflanzungsgesetz zu falschen Vertrauensintervallen führen kann.

Der dazugehörige Octavecode lautet:

```
function non_belldistri_linmodel (sigma=0.05)
% sigma = 0.05, d.h. der Modellparameter y wird hier mit gaussförmigem
% Modellprior, aber sehr schmal, fast dirac'sche delta-Distribution angenommen.

close all

% Falls das Statistikpaket noch nicht geladen ist, kann es wie folgt geladen werden:
% Entsprechend die folgende Zeile ein- oder auskommentieren
pkg load statistics

boxflag = 0; % 0: U-Verteilung; 1: Rechteck-Verteilung;
```

```

pltflag1 = 1;
pltflag1a = 0;
pltflag2 = 1;
pltflag3 = 1;

a = 1.0;
b = 1.0;
x1_quer = 3.0;
% Spanne
s1 = 2.0;
%
x2_quer = 3.5;
% Standardabweichung
s2 = 0.5;
%
% Modellgleichung y = a x1 + b x2
%
% Zeilenvektor
x1 = [-s1:0.05:s1] + x1_quer;
if boxflag
p1 = ones(size(x1));
else
p1 = asin((x1 - x1_quer)/s1).^2;
end
p1 = p1/sum(p1);
figure(1);
plot(x1, p1, 'k-');
grid on;
xlabel('direkte Messgroesse x1', 'fontsize', 14);
ylabel('Wahrscheinlichkeitsdichte p(x1)', 'fontsize', 14);
set(gca, 'fontsize', 12);
if pltflag1
print(1,'direkte_x1_box.png', '-dpng');
end
%
% Spaltenvektor
x2 = [-3*s2:0.05:3*s2]' + x2_quer;
p2 = normpdf(x2, x2_quer, s2);
p2chk = exp(-0.5 * ((x2 - x2_quer)/s2).^2) / (sqrt(2*pi) * s2);
figure(2);
plot(x2, p2, 'k-', x2, p2chk, 'r--');
grid on;
xlabel('direkte Messgroesse x2', 'fontsize', 14);

```

```

ylabel('Wahrscheinlichkeitsdichte p(x2)', 'fontsize', 14);
set(gca, 'fontsize', 12);
if pltflag2
print(2,'direkte_x2.png', '-dpng');
end
%
% Spalte mal Zeile liefert Matrix
pgesamt = p2*p1;
pgesamt = pgesamt / sum(pgesamt(:));
%
figure(3);
imagesc(x1, x2, pgesamt);
colorbar();
xlabel('direkte Messgroesse x1', 'fontsize', 14);
ylabel('direkte Messgroesse x2', 'fontsize', 14);
title('WahrscheinlichkeitsdichteVerteilung p(x1, x2)', 'fontsize', 14);
set(gca, 'fontsize', 12);
if pltflag1
print(3,'direkte_x1_x2_box.png', '-dpng');
end
%
% ----
% x1 und x2 in Matrizen fuellen
n2_r = length(x2)
n1_c = length(x1)
X_1 = ones(n2_r,1) * x1;
X_2 = x2 * ones(1, n1_c);
Y_matrix = a*X_1 + b*X_2;
%
% nun als 1D arrays
y = Y_matrix(:);
p_y = pgesamt(:);
figure(4);
plot(y, p_y);
axis([2 11]);
title('scharfer Modellprior - GUM S1', 'fontsize', 14);
xlabel('indirekte Messgroesse y', 'fontsize', 14);
ylabel('Wahrscheinlichkeitsdichte', 'fontsize', 14);
set(gca, 'fontsize', 12);
if pltflag1a
print(4,'indirekte_y_GUMS1.png', '-dpng');
end
%
% sortieren, so dass y monoton steigend ist

```

```

[ys, isort] = sort(y);
p_sort = p_y(isort);
figure(5);
plot(ys, p_sort, 'x-');
axis([2 11]);
grid on;
title('scharfer Modellprior - GUM S1', 'fontsize', 14);
xlabel('indirekte Messgroesse y', 'fontsize', 14);
ylabel('Wahrscheinlichkeitsdichte', 'fontsize', 14);
set(gca, 'fontsize', 12);
if pltflag1
print(5,'indirekte_ysorted_GUMS1_box.png', '-dpng');
end

%
% mit gaussfoermigem Modellprior, aber sehr schmal, fast dirac'sche delta-Distribution
%
dy = 0.01;
y_model = [2:dy:11];
n_model = length(y_model);
% sigma = 0.5;
p_mod = zeros(1, n_model);
for k = 1:n_model
%
% Summationen entsprechen der Integration ueber x1 und x2
% zum Berechnen der Marginalverteilung
p_mod(k) = sum(sum(exp(-0.5 * ((y_model(k) - Y_matrix)/sigma).^2) .* pgesamt));
end
Pkum = cumsum(p_mod);
psum = Pkum(n_model);
Pkum = Pkum/psum;
p_mod = p_mod/psum;

%
% Signifikanzniveau
alpha = 0.05;
%
% Quantile
[left, ileft] = min(abs(Pkum - alpha/2))
[right, iright] = min(abs(Pkum - (1-alpha/2)))
%
erw_y = sum(y_model .* p_mod);
std_y = sqrt(sum((y_model - erw_y).^2 .* p_mod));
%
```

```
% Vergleich mit Fortpflanzungsgesetz, das eigentlich nur fuer
% glockenfoermige Verteilungen gilt, bei denen die Wurzel des zweiten statistischen
% Moments, also die Varianz die Breite der Verteilung definiert, bzgl der sie
% sich normieren laesst.
y_quer = a*x1_quer + b*x2_quer;
if boxflag
u_y = sqrt((a * s1)^2/3 + (b * s2)^2);
else
u_y = sqrt((a * s1)^2/2 + (b * s2)^2);
end
%
figure(6);
sigma_str = num2str(sigma, '%1.3f');
plot(y_model, p_mod, 'kx-', ...
[y_quer-1.96*u_y, y_quer-1.96*u_y], [0, p_mod(iright)], 'b-.', 'linewidth', 2.5, ...
[y_quer+1.96*u_y, y_quer+1.96*u_y], [0, p_mod(iright)], 'b-.', 'linewidth', 2.5, ...
[y_model(ileft), y_model(ileft)], [0, p_mod(iright)], 'r--', 'linewidth', 3, ...
[y_model(iright), y_model(iright)], [0, p_mod(iright)], 'r--', 'linewidth', 3);
% plot(y_model, p_mod, 'kx-', ...
% [y_quer-1.96*u_y, y_quer-1.96*u_y], [0, p_mod(iright)], 'b-.', 'linewidth', 2.5, ...
% [y_quer+1.96*u_y, y_quer+1.96*u_y], [0, p_mod(iright)], 'b-.', 'linewidth', 2.5, ...
% [erw_y-1.96*std_y, erw_y-1.96*std_y], [0, p_mod(iright)], 'g-', 'linewidth', 2.5, ...
% [erw_y+1.96*std_y, erw_y+1.96*std_y], [0, p_mod(iright)], 'g-', 'linewidth', 2.5, ...
% [y_model(ileft), y_model(ileft)], [0, p_mod(iright)], 'r--', 'linewidth', 3, ...
% [y_model(iright), y_model(iright)], [0, p_mod(iright)], 'r--', 'linewidth', 3);
grid on;
axis([2 11]);
title(['normalverteilter Modellprior mit \sigma = ', sigma_str], 'fontsize', 14);
xlabel('indirekte Messgroesse y', 'fontsize', 14);
ylabel('Wahrscheinlichkeitsdichte', 'fontsize', 14);
set(gca, 'fontsize', 12);
if pltflag3
print(6,['indirekte_y_modellGauss_sigma', strrep(sigma_str,'.', 'p'), '_box.png'], '-dpng');
end

%
figure(7);
plot(y_model, Pkum, 'ko-', ...
[y_model(ileft), y_model(ileft)], [0, Pkum(ileft)], 'r--', 'linewidth', 3, ...
[y_model(iright), y_model(iright)], [0, Pkum(iright)], 'r--', 'linewidth', 3, ...
[y_model(1), y_model(ileft)], [Pkum(ileft), Pkum(ileft)], 'r--', 'linewidth', 3, ...
[y_model(1), y_model(iright)], [Pkum(iright), Pkum(iright)], 'r--', 'linewidth', 3);
grid on;
axis([2 11]);
```

```
title(['normalverteilter Modellprior mit \sigma = ', sigma_str], 'fontsize', 14);
xlabel('indirekte Messgroesse y', 'fontsize', 14);
ylabel('kumulierte Wahrscheinlichkeit', 'fontsize', 14);
set(gca, 'fontsize', 12);
if pltflag3
print(7,['Pkum_indirekte_y_Gauss_sigma', strrep(sigma_str,'.','p'), '_box.png'], '-dpng');
end
%
printf('Modellprior mit sigma = %1.3f\n', sigma);
printf('Erwartungswerte:\n');
printf('aus Erw von x1 und x2:  %1.2f\n', y_quer);
printf('aus Verteilung p_model: %1.2f\n', erw_y);
printf('Standardabweichungen als Wurzel der Varianz:\n');
printf('aus Fortpflanzung klassisch: %1.2f\n', u_y);
printf('aus Verteilung p_model:      %1.2f\n', std_y);
printf('Ueberdeckungsintervall fuer 95 Prozent\n');
printf('aus Fortpflanzung klassisch mit Quantil aus Gaussvert 1.96:  [%1.2f, %1.2f]\n',
y_quer - 1.96*u_y, y_quer + 1.96*u_y);
printf('aus Verteilung p_model, ueber die Umkehrfkt der kumulativen: [%1.2f, %1.2f]\n',
y_model(ileft), y_model(iright));
end
```

10.3 Anhang

10.3.1 A1: Matlab-Skript zum Delta-Peaken mit Bayes

```

1 function bayes_indirect_quantity()
2 % Hueser, 2017-11-25
3 % Kalibrierfaktor
4 K_faktor_0 = 0.0925;
5 sigma_K = 0.0090;
6 % Unsicherheit Messgeraet
7 sigma_M = 0.3; % Volt
8 %
9 %
10 JM = 11;
11 sig = 12;
12 mue = 500;
13 data_M = mue + sig*randn( JM,1 );
14 %
15 %
16 xK = [-4*sigma_K:0.001:4*sigma_K] + K_faktor_0;
17 nK = length(xK)
18 xM = [-4*sig:0.1:4*sig] + mue;
19 nM = length(xM)
20 y = [20:0.2:80];
21 nY = length(y)
22 %
23 std_M = std(data_M)*1.7;
24 std_KM = K_faktor_0 * std_M;
25 %
26 % Messungen in Volt
27 XM = data_M * ones(1,nM) - ones(JM,1) * xM;
28 %
29 % Das Modell: y = f(xK, xM)
30 % fuer jedes xK und jedes xM kombiniert
31 % y_{KM,i,j} = xK_i * xM_j fuer alle i=1,...,nK und j=1,...,nM
32 y_KM_matrix = xK' * xM;
33 % in einen langen Spaltenvektor gebracht
34 y_KM = y_KM_matrix(:);
35 nKM = length(y_KM);
36 delta_y = y_KM * ones(1,nY) - ones(nKM, 1)*y;
37 %
38 % direkte Groessen
39 % Kalibrierfaktor
40 p_K = exp(-0.5 * ( (xK - K_faktor_0)/sigma_K ).^2 );

```

```

41 p_K = p_K / sum(p_K);
42 figure(110);
43 plot( xK, p_K, 'linewidth',2);
44 xlabel('Kalibrierfaktor / (Einheit/Volt)', 'fontsize', 14);
45 ylabel('pdf', 'fontsize', 14);
46 set(gca, 'fontsize', 14);
47 set(gcf, 'PaperUnits', 'centimeters');
48 x_width=15 ;y_width= 8;
49 set(gcf, 'PaperPosition', [0 0 x_width y_width]);
50 print(110, 'Kalibrierfaktor.png', '-dpng');
51 %
52 % Messungen mit der kleinen Streuung
53 p_M_matrix = exp(-0.5 * ( XM/sigma_M ).^2 );
54 % Summe ueber alle Messungen
55 p_M_1 = sum( p_M_matrix );
56 p_M_1 = p_M_1 / sum(p_M_1);
57 % mit der grossen Streuung
58 p_M_matrix = exp(-0.5 * ( XM/std_M ).^2 );
59 % Summe ueber alle Messungen
60 p_M_2 = sum( p_M_matrix );
61 p_M_2 = p_M_2 / sum(p_M_2);
62 figure(111);
63 plot(xM, p_M_1,'k-', ...
64      xM, p_M_2,'r--','linewidth',2);
65 xlabel('Rohdaten / Volt', 'fontsize', 14);
66 ylabel('pdf', 'fontsize', 14);
67 legend('Geraet genau','Geraet ungenau','fontsize',14)
68 set(gca, 'fontsize', 14);
69 set(gcf, 'PaperUnits', 'centimeters');
70 x_width=15 ;y_width= 8;
71 set(gcf, 'PaperPosition', [0 0 x_width y_width]);
72 print(111, 'Rohdaten.png', '-dpng');
73 %
74 % PWahrscheinlichkeit als Funktion der beiden
75 % direkten Groessen xK und xM
76 p_KM_matrix = p_K' * p_M_1;
77 % in einen langen Spaltenvektor gebracht
78 p_KM_1 = p_KM_matrix(:);
79 %
80 p_KM_matrix = p_K' * p_M_2;
81 % in einen langen Spaltenvektor gebracht
82 p_KM_2 = p_KM_matrix(:);
83 %
84 % WahrscheinlichkeitsdichteVerteilung des Modells

```

```

85 % zur Vereinfachung nehme ich jetzt keine
86 % Verteilung um die std_m sondern nur den einen Schätzwert
87 p_y = exp(-0.5 * ( delta_y/std_KM ).^2 );
88 %
89 % posterior
90 post_matrix = p_y .* (p_KM_1 * ones(1,nY));
91 sumpost = sum(post_matrix(:));
92 post_matrix = post_matrix/sumpost;
93 %
94 % Marginalverteilung durch Summation über xM, xK
95 posterior1 = sum(post_matrix);
96 %
97 %
98 % posterior
99 post_matrix = p_y .* (p_KM_2 * ones(1,nY));
100 sumpost = sum(post_matrix(:));
101 post_matrix = post_matrix/sumpost;
102 %
103 % Marginalverteilung durch Summation über xM, xK
104 posterior2 = sum(post_matrix);
105 %
106 %
107 p_y2 = exp(-0.5 * ( delta_y/(std_KM*0.003) ).^2 );
108 % posterior
109 post_matrix = p_y2 .* (p_KM_2 * ones(1,nY));
110 sumpost = sum(post_matrix(:));
111 post_matrix = post_matrix/sumpost;
112 %
113 % Marginalverteilung durch Summation über xM, xK
114 posterior3 = sum(post_matrix);
115 %
116 [y_KMsort, km_sort] = sort(y_KM);
117 p_2 = p_KM_2(km_sort);
118 diff_y = diff(y_KMsort);
119 mid_p_KM_2 = (p_2(2:nKM)+p_2(1:nKM-1))*0.5;
120 iuse = find(diff_y>0);
121 meandiff = mean(diff_y(iuse))
122 diff_y = diff_y/meandiff;
123 pdf = mid_p_KM_2(iuse)./diff_y(iuse);
124 %
125 figure(120);
126 plot(diff_y);
127 figure(121);
128 plot(y_KMsort(iuse),pdf);

```

```

129 sumpdf = 5.2e6; max(pdf)
130 figure(100);
131 plot( y_KMsort(iuse), max(posterior1)*pdf/sumpdf, 'g.', ...
132     y, posterior1, 'k-', ...
133     y, posterior2, 'r--', ...
134     y, posterior3, 'b-', 'linewidth', 2);
135 xlabel('indirekte Groesse / Einheit', 'fontsize', 14);
136 ylabel('pdf', 'fontsize', 14);
137 legend('Dirac: Weise/Wuebbler', 'Geraet genau, Messobj schlecht', ...
138     'beides ungenau', 'Geraet ungenau, Modell duenner Peak')
139 % axis([10 100]);
140 % set(gca, 'fontsize', 14);
141 set(gcf, 'PaperUnits', 'centimeters');
142 x_width=15 ;y_width= 8;
143 set(gcf, 'PaperPosition', [0 0 x_width y_width]);
144 print(100, 'Posterior_all.png', '-dpng');
145 end

```

10.3.2 A2: Octave-Skript zur Fehlerfortpflanzung mit U- und Rechteck-Verteilungen

```

1 function non_belldistri_linmodel (sigma=0.05)
2 % Hueser, 2019-12-19
3
4 % sigma = 0.05, d.h. der Modellparameter y wird hier mit gaussfoermigem
5 % Modellprior, aber sehr schmal, fast dirac'sche delta-Distribution angenommen.
6
7 close all
8
9 % Falls das Statistikpaket noch nicht geladen ist, kann es wie folgt geladen werden:
10 % Entsprechend die folgende Zeile ein- oder auskommentieren
11 pkg load statistics
12
13 boxflag = 0; % 0: U-Verteilung; 1: Rechteck-Verteilung;
14
15 pltflag1 = 1;
16 pltflag1a = 0;
17 pltflag2 = 1;
18 pltflag3 = 1;
19
20 a = 1.0;
21 b = 1.0;
22 x1_quer = 3.0;
23 % Spanne

```

```

24 s1 = 2.0;
25 %
26 x2_quer = 3.5;
27 % Standardabweichung
28 s2 = 0.5;
29 %
30 % Modellgleichung  $y = a \cdot x1 + b \cdot x2$ 
31 %
32 % Zeilenvektor
33 x1 = [-s1:0.05:s1] + x1_quer;
34 if boxflag
35 p1 = ones(size(x1));
36 else
37 p1 = asin((x1 - x1_quer)/s1).^2;
38 end
39 p1 = p1/sum(p1);
40 figure(1);
41 plot(x1, p1, 'k-');
42 grid on;
43 xlabel('direkte Messgroesse x1', 'fontsize', 14);
44 ylabel('Wahrscheinlichkeitsdichte p(x1)', 'fontsize', 14);
45 set(gca, 'fontsize', 12);
46 if pltflag1
47 print(1,'direkte_x1_box.png', '-dpng');
48 end
49 %
50 % Spaltenvektor
51 x2 = [-3*s2:0.05:3*s2]' + x2_quer;
52 p2 = normpdf(x2, x2_quer, s2);
53 p2chk = exp(-0.5 * ((x2 - x2_quer)/s2).^2) / (sqrt(2*pi) * s2);
54 figure(2);
55 plot(x2, p2, 'k-', x2, p2chk, 'r--');
56 grid on;
57 xlabel('direkte Messgroesse x2', 'fontsize', 14);
58 ylabel('Wahrscheinlichkeitsdichte p(x2)', 'fontsize', 14);
59 set(gca, 'fontsize', 12);
60 if pltflag2
61 print(2,'direkte_x2.png', '-dpng');
62 end
63 %
64 % Spalte mal Zeile liefert Matrix
65 pgesamt = p2*p1;
66 pgesamt = pgesamt / sum(pgesamt(:));
67 %

```

```

68 figure(3);
69 imagesc(x1, x2, pgesamt);
70 colorbar();
71 xlabel('direkte Messgroesse x1', 'fontsize', 14);
72 ylabel('direkte Messgroesse x2', 'fontsize', 14);
73 title('Wahrscheinlichkeitsdichteverteilung p(x1, x2)', 'fontsize', 14);
74 set(gca, 'fontsize', 12);
75 if pltflag1
76   print(3,'direkte_x1_x2_box.png', '-dpng');
77 end
78 %
79 % _____
80 % x1 und x2 in Matrizen fuellen
81 n2_r = length(x2)
82 n1_c = length(x1)
83 X_1 = ones(n2_r,1) * x1;
84 X_2 = x2 * ones(1, n1_c);
85 Y_matrix = a*X_1 + b*X_2;
86 %
87 % nun als 1D arrays
88 y = Y_matrix(:);
89 p_y = pgesamt(:);
90 figure(4);
91 plot(y, p_y);
92 axis([2 11]);
93 title('scharfer Modellprior – GUM S1', 'fontsize', 14);
94 xlabel('indirekte Messgroesse y', 'fontsize', 14);
95 ylabel('Wahrscheinlichkeitsdichte', 'fontsize', 14);
96 set(gca, 'fontsize', 12);
97 if pltflag1a
98   print(4,'indirekte_y_GUMS1.png', '-dpng');
99 end
100 %
101 % sortieren, so dass y monoton steigend ist
102 [ys, isort] = sort(y);
103 p_sort = p_y(isort);
104 figure(5);
105 plot(ys, p_sort, 'x-');
106 axis([2 11]);
107 grid on;
108 title('scharfer Modellprior – GUM S1', 'fontsize', 14);
109 xlabel('indirekte Messgroesse y', 'fontsize', 14);
110 ylabel('Wahrscheinlichkeitsdichte', 'fontsize', 14);
111 set(gca, 'fontsize', 12);

```

```

112 if pltflag1
113   print(5,'indirekte_ysorted_GUMS1_box.png', '-dpng');
114 end
115 %
116 %
117 % mit gaussfoermigem Modellprior, aber sehr schmal, fast dirac'sche delta-Distribution
118 %
119 dy = 0.01;
120 y_model = [2:dy:11];
121 n_model = length(y_model);
122 % sigma = 0.5;
123 p_mod = zeros(1, n_model);
124 for k = 1:n_model
125 %
126 % Summationen entsprechen der Integration ueber x1 und x2
127 % zum Berechnen der Marginalverteilung
128 p_mod(k) = sum(sum(exp(-0.5 * ((y_model(k) - Y_matrix)/sigma).^2) .* pgesamt));
129 end
130 Pkum = cumsum(p_mod);
131 psum = Pkum(n_model);
132 Pkum = Pkum/psum;
133 p_mod = p_mod/psum;
134 %
135 %
136 % Signifikanzniveau
137 alpha = 0.05;
138 %
139 % Quantile
140 [left, ileft] = min(abs(Pkum - alpha/2))
141 [right, iright] = min(abs(Pkum - (1-alpha/2)))
142 %
143 erw_y = sum(y_model .* p_mod);
144 std_y = sqrt(sum((y_model - erw_y).^2 .* p_mod));
145 %
146 % Vergleich mit Fortpflanzungsgesetz, das eigentlich nur fuer
147 % glockenfoermige Verteilungen gilt, bei denen die Wurzel des zweiten statistischen
148 % Moments, also die Varianz die Breite der Verteilung definiert, bzgl der sie
149 % sich normieren laesst.
150 y_quer = a*x1_quer + b*x2_quer;
151 if boxflag
152   u_y = sqrt((a * s1)^2/3 + (b * s2)^2);
153 else
154   u_y = sqrt((a * s1)^2/2 + (b * s2)^2);
155 end

```

```

156 %
157 figure(6);
158 sigma_str = num2str(sigma, '%1.3f');
159 plot(y_model, p_mod, 'kx-', ...
160       [y_quer-1.96*u_y, y_quer-1.96*u_y], [0, p_mod(iright)], 'b-', 'linewidth', 2.5,
161       ...
162       [y_quer+1.96*u_y, y_quer+1.96*u_y], [0, p_mod(iright)], 'b-', 'linewidth', 2.5,
163       ...
164       [y_model(ileft), y_model(ileft)], [0, p_mod(iright)], 'r-', 'linewidth', 3, ...
165       [y_model(iright), y_model(iright)], [0, p_mod(iright)], 'r-', 'linewidth', 3);
166 grid on;
167 axis([2 11]);
168 title(['normalverteilter Modellprior mit \sigma = ', sigma_str], 'fontsize', 14);
169 xlabel('indirekte Messgroesse y', 'fontsize', 14);
170 ylabel('Wahrscheinlichkeitsdichte', 'fontsize', 14);
171 set(gca, 'fontsize', 12);
172 if pltflag3
173   print(6,['indirekte_y_modellGauss_sigma', strrep(sigma_str,'.', 'p'), '_box.png'], '-dpng');
174 end

175 %
176 figure(7);
177 plot(y_model, Pkum, 'ko-', ...
178       [y_model(ileft), y_model(ileft)], [0, Pkum(ileft)], 'r-', 'linewidth', 3, ...
179       [y_model(iright), y_model(iright)], [0, Pkum(iright)], 'r-', 'linewidth', 3, ...
180       [y_model(1), y_model(ileft)], [Pkum(ileft), Pkum(ileft)], 'r-', 'linewidth', 3,
181       ...
182       [y_model(1), y_model(iright)], [Pkum(iright), Pkum(iright)], 'r-', 'linewidth',
183       3);
184 grid on;
185 axis([2 11]);
186 title(['normalverteilter Modellprior mit \sigma = ', sigma_str], 'fontsize', 14);
187 xlabel('indirekte Messgroesse y', 'fontsize', 14);
188 ylabel('kumulierte Wahrscheinlichkeit', 'fontsize', 14);
189 set(gca, 'fontsize', 12);
190 if pltflag3
191   print(7,['Pkum_indirekte_y_Gauss_sigma', strrep(sigma_str,'.', 'p'), '_box.png'], '-dpng');
192 end

193 %
194 printf('Modellprior mit sigma = %1.3f\n', sigma);
195 printf('Erwartungswerte:\n');
196 printf('aus Erw von x1 und x2: %1.2f\n', y_quer);

```

```
194 printf('aus Verteilung p_model: %1.2f\n', erw_y);
195 printf('Standardabweichungen als Wurzel der Varianz:\n');
196 printf('aus Fortpflanzung klassisch: %1.2f\n', u_y);
197 printf('aus Verteilung p_model:      %1.2f\n', std_y);
198 printf('Ueberdeckungsintervall fuer 95 Prozent\n');
199 printf('aus Fortpflanzung klassisch mit Quantil aus Gaussvert 1.96: [%1.2f, %1.2f]\n',
200      n',
201      y_quer - 1.96*u_y, y_quer + 1.96*u_y);
202 printf('aus Verteilung p_model, ueber die Umkehrfkt der kumulativen: [%1.2f, %1.2f]\n',
203      n',
204      y_model(ileft), y_model(iright));
205 end
```

10.4 Übungsaufgaben

Aufgabe 10-1: Bestimmung des Schätzwertes und der Messunsicherheit bei einem linearen Modell und Normalverteilungen

Gegeben ist ein linearer Zusammenhang zwischen der Ausgangsgröße (indirekte Messgröße) Y und den Eingangsgrößen X_j (direkte Messgrößen) wie folgt:

$$Y = c_1 \cdot X_1 + c_2 \cdot X_2$$

Die direkten Messgrößen X_1 und X_2 sind nicht korreliert und normalverteilt mit:

$$X_1 \sim N_1(\mu_1 = x_1 = 3, \sigma_1^2 = 0.5^2) \quad X_2 \sim N_2(\mu_2 = x_2 = 5, \sigma_2^2 = 0.2^2)$$

Nehmen Sie für $c_1 = 2$ und für $c_2 = 3$ an.

Als Priorinformation ist bekannt, dass die indirekte Messgröße normalverteilt ist mit $N_0(y_0 = 20, \sigma_0^2 = 3)$

Bestimmen Sie den Schätzwert y und die dazugehörige Unsicherheit $u(y)$ mit den Gl.(10.37) und (10.38)

Lösung zu 10-1

Die Kovarianzmatrix ist gegeben durch

$$\mathbf{V}(\mathbf{X}) = \begin{pmatrix} \text{Var}(X_1) & 0 \\ 0 & \text{Var}(X_2) \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \quad (10.54)$$

Für die Unsicherheit von y erhält man mit Vorwissen nach Gl. (10.38):

$$u^2(y) = \left[\frac{1}{c_1^2 \cdot \sigma_1^2 + c_2^2 \cdot \sigma_2^2} + \frac{1}{\sigma_0^2} \right]^{-1} = \left[\frac{1}{2^2 \cdot 0.5^2 + 3^2 \cdot 0.2^2} + \frac{1}{3^2} \right]^{-1} = 1.1815$$

Für den Schätzwert y ergibt sich nach Gl.(10.37):

$$y = u^2(y) \left[\frac{c_1 \cdot x_1 + c_2 \cdot x_2}{c_1^2 \cdot \sigma_1^2 + c_2^2 \cdot \sigma_2^2} + \frac{y_0}{\sigma_0^2} \right] = 1.1815 \cdot \left[\frac{2 \cdot 3 + 3 \cdot 5}{2^2 \cdot 0.5^2 + 3^2 \cdot 0.2^2} + \frac{20}{3^2} \right] = 20.8867$$

Als **Schätzwert** erhält man $y = 20.89$ und als **Unsicherheit** $u(y) = 1.09$

Literaturverzeichnis

zu Kapitel 10:

- [Els07] C. Elster: Calculation of uncertainty in the presence of prior knowledge, Metrologia 44, 111-116, (2007)
- [Cox06] M.G. Cox, B.R.L. Siebert: The use of a Monte Carlo method for evaluating ..., Metrologie 43, S178-88, (2006)
- [GUMS1] 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 (2008);
https://www.bipm.org/utils/common/documents/jcgm/JCGM_101_2008_E.pdf
- [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 (2008) 084009 (4pp) doi:10.1088/0957-0233/19/8/084009

Kapitel 11

Infrastrukturen der Metrologie

11.1 Internationale Metrologieorganisationen (BIPM,NMIs)

...

11.2 Akkreditierung, Zertifizierung, DAkkS

11.2.1 Messen, Prüfen, Lehren

Messen: Ermittlung von Werten einer Messgröße i. Allg. unter festgelegten Bedingungen (Messbedingungen), siehe z. B. auch die Definition von *Messung* im [VIM08]

Prüfen: Ermittlung von Eigenschaften (quantitative, qualitativ) eines Objektes nach einem vorgegebenen Verfahren - häufig auch einschl. Vergleich mit vorgegebenen Spezifikationen (*Konformität*) Prüfen bedeutet nach DIN 1319 das Feststellen, inwieweit ein Prüfobjekt eine Forderung erfüllt [DIN1319]. Wird nur mit dem menschlichen Sinnen – ohne Hilfsmittel – geprüft, spricht man vom subjektiven Prüfen. Die Prüfergebnisse sind nur schlecht miteinander vergleichbar. Werden Hilfsmittel – sogenannte Prüfmittel – verwendet, so spricht man vom objektiven Prüfen.

Lehren: Objektive Prüfverfahren werden in die Arten Messen und Lehren unterteilt. Beim Messen wird eine physikalische Größe mit einem Messgerät erfasst und ein Messwert ermittelt. Der Messwert setzt sich aus Zahlenwert mit Einheit, welche die physikalische Größe repräsentiert, zusammen. Beim Lehren wird festgestellt, ob das zu prüfende Objekt innerhalb vorgegebener Grenzen liegt oder nicht. Das *Prüfergebnis ist kein Zahlenwert, sondern eine Gut- / Schlecht-Aussage*. Oft lässt sich beim Lehren erkennen, in welche Richtung die Grenze überschritten wurde.

11.2.2 Akkreditierung (Wirtschaft)

Die Anforderungen an die Qualität von Waren und Dienstleistungen nehmen angesichts der Liberalisierung des Welthandels sowie der steigenden Ansprüche von Verbrauchern, Unternehmen und Gesetzgebern stetig zu. Ob im Umweltschutz, in der Lebensmittel- oder Elektroindustrie, im Gesundheitswesen oder im Bereich Erneuerbarer Energien – in diesen wie in vielen anderen Wirtschaftsbereichen sind objektive Prüfungen, Kalibrierungen, Inspektionen oder Zertifizierungen daher von großer Bedeutung [wiki01].

Diese Bewertungen stellen sicher, dass die überprüften Produkte, Verfahren, Dienstleistungen oder Systeme hinsichtlich ihrer Qualität und Sicherheit *verlässlich* sind, sie einem technischen Mindestniveau entsprechen und mit den Vorgaben entsprechender Normen, Richtlinien und Gesetze konform sind. Daher werden diese objektiven Bestätigungen auch als Konformitätsbewertung bezeichnet. Vertrauen durch Akkreditierung

Das Vertrauen in Zertifikate, Inspektionen, Prüfungen oder Kalibrierungen steht und fällt jedoch mit der Kompetenz desjenigen, der die Bewertungsleistung erbringt. Viele dieser sogenannten **Konformitätsbewertungsstellen** belegen die Qualität ihrer eigenen Arbeit daher durch eine Akkredi-

tierung.

In diesem Verfahren weisen sie gegenüber einer unabhängigen Akkreditierungsstelle nach, dass sie ihre Tätigkeiten fachlich kompetent, unter Beachtung gesetzlicher sowie normativer Anforderungen und auf international vergleichbarem Niveau erbringen. Die Akkreditierungsstelle begutachtet und überwacht dabei das Managementsystem und die Kompetenz des eingesetzten Personals der Konformitätsbewertungsstelle.

Akkreditierungen tragen deshalb somit entscheidend dazu bei, die Vergleichbarkeit von Konformitätsbewertungsergebnissen zu gewährleisten und *Vertrauen in die Qualität und Sicherheit von Produkten und Dienstleistungen zu erzeugen*.

Der Begriff „Akkreditierung“ (Latein: „accredere“, Glauben schenken) wird in verschiedenen Bereichen benutzt, um den Umstand zu beschreiben, dass eine allgemein anerkannte Instanz einer anderen das Erfüllen einer besonderen (nützlichen) Eigenschaft bescheinigt.

Akkreditierung im hier gemeinten Sinne ist gemäß DIN EN ISO/IEC 17011 - 2018-03 (Konformitätsbewertung - Anforderungen an Akkreditierungsstellen, die Konformitätsbewertungsstellen akkreditieren) definiert. Sie ist die Bestätigung durch eine dritte Stelle, die formal darlegt, dass eine Konformitätsbewertungsstelle (z.B. Laboratorien, Inspektions- und Zertifizierungsstellen) die Kompetenz besitzt, bestimmte Konformitätsbewertungsaufgaben durchzuführen.

Konformitätsbewertungsstellen sind Organisationen, die folgende Dienstleistungen zur Konformitätsbewertung bereitstellen: Prüfung, Inspektion, Zertifizierung von Managementsystemen, Personen Personenzertifizierung im Sinne eines Qualifikationsnachweises und Produkten. **Kalibrierlabors** werden ebenfalls dem Begriff Konformitätsbewertungsstellen untergeordnet (obwohl sie nicht in erster Linie eine Konformität bewerten).

Eine **Zertifizierungsstelle** ist akkreditiert, wenn eine Akkreditierungsstelle formell bestätigt, dass die Zertifizierungsstelle die z.B. in der Normenreihe EN 45011–45013 aufgeführten Voraussetzungen zur Durchführung von Bewertungen von Managementsystemen (zum Beispiel Qualitätsmanagement oder Umweltmanagement), Produkten oder Personen erfüllt. Ein **Prüf- bzw. Kalibrierlaboratorium** ist akkreditiert, wenn es die z. B. Anforderungen der Norm DIN EN ISO/IEC 17025:2018 (Allgemeine Anforderungen an die Kompetenz von Prüf- und Kalibrierlaboratorien) erfüllt, die den Qualitätsstandard (DIN EN ISO 9001:2008) einschließt und darüber hinausgehende Anforderungen enthält. Hierzu wird das Labor durch eine **Expertengruppe** einer unabhängigen Akkreditierungsstelle, welche die Norm DIN EN ISO/IEC 17011 erfüllt, begutachtet und durch meist jährliche Begehungen überwacht. Eine **Konformitätsbewertungsstelle** ist akkreditiert, wenn sie den Anforderungen z. B. der Norm ISO/IEC 17021 (vormals ISO/IEC 45012) entspricht und eine Akkreditierungsstelle dies formell bestätigt. Eine **Inspektionsstelle** ist akkreditiert, wenn sie die Anforderungen der Norm z. B. ISO/IEC 17020 erfüllt und eine Akkreditierungsstelle dies formell bestätigt.

In Deutschland ist seit dem 1. Januar 2010 ausschließlich die Deutsche Akkreditierungsstelle GmbH (DAkkS) für die Erteilung und Aufrechterhaltung von Akkreditierungen zuständig. Die Errichtung einer nationalen Akkreditierungsstelle erfolgte nach der Vorgabe der Verordnung (EG) Nr. 765/2008 und nach Maßgabe des deutschen Akkreditierungsstellengesetzes. Gesellschafter der DAkkS sind zu gleichen Teilen die Bundesrepublik Deutschland, die Bundesländer und die durch den Bundesverband

der Deutschen Industrie (BDI) vertretene Wirtschaft (siehe [DAkkS]). Die DAkkS ist ein zentraler Baustein der Qualitätsinfrastruktur und erteilen Akkreditierungen im Bereich der Konformitätsbewertung. Sie hat einen gesetzlichen Auftrag und handelt im Interesse des Staates, der Wirtschaft sowie zum Schutz von Gesellschaft und Umwelt. Mit einer Akkreditierung wird bestätigt die DAkkS, das Organisationen wie Prüf-, Inspektions-, und Zertifizierungsstellen ihre Tätigkeiten nach international gültigen Maßstäben kompetent erbringen können. Für die Durchführung von Akkreditierungen greift die DAkkS auf eine breite Basis von Experten zurück (Gutachter und Fachexperten). DAkkS muss *unabhängig* und *unparteiisch* sein. Die DAkkS selbst hat ca. 190 Mitarbeiter in 49 Fachbereichen und hat ca. 4250 gültige Akkreditierungen im Bestand. Jährlich werden ca. 4000 Begutachtungen durchgeführt. Sie wurde entsprechend ISO/IEC 17011:2017 (Conformity assessment – Requirements for accreditation bodies accrediting conformity assessment bodies) umstrukturiert. Die DAkkS wurde dementsprechend angepasst. **Aktuell werden alle akkreditierten Stellen auf die DIN EN ISO/IEC 17025:2018 (Allgemeine Anforderungen an die Kompetenz von Prüf- und Kalibrierlaboratorien) umgestellt.** Im Vergleich zur Ausgabe aus dem Jahr 2005, sind die wesentlichen Änderungen von ISO/IEC 17025 folgende: Die Hinzunahme eines risikobasierten Ansatzes, größere Flexibilität bei den Anforderungen für Prozesse, Verfahren, dokumentierte Informationen und organisatorische Verantwortlichkeiten sowie die Einführung von Definitionen. Die dreijährige Übergangsfrist endet im November 2020. Akkreditierungen auf Basis der alten DIN EN ISO/IEC 17025:2005 verlieren Ende 2020 ihre Gültigkeit.

Beispiel: Die Akkreditierung von TÜV, DEKRA und Co

Die TÜV-Gesellschaften, DEKRA und Co bekommen für ihre Arbeit eine sogenannte Akkreditierung von einer staatlichen Stelle namens DAkkS (Deutsche Akkreditierungsstelle), die bestätigt, dass die Messungen korrekt durchgeführt werden. Es kann durchaus mal passieren wie im Jahre 2015, dass diese Akkreditierung bei Vorliegen von Fehlverhalten (hier: keine vollständige Dokumentation) die Akkreditierung entzogen werden kann (siehe z. B. Artikel: <http://www.autobild.de/artikel/tuev-dekra-und-co-ohne-zulassung- -8471909.html>)

Wer kann sich akkreditieren lassen?

Die DAkkS akkreditiert Konformitätsbewertungsstellen (KBS). Als solche Stellen gelten Laboratorien, Zertifizierungs- und Inspektionsstellen, Anbieter von Eignungsprüfungen und Referenzmaterialhersteller (siehe Abb.11.1). Die Akkreditierung einer KBS durch die DAkkS ist möglich, wenn diese die Anforderungen der entsprechenden internationalen Normen erfüllt.

Seit 1. September 2019 sind Akkreditierungen von Biobanken auf Basis der ISO 20387 möglich, da diese eine zunehmende Relevanz spielen und es eine hohe Nachfrage gibt. In Biodatenbanken werden DNA-, Blut- oder Gewebeproben zusammen mit Hintergrundinformationen gespeichert. Bislang konnten nur spezifische Teile der Tätigkeit von Biobanken akkreditiert werden. Nun kann der gesamte Prozess - von der Entnahme der Probe bis zur Verteilung der Materialien inklusive der assoziierten Daten akkreditiert werden.

Das Regelwerk ist zu finden unter: https://www.dakks.de/sites/default/files/dokumente/71_sd_0_000_liste-dakks-regelwerk_20191218.pdf

Die internationale Harmonisierung dieser Normen gewährleistet, dass die Akkreditierung weltweit nach gleichen Voraussetzungen erfolgt. Durch diese harmonisierten Normen und dank internationaler

Typen von Konformitätsbewertungsstellen

Laboratorien	Prüf- und Kalibrierlaboratorien DIN EN ISO/IEC 17025
	Medizinische Laboratorien DIN EN ISO 15189
Inspektionsstellen	DIN EN ISO/IEC 17020
	für Personen DIN EN ISO/IEC 17024
Zertifizierungsstellen	für Managementsysteme DIN EN ISO/IEC 17021-1
	für Produkte, Prozesse und Dienstleistungen DIN EN ISO/IEC 17065
Validierungs- und Verifizierungsstellen	DIN EN ISO 14065
Anbieter von Eignungsprüfungen	DIN EN ISO/IEC 17043
Referenzmaterialhersteller	ISO Guide 34 i.V.m. DIN EN ISO/IEC 17025

Abbildung 11.1: Typen von Konformitätsstellen, Quelle:[DAkkS]

Abkommen werden die Bewertungsleistungen der in Deutschland akkreditierten Stellen in vielen anderen Ländern Europas und der Welt anerkannt.

Diese Überwindung technischer Handelshemmisse erleichtert den Handel über Grenzen hinweg und stellt sicher, dass die Ergebnisse von Konformitätsbewertungen ohne eine erneute Überprüfung *international akzeptiert* werden.

Vorteile von Akkreditierungen

Für Unternehmen:

- Höhere Akzeptanz von Produkten und Dienstleistungen erleichtert den Marktzugang bzw. ermöglicht diesen erst
- Einmal geprüft, überall akzeptiert: Internationale Vergleichbarkeit und Anerkennung von Zertifikaten, Inspektionen, Prüfungen oder Kalibrierungen vermeidet Kosten durch mehrfache Bewertungen
- Kompetenznachweis erleichtert die Auswahl eines passenden Dienstleisters für die Konformitätsbewertung von Waren und Dienstleistungen

Für akkreditierte Stellen:

- objektiver Beleg für die Güte und Kompetenz der Tätigkeit einer Konformitätsbewertungsstelle nach internationalen Standards
- Wettbewerbsvorteile gegenüber nicht akkreditierten Marktteilnehmern

Für Verbraucher:

- größeres Vertrauen der Verbraucher in die Qualität von Produkten und Dienstleistungen – trotz eines komplexen Weltmarkts
- weniger Produktfehler oder Rückrufaktionen

Für den Gesetzgeber:

- verbesserte Wettbewerbsfähigkeit der Wirtschaft durch den Abbau
- technischer Handelshemisse

Akkreditierungsprozess

- Antragsphase

Der Akkreditierungsprozess beginnt mit der Einsendung des Akkreditierungsantrages und der fachspezifischen Anlagen an die Zentrale Antragsbearbeitung (ZAB) der DAkkS in Berlin. Optional können sich Kunden auch vorab in einem Vorgespräch in den Geschäftsstellen der DAkkS über ihre angestrebte Akkreditierung oder das Akkreditierungsverfahren im konkreten Fall informieren. Die ZAB überprüft in Abstimmung mit der zuständigen Fachabteilung der DAkkS den Antrag. Dabei wird auch geprüft, ob eine Befugnis erteilende Behörde (BeB) in das Akkreditierungsverfahren eingebunden werden muss. Nach einer erfolgreichen Antragsprüfung informiert der zugeteilte Verfahrensmanager die antragstellende Konformitätsbewertungsstelle (KBS) über das weitere Vorgehen.

- Begutachtungsphase

In der zweiten Phase des Akkreditierungsprozesses begutachtet die DAkkS durch ein Begutachterteam die technische Kompetenz und das Managementsystem der KBS. Zunächst prüfen die Begutachter die eingereichten Dokumente, dann findet zum vereinbarten Termin die Begehung vor Ort statt. Der Umfang und die Dauer der Begutachtung sind von der Größe der KBS, dem beauftragten Geltungsbereich (Scope) der Akkreditierung und der Komplexität des Verfahrens abhängig. Die Ergebnisse werden in einem Begutachtungsbericht dokumentiert. Festgestellte Abweichungen kann der Kunde durch entsprechende Korrekturmaßnahmen im Anschluss an den Begutachtungstermin innerhalb von zwei Monaten beheben. Diese werden nochmals überprüft und bewertet.

- Akkreditierungsphase

In dieser Phase bewertet ein Akkreditierungsausschuss (AkA) die Begutachtungsergebnisse und entscheidet über die Erteilung der Akkreditierung. Die DAkkS bescheinigt den erfolgreichen Abschluss der Akkreditierungsphase durch den Akkreditierungsbescheid und die Akkreditierungsurkunde. Damit bestätigt die DAkkS der überprüften KBS die Erfüllung der entsprechenden Normen, Standards oder Gesetzen im Hinblick auf ihre Konformitätbewertungstätigkeiten – und damit ihre technische Kompetenz. Die Akkreditierung wird anschließend im Verzeichnis der akkreditierten Stellen gelistet.

- Überwachungsphase

Eine Akkreditierung wird in gewissen zeitlichen Abständen überwacht, um die Kompetenz sicherzustellen. Je nachdem, ob es sich um ein Laboratorium, eine Inspektions-, Zertifizierungs- oder Verifizierungsstelle handelt, gibt es unterschiedliche Zeitintervalle. Der Akkreditierungszyklus endet in der Regel spätestens nach fünf Jahren und bedarf dann einer Reakkreditierung.

Akkreditierungssymbol

Auf Antrag gestattet die DAkkS einer akkreditierten Stelle die Verwendung des DAkkS-Akkreditierungssymbols (siehe Abb.11.2). Zusammen mit der eindeutigen Registrierungsnummer weist es auf die erfolgreiche Akkreditierung hin. Durch die genehmigte Verwendung des Symbols können akkreditierte Stellen ihren Kunden die nachgewiesene Kompetenz signalisieren.



Abbildung 11.2: Muster des Akkreditierungssymbols der DAkkS für akkreditierte Stellen,
Quelle:[DAkkS]

Kosten der Akkreditierung

Als Deutschlands nationale Akkreditierungsstelle handelt die DAkkS zum Wohle des Staates, der Wirtschaft und der Verbraucher: Sie arbeitet im hoheitlichen Bereich kostendeckend, aber nicht gewinorientiert.

Gültigkeit von Akkreditierungen

Akkreditierungen sind üblicherweise fünf Jahre lang gültig, müssen jedoch in regelmäßigen Abständen durch die DAkkS überwacht werden.

11.2.3 Zertifizierung

Als „Zertifizierung“ (von lat. „certe“ = bestimmt, gewiss, sicher und „facere“ = machen, schaffen, verfertigen) bezeichnet man ein Verfahren, mit dessen Hilfe die Einhaltung bestimmter Anforderungen nachgewiesen wird.

Zertifizierung ist ein Teilprozess der Konformitätsbewertung. Zertifizierungen werden oft zeitlich befristet von unabhängigen Zertifizierungsstellen wie z. B. DQS, TÜV oder DEKRA vergeben und die Standards unabhängig kontrolliert.

Anforderungsbereiche

Die Bereiche, in denen Anforderungen gestellt werden, die zertifiziert werden können, umfassen im Allgemeinen:

- Produkte und Dienstleistungen und ihre jeweiligen Herstellungsverfahren einschließlich der Handelsbeziehungen
- Personen
- Systeme
- Unternehmen

Arten der Zertifizierung

- Nachweis von Ausbildungsstandards oder besonders ausgearbeiteten Fachnormen bei Personenzertifizierungen.
- Nachweis von Ausbildungsstandards bei der Anerkennung von Ausbildungsinstituten, wie beispielsweise durch Berufsverband|Berufsverbände durchgeführt wird (bei nichtuniversitären Ausbildungen wird teils von „zertifizierten“ Ausbildungsinstituten, teils von „akkreditierten“ Ausbildungsinstituten gesprochen, wobei Letztere zugleich befugt sind, Personenzertifikationen oder Teile davon durchzuführen).
- International anerkannter Nachweis der persönlichen Befähigung, zum Beispiel als PMP (Project Management Professional) durch das PMI (Project Management Institute) IPMA-Zertifikate Level D-A für Projektmanager.
- Zertifizierung eines Managementsystems (zum Beispiel nach ISO 9001, ISO 14001).
- Zertifizierung von Produkten oder Dienstleistungen. Für Zertifizierungsstellen, die Zertifizierungssysteme für Produkte oder Dienstleistungen betreiben, besteht die EN ISO/IEC 17065 (früher EN 45011 bzw. ISO/IEC Guide 65).

- Zertifizierung der Herkunftsregion eines Produktes (DOC).
- Zertifizierung der Informationssicherheit nach BS 7799 oder ISO 27001|ISO/IEC 27001.
- Zertifizierung zum Nachweis der Einhaltung von Umwelt- und Sozialstandards, zum Beispiel bei der Zertifizierung von nachhaltig erzeugtem Holz (siehe Zertifizierung (Forstwirtschaft)|FSC) oder von Produkten aus Entwicklungsländern, die bessere Konditionen für die dortigen Produzenten garantieren nach Fairer Handel|Fair-Trade-Kriterien.
- Zertifizierung zum Nachweis der Einhaltung von Anforderungen an den Arbeitssicherheit, Arbeits- und Umweltschutz gem. Occupational Safety and Health Administration|OHSAS bzw. ISO 14001
- Zertifizierung zum Nachweis von Arbeitsbedingungen gemäß SA8000 und ähnlichen Regelwerken (Beispiele: Sedex und Business Social Compliance Initiative|BSCI).
- In der Software-Industrie ist die Zertifizierung insbesondere im Hinblick auf die Computersicherheit wichtig:
- Zertifizierung der Mitarbeiter zur Dokumentation von Fähigkeiten, Qualifikation und Kompetenz. Siehe dazu Liste der IT-Zertifikate.
- Zertifizierung von Softwareprodukten in Hinblick auf Funktionalität und Qualität. Besonders wichtig sind hier der amerikanische TCSEC- und der Europäische Information Technology Security Evaluation Criteria-Standard sowie im Hinblick auf die internationale Anerkennung die Common Criteria (CC). In Deutschland erfolgt die Zertifizierung durch das Bundesamt für Sicherheit in der Informationstechnik|BSI.
- Zertifizierung der IT-Umgebung nach IT-Grundschutz. Um in einem Unternehmen einen solchen Prozess zu begleiten, werden vom BSI Grundschutz-Auditoren lizenziert. Diese sind autorisiert, Testate als Vorbereitung auf die Zertifizierung auszugeben. Die eigentliche Zertifizierung erfolgt durch das BSI.
- Im Bereich Linux und freie Software ist ein wichtiges zertifizierendes Institut das kanadische Linux Professional Institute|LPI.
- In der Lebensmittelindustrie gibt es heute verschiedene Normen, angelehnt an die weltweit bekannte ISO 9001. Diese wurden zugeschnitten auf die Bedürfnisse der Lebensmittelindustrie. Weit verbreitete Standards sind zum Beispiel der International Food Standard (IFS), die Anforderungen des British Retail Consortium, Good Manufacturing Practice|GMP, Hazard Analysis and Critical Control Points und die ISO 22000. Zunehmender Akzeptanz erfreut sich für Fische auch der MSC-Standard des Marine Stewardship Council.
- Packmittelhersteller, die Packmittel für den direkten Kontakt mit Lebensmitteln (Primärpackmittel) herstellen, sind zunehmend nach der Norm BRC-IoP zertifiziert.
- Zertifizierung von Altersvorsorgeprodukten, z. B. Riesterrente.
- Zertifizierung von Energiemanagementsystemen gemäß ISO 50001.
- Nachhaltigkeitszertifizierung von Biomasse und Biokraftstoffen.

- Unternehmenszertifizierung, z.B. für den Bereich Nachhaltigkeit. Hier werden Standards vom Standardgeber vorgeschrieben. In diesem Fall wird vom Standardgeber definiert, was Nachhaltigkeit im unternehmerischen Kontext bedeutet. Das gesamte Unternehmen (Management, Lieferketten, Produkte usw.) wird im Zertifizierungsprozess überprüft.

11.2.4 Konformitätsbewertung

„Konformitätsbewertung“ ist in der internationalen Norm ISO/IEC 17000:2004 „Konformitätsbewertung – Begriffe und allgemeine Grundlagen“ definiert als „ Darlegung, dass festgelegte Anforderungen bezogen auf ein Produkt, einen Prozess, ein System, eine Person oder eine Stelle erfüllt sind“.

Konformitätsbewertung ist ein Überbegriff für Tätigkeiten des Auswählens, Ermittelns (von Eigenschaften), Bewertens (etwa auf Einhaltung vorgegebener oder allgemeiner Anforderungen) und Bestätigens (etwa durch Konformitätserklärung|Erklärung des Herstellers, oder ein Zertifizierung|Zertifikat einer Zertifizierungsstelle, dass ein Produkt bestimmte Normen einhält). Solche Tätigkeiten sind beispielsweise Stichprobennahme, Prüfen, Inspizieren, Erklären, Zertifizieren, Akkreditieren. Die Objekte der Konformitätsbewertung sind nicht eingeschränkt.

Konformitätsbewertung findet auf vielfältige Weise und allen Ebenen statt:

- im Betrieb (etwa Endprüfung, Auditierung eines Qualitätsmanagementsystems durch eigene Auditoren):
- durch Stellen / Personen des Kunden oder Abnehmers:
- durch kommerzielle, vom Auftraggeber unabhängige Konformitätsbewertungsstellen (etwa Laboratorien; Zertifizierungsstellen; Inspektionsstellen; Akkreditierung (Wirtschaft), Akkreditierungsstellen):

EU-Konformitätsbewertung

Eine besondere Bedeutung hat die Konformitätsbewertung durch Benannte Stellen bei der Bewertung von Produkten auf ihre Übereinstimmung mit den Anforderungen einer Richtlinie (EU)|EU-Richtlinie. EU-Richtlinien gemäß Art. 95 EG-Vertrag für den Europäischer Binnenmarkt|Europäischen Binnenmarkt legen für zahlreiche Produkte Mindestanforderungen an die Sicherheit fest, die vom Hersteller erfüllt werden müssen.

Durch ein „Konformitätsbewertungsverfahren“ muss der Hersteller nachweisen, dass er die in der Richtlinie oder den Richtlinien enthaltenen grundlegenden Sicherheitsanforderungen eingehalten hat. Das Konformitätsbewertungsverfahren muss vom Hersteller für jedes Produkt vor dem erstmaligen Inverkehrbringen durchgeführt werden. Am Ende des Konformitätsbewertungsverfahrens stellt der Hersteller eine EU-Konformitätserklärung für sein Produkt aus, in der er erklärt, dass das Produkt zu den Anforderungen der entsprechenden Richtlinie(n) konform ist. Am Produkt bringt der Hersteller dann die CE-Kennzeichnung an, falls die angewandte Richtlinie dies vorsieht.

Nur im Sektor „Medizinprodukte“ besteht die Besonderheit, dass im Rahmen der Konformitätsbewertung nicht nur die Produktsicherheit nachgewiesen werden muss, sondern zusätzlich auch die

medizinisch-technische Leistungsfähigkeit von Medizinprodukten, so wie sie vom Hersteller in der Produktkennzeichnung einschließlich der Werbung als medizinische Indikation ausgelobt ist. Das entsprechende Nachweisverfahren nennt sich klinische Bewertung. Erst der (je nach der Produktklasse) extern durch Benannte Stelle zertifizierte Nachweis der Produktsicherheit und der Leistungsfähigkeit berechtigt Hersteller von Medizinprodukten zur Anbringung der **CE-Kennzeichnung**.

In den Anhängen der Richtlinien werden verschiedene Module für die Durchführung eines Konformitätsbewertungsverfahrens genannt. Welche Module gewählt werden können, hängt von der Klassifizierung des Produktes ab. Für Produkte mit höherem Risiko ist die Einbeziehung einer Benannten Stelle bei der Durchführung des Konformitätsbewertungsverfahrens obligatorisch. Beispiele für Module:

- Modul A2: Interne Fertigungskontrolle durch den Hersteller und Überwachung der Abnahme durch benannte Stelle
- Modul B *Baumusterprüfung* EG-Baumusterprüfung (Feststellung der Übereinstimmung mit den einschlägigen internationalen und nationalen Fachnormen (DIN, EN, ISO, IEC etc.), IMO-Resolutionen und SOLAS-Bestimmungen)
- Modul D : Qualitätssicherung Produktion (Darlegung der Qualitätssicherungsmaßnahmen vor, während und nach der Produktion einschließlich deren Häufigkeit)
- Modul E : Qualitätssicherung Produkt (Darlegung der Qualitätssicherungsmaßnahmen nach der Produktion einschließlich deren Häufigkeit (Endabnahme und Prüfung))
- Modul F : Prüfung der Produkte (Prüfung der Produkte auf Übereinstimmung mit dem Baumuster entweder durch statistische Kontrollen oder Prüfung jedes einzelnen Produkts durch die Benannte Stelle)
- Modul G : Einzelprüfung (anwendbar auf Produkte, die nicht in Serie produziert werden)

Je höher das Gefahrenpotential eines Produktes ist, umso mehr Prüfungsumfang muss auf eine „Benannte Stelle“ übertragen werden. Diese wird durch eine vierstellige Kennziffer hinter der CE-Kennzeichnung angegeben. Die Konformitätsbewertung für Produkte mit einem sehr geringen Gefährdungspotenzial kann der Hersteller ohne Einschaltung einer Benannten Stelle selbst durchführen.

In jedem Fall - also auch bei Einschaltung einer Drittstelle - muss eine EU-Konformitätserklärung (in der Regel) durch den Hersteller ausgestellt werden. Dies unterstreicht dessen alleinige Verantwortung (= Haftung) für das Produkt.

Bei der Konformitätsbewertung müssen die Sicherheitsanforderungen der EU-Richtlinien eingehalten werden.

11.2.5 Audit

Ein „Audit“ untersucht, ob Geschäftsprozesse, Anforderungen und Richtlinien die geforderten Standards erfüllen. Ein solches Untersuchungsverfahren erfolgt häufig im Rahmen eines Qualitätsmanagements. Die Audits werden von einem speziell hierfür geschulten Auditor durchgeführt.

Innerhalb des Qualitätsmanagements werden zwei Arten von Audits unterschieden: Im Bereich des Qualitätssicherung, statischen Qualitätsmanagements haben die Audits Prüfungscharakter, da sie Nachweise über vertragsmäßige Vereinbarungen liefern. Sie werden daher pro Überprüfungszyklus nur einmalig durchgeführt. In der Qualitätssicherung, dynamischen Qualitätssicherung (oder Qualitätsmanagement) kommt den Audits eine erweiterte Bedeutung zu: Sie dienen der Erfassung von Entwicklungstrends und geben den Initiatoren von Veränderungen wichtige Rückmeldungen über die Wirksamkeit ihrer eingeleiteten Maßnahmen. Die Aussagekraft dieser begleitenden Audits steigt mit der Wiederholungsrate, mit der der identische Fragenkatalog der identischen Betroffenengruppe zum identischen Thema vorgelegt wird. Vorgaben macht die „DIN/ISO 19011, Leitfaden zur Auditierung von Managementsystemen“.

11.2.6 DAkkS

Wie bereits oben erwähnt ist die die „Deutsche Akkreditierungsstelle GmbH“ („DAkkS“) zuständig für Akkreditierungen. Sie ist die nationale Akkreditierungsstelle mit Sitz in Berlin.

Gründungsgeschichte

Im Zuge der europäischen Verordnung (EG) Nr. 765/2008 (Artikel 4 Absatz 1) müssen alle EU-Mitgliedstaaten ab 1. Januar 2010 eine einzige nationale Akkreditierungsstelle benennen. Hierzu musste die Dachorganisation Deutscher Akkreditierungsrat (DAR) mit den folgenden vier Fachgesellschaften für bestimmte Gebiete zur DAkkS fusionieren:

- Deutsche Akkreditierungsstelle Chemie (DACH)
- Deutscher Kalibrierdienst (DKD)
- Deutsches Akkreditierungssystem Prüfwesen (DAP)
- Trägergemeinschaft für Akkreditierung (TGA) Über eine vorweggenommene Fusion enthält die TGA auch die Deutsche Akkreditierungsstelle Technik (DATech)

Organisation

Die DAkkS ist eine privatwirtschaftliche Organisation, die beliebte hoheitliche Aufgaben wahrnimmt.

Die GmbH-Anteilseigner der DAkkS sind jeweils zu einem Drittel:

- die Bundesländer, vertreten durch die Länder Bayern, Hamburg, Niedersachsen, Nordrhein-Westfalen und Sachsen-Anhalt.
- die Bundesrepublik Deutschland, vertretend durch das Bundesministerium für Wirtschaft und Energie
- die Deutsche Wirtschaft, vertretend durch den Bundesverband der Deutschen Industrie (BDI).

Die Bundesländer wurden primär beteiligt, um die bestehenden Organisationen der Länder leichter in die DAkkS zu überführen, „wodurch parallele Strukturen und Aktivitäten auf Landesebene verzichtbar werden“.

Kritik an der DAkkS: *In einer gemeinsamen Erklärung kritisierten Eurolab-D, der Verband der Materialprüfungsanstalten (VMPA) e.V. und der Deutsche Verband Unabhängiger Prüflaboratorien (VUP) die vom Bundeswirtschaftsministerium (BMWi) vorgelegte Reform der Gebührenverordnung für die Akkreditierungsstelle. Die Preissteigerungsrate schade dem Mittelstand und somit dem Wirtschaftsstandort Deutschland.*

Literaturverzeichnis

zu Kapitel 11:

[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.

[ISO17011] DIN EN ISO/IEC 17011 - 2018-03 Konformitätsbewertung - Anforderungen an Akkreditierungsstellen, die Konformitätsbewertungsstellen akkreditieren

[VIM08] VIM International vocabulary of metrology – Basic and general concepts and associated terms (VIM) 3rd edition (2008)

[wiki01] [https://de.wikipedia.org/wiki/Akkreditierung_\(Wirtschaft\)](https://de.wikipedia.org/wiki/Akkreditierung_(Wirtschaft))

[DAkkS] <http://www.dakks.de>

Kapitel 12

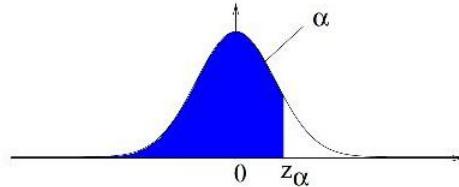
Anhang: Quantiltabellen

12.1 Quantile der Normalverteilung

Ablesebeispiel: $z_{0.95} = 1.6449$.

Erweiterung der Tafel: $z_{1-\alpha} = -z_\alpha$

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-z^2/2} dz \quad \text{Verteilungsfunktion}$$



α	z_α	α	z_α	α	z_α	α	z_α
0.9999	3.7190	0.9955	2.6121	0.975	1.9600	0.780	0.7722
0.9998	3.5401	0.9950	2.5758	0.970	1.8808	0.770	0.7388
0.9997	3.4316	0.9945	2.5427	0.965	1.8119	0.760	0.7063
0.9996	3.3528	0.9940	2.5121	0.960	1.7507	0.750	0.6745
0.9995	3.2905	0.9935	2.4838	0.955	1.6954	0.740	0.6433
0.9994	3.2389	0.9930	2.4573	0.950	1.6449	0.730	0.6128
0.9993	3.1946	0.9925	2.4324	0.945	1.5982	0.720	0.5828
0.9992	3.1559	0.9920	2.4089	0.940	1.5548	0.710	0.5534
0.9991	3.1214	0.9915	2.3867	0.935	1.5141	0.700	0.5244
0.9990	3.0902	0.9910	2.3656	0.930	1.4758	0.690	0.4959
0.9989	3.0618	0.9905	2.3455	0.925	1.4395	0.680	0.4677
0.9988	3.0357	0.9900	2.3263	0.920	1.4051	0.670	0.4399
0.9987	3.0115	0.9895	2.3080	0.915	1.3722	0.660	0.4125
0.9986	2.9889	0.9890	2.2904	0.910	1.3408	0.650	0.3853
0.9985	2.9677	0.9885	2.2734	0.905	1.3106	0.640	0.3585
0.9984	2.9478	0.9880	2.2571	0.900	1.2816	0.630	0.3319
0.9983	2.9290	0.9875	2.2414	0.895	1.2536	0.620	0.3055
0.9982	2.9112	0.9870	2.2262	0.890	1.2265	0.610	0.2793
0.9981	2.8943	0.9865	2.2115	0.885	1.2004	0.600	0.2533
0.9980	2.8782	0.9860	2.1973	0.880	1.1750	0.590	0.2275
0.9979	2.8627	0.9855	2.1835	0.875	1.1503	0.580	0.2019
0.9978	2.8480	0.9850	2.1701	0.870	1.1264	0.570	0.1764
0.9977	2.8338	0.9845	2.1571	0.865	1.1031	0.560	0.1510
0.9976	2.8202	0.9840	2.1444	0.860	1.0803	0.550	0.1257
0.9975	2.8070	0.9835	2.1321	0.855	1.0581	0.540	0.1004
0.9974	2.7944	0.9830	2.1201	0.850	1.0364	0.530	0.0753
0.9973	2.7821	0.9825	2.1084	0.845	1.0152	0.520	0.0502
0.9972	2.7703	0.9820	2.0969	0.840	0.9945	0.510	0.0251
0.9971	2.7589	0.9815	2.0858	0.835	0.9741	0.500	0.0000
0.9970	2.7478	0.9810	2.0749	0.830	0.9542		
0.9969	2.7370	0.9805	2.0642	0.825	0.9346		
0.9968	2.7266	0.9800	2.0537	0.820	0.9154		
0.9967	2.7164	0.9795	2.0435	0.815	0.8965		
0.9966	2.7065	0.9790	2.0335	0.810	0.8779		
0.9965	2.6968	0.9785	2.0237	0.805	0.8596		
0.9964	2.6874	0.9780	2.0141	0.800	0.8416		
0.9963	2.6783	0.9775	2.0047	0.795	0.8239		
0.9962	2.6693	0.9770	1.9954	0.790	0.8064		
0.9961	2.6606	0.9765	1.9863	0.785	0.7892		
0.9960	2.6521	0.9760	1.9774	0.780	0.7722		

12.2 Quantile der 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

12.3 Quantile der χ^2 -Verteilung

ν	Wahrscheinlichkeit				
	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

ν	Wahrscheinlichkeit				
	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 13

Lösungen zu den Aufgaben

13.1 Lösungen zu den Aufgaben aus Vorl 2

13.1.1 Lösung zur 1. Aufgabe: Lineare Regression

siehe 2.8.1

zu (a) Stellen Sie das lineare Gleichungssystem auf, das sich durch partielle 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

$$z_j = a x_j + c + h \delta_{j \in C} + \varepsilon_j \quad (13.1)$$

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 (13.1) wie folgt aus

$$\mathbf{z} = c \mathbf{x}_0 + h \mathbf{x}_1 + a \mathbf{x}_2 + \varepsilon$$

nun schreiben wir alle Spaltenvektoren mit x in eine gemeinsame Matrix, die dann 3 Spalten hat und $J_T = 21$ Zeilen

$$\mathbf{X} = (\mathbf{x}_0 \ \mathbf{x}_1 \ \mathbf{x}_2) \quad (13.2)$$

so dass die Modellgleichung wie folgt aussieht

$$\boldsymbol{\varepsilon} = \mathbf{z} - \mathbf{X} \begin{pmatrix} c \\ h \\ a \end{pmatrix}.$$

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. (13.2) aufstellen zu können und Gl. (2.78) aus dieser Vorlesung zu kennen:

$$(\mathbf{X}^\top \mathbf{X}) \begin{pmatrix} c \\ h \\ a \end{pmatrix} = \mathbf{X}^\top \mathbf{z} \quad (13.3)$$

wobei \mathbf{X}^\top 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.

$$d = \frac{h}{\sqrt{1+a^2}} \quad (13.4)$$

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.

$$\sigma_\varepsilon^2 = \frac{1}{J_T - 3} \boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon}$$

zu (d) Schreiben Sie die Formel für die Kovarianzmatrix der Modellparameter a, c, h auf.

Wir verwenden dazu Gl. (2.67) dieser Vorlesung

$$\boldsymbol{\Sigma} = (\mathbf{X}^\top \mathbf{X})^{-1} \sigma_\varepsilon^2 \quad (13.5)$$

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 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_c,*indx_r,*ipiv;
int i,icol,irow,j,k,l,ll;
double big,tmp,pivinv;

// The integer arrays ipiv, indx_r, and indx_c are
// used for bookkeeping on the pivoting.
indx_c = (int*)calloc( n, sizeof(int));
indx_r = (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;

```

```

indx[icol]=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 (indx[1] != indx[1])
for (k=0; k<n; k++) SWAP(a[indx[1]][k],a[indx[1]][k]);
}
// And we are done.
free(ipiv);
free(indxr);
free(indxc);
}

```

13.2 Hilfe zu den Übungsaufgaben 5.9 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 $H_0: \mu_1 = \mu_2$

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2 + s_2^2}}$$

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

13.3 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:

$$d = (4.997 \pm 0.011) \mu\text{m} \quad \text{mit} \quad k = 2.1, \nu_d = 26 \quad (13.6)$$

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

$$d_W = \frac{d}{U_B} U_W \quad (13.7)$$

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.

$$c_B = \left. \frac{\partial}{\partial U_B} d_W \right|_{\bar{U}_B, \bar{d}, \bar{U}_W} \quad (13.8)$$

Die Sensitivitäten c_d und c_w von d_W bezüglich Änderungen von d und U_W sind aufgrund des linearen Zusammenhangs genau

$$c_d = \left. \frac{U_W}{U_B} \right|_{\bar{U}_B, \bar{U}_W} \quad c_w = \left. \frac{d}{U_B} \right|_{\bar{U}_B, \bar{d}} \quad (13.9)$$

definiert. Die kombinierte Varianz unter der Voraussetzung, dass die gemessenen Größen und die Angabe aus dem Kalbrierschein unkorreliert sind, ist

$$s_{dW}^2 = (c_d s_d)^2 + (c_B s_B)^2 + (c_W s_W)^2 \quad (13.10)$$

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 Kalbrierschein.

- a) Ermitteln Sie aus den Angaben des Kalbierscheins (hier Gl. (13.6)) 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.

$$c_d = \frac{\partial d_W}{\partial d} = \frac{U_W}{U_B} = 0.9306$$

$$c_W = \frac{\partial d_W}{\partial U_W} = \frac{d}{U_B} = 0.0257 \frac{\mu\text{m}}{\text{mV}}$$

$$c_B = \frac{\partial d_W}{\partial U_B} = -\frac{dU_W}{U_B^2} = -0.0239 \mu\text{m}$$

- 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 CCCR Key Comparison Report Preparation,
<https://www.bipm.org/utils/common/pdf/CC/CCPR/CCPR-G2.pdf>
- [ISO17011] ISO/IEC 17011:2005
- [Pau82] R.C. Paule, J. Mandel: Consensus Values and Weighting Factors, *J. Res. NBS*, Vol.87, No.5, 377-385 (1982)