

Michaela Unterweger, BSc

Faktorielle Versuchspläne mit mehreren Zielgrößen

MASTERARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieurin

Masterstudium Finanz- und Versicherungsmathematik

eingereicht an der

Technischen Universität Graz

Betreuer:

Univ.-Prof. Dipl.-Ing. Dr.techn. Ernst Stadlober

Institut für Statistik

Graz, März 2015

EIDESSTATTLICHE ERKLÄRUNG

AFFIDAVIT

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Das in TUGRAZonline hochgeladene Textdokument ist mit der vorliegenden Masterarbeit identisch.

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis.

Datum/Date

Unterschrift/Signature

DANKSAGUNG

Ich möchte mich bei allen bedanken, die mich bei der Erstellung dieser Arbeit unterstützt haben. Besonderen Dank gebührt Herrn Professor Stadlober für die Betreuung der Arbeit. Er hatte immer ehestmöglich einen Termin für mich frei und half mir bei auftretenden Schwierigkeiten.

Ein großes Dankeschön gilt meinen Freundinnen Judith und Katrin, die mir mit hilfreichen L^AT_EX-Tipps zur Seite standen und meine Arbeit Korrektur gelesen haben.

Weiters möchte ich meinen Eltern für die finanzielle und seelische Unterstützung während der ganzen Studienzzeit danken und meiner Schwester, die immer für mich da war.

ZUSAMMENFASSUNG

Wir beschäftigen uns mit der Analyse und Auswertung von faktoriellen Versuchsplänen. Zu Beginn wird ein Einblick in die Methodik der Regressionsflächen gegeben. Es werden drei Methoden vorgestellt, mit welchen die simultane robuste Optimierung mehrerer Response-Variablen möglich ist. Bei zwei dieser Methoden sind die Response-Variablen unterschiedlich gewichtet. Diese Verfahren werden mit Daten eines Ätzprozesses getestet und miteinander verglichen. Dabei kann der optimale Punkt vom robusten optimalen Punkt abweichen. Unterschiede ergeben sich auch, wenn man die Response-Variablen verschieden gewichtet.

Eine der Response-Variablen ist der Prozessfähigkeitsindex C_p . Als erstes wird der klassische Ansatz, bei welchem man normalverteilte Daten voraussetzt, beschrieben. Dann werden Möglichkeiten vorgestellt, wie man die Prozessfähigkeit bei nicht normalverteilten Populationen bewerten kann. Alle Methoden werden an simulierten, nicht normalverteilten Daten getestet. Für die Daten des Ätzprozesses werden passende Verteilungen gesucht und daraus die Prozessfähigkeit berechnet. Die Abweichungen der Methoden für nicht normalverteilte Populationen von der Standardmethode werden bei steigender Schiefe größer.

Abstract

We consider the analysis and interpretation of factorial designs. A short introduction into the response surface methodology is given. Three methods are examined which optimize several response variables simultaneously and robustly. Two of these methods permit assigning different weights to each response during the optimization process. The procedures are tested on real data from an etching process and the results are compared. It turns out that the robust optimal point may differ from the optimal point. Additionally, the optimal and the robust optimal point can change using different weights.

The process capability index C_p is one of the responses. At first the standard approach assuming normal distribution is explained. Then several possibilities to deal with non-normal population are covered. All methods are tested on simulated, non-normally distributed data. We propose distributions which adequately fit the data from the etching process and process capability indices are calculated using these distributions. Differences between the estimated capability indices for non-normal populations and the estimated capability indices for normal populations become larger as skewness increases.

Inhaltsverzeichnis

1. Einleitung	1
1.1. Faktorielle Versuchspläne	1
1.2. Inhalt	2
1.3. Notation	2
2. Regressionsflächen Methodik	3
2.1. Konturplot	3
2.2. Approximation der Response-Funktion	4
2.3. Drei Phasen in RSM	5
2.4. Methode des steilsten Anstiegs	6
2.4.1. Bestimmung der Punkte auf dem Pfad des steilsten Anstiegs .	7
2.5. Analyse der Regressionsfläche 2. Ordnung	7
2.5.1. Bestimmung des stationären Punktes	7
2.5.2. Charakterisierung der Regressionsfläche	8
2.5.3. Ridge Systeme	9
3. Mehrere Zielgrößen	11
3.1. Das multivariate lineare Modell und die gefittete Response-Funktion .	11
3.2. Überlappender Konturplot	12
3.3. Optimierungsproblem	14
3.4. Desirability-Funktion	14
3.4.1. Einseitige Transformation	15
3.4.2. Zweiseitige Transformation	15
3.5. Dual Response	19
3.5.1. \widehat{C} definit	20
3.5.2. \widehat{C} indefinit	22
3.5.3. Weitere Nebenbedingung	24
3.6. Distanzfunktion	24
3.6.1. Lineare Abhängigkeiten zwischen den Response-Variablen . . .	25
3.6.2. Simultane Optimierung der multiplen Response-Funktion . . .	25
3.6.3. Definition der Distanzfunktion	26
3.6.4. Zufälligkeit von ϕ	26
3.6.5. Vorgangsweise	27
3.7. Verlustfunktion	27

4. Robuste Optimierung	29
4.1. RPD-Problem	29
4.1.1. Produkt-Versuchsplan	30
4.1.2. Analyse des Produkt-Versuchsplans	32
4.1.3. Kombiniertes Versuchsplan	33
4.2. Robustes E-Modell	34
4.2.1. Formulierung des Problems	34
4.2.2. Robustes E-Modell	36
4.3. Robuste Optimierung mit einem Konfidenzbereich	37
4.3.1. Methode	37
4.3.2. Optimierung	39
4.4. Robuste Optimierung mit Desirability-Funktionen	39
4.4.1. Robustheitsmaß	39
4.4.2. Optimierungs-Desirability-Funktion	41
4.4.3. Vorgangsweise	41
5. Anwendung	43
5.1. Robustes E-Modell	46
5.2. Robuste Optimierung mit einem Konfidenzbereich	47
5.3. Robuste Optimierung mit Desirability-Funktionen	49
6. Prozessfähigkeit	51
6.1. Normalverteilte und nicht normalverteilte Daten	51
6.1.1. Normalverteilung und der 6σ -Bereich	51
6.1.2. Ansatz für nicht normalverteilte Populationen	52
6.2. C_p -Index	52
6.2.1. Der C_p -Index und der Prozentsatz des fehlerhaften Materials	53
6.2.2. Wann wird ein Prozess als „fähig“ bezeichnet?	54
6.2.3. Schätzer für den C_p -Index	54
6.2.4. Konfidenzintervall für den C_p -Index	55
6.2.5. Hypothesentests	55
6.2.6. Schwächen des C_p -Index	56
6.3. C_{pk} -Index	57
6.3.1. Messen der Prozessfähigkeit mithilfe des C_{pk} -Index	58
6.3.2. Schätzer für den C_{pk} -Index	59
6.3.3. Konfidenzintervall für den C_{pk} -Index	59
6.3.4. Hypothesentests	60
6.3.5. Schwächen des C_{pk} -Index	60
6.4. Der k -Index	61
6.4.1. Schätzer für den k -Index	61
6.5. Weitere Indizes	61
6.6. Nicht normalverteilte Populationen	63
6.6.1. Datentransformation und graphische Methoden	63
6.6.2. Quantilschätzung für nicht normalverteilte Populationen	64
6.6.3. Munechika's Annäherung	66

6.6.4.	S_{pmk} - Index	68
6.6.5.	Konfidenzintervalle für nicht normalverteilte Populationen . .	70
6.7.	Simulierte Zufallsvariablen	71
6.7.1.	Gammaverteilung	72
6.7.2.	Lognormalverteilung	76
6.7.3.	Weibullverteilung	79
6.8.	Datenbeispiel	82
6.8.1.	Fitten der Verteilung	82
6.8.2.	Ergebnisse	94
A.	Anhang	95
A.1.	Kanonische Reduktion von Matrizen	95
A.2.	Die Varianz von λ^0	96
A.3.	Entfernen von linear abhängigen Response-Variablen	98
A.4.	Simplex-Suchverfahren	98
A.5.	R-Routinen	100
A.5.1.	Kapitel 5	100
A.5.2.	Kapitel 6	104

Abbildungsverzeichnis

2.1. Konturplot - $y = \text{Viskosität}$ in Abhängigkeit von $\xi_1 = \text{Reaktionszeit}$ und $\xi_2 = \text{Temperatur}$	4
3.1. Konturplot - $y_1 = \text{Viskosität}$ (grün) und $y_2 = \text{Molekulargewicht}$ (schwarz) in Abhängigkeit von $\xi_1 = \text{Reaktionszeit}$ und $\xi_2 = \text{Temperatur}$	13
3.2. Desirability-Funktion - Einseitige Transformation	17
3.3. Desirability-Funktion - Zweiseitige Transformation ($c_i = 4$)	18
4.1. Signifikante Kontroll \times Stör Interaktionen	31
4.2. Nicht signifikante Kontroll \times Stör Interaktionen	32
4.3. Grafische Darstellung der sechs Fälle	41
6.1. Der C_p -Index vergleicht den tatsächlichen mit dem erlaubten Bereich, vgl. Kotz & Lovelace (1998)	53
6.2. C_{pk} -Index, vgl. Kotz & Lovelace (1998)	58
6.3. Histogramm - Bottom	83
6.4. Skewness-Kurtosis Plot - Bottom	84
6.5. Verteilungsfunktion - Bottom	84
6.6. Dichtefunktion - Bottom	85
6.7. PP-Plot - Bottom	85
6.8. QQ-Plot - Bottom	86
6.9. Plot der gefitteten Weibullverteilung	87
6.10. Histogramm - Top	88
6.11. Histogramm - Top (o.A.)	89
6.12. Skewness-Kurtosis Plot - Top	89
6.13. Plot der gefitteten Normalverteilung für X^2	90
6.14. Verteilungsfunktion - Top	91
6.15. Dichtefunktion - Top	91
6.16. PP-Plot - Top	92
6.17. QQ-Plot - Top	92
6.18. Plot der gefitteten Gompertzverteilung	93

Tabellenverzeichnis

3.1. Überblick für \widehat{C} definit	22
3.2. Überblick für \widehat{C} indefinit	24
4.1. Beispiel eines Crossed-Array Versuchsplans	30
4.2. Beispiel eines Crossed-Array Versuchsplans	32
5.1. Ergebnisse des Experiments	43
5.2. Individuelle Optima	45
5.3. Robustes E-Modell	46
5.4. Robuste Optimierung mit Konfidenzbereich	48
5.5. Konfidenzbereiche der robusten Lösungen	49
5.6. Robuste Optimierung mit Desirability-Funktionen	49
6.1. Inverse Verteilungsfunktionen	66
6.2. Gammaverteilung - Beispiel 1	73
6.3. Konfidenzintervalle Gammaverteilung - Beispiel 1	73
6.4. Gammaverteilung - Beispiel 2	74
6.5. Konfidenzintervalle Gammaverteilung - Beispiel 2	74
6.6. Gammaverteilung - Beispiel 3	75
6.7. Konfidenzintervalle Gammaverteilung - Beispiel 3	75
6.8. Lognormalverteilung - Beispiel 1	76
6.9. Konfidenzintervalle Lognormalverteilung - Beispiel 1	76
6.10. Lognormalverteilung - Beispiel 2	77
6.11. Konfidenzintervalle Lognormalverteilung - Beispiel 2	77
6.12. Lognormalverteilung - Beispiel 3	78
6.13. Konfidenzintervalle Lognormalverteilung - Beispiel 3	78
6.14. Weibullverteilung - Beispiel 1	79
6.15. Konfidenzintervalle Weibullverteilung - Beispiel 1	79
6.16. Weibullverteilung - Beispiel 2	80
6.17. Konfidenzintervalle Weibullverteilung - Beispiel 2	80
6.18. Weibullverteilung - Beispiel 3	81
6.19. Konfidenzintervalle Weibullverteilung - Beispiel 3	81
6.20. Wichtige Kenngrößen	82
6.21. Goodness of Fit Tests - Bottom	83
6.22. Vergleich der Prozessfähigkeitsindizes - Bottom	86
6.23. Goodness of Fit Tests - Top	88
6.24. Vergleich der Prozessfähigkeitsindizes - Top	94
6.25. Vergleich der C_p -Werte für Versuch 4 Wd. 1	94

1. Einleitung

Aufbauend auf Montgomery (2005) gibt es zu Beginn eine kurze Einleitung zu den faktoriellen Versuchsplänen. Anschließend wird der Inhalt der Arbeit und die Notation kurz beschrieben.

1.1. Faktorielle Versuchspläne

Faktorielle Versuchspläne werden verwendet um die Auswirkungen von mehreren Faktor auf eine Response-Variable y zu untersuchen. Im einfachsten Fall setzt man für jeden Faktor eine niedrige und eine hohe Stufe fest. In einem faktoriellen Versuchsplan werden Versuche mit allen möglichen Einstellungen dieser Faktoren in zufälliger Reihenfolge durchgeführt. Sind k Variablen vorhanden, welche jeweils auf zwei verschiedene Stufen gestellt werden können, besteht der Versuchsplan aus 2^k Experimenten mit oder ohne Wiederholung. Für den einzelnen Faktor vergleicht man dann alle Versuche, bei welchen dieser auf der niedrigen Stufe eingestellt ist, mit allen Versuchen, bei welchen er auf der hohen Stufe eingestellt ist. Dabei kann man herausfinden, welcher Effekt auftritt bzw. wie sich die Response-Variable verändert, wenn man den Faktor von der niedrigen auf die hohe Stufe stellt. Dieser Effekt wird Haupteffekt genannt. Es ist allerdings möglich, dass der Effekt eines Faktors auf den unterschiedlichen Stufen eines anderen Faktors verschieden ist. In diesem Fall liegt eine Interaktion zwischen diesen beiden vor. Auch Interaktionen zwischen mehr als zwei Faktoren können auftreten. Diese werden Mehrfach-Interaktionen genannt. Oft werden in der Praxis faktorielle Versuchspläne durchgeführt um Aufschluss über diese Zweifach- und Mehrfach-Interaktionen zu bekommen.

Wie bereits erwähnt, betrachtet man im einfachsten Fall die Faktoren auf einer niedrigen und einer hohen Stufe. Mit diesem 2^k -Versuchsplan kann man den linearen Zusammenhang zwischen der Response-Variable und den Faktoren modellieren. Sollte es aber Grund zur Annahme geben, dass ein quadratischer Zusammenhang zwischen der Response-Variable und den Faktoren besteht, kann man die Faktoren auch auf drei Stufen untersuchen. Die Anzahl der Experimente vergrößert sich dann von 2^k auf 3^k . Wird die Anzahl der Versuche so groß, dass es zu teuer bzw. zu aufwändig wird, diesen Versuchsplan durchzuführen, hat man die Möglichkeit mit fraktionierten faktoriellen Versuchsplänen zu arbeiten. Die Anzahl der Versuche wird bei diesen Versuchsplänen mit Hilfe von Aliasbildung verringert. Eine weitere Möglichkeit quadratische Zusammenhänge zu modellieren, ohne dass die Anzahl der Versuche zu groß wird, ist das Hinzufügen von Zentrums- und Axialpunkten zum 2^k -Versuchsplan. Man geht dabei allerdings davon aus, dass die Faktoren quantitativ sind. Weitere Informationen zu fraktionierten Versuchsplänen und zum Hinzufügen von Zentrums- und

Axialpunkten findet man bei Montgomery (2005).

Um die Beobachtungen eines faktoriellen Experiments zu analysieren, kann man diese mit einem Modell beschreiben. Montgomery (2005) schlägt drei verschiedene Modelle vor: das Effekt-, das Mittelwert- und das Regressionsmodell. Wenn ein oder mehrere Faktoren quantitativ sind, ist es sinnvoll mit letzterem zu arbeiten. In dieser Arbeit werden wir die Beobachtungen immer mit einem Regressionsmodell beschreiben.

1.2. Inhalt

In Kapitel 2 beschäftigen wir uns mit der Regressionsflächen Methodik. Es wird erklärt, wie man die Beziehung zwischen der Response-Variable und den Inputvariablen mit einem Regressionsmodell beschreibt, wie man einen stationären Punkt findet und wie man diesen im Weiteren näher untersucht.

Da es in der Praxis oft vorkommt, dass man mehrere Zielgrößen optimieren will, diese aber für unterschiedliche Einstellungen optimal werden, beschäftigt sich Kapitel 3 mit der simultanen Optimierung mehrerer Zielgrößen. Nach der Formulierung des multivariaten Problems betrachten wir Methoden, welche mehrere Zielgrößen simultan optimieren.

Anschließend wird in Kapitel 4 die robuste Optimierung vorgestellt. Nach einer Erklärung des Problems wird der Produkt-Versuchsplan und der kombinierte Versuchsplan beschrieben. Im Anschluss werden drei Methoden angeführt, mit welchen man ein Problem mit mehreren Zielgrößen robust optimieren kann. Diese Methoden werden in Kapitel 5 anhand eines Datenbeispiels diskutiert.

Eine Zielgröße dieses Datensatzes ist der C_p -Wert. Aus diesem Grund beschäftigt sich Kapitel 6 mit Prozessfähigkeitsindizes. Es wird dabei näher auf den Unterschied zwischen normalverteilten und nicht normalverteilten Populationen eingegangen.

Alle Auswertungen, die in dieser Arbeit präsentiert werden, sind im Statistikprogramm R 3.1.2 umgesetzt worden.

1.3. Notation

In dieser Arbeit werden Matrizen und Vektoren mit fett gedruckten Buchstaben bezeichnet. Die Vektoren werden immer als Spaltenvektoren und die n -dimensionale Einheitsmatrix wird durch \mathbf{I}_n angegeben. Mit $\mathbf{1}_n$ bezeichnen wir einen n -dimensionalen Vektor, der Einsen enthält und die Schätzer von Parametern werden mit einem Dach gekennzeichnet. Das Zeichen \otimes steht für das Kronecker Produkt von Matrizen. Für Zufallsvariablen werden Großbuchstaben verwendet. Die Abkürzung u.d.N. bei Optimierungsproblemen steht für „unter der/den Nebenbedingung/-en“.

2. Regressionsflächen Methodik

Dieses Kapitel basiert auf den Beschreibungen von Montgomery (2005) und Myers & Montgomery (2002). Unter Regressionsflächen Methodik (Response Surface Methodology, kurz: RSM) versteht man mathematische und statistische Methoden um einen Prozess zu analysieren und zu modellieren. Dieser Prozess besteht aus Response-Variablen und mehreren Inputvariablen. Man will eine Beziehung zwischen der Response-Variable und den Inputvariablen, die oft auch unabhängige Variablen genannt werden, herstellen und diese analysieren. Dies geschieht durch ein Regressionsmodell, grafisch dargestellt durch eine Regressionsfläche (Response Surface).

Am häufigsten wird RSM in der Industrie angewendet, wenn ein Leistungs- oder Qualitätsmaß eines Prozesses oder eines Produkts möglicherweise von einigen Inputvariablen beeinflusst wird. Die Inputvariablen werden dann vom Ingenieur oder Wissenschaftler kontrolliert bzw. für Testzwecke verändert. RSM wird sowohl bei der Entwicklung und Herstellung neuer Produkte oder Prozesse eingesetzt, als auch für die Verbesserung bereits verwendeter Produkte und Prozesse. Man kann mit Hilfe von RSM herausfinden, welche Parametereinstellungen man wählen muss um die Response-Variable zu optimieren oder auch um gewisse Kundenanforderungen zu erfüllen.

Die Namensgebung geht auf die grafische Ansicht des Problems zurück. Sei Y die Response-Variable und ξ_1 und ξ_2 zwei Inputvariablen. Für jeden Wert von ξ_1 und ξ_2 gibt es einen entsprechenden Wert der Response-Variable. Diese Werte liegen wie eine Oberfläche auf der ξ_1 - ξ_2 Fläche.

2.1. Konturplot

Ein wichtiges grafisches Hilfsmittel bei der Arbeit mit RSM ist der Konturplot. Man sieht sich die ξ_1 - ξ_2 Fläche an und verbindet alle Punkte, die den gleichen Wert der Response-Variable haben. Diese Linien von konstanten Werten der Response-Variable werden Konturlinien genannt.

In nachfolgender Abbildung sieht man einen Konturplot eines Beispiels aus Myers & Montgomery (2002), das Werte von einem chemischen Prozess analysiert. Die Response-Variable $y = \text{Viskosität}$ hängt dabei von den zwei unabhängigen Variablen $\xi_1 = \text{Reaktionszeit}$ und $\xi_2 = \text{Temperatur}$ ab. Untersucht man einen Prozess mit mehr als zwei Inputvariablen, kann man einen Konturplot erstellen, indem man alle anderen Inputvariablen konstant hält, d.h. bei k Inputvariablen werden $k - 2$ Variablen konstant gehalten.

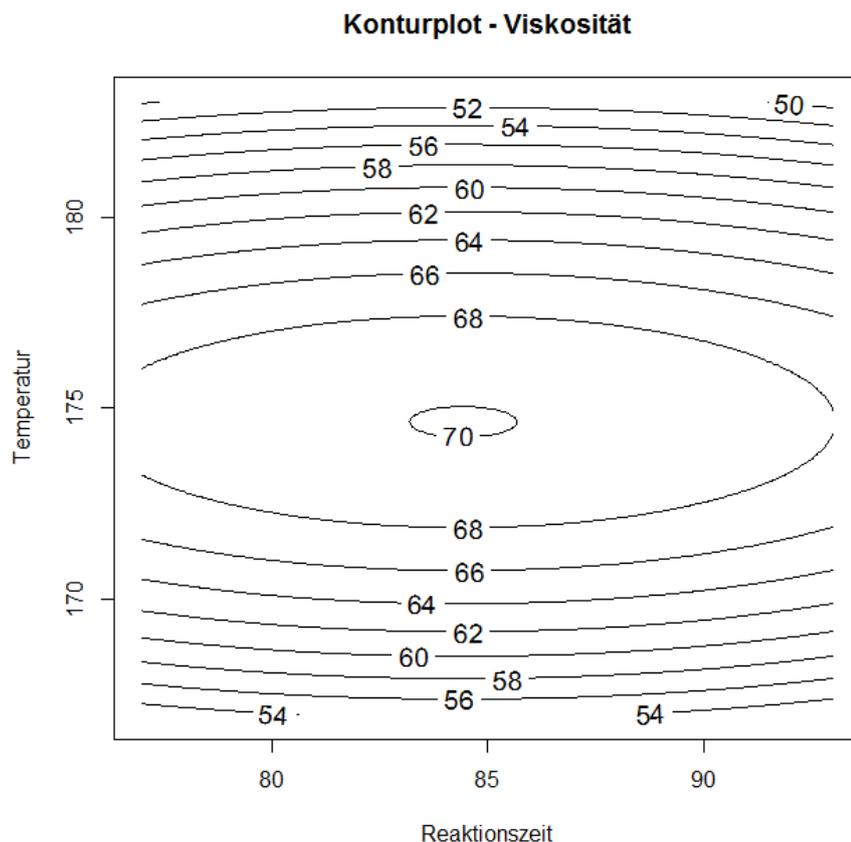


Abbildung 2.1.: Konturplot - $y = \text{Viskosität}$ in Abhängigkeit von $\xi_1 = \text{Reaktionszeit}$ und $\xi_2 = \text{Temperatur}$

2.2. Approximation der Response-Funktion

In vielen Problemstellungen ist die Beziehung der Response-Variable zu den Inputvariablen unbekannt. Es ist daher sehr wichtig eine angemessene Verbindung zwischen der Response-Variable und den Prozessvariablen herzustellen. Die Response-Variable wird durch

$$Y = f(\xi_1, \xi_2, \dots, \xi_k) + \epsilon$$

dargestellt, wobei f eine unbekannte Funktion ist und ϵ ein Fehlerterm, der die Variabilität beschreibt, die mit der Funktion f nicht beschrieben werden kann. Oft wird angenommen, dass $\epsilon \sim N(0, \sigma^2)$.

Unter der Annahme, dass ϵ Erwartungswert 0 hat, erhält man

$$\begin{aligned} E(Y) &\equiv \mu = E[(f(\xi_1, \xi_2, \dots, \xi_k))] + E(\epsilon) \\ &= f(\xi_1, \xi_2, \dots, \xi_k). \end{aligned}$$

Die Variablen $\xi_1, \xi_2, \dots, \xi_k$ sind in ihren natürlichen Maßeinheiten gegeben und wer-

den daher auch natürliche Variablen genannt. Oft ist es praktisch, diese in sogenannte kodierte Variablen x_1, x_2, \dots, x_k umzuwandeln. Die kodierte Variablen befinden sich üblicherweise im Intervall $[-1, 1]$ und sind so definiert, dass sie alle den gleichen Abstand zueinander haben. Der Erwartungswert μ wird durch die Response-Funktion

$$\mu = f(x_1, x_2, \dots, x_k)$$

beschrieben.

Eine wichtige Aufgabe ist diese unbekannt Funktion f passend zu approximieren. In den häufigsten Fällen wird dafür ein Modell 1. Ordnung oder ein Modell 2. Ordnung verwendet. Wenn die Response-Variable mit einer linearen Funktion ausreichend genau modelliert werden kann, verwendet man zur Approximation ein Modell 1. Ordnung

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon.$$

Befindet sich aber eine Krümmung im Prozess, benötigen wir eine Modell höherer Ordnung. Oft wird ein Modell 2. Ordnung

$$Y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ii} x_i^2 + \sum_{i < j} \beta_{ij} x_i x_j + \epsilon$$

verwendet. Natürlich kann man mit einem polynomialen Modell den wahren und meistens auch sehr komplizierten Zusammenhang zwischen der Response-Variable und den Inputvariablen nie auf dem ganzen Bereich der Inputvariablen richtig modellieren. Auf einem kleinen Bereich funktioniert diese Annäherung allerdings gut.

Um die Parameter dieser Modelle zu schätzen, wird die Methode der kleinsten Quadrate verwendet. Die Response Surface-Analyse wird dann auf diese gefittete Funktion angewendet. Kann die gefittete Funktion den wahren Zusammenhang angemessen wiedergeben, ist die Analyse der gefitteten Funktion approximativ äquivalent zur Analyse des Problems. Die Qualität der Analyse hängt also stark davon ab, wie gut die Approximation der Funktion f ist.

2.3. Drei Phasen in RSM

In einem ersten Schritt wird ein Screening-Versuch durchgeführt. Dieser dient dazu sehr viele allgemeine Informationen über den Prozess zu sammeln, z.B. welche Inputvariablen auf die Response-Variable einen Einfluss haben und welche einen sehr geringen bzw. gar keinen Einfluss haben. Letztere können dann in der zweiten Phase weggelassen werden. Hierbei reicht meistens ein Modell niedriger Ordnung aus.

In der zweiten Phase will man herausfinden, ob die aktuelle Einstellung der Inputvariablen zu einem Response-Wert führt, der in der Nähe des wahren Optimums liegt. Es ist nämlich durchaus möglich, dass der Prozess in einem Bereich arbeitet, der weit entfernt von diesem Optimum ist. In dieser Phase wird oft die Methode des steilsten Anstiegs angewendet, die den Prozess nach und nach zum wahren Optimum führt.

Wenn der Prozess dann in der Nähe des wahren Optimums ist, beginnt die dritte Phase. Nun wird die wahre Response Surface mit einem Modell höherer Ordnung genauer analysiert. Wenn eine passende Approximation gefunden worden ist, kann man damit die optimale Einstellung für den Prozess finden.

2.4. Methode des steilsten Anstiegs

In vielen Experimenten wird der Arbeitsbereich aufgrund einer bestimmten Vermutung gewählt. Diese ist zwar meistens gut begründet, dennoch kann es vorkommen, dass die erste Schätzung des Bereichs für das Optimum sehr weit vom wahren Optimum entfernt ist. Um jetzt schrittweise zum wahren Optimum zu gelangen, verwendet man die Methode des steilsten Anstiegs.

Man beginnt mit der Annahme, dass ein Modell 1. Ordnung zumindest eine passende Approximation für den Prozess in der Anfangsregion x_1, x_2, \dots, x_k ist. Wenn ein gefittetes Modell 1. Ordnung

$$\hat{y} = \hat{\beta}_0 + \sum_{i=1}^k \hat{\beta}_i x_i$$

verwendet wird, erhält man für die Konturen von \hat{y} eine Reihe von parallelen Linien. Die Richtung des steilsten Anstiegs ist die Richtung, in welche sich der Wert \hat{y} am meisten vergrößert. Diese Richtung steht normal auf den Konturen von \hat{y} . Als Pfad des steilsten Anstiegs wird üblicherweise die Linie verwendet, welche durch das Zentrum des Anfangsbereiches geht und normal auf den Konturlinien steht. Hat man den Pfad des steilsten Anstiegs bestimmt, kann man entlang dieses Pfades weitere Versuche durchführen. Die Bewegung von x_j auf dem Pfad des steilsten Anstiegs ist proportional zur Größe des Regressionskoeffizienten β_j , wobei die Richtung aus dem Vorzeichen des Koeffizienten ablesbar ist.

Die Response-Werte sollten sich entlang des Pfades immer weiter verbessern. Irgendwann wird dies aber nicht mehr der Fall sein. Dies tritt ein, wenn man in einem Bereich angekommen ist, der sehr weit weg vom Anfangsbereich ist. Das verwendete Modell 1. Ordnung ist in diesem Bereich keine gute Approximation mehr. Man kann sich dann ein neues Modell 1. Ordnung wählen, und mit Hilfe dieses Modells einen neuen Pfad des steilsten Anstiegs berechnen. Diese Abfolge wird wiederholt bis man in der unmittelbaren Umgebung des Optimums angelangt ist. Erkennbar ist dies dadurch, dass hier meistens ein Modell 1. Ordnung aufgrund der Signifikanz der Terme 2. Ordnung nicht mehr passend ist.

Will man ein Minimum der Response-Variable finden, spricht man von der Methode des steilsten Abstiegs. Um die Richtung des steilsten Abstiegs ablesen zu können, muss man das Vorzeichen des Regressionskoeffizienten umdrehen.

Im Folgenden wird erklärt, wie man einen Punkt auf dem Pfad des steilsten Anstiegs bestimmt.

2.4.1. Bestimmung der Punkte auf dem Pfad des steilsten Anstiegs

Man nimmt an, dass der Ursprung bei $x_1 = x_2 = \dots = x_k = 0$ ist. Um sich die Punkte auf dem Pfad des steilsten Anstiegs zu bestimmen, geht man folgendermaßen vor.

1. Man wählt sich eine Schrittweite in einer Variable aus (Δx_j). Meistens wird dafür jene Variable gewählt, die den größten absoluten Regressionskoeffizienten hat.
2. Die Schrittweite wird in den anderen Variablen mit

$$\Delta x_i = \frac{\hat{\beta}_i}{\hat{\beta}_j} \Delta x_j$$

ausgedrückt.

3. Die Δx_i werden von den kodierten Variablen in die natürlichen Variablen umgewandelt.

2.5. Analyse der Regressionsfläche 2. Ordnung

Wenn man in der Nähe des wahren Optimums angelangt ist, wird nun ein Modell 2. Ordnung verwendet um die Response-Variable zu beschreiben. Um die Response-Variable optimieren zu können, muss man die stationären Punkte finden, sofern diese existieren. Man erhält sie, indem man die partiellen Ableitungen von \hat{y} nach allen x_i gleich 0 setzt.

Hat man einen stationären Punkt gefunden, können drei verschiedene Möglichkeiten auftreten. Der stationäre Punkt ist ein Punkt mit maximalem Response-Wert, ein Punkt mit minimalem Response-Wert oder ein Sattelpunkt.

2.5.1. Bestimmung des stationären Punktes

Wir betrachten in diesem Abschnitt ein gefittetes Response Surface-Modell 2. Ordnung in Matrixschreibweise:

$$\hat{y} = b_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\widehat{\mathbf{B}}\mathbf{x}, \tag{2.1}$$

wobei b_0 , \mathbf{b} und $\widehat{\mathbf{B}}$ die Schätzer des Intercepts, der linearen Koeffizienten und der Koeffizienten 2. Ordnung sind. \mathbf{x} und \mathbf{b} sind k -dimensionale Vektoren und $\widehat{\mathbf{B}}$ ist eine $k \times k$ symmetrische Matrix

$$\widehat{\mathbf{B}} = \begin{bmatrix} b_{11} & b_{12}/2 & \cdots & b_{1k}/2 \\ & b_{22} & \cdots & b_{2k}/2 \\ & & \ddots & \vdots \\ sym. & & & b_{kk} \end{bmatrix}.$$

Wenn man die Gleichung (2.1) nach \mathbf{x} ableitet, erhält man

$$\frac{\partial \hat{y}}{\partial \mathbf{x}} = \mathbf{b} + 2\widehat{\mathbf{B}}\mathbf{x}.$$

Diese Ableitung wird 0 gesetzt, und es ergibt sich daraus für den stationären Punkt

$$\mathbf{x}_s = -\frac{1}{2}\widehat{\mathbf{B}}^{-1}\mathbf{b}. \quad (2.2)$$

Die geschätzte Response in diesem Punkt ist

$$\begin{aligned} \hat{y}_s &= b_0 + \mathbf{x}'_s \mathbf{b} + \mathbf{x}'_s \widehat{\mathbf{B}} \mathbf{x}_s \\ &= b_0 + \frac{1}{2} \mathbf{x}'_s \mathbf{b}. \end{aligned}$$

2.5.2. Charakterisierung der Regressionsfläche

Wie bereits erwähnt, kann ein stationärer Punkt ein Punkt mit maximalem Response-Wert, ein Punkt mit minimalem Response-Wert oder ein Sattelpunkt sein. Um dies zu untersuchen, kann man für wenige Variablen sehr gut einen Konturplot verwenden. Je mehr Inputvariablen man allerdings hat, desto komplizierter wird die Analyse des Konturplots und daher greift man in diesem Fall auf die sogenannte kanonische Analyse zurück.

Welche der drei Möglichkeiten vorliegt, kann man an den Vorzeichen der Eigenwerte der Matrix $\widehat{\mathbf{B}}$ erkennen. Sei \mathbf{P} die $k \times k$ Matrix, in deren Spalten die normierten Eigenvektoren zu den Eigenwerten der Matrix $\widehat{\mathbf{B}}$ stehen. Es gilt die Gleichung

$$\mathbf{P}' \widehat{\mathbf{B}} \mathbf{P} = \mathbf{\Lambda},$$

wobei $\mathbf{\Lambda}$ die Diagonalmatrix ist, welche die Eigenwerte von $\widehat{\mathbf{B}}$ in den Diagonalelementen enthält.

Auf das Modell (2.1) wird zuerst eine Translation angewendet. Das neue Zentrum soll der stationäre Punkt sein. Wir erhalten wegen

$$\mathbf{z} = \mathbf{x} - \mathbf{x}_s$$

und (2.2) für das Modell

$$\begin{aligned} \hat{y} &= b_0 + (\mathbf{z} + \mathbf{x}_s)' \mathbf{b} + (\mathbf{z} + \mathbf{x}_s)' \widehat{\mathbf{B}} (\mathbf{z} + \mathbf{x}_s) \\ &= [b_0 + \mathbf{x}'_s \mathbf{b} + \mathbf{x}'_s \widehat{\mathbf{B}} \mathbf{x}_s] + \mathbf{z}' \mathbf{b} + \mathbf{z}' \widehat{\mathbf{B}} \mathbf{z} + 2\mathbf{x}'_s \widehat{\mathbf{B}} \mathbf{z} \\ &= \hat{y}_s + \mathbf{z}' \widehat{\mathbf{B}} \mathbf{z}. \end{aligned}$$

Darauf wendet man eine Drehung an. Man will erreichen, dass die Achsen dieses Systems parallel zur Hauptachse der gefitteten Regressionsfläche sind. Man erhält mit der Projektionsmatrix \mathbf{P} , die idempotent und symmetrisch ist, d.h.

$$\mathbf{P}' \mathbf{P} = \mathbf{P} \quad \text{und} \quad \mathbf{P}' = \mathbf{P}^{-1}$$

mit

$$\mathbf{w} = \mathbf{P}'\mathbf{z}$$

die Gleichung

$$\begin{aligned}\hat{y} &= \hat{y}_s + \mathbf{w}'\mathbf{P}'\widehat{\mathbf{B}}\mathbf{P}\mathbf{w} \\ &= \hat{y}_s + \mathbf{w}'\mathbf{\Lambda}\mathbf{w}.\end{aligned}$$

Das gefittete Modell ist durch

$$\hat{y} = \hat{y}_s + \sum_{i=1}^k \lambda_i w_i^2 \quad (2.3)$$

gegeben, wobei $\{w_i\}$ die unabhängigen transformierten Variablen sind, die auch kanonische Variablen genannt werden, und $\{\lambda_i\}$ die Eigenwerte der Matrix $\widehat{\mathbf{B}}$ sind. Die obige Gleichung wird als kanonische Form des gefitteten Modells bezeichnet. Aus der Gleichung (2.3) kann man sehr leicht ablesen, welche Vorzeichen die Eigenwerte haben und daher, welcher der drei Fälle vorliegt.

1. Sind alle Eigenwerte negativ, ist der stationäre Punkt ein Punkt mit maximalem Response-Wert.
2. Sind alle Eigenwerte positiv, ist der stationäre Punkt ein Punkt mit minimalem Response-Wert.
3. Haben die Eigenwerte unterschiedliche Vorzeichen, liegt ein Sattelpunkt vor.

2.5.3. Ridge Systeme

Man betrachtet erneut die kanonische Form des Modells 2. Ordnung

$$\hat{y} = \hat{y}_s + \lambda_1 w_1^2 + \lambda_2 w_2^2 + \cdots + \lambda_k w_k^2.$$

Sei jetzt eines (oder mehrere) der λ_i sehr klein (d.h. $\lambda_i \simeq 0$), dann unterscheidet man zwei Fälle.

1. Der Punkt \mathbf{x}_s befindet sich im betrachteten Bereich: In diesem Fall ändert sich die Response-Variable fast gar nicht in z_i -Richtung. Es handelt sich hierbei um ein stationäres Ridge-System. Das Optimum kann irgendwo entlang der z_i -Gerade durch den Ursprung angenommen werden. Sind alle anderen λ_j negativ, handelt es sich um ein Maximum und sind alle anderen λ_j positiv, handelt es sich um ein Minimum. In diesem Fall hat man sehr viel Spielraum bei der Wahl der Einstellung der Inputvariablen. (Ist $\lambda_i = 0$ sind die Konturen parallele Linien.)
2. Der Punkt \mathbf{x}_s befindet sich außerhalb des betrachteten Bereichs: Diese Art der Regressionsfläche nennt man steigendes oder fallendes Ridge-System. Der Response-Wert \hat{y} steigt oder fällt langsam in z_i -Richtung. Hier kann man keine

Schlussfolgerungen für die wahre Regressionsfläche ziehen, da das Modell außerhalb des betrachteten Bereichs nicht passend ist. Sind alle anderen λ_j positiv, handelt es sich um ein fallendes Ridge-System, sind alle anderen λ_j negativ, handelt es sich um ein steigendes Ridge-System. Tritt dieser Fall ein, ist es meistens ein Zeichen dafür, dass der Bereich des Experiments nicht adäquat gewählt wurde.

3. Mehrere Zielgrößen

Dieses Kapitel ist aufbauend auf Montgomery (2005) und Myers & Montgomery (2002) verfasst. Üblicherweise wird in der Versuchsplanung mit einer Response-Variable gearbeitet. In der Praxis muss man aber meistens auf mehrere Zielgrößen Rücksicht nehmen. Bei der gleichzeitigen Betrachtung von mehreren Zielgrößen kann es natürlich möglich sein, dass diese für unterschiedliche Parametereinstellungen optimal werden. In solchen Fällen müssen Kompromisse eingegangen werden, um zwar nicht zu einem optimalen, aber wenigstens zu einem akzeptablen Response-Wert zu kommen. Zunächst wird für jede einzelne Response-Variable ein passendes Regressionsmodell aufgestellt. Mit diesen Modellen kann man dann eine Einstellung finden, die alle Response-Variablen optimiert bzw. zumindest in vorher festgelegte Bereiche bringt. In diesem Kapitel werden nach der Formulierung des multivariaten Modells verschiedene Methoden betrachtet, die es erlauben Probleme mit mehreren Zielgrößen zu lösen.

3.1. Das multivariate lineare Modell und die gefittete Response-Funktion

Wir definieren in diesem Abschnitt das multivariate lineare Modell (vgl. dazu beispielsweise Khuri & Conlon (1981) und Hejazi et al. (2012)). Seien x_1, x_2, \dots, x_k die Inputvariablen, sei l die Anzahl der Response-Variablen und n die Anzahl der Experimente, die durchgeführt wurden. Man erhält als univariates Regressionsmodell für die i -te Response-Variable

$$\mathbf{y}_i = \mathbf{X}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i \quad \text{für } i = 1, 2, \dots, l,$$

wobei \mathbf{y}_i der n -dimensionale Beobachtungsvektor der i -ten Response-Variable, $\boldsymbol{\beta}_i$ der Vektor der p unbekanntenen Regressionsparameter und $\boldsymbol{\epsilon}_i$ der n -dimensionale Vektor der Fehlerterme der i -ten Response-Variable ist. Die konstante $n \times p$ Regressionsmatrix \mathbf{X} hat vollen Rang $p \ll n$.

Um das multivariate Modell zu definieren, wird angenommen, dass \mathbf{X} für alle Response-Variablen gleich ist und dass jede Response-Variable im betrachteten Bereich R durch eine polynomiale Funktion von Grad s angenähert werden kann. Wir erhalten dadurch das multivariate Modell

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{3.1}$$

wobei $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_l]'$, $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_l]'$ und $\boldsymbol{\epsilon} = [\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \dots, \boldsymbol{\epsilon}_l]'$ ist. Das heißt, $\boldsymbol{\beta}$ ist eine $p \times l$ Matrix und $\boldsymbol{\epsilon}$ eine $n \times l$ Matrix, deren Zeilen voneinander unabhängig

sind und jede Zeile den Erwartungsvektor $\mathbf{0}$ und die Varianz-Kovarianz Matrix Σ hat ($\epsilon \sim \mathcal{N}_{n \times l}(\mathbf{0}, \mathbf{I}_n \otimes \Sigma)$). Der Schätzer

$$\hat{\Sigma} = \mathbf{Y}'[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Y}/(n-p) \quad (3.2)$$

ist ein unverzerrter Schätzer für Σ . Für diesen Schätzer gilt, dass $(n-p)\hat{\Sigma} \sim \mathcal{W}_l(n-p, \Sigma)$, d.h. der Schätzer multipliziert mit $(n-p)$ folgt einer Wishart-Verteilung mit $n-p$ Freiheitsgraden und Parameter Σ .

Da in den meisten Fällen die Response-Variable mit einem Modell 2. Ordnung approximiert wird, betrachten wir die gefittete i -te Response-Funktion 2. Ordnung, die gegeben ist durch

$$\hat{y}_i(\mathbf{x}) = \mathbf{z}'(\mathbf{x})\hat{\beta}_i,$$

wobei $\mathbf{x} = [x_1, x_2, \dots, x_k]'$, $\mathbf{z}'(\mathbf{x}) = (1, x_1, x_2, \dots, x_k, x_1^2, x_2^2, \dots, x_k^2, x_1x_2, x_1x_3, \dots, x_{k-1}x_k)$ ein p -dimensionaler Zeilenvektor und $\hat{\beta}_i = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_i$ der Kleinste-Quadrate-Schätzer der i -ten Response-Variable ist. Für $\hat{\beta}$ gilt unter obigen Annahmen für ϵ , dass $\hat{\beta} \sim \mathcal{N}_{p \times l}(\beta, (\mathbf{X}'\mathbf{X})^{-1} \otimes \Sigma)$. Für $i = 1, 2, \dots, l$ erhalten wir

$$\begin{aligned} \text{var}[\hat{y}_i(\mathbf{x})] &= \mathbf{z}'(\mathbf{x})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{z}(\mathbf{x})\sigma_{ii} \\ \text{cov}[\hat{y}_i(\mathbf{x}), \hat{y}_j(\mathbf{x})] &= \mathbf{z}'(\mathbf{x})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{z}(\mathbf{x})\sigma_{ij} \quad \text{für } i \neq j, \end{aligned}$$

wobei σ_{ij} das (i, j) -te Element der Matrix Σ ist. Zusammen ergibt das für $\hat{\mathbf{y}}(\mathbf{x}) = [\hat{y}_1(\mathbf{x}), \hat{y}_2(\mathbf{x}), \dots, \hat{y}_l(\mathbf{x})]'$, dass

$$\text{var}[\hat{\mathbf{y}}(\mathbf{x})] = \mathbf{z}'(\mathbf{x})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{z}(\mathbf{x})\Sigma. \quad (3.3)$$

Setzt man in Gleichung (3.3) für Σ den Schätzer $\hat{\Sigma}$ ein, so erhält man den unverzerrten Schätzer

$$\widehat{\text{var}}[\hat{\mathbf{y}}(\mathbf{x})] = \mathbf{z}'(\mathbf{x})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{z}(\mathbf{x})\hat{\Sigma}. \quad (3.4)$$

3.2. Überlappender Konturplot

Sollte es nur wenige Inputvariablen geben, kann man auch im multivariaten Fall auf die grafische Methode der Konturplots zurückgreifen. Man legt die Konturplots der verschiedenen Response-Variablen übereinander und erhält einen sogenannten überlappenden Konturplot. Vergleichbar mit dem Fall einer Response-Variable wird diese Methode aber für mehrere Inputvariablen sehr kompliziert.

In Abbildung 3.1 sieht man einen überlappenden Konturplot des Beispiels aus Abschnitt 2.1. Es werden die Konturen der Response-Variablen $y_1 = \text{Viskosität}$ und $y_2 = \text{Molekulargewicht}$ übereinander gelegt. In diesem Beispiel soll die Viskosität maximiert und das Molekulargewicht minimiert werden. Die optimale Einstellung für die Viskosität ist $\xi_1 = 85$ und $\xi_2 = 175$. Es wird eine Viskosität y_1 über 70 erreicht. Bei dieser Einstellung erhalten wir für das Molekulargewicht y_2 einen Wert zwischen 33 und 34. Für das Molekulargewicht erhält man als optimale Einstellung $\xi_1 = 77$ und $\xi_2 = 167$, welche in einem Wert unter 29 resultiert. Diese Einstellung erreicht

bei der Viskosität nur einen Wert unter 54. Man sieht, dass diese zwei Response-Variablen für unterschiedliche Parametereinstellungen optimal werden. Die optimale Einstellung für die eine Response-Variable führt zu einem schlechten Wert der anderen Response-Variablen. Bei der simultanen Optimierung müssen wir daher Kompromisse eingehen. Eine passende Einstellung für beide Variablen ist $\xi_1 = 77$ und $\xi_2 = 175$. Diese führt zu einer Viskosität y_1 zwischen 68 und 70 und zu einem Molekulargewicht y_2 zwischen 29 und 30. Mit dieser Einstellung erreicht zwar keine der Variablen ihr individuelles Optimum, dennoch sind die beiden Werte akzeptabel.

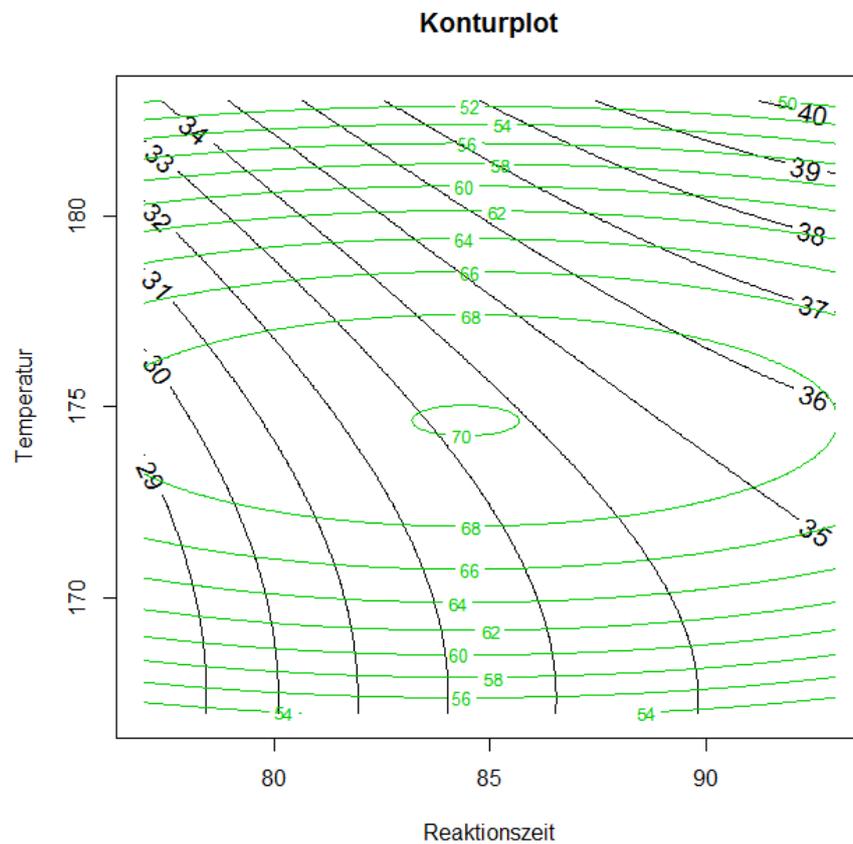


Abbildung 3.1.: Konturplot - $y_1 = \text{Viskosität}$ (grün) und $y_2 = \text{Molekulargewicht}$ (schwarz) in Abhängigkeit von $\xi_1 = \text{Reaktionszeit}$ und $\xi_2 = \text{Temperatur}$

3.3. Optimierungsproblem

Eine weitere Möglichkeit dieses Problem zu behandeln, ist ein entsprechendes Optimierungsproblem aufzustellen. Wir erhalten für l Response-Variablen

$$\begin{aligned} & \text{Max } y_1 \\ & \text{u.d.N. } U_2 \leq y_2 \leq O_2 \\ & \quad \vdots \\ & U_l \leq y_l \leq O_l, \end{aligned}$$

wobei U_i die untere und O_i die obere Grenze des Bereichs darstellt, der im Vorfeld für die Response-Variable i als akzeptabel festgelegt wurde. Lösungen für dieses Problem findet man mit Hilfe von Methoden der Nichtlinearen Optimierung.

3.4. Desirability-Funktion

Sehr hilfreich ist auch die Methode von Derringer & Suich (1980), bei der man mit sogenannten Desirability-Funktionen arbeitet. In Abschnitt 3.3 haben wir einen Ansatz gesehen, der eine Response-Variable optimiert und für die anderen nur Bedingungen aufstellt, die diese erfüllen müssen. Bei dieser Methode versucht man aber eine Balance zwischen der Optimierung der verschiedenen Response-Variablen zu finden. Das heißt, dass man jeder Response-Variable in der Optimierung die gleiche Bedeutung zukommen lassen will.

Die Desirability-Funktion ordnet jeder geschätzten Response-Variable \hat{y}_i einen bestimmten Desirability-Wert d_i zu, wobei $0 \leq d_i \leq 1$. Der Wert der Desirability-Funktion steigt, wenn der Response-Wert ein besseres Ergebnis erzielt. Die einzelnen Desirability-Funktionen werden dann mit Hilfe des geometrischen Mittels zusammengesetzt und man erhält die Gesamt-Desirability-Funktion als

$$D = (d_1 \times d_2 \times \dots \times d_l)^{1/l} = \prod_{i=1}^l d_i^{1/l},$$

welche in das Intervall $[0, 1]$ fällt. Dieser Wert gibt die Gesamtsituation aller Response-Werte wieder.

Die Wahl des geometrischen Mittels als Gesamt-Desirability-Funktion geht darauf zurück, dass man - wie vorher erwähnt - jeder Response-Variable in der Optimierung die gleiche Bedeutung zukommen lassen will. Sollte nur ein einziger Response-Wert in einen nicht akzeptablen Bereich fallen, und daher eine Desirability-Funktion von 0 haben, folgt daraus automatisch, dass die Gesamt-Desirability-Funktion auch 0 ist. Im Folgenden werden Möglichkeiten erklärt, wie man die Variable \hat{y}_i in d_i transformieren kann.

3.4.1. Einseitige Transformation

Larger-the-better

Wie der Name schon verrät, verwenden wir diese Transformation, wenn wir \hat{y}_i maximieren wollen. Die Funktion d_i wächst, wenn der Wert \hat{y}_i wächst. Es wird die Desirability-Funktion

$$d_i = \begin{cases} 0 & \hat{y}_i \leq U_i \\ \left[\frac{\hat{y}_i - U_i}{O_i - U_i} \right]^r & U_i < \hat{y}_i < O_i \\ 1 & \hat{y}_i \geq O_i \end{cases}$$

verwendet. Der Wert U_i gibt die untere Grenze des akzeptablen Bereichs von \hat{y}_i an. Diese wird so festgelegt, dass jeder Wert, der kleiner als diese Grenze ist, einen Desirability-Wert von 0 zugeordnet bekommt, und dadurch auch die Gesamt-Desirability-Funktion auf 0 setzt. Der Wert O_i gibt die obere Grenze des akzeptablen Bereichs an. Aufgrund der Maximierung, gibt es diese Grenze eigentlich nicht. Man könnte sie aber als Wert ansehen, ab welchem man fast keinen zusätzlichen Nutzen mehr aus der Erhöhung von \hat{y}_i ziehen kann.

Auch die Variable r kann man für jedes Problem individuell festsetzen. Ist r groß, bedeutet dies, dass es erwünscht ist, dass der Wert von \hat{y}_i sehr schnell über U_i steigt. Umgekehrt bedeutet ein kleiner Wert von r , dass der Wert von \hat{y}_i nur sehr langsam über U_i steigt. Um d_i beachtlich zu vergrößern, muss \hat{y}_i sehr viel größer als U_i sein. In Abbildung 3.2 kann man sehen, wie die Desirability-Funktion für verschiedene Werte von r aussieht.

Smaller-the-better

Wenn wir \hat{y}_i minimieren wollen, setzen wir in die vorher beschriebene Desirability-Funktion den Wert $-\hat{y}_i$ ein, da die Minimierung von \hat{y}_i äquivalent zur Maximierung von $-\hat{y}_i$ ist. Man erhält also die Desirability-Funktion

$$d_i = \begin{cases} 0 & \hat{y}_i \geq O_i \\ \left[\frac{O_i - \hat{y}_i}{O_i - U_i} \right]^r & U_i < \hat{y}_i < O_i \\ 1 & \hat{y}_i \leq U_i \end{cases} .$$

3.4.2. Zweiseitige Transformation

Nominal-the-best

Diese Transformation wird angewendet, wenn wir sowohl eine Maximum- als auch eine Minimum-Bedingung haben. Es ist wünschenswert, dass \hat{y}_i genau den Wert T_i annimmt. Dieser wird oft als Zielwert (Target) bezeichnet. Die Funktion nimmt bei diesem Wert die höchste Desirability, also $d_i = 1$, an. Der Wert T_i liegt zwischen U_i

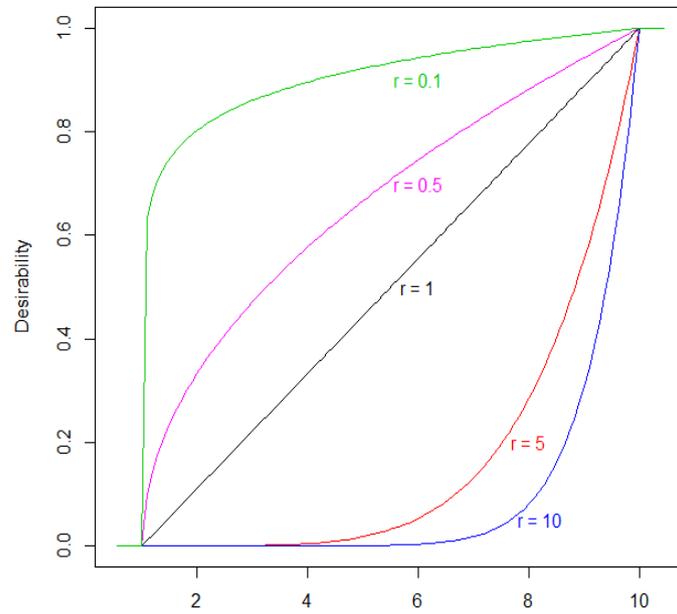
und O_i und die Desirability-Funktion sieht folgendermaßen aus

$$d_i = \begin{cases} \left[\frac{\hat{y}_i - U_i}{T_i - U_i} \right]^s & U_i \leq \hat{y}_i \leq T_i \\ \left[\frac{\hat{y}_i - O_i}{T_i - O_i} \right]^t & T_i < \hat{y}_i \leq O_i \\ 0 & \text{sonst} \end{cases} .$$

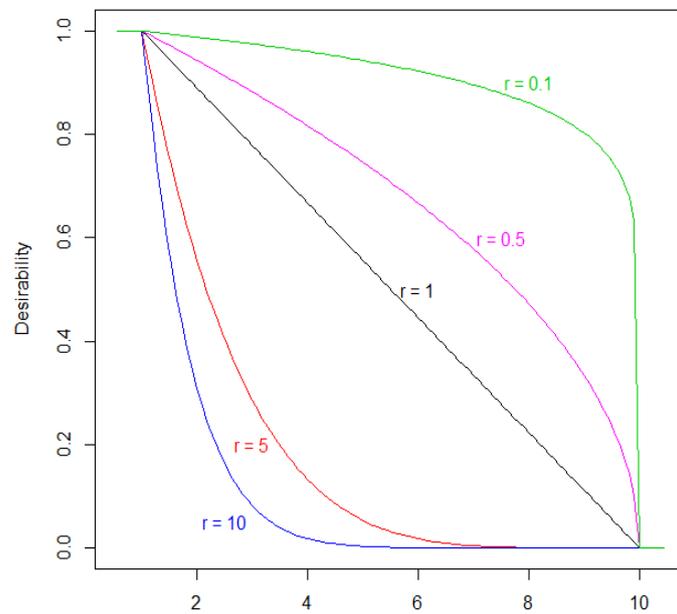
Die Variablen s und t sind vergleichbar mit der Variablen r in der einseitigen Transformation und können vom Anwender auf einen bestimmten Wert gesetzt werden. Große Werte von s und t „belohnen“ Werte, die sehr nahe am Wert T_i sind. Wählt man s und t klein, ist fast jeder Wert zwischen U_i und O_i gleich akzeptabel. Natürlich kann man auch einen Wert groß und den anderen Wert klein wählen. Man erhält eine Kombination aus den vorher beschriebenen Situationen. In Abbildung 3.3 sieht man Desirability-Funktionen mit unterschiedlichen Werten für s und t .

Optimierung

Da die Desirability-Funktion d_i nicht differenzierbar ist, kann man keine numerischen Standardmethoden der Optimierung anwenden, um dieses Problem zu lösen. Die Funktion \hat{y}_i wird durch eine polynomiale Funktion dargestellt und ist daher eine stetige Funktion der x_i . Die Desirability-Funktion d_i ist in jedem der drei Fälle eine stetige Funktion von \hat{y}_i und auch die Gesamt-Desirability-Funktion D ist eine stetige Funktion in den d_i . Daraus folgt, dass D eine stetige Funktion in den x_i ist. Durch Verwendung der Desirability-Funktionen wird ein multivariates in ein univariates Optimierungsproblem transformiert. Derringer & Suich (1980) schlagen aufgrund dieser Argumente die Verwendung von univariaten direkten Suchmethoden vor, um die Gesamt-Desirability-Funktion zu optimieren.



(a) Larger-the-better



(b) Smaller-the-better

Abbildung 3.2.: Desirability-Funktion - Einseitige Transformation

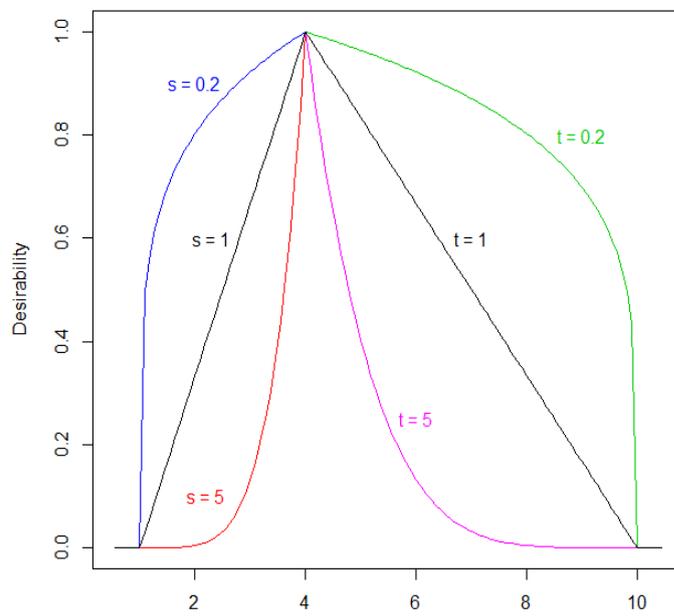


Abbildung 3.3.: Desirability-Funktion - Zweiseitige Transformation ($c_i = 4$)

3.5. Dual Response

In diesem Abschnitt wird die Methode von Myers & Carter (1973) vorgestellt. Diese Methode wird - wie der Name sagt - beim Auftreten von nur zwei Response-Variablen verwendet. Man muss hierbei eine der Response-Variablen als erste (primary) und eine als zweite (secondary) Response-Variable festlegen. Das Ziel ist es dann, die gefittete Response-Funktion \hat{y}_p der ersten Response-Variable zu optimieren, unter der Bedingung, dass die gefittete Response-Funktion \hat{y}_s der zweiten Response-Variable einen bestimmten Wert t annimmt. Die Voraussetzung für diese Methode ist, dass ein Problem mit quadratischen Response-Funktionen vorliegt. Man erhält aus

$$\hat{y}_p = b_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\widehat{\mathbf{B}}\mathbf{x} \quad (3.5)$$

und

$$\hat{y}_s = c_0 + \mathbf{x}'\mathbf{c} + \mathbf{x}'\widehat{\mathbf{C}}\mathbf{x} \quad (3.6)$$

unter Verwendung des Lagrange-Multiplikators die Funktion

$$L = b_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\widehat{\mathbf{B}}\mathbf{x} - \mu(c_0 + \mathbf{x}'\mathbf{c} + \mathbf{x}'\widehat{\mathbf{C}}\mathbf{x} - t).$$

Die mögliche Lösung des Optimierungsproblems findet man, indem man die Ableitungen dieser Funktion nach \mathbf{x}

$$\frac{\partial L}{\partial \mathbf{x}} = \mathbf{b} + 2\widehat{\mathbf{B}}\mathbf{x} - \mu\mathbf{c} - 2\mu\widehat{\mathbf{C}}\mathbf{x}$$

gleich 0 setzt. Man erhält die Gleichung

$$(\widehat{\mathbf{B}} - \mu\widehat{\mathbf{C}})\mathbf{x} = \frac{1}{2}(\mu\mathbf{c} - \mathbf{b}). \quad (3.7)$$

Um den stationären Punkt - der Punkt welcher Gleichung (3.7) erfüllt - genauer zu untersuchen, muss man die Matrix der zweiten partiellen Ableitungen $\mathbf{M}(\mathbf{x})$ - das (i, j) -te Element ist $(\partial^2 L / \partial x_i \partial x_j)$ - betrachten. Diese ist durch

$$\mathbf{M}(\mathbf{x}) = 2(\widehat{\mathbf{B}} - \mu\widehat{\mathbf{C}})$$

gegeben. An dieser Gleichung kann man erkennen, dass die Lösung des Problems von den Matrizen $\widehat{\mathbf{B}}$ und $\widehat{\mathbf{C}}$ abhängig ist. Dies werden wir im Folgenden genauer betrachten.

Ist die Matrix $\mathbf{M}(\mathbf{x})$ negativ definit, ist der stationäre Punkt \mathbf{x}_s ein lokales Maximum. Umgekehrt impliziert positive Definitheit der Matrix $\mathbf{M}(\mathbf{x})$ ein lokales Minimum. Myers & Carter (1973) schlagen vor, dass man, statt $\hat{y}_s = t$ zu fixieren, den Wert des Lagrange-Multiplikators im Bereich variieren kann, in welchem die \mathbf{x} Gleichung (3.7) erfüllen und in absoluten Maxima von \hat{y}_p resultieren. Dies führt man unter der Bedingung durch, dass man auf der Regressionsfläche der zweiten Response-Variablen bleibt, welche durch (3.6) gegeben ist.

Wegen nachfolgendem Satz 3.5.1 kann man sich bei der Suche nach \mathbf{x} , die in Maxima (Minima) resultieren, auf Lagrange-Multiplikatoren μ beschränken, welche $(\widehat{\mathbf{B}} - \mu\widehat{\mathbf{C}})$ negativ (positiv) definit machen, sofern diese existieren.

Satz 3.5.1. Seien \mathbf{x}_1 und \mathbf{x}_2 zwei Lösungen der Gleichung (3.7), μ_1 und μ_2 die zugehörigen Lagrange-Multiplikatoren und $\hat{y}_{s,1} = \hat{y}_{s,2}$. Ist die Matrix $(\widehat{\mathbf{B}} - \mu_1 \widehat{\mathbf{C}})$ negativ definit, dann gilt $\hat{y}_{p,1} > \hat{y}_{p,2}$. Ist die Matrix $(\widehat{\mathbf{B}} - \mu_1 \widehat{\mathbf{C}})$ positiv definit, dann gilt $\hat{y}_{p,1} < \hat{y}_{p,2}$. Außerdem gilt: ist $(\widehat{\mathbf{B}} - \mu_1 \widehat{\mathbf{C}})$ negativ definit, kann $(\widehat{\mathbf{B}} - \mu_2 \widehat{\mathbf{C}})$ nicht negativ definit sein.

Beweis: Siehe Myers & Carter (1973).

Man wird feststellen, dass dieser Bereich für μ sehr oft existiert und es von den Matrizen $\widehat{\mathbf{B}}$ und $\widehat{\mathbf{C}}$ abhängt, wo sich dieser befindet. Man unterscheidet jetzt im Groben zwei Möglichkeiten.

1. $\widehat{\mathbf{C}}$ definit
 - a) $\widehat{\mathbf{C}}$ positiv definit
 - b) $\widehat{\mathbf{C}}$ negativ definit
2. $\widehat{\mathbf{C}}$ indefinit

3.5.1. $\widehat{\mathbf{C}}$ definit

$\widehat{\mathbf{C}}$ positiv definit

In diesem Fall ist der stationäre Punkt der zweiten Response-Variable ein Minimum. Da $\widehat{\mathbf{C}}$ positiv definit ist, existiert laut Rao (1973) eine reguläre Matrix \mathbf{S} , sodass

$$\mathbf{S}' \widehat{\mathbf{B}} \mathbf{S} = \text{diag} (\lambda_1, \lambda_2, \dots, \lambda_k)$$

und

$$\mathbf{S}' \widehat{\mathbf{C}} \mathbf{S} = \mathbf{I}_k.$$

Den entsprechenden Satz mit Beweis findet man im Anhang A.1.

Man betrachtet die quadratische Form mit der Matrix $\mathbf{M}(\mathbf{x})$

$$q = \mathbf{w}' (\widehat{\mathbf{B}} - \mu \widehat{\mathbf{C}}) \mathbf{w}.$$

Wenn man hier für

$$\mathbf{w}' = \mathbf{v}' \mathbf{S}'$$

einsetzt, erhält man

$$\begin{aligned} q &= \mathbf{v}' \mathbf{S}' (\widehat{\mathbf{B}} - \mu \widehat{\mathbf{C}}) \mathbf{S} \mathbf{v} \\ &= \mathbf{v}' (\mathbf{S}' \widehat{\mathbf{B}} \mathbf{S} - \mu \mathbf{S}' \widehat{\mathbf{C}} \mathbf{S}) \mathbf{v} \\ &= \mathbf{v}' (\text{diag} (\lambda_1, \lambda_2, \dots, \lambda_k) - \mu \mathbf{I}_k) \mathbf{v} \\ &= \mathbf{v}' \text{diag} (\lambda_1 - \mu, \lambda_2 - \mu, \dots, \lambda_k - \mu) \mathbf{v}. \end{aligned} \tag{3.8}$$

Die λ_i sind die Eigenwerte der reellen und symmetrischen Matrix

$$\mathbf{T} = \mathbf{D}_2^{-1/2} \mathbf{Q}' \widehat{\mathbf{B}} \mathbf{Q} \mathbf{D}_2^{-1/2},$$

wobei hier \mathbf{Q} eine orthogonale Matrix ist, für die gilt, dass

$$\mathbf{Q}' \widehat{\mathbf{C}} \mathbf{Q} = \mathbf{D}_2$$

und \mathbf{D}_2 eine Diagonalmatrix ist, welche in der Diagonale die Eigenwerte von $\widehat{\mathbf{C}}$ enthält. Die Matrix $\mathbf{D}_2^{-1/2}$ ist die Diagonalmatrix, welche in der Diagonale die Kehrwerte der Wurzel von den Eigenwerten von $\widehat{\mathbf{C}}$ enthält.

Aus (3.8) kann man erkennen, dass $\mu > \max_i \lambda_i$ und $\mu < \min_i \lambda_i$ die Bereiche darstellen, in welchen die \mathbf{x}_i ein absolutes Optimum von $\hat{y}_{p,i}$ ergeben. Der Bereich $\mu > \max_i \lambda_i$ impliziert, dass die Matrix $\mathbf{M}(\mathbf{x})$ negativ definit ist und wir erhalten ein Maximum. Umgekehrt impliziert der Bereich $\mu < \min_i \lambda_i$ positive Definitheit und damit ein Minimum. Man kann sich in Folge bei der Suche nach Optima auf diese Bereiche beschränken unter der Bedingung, dass man auf der Response Surface der zweiten Response-Variable bleibt. Die gefundenen Vektoren \mathbf{x}_i führen zu allen möglichen Werten von \hat{y}_s . Folgender Satz erklärt die Beziehung von μ zum geschätzten Response-Wert der zweiten Response-Variable \hat{y}_s .

Satz 3.5.2. *Sei \mathbf{x} eine Lösung von (3.7) und $\widehat{\mathbf{C}}$ positiv definit. Dann gilt $\frac{\partial^2 \hat{y}_s}{\partial \mu^2} \geq 0$. Gleichheit gilt, wenn $\mu \rightarrow \pm \infty$.*

Beweis: Siehe Myers & Carter (1973).

Aus Gleichung (3.7) erkennt man, dass

$$\lim_{\mu \rightarrow \infty} \mathbf{x} = -\widehat{\mathbf{C}}^{-1} \mathbf{c}/2 = \mathbf{x}_{s,o},$$

wobei $\mathbf{x}_{s,o}$ der stationäre Punkt der zweiten Response-Variable ist. Für Werte im Bereich $\mu > \max_i \lambda_i$, fällt \hat{y}_s für steigendes μ und nähert sich dem Minimalwert $\hat{y}_{s,o}$ an. Dies ist der Wert der geschätzten Response-Funktion der zweiten Response-Variablen an der Stelle $\mathbf{x}_{s,o}$. Weiters gilt, dass

$$\lim_{\mu \rightarrow \lambda_i} \hat{y}_s(\mu) = \infty,$$

weil $|\widehat{\mathbf{B}} - \lambda_i \widehat{\mathbf{C}}| = 0$ für $i = 1, 2, \dots, k$. Die Asymptoten der Funktion $\hat{y}_s(\mu)$ liegen also bei $\mu = \lambda_i$.

$\widehat{\mathbf{C}}$ negativ definit

Betrachtet man den Fall, dass $\widehat{\mathbf{C}}$ negativ definit ist, ist der stationäre Punkt der zweiten Response-Variable ein Maximum. Man kann für diesen Fall die gleichen Schritte wie im Fall $\widehat{\mathbf{C}}$ positiv definit anwenden. Der einzige Unterschied ist, dass man hier, wenn man die Response-Funktion der ersten Variable unter einer Bedingung für \hat{y}_s maximieren, bzw. die Matrix $\mathbf{M}(\mathbf{x})$ negativ definit machen will, im Bereich $\mu < \min_i \lambda_i$ suchen muss. Umgekehrt sucht man für die Minimierung von \hat{y}_p im Bereich $\mu > \max_i \lambda_i$. Es gilt analog zum vorigen Abschnitt folgender Satz.

Satz 3.5.3. Sei \mathbf{x} eine Lösung von Gleichung (3.7) und sei $\widehat{\mathbf{C}}$ negativ definit, so gilt $\frac{\partial^2 \hat{y}_s}{\partial \mu^2} \leq 0$. Die Gleichheit ist erfüllt, wenn $\mu \rightarrow \pm \infty$.

In diesem Fall gilt, dass

$$\lim_{\mu \rightarrow \lambda_i} \hat{y}_s(\mu) = -\infty$$

und dass \hat{y}_s für Werte im Bereich $\mu < \min_i \lambda_i$ für kleiner werdendes μ steigt und sich dem Maximalwert $\hat{y}_{s,o}$ annähert.

Überblick für den Fall $\widehat{\mathbf{C}}$ definit

In Tabelle 3.1 ist ersichtlich, in welchem Bereich man in den unterschiedlichen Fällen jene \mathbf{x} suchen muss, welche die geschätzte Response-Funktion \hat{y}_p optimieren.

	Maximiere \hat{y}_p	Minimiere \hat{y}_p
$\widehat{\mathbf{C}}$ positiv definit	$\mu > \max_i \lambda_i$	$\mu < \min_i \lambda_i$
$\widehat{\mathbf{C}}$ negativ definit	$\mu < \min_i \lambda_i$	$\mu > \max_i \lambda_i$

Tabelle 3.1.: Überblick für $\widehat{\mathbf{C}}$ definit

Sehr hilfreich ist es auch, sich Plots von x_1 gg. \hat{y}_s , x_2 gg. \hat{y}_s , \dots , x_k gg. \hat{y}_s und \hat{y}_s gg. \hat{y}_p des zugrundeliegenden Problems anzusehen.

3.5.2. $\widehat{\mathbf{C}}$ indefinit

Auch für den Fall, dass die Matrix $\widehat{\mathbf{C}}$ indefinit ist, kann es eine Lösung für das Dual-Response-Problem geben. Will man \hat{y}_p unter einer Bedingung für \hat{y}_s maximieren, muss dafür die Matrix $\widehat{\mathbf{B}}$ negativ definit sein. Umgekehrt muss für die Minimierung von \hat{y}_p , die Matrix $\widehat{\mathbf{B}}$ positiv definit sein.

$\widehat{\mathbf{B}}$ negativ definit

Hier will man, dass die quadratische Form

$$q = \mathbf{w}'(\mu\widehat{\mathbf{C}} + (-\widehat{\mathbf{B}}))\mathbf{w}$$

positiv definit wird, um zu erreichen, dass $\mathbf{M}(\mathbf{x})$ negativ definit wird. Auch in diesem Fall verwenden wir Satz A.1.1. Der Unterschied dabei ist, dass die Matrizen $\widehat{\mathbf{B}}$ und $\widehat{\mathbf{C}}$ vertauscht werden. Man erhält

$$\mathbf{S}'\widehat{\mathbf{C}}\mathbf{S} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k)$$

und

$$\mathbf{S}'(-\widehat{\mathbf{B}})\mathbf{S} = \mathbf{I}_k.$$

Die λ_i sind die Eigenwerte der Matrix \mathbf{T}^*

$$\mathbf{T}^* = \mathbf{D}_1^{-1/2} \mathbf{P}' \widehat{\mathbf{C}} \mathbf{P} \mathbf{D}_1^{-1/2}, \quad (3.9)$$

wobei hier \mathbf{P} die orthogonale Matrix ist, für welche

$$\mathbf{P}'(-\widehat{\mathbf{B}})\mathbf{P} = \mathbf{D}_1$$

gilt. $\mathbf{D}_1^{-1/2}$ ist die Diagonalmatrix, welche in der Diagonale die Kehrwerte der Wurzel der Eigenwerte von $-\widehat{\mathbf{B}}$ enthält.

Durch die Transformation $\mathbf{w} = \mathbf{S}\mathbf{v}$ erhält man

$$q = \mathbf{v}'(\mu \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k) + \mathbf{I}_k)\mathbf{v}.$$

Daraus kann man ablesen, dass für den Bereich, indem wir nach den \mathbf{x} suchen, welche die Funktion \hat{y}_p maximieren, die Ungleichung $\mu\lambda_i > -1$ für $i = 1, 2, \dots, k$ gelten muss. Man weiß aus der Definition von \mathbf{T}^* , dass die λ_i verschiedene Vorzeichen haben. Es ergibt sich daraus der Bereich

$$-\frac{1}{\min_i \lambda_i} > \mu > -\frac{1}{\max_i \lambda_i}.$$

Mit Hilfe eines Plots der Funktion $\hat{y}_s(\mu)$ kann man erkennen, dass in diesem Fall die Asymptoten bei $\mu = -1/\lambda_i$ ($i = 1, 2, \dots, k$) liegen. Folgender Satz beschreibt die Ableitung der Funktion \hat{y}_s nach μ im betrachteten Bereich.

Satz 3.5.4. *Sei $\widehat{\mathbf{C}}$ indefinit und $\widehat{\mathbf{B}}$ negativ definit, dann ist im Bereich $-1/\min_i \lambda_i > \mu > -1/\max_i \lambda_i$ die Ableitung $\frac{\partial \hat{y}_s}{\partial \mu} < 0$.*

Beweis: Siehe Myers & Carter (1973).

$\widehat{\mathbf{B}}$ positiv definit

Es wird noch der Fall betrachtet, dass $\widehat{\mathbf{B}}$ positiv definit ist. Man kann in diesem Fall eine Lösung finden, die \hat{y}_p für einen festgesetzten Wert von \hat{y}_s minimiert. Es muss gelten, dass $\mu\lambda_i < 1$ für $i = 1, 2, \dots, k$, wobei hier die λ_i die Eigenwerte der Matrix \mathbf{T}^* aus Gleichung (3.9) sind. Für die orthogonale Matrix \mathbf{P} gilt hier jedoch $\mathbf{P}'\widehat{\mathbf{B}}\mathbf{P} = \mathbf{D}_1$. Man sucht in diesem Fall im Bereich

$$\frac{1}{\min_i \lambda_i} < \mu < \frac{1}{\max_i \lambda_i}.$$

Überblick für den Fall $\widehat{\mathbf{C}}$ indefinit

In Tabelle 3.2 ist ersichtlich, in welchem Bereich der μ man im Fall, dass die Matrix $\widehat{\mathbf{C}}$ indefinit ist, jene \mathbf{x} suchen muss, welche die geschätzte Response-Funktion \hat{y}_p optimieren.

	Maximiere \hat{y}_p	Minimiere \hat{y}_p
$\widehat{\mathbf{B}}$ negativ definit	$-\frac{1}{\min_i \lambda_i} > \mu > -\frac{1}{\max_i \lambda_i}$	keine Lösung (siehe 3.5.3)
$\widehat{\mathbf{B}}$ positiv definit	keine Lösung (siehe 3.5.3)	$\frac{1}{\min_i \lambda_i} < \mu < \frac{1}{\max_i \lambda_i}$

Tabelle 3.2.: Überblick für $\widehat{\mathbf{C}}$ indefinit

3.5.3. Weitere Nebenbedingung

Wie im vorigen Abschnitt gezeigt, ist es auch möglich, dass man keine Lösung findet. Zum Beispiel tritt dies ein, wenn $\widehat{\mathbf{C}}$ indefinit und $\widehat{\mathbf{B}}$ positiv definit ist und man die Response-Funktion \hat{y}_p maximieren will, oder umgekehrt $\widehat{\mathbf{C}}$ indefinit und $\widehat{\mathbf{B}}$ negativ definit ist, und man ein Minimum für die Response-Funktion \hat{y}_p finden will. Weiters kann das gefundene \mathbf{x} auch außerhalb des Arbeitsbereiches liegen.

Tritt einer der vorher genannten Fälle ein, hat man die Möglichkeit die zusätzliche Bedingung $\sum_{i=1}^k x_i^2 = R^2$ einzuführen ($\mathbf{x} = \mathbf{0}$ ist hier der Ursprung der Inputvariablen). Der Wert R wird so festgelegt, dass man sich noch im Bereich, in welchem das Experiment durchgeführt wurde, befindet. Unter Verwendung des Lagrange-Multiplikators erhält man

$$L = \hat{y}_p - \mu(\hat{y}_s - t) - \gamma(\mathbf{x}'\mathbf{x} - R^2).$$

Wenn man die Ableitung dieser Funktion nach \mathbf{x} gleich 0 setzt, erhält man

$$(\widehat{\mathbf{B}} - \mu\widehat{\mathbf{C}} - \gamma\mathbf{I})\mathbf{x} = \frac{1}{2}(\mu\mathbf{c} - \mathbf{b}). \quad (3.10)$$

Man wählt jetzt μ und γ und sucht dann die Vektoren \mathbf{x} , die die Gleichung (3.10) erfüllen. Für ein gegebenes μ , wählt man $\gamma > \max_i \lambda_i$, wobei λ_i die Eigenwerte der Matrix $(\widehat{\mathbf{B}} - \mu\widehat{\mathbf{C}})$ sind, wenn man \hat{y}_p maximieren will. Aus dieser Bedingung folgt, dass die Matrix $\mathbf{M}(\mathbf{x}) = 2(\widehat{\mathbf{B}} - \mu\widehat{\mathbf{C}} - \gamma\mathbf{I})$ negativ definit wird. Um ein Minimum zu bekommen, wählt man $\gamma < \min_i \lambda_i$. Setzt man $\mu = 0$, optimiert man die Funktion \hat{y}_p ohne eine Bedingung für \hat{y}_s aufzustellen. Das Problem reduziert sich auf die sogenannte Ridge-Analyse. Der interessierte Leser findet weitere Details dazu bei Myers & Montgomery (2002).

3.6. Distanzfunktion

Da keine der bereits vorgestellten Methoden auf Varianzen und Korrelationen zwischen den Response-Variablen Rücksicht nimmt, sehen wir uns in diesem Abschnitt die Methode von Khuri & Conlon (1981) an, die mit einer Distanzfunktion arbeitet. Die Funktion misst den Abstand, den die geschätzten Response-Werte von ihren individuellen Optima haben. Weiters fließen in diese Funktion auch die Varianzen und Korrelationen mit ein, die zwischen den Response-Variablen bestehen sowie der Fehler, der durch das Schätzen des Vektors der individuellen Optima entsteht.

Wir betrachten das Modell (3.1) aus Abschnitt 3.1. Wenn $l \leq n-p$ ist, ist die Matrix $\widehat{\Sigma}$ aus (3.2) invertierbar, wenn \mathbf{Y} vollen Rang hat. Da wir später bei der Definition des Distanzmaßes die Invertierbarkeit dieser Matrix voraussetzen, ist es wichtig zu überprüfen, ob es lineare Abhängigkeiten zwischen den l Response-Variablen gibt. Dies werden wir im nächsten Abschnitt genauer analysieren.

3.6.1. Lineare Abhängigkeiten zwischen den Response-Variablen

Um lineare Abhängigkeiten zu untersuchen, betrachtet man die Eigenwerte der Matrix $\mathbf{D}\mathbf{D}'$, wobei

$$\mathbf{D} = \mathbf{Y}'(\mathbf{I}_n - \mathbf{1}_n\mathbf{1}'_n/n) \quad (3.11)$$

eine $l \times n$ Matrix und $\mathbf{1}_n$ ein n -dimensionaler Vektor ist, der Einsen enthält. Die Matrix $\mathbf{D}\mathbf{D}'$ hat genau dann einen m -fachen Eigenwert 0, wenn es m linear abhängige Beziehungen zwischen den Response-Variablen gibt (Box et al., 1973).

Vorsichtig muss man aufgrund auftretender Rundungsfehler sein. Wenn man bei den Berechnungen Rundungen durchführt, kann es durchaus passieren, dass kein Eigenwert exakt 0 ist, es aber dennoch lineare Abhängigkeiten in den Daten gibt. Man muss also bei jedem sehr kleinen Eigenwert (λ^0), die erwartete Größe und die Varianz genau überprüfen. Unter Annahme, dass die Rundungsfehler unabhängig und identisch verteilte Zufallsvariablen sind, die einer Gleichverteilung $U(-\delta, \delta)$ folgen, gilt laut Box et al. (1973) approximativ, dass für δ ausreichend klein,

$$E(\lambda^0) = (n-1)\sigma_{rf}^2. \quad (3.12)$$

Für die Varianz des Eigenwerts erhalten wir

$$\text{var}(\lambda^0) \leq q\sigma_{rf}^4, \quad (3.13)$$

wobei $\sigma_{rf}^2 = \delta^2/3$ die Varianz des Rundungsfehlers und $q = 9nl/5 + nl(nl-1) - (n-1)^2$ ist. Die Herleitung dieser Aussage findet man in Anhang A.2. Mit Hilfe der Gleichungen (3.12) und (3.13) überprüft man, ob bei einem sehr kleinen Eigenwert von linearer Abhängigkeit ausgegangen werden muss.

Bei einem m -fachen Eigenwert 0 kann man von $l-m$ unabhängigen Response-Variablen ausgehen. Die m voneinander abhängigen Response-Variablen müssen allerdings entfernt werden. Dies kann mit Hilfe der Methode aus Abschnitt A.3 durchgeführt werden.

3.6.2. Simultane Optimierung der multiplen Response-Funktion

Für die Optimierung muss angenommen werden, dass zwischen den l Response-Variablen keine linearen Abhängigkeiten bestehen. Würde es welche geben, wären diese Response-Variablen - wie im vorigen Abschnitt beschrieben - entfernt worden. Sei jetzt ϕ_i das Optimum von \hat{y}_i im betrachteten Bereich R und $\phi = [\phi_1, \phi_2, \dots, \phi_l]'$, dann wird die Distanz von $\hat{\mathbf{y}}(\mathbf{x})$ zu ϕ mit $\rho[\hat{\mathbf{y}}(\mathbf{x}), \phi]$ bezeichnet. Da man die Einstellung mit der geringsten Abweichung von diesem Optimumsvektor finden will, muss man die Distanz $\rho[\hat{\mathbf{y}}(\mathbf{x}), \phi]$ auf dem betrachteten Bereich R minimieren.

3.6.3. Definition der Distanzfunktion

Khuri & Conlon (1981) schlagen als Distanzfunktion

$$\rho_1[\hat{\mathbf{y}}(\mathbf{x}), \boldsymbol{\phi}] = [(\hat{\mathbf{y}}(\mathbf{x}) - \boldsymbol{\phi})' \widehat{\boldsymbol{\Sigma}}^{-1} (\hat{\mathbf{y}}(\mathbf{x}) - \boldsymbol{\phi}) / \mathbf{z}'(\mathbf{x}) (\mathbf{X}' \mathbf{X})^{-1} \mathbf{z}(\mathbf{x})]^{1/2} \quad (3.14)$$

vor. Diese erhält man, indem man in der Gleichung

$$\rho[\hat{\mathbf{y}}(\mathbf{x}), \boldsymbol{\phi}] = [(\hat{\mathbf{y}}(\mathbf{x}) - \boldsymbol{\phi})' \text{var}([\hat{\mathbf{y}}(\mathbf{x})])^{-1} (\hat{\mathbf{y}}(\mathbf{x}) - \boldsymbol{\phi})]^{1/2}$$

den Schätzer von $\text{var}([\hat{\mathbf{y}}(\mathbf{x})])$ aus Gleichung (3.3) einsetzt. Ist die Matrix $\boldsymbol{\Sigma}$ diagonal, kann man die Distanzfunktion vereinfachen und man erhält

$$\rho_2[\hat{\mathbf{y}}(\mathbf{x}), \boldsymbol{\phi}] = \left[\sum_{i=1}^l (\hat{y}_i(\mathbf{x}) - \phi_i)^2 / \hat{\sigma}_{ii} \mathbf{z}'(\mathbf{x}) (\mathbf{X}' \mathbf{X})^{-1} \mathbf{z}(\mathbf{x}) \right]^{1/2},$$

wobei σ_{ii} das i -te Diagonalelement der Matrix $\widehat{\boldsymbol{\Sigma}}$ ist.

Wir haben also gesehen, dass die Annahme, dass alle Response-Variablen voneinander unabhängig sind, sehr wichtig ist. Die Matrix $\widehat{\boldsymbol{\Sigma}}$ wäre im Falle einer linearen Abhängigkeit singular und die Distanzfunktion nicht definiert.

3.6.4. Zufälligkeit von $\boldsymbol{\phi}$

Da $\boldsymbol{\phi}$ ein Zufallsvektor ist, aber $\boldsymbol{\phi}$ bei der Minimierung von ρ als fixer deterministischer Punkt im l -dimensionalen Raum betrachtet wird, kann das zu einem falschen Ergebnis führen. Man betrachtet den Vektor $\boldsymbol{\zeta} = [\zeta_1, \zeta_2, \dots, \zeta_l]'$, wobei ζ_i das Optimum des Erwartungswerts der i -ten Response-Variablen ist. Da $\boldsymbol{\phi}$ ein Zufallsvektor ist, ist das wahre Optimum durch $\boldsymbol{\zeta}$ gegeben. Um die Variabilität von $\boldsymbol{\phi}$ zu berücksichtigen, ist es sinnvoll im betrachteten Bereich R ein \mathbf{x} zu suchen, welches ein absolutes Minimum von $\rho[\hat{\mathbf{y}}(\mathbf{x}), \boldsymbol{\zeta}]$ ist. Da der Wert von $\boldsymbol{\zeta}$ unbekannt ist, schlagen Khuri & Conlon (1981) vor sich einen Konfidenzbereich $D_{\boldsymbol{\zeta}}$ zu berechnen. Mit Hilfe dieses Konfidenzbereiches legen sie eine obere Grenze für die Distanz fest. Man erhält für ein fixes \mathbf{x} in R die Ungleichung

$$\rho[\hat{\mathbf{y}}(\mathbf{x}), \boldsymbol{\zeta}] \leq \max_{\boldsymbol{\eta} \in D_{\boldsymbol{\zeta}}} \rho[\hat{\mathbf{y}}(\mathbf{x}), \boldsymbol{\eta}]. \quad (3.15)$$

Da die rechte Seite die linke Seite in (3.15) überschätzt, folgt daraus

$$\min_{\mathbf{x} \in R} \rho[\hat{\mathbf{y}}(\mathbf{x}), \boldsymbol{\zeta}] \leq \min_{\mathbf{x} \in R} \max_{\boldsymbol{\eta} \in D_{\boldsymbol{\zeta}}} \rho[\hat{\mathbf{y}}(\mathbf{x}), \boldsymbol{\eta}]. \quad (3.16)$$

Die rechte Seite der Gleichung (3.16) gibt ein Distanzmaß zwischen dem individuellen und dem Gesamt-Optimum an.

Der Konfidenzbereich $D_{\boldsymbol{\zeta}}$ wird durch die folgenden Ungleichungen beschrieben

$$\phi_i - g_i(\mathbf{X}, \hat{\boldsymbol{\xi}}_i) \sqrt{MS_i t_{n-p, 1-\alpha/2}} \leq \zeta_i \leq \phi_i + g_i(\mathbf{X}, \hat{\boldsymbol{\xi}}_i) \sqrt{MS_i t_{n-p, 1-\alpha/2}}$$

für $i = 1, 2, \dots, l$. Diese Ungleichungen haben ein approximatives Signifikanzniveau von zumindest $1 - \alpha^*$, wobei $\alpha^* = 1 - (1 - \alpha)^l$ ist. Der Wert MS_i ist der mittlere quadratische Fehler der i -ten Response-Variable, $\hat{\boldsymbol{\xi}}_i$ ist der Punkt in R , bei welchem $\hat{y}_i(\mathbf{x})$ das individuelle Optimum ϕ_i annimmt und $g_i(\mathbf{X}, \hat{\boldsymbol{\xi}}_i) = [\mathbf{z}'(\hat{\boldsymbol{\xi}}_i) (\mathbf{X}' \mathbf{X})^{-1} \mathbf{z}(\hat{\boldsymbol{\xi}}_i)]^{1/2}$. Die Herleitung des Konfidenzbereiches findet man bei Khuri & Conlon (1981).

3.6.5. Vorgangsweise

Zusammengefasst geht man bei dieser Methode folgendermaßen vor. Man untersucht die Response-Variablen auf lineare Abhängigkeiten mit Hilfe der Matrix \mathbf{D} . Alle Response-Variablen, die in einer linearen Beziehung zu einer anderen Response-Variable stehen, werden entfernt, sodass nur noch linear unabhängige Response-Variablen übrig bleiben. Dann wählt man sich ein Distanzmaß ρ und schätzt mit Hilfe der Kleinste-Quadrate-Methode die Regressionskoeffizienten. Die geschätzten Response-Variablen werden individuell im betrachteten Bereich R optimiert, um den Vektor ϕ zu erhalten. Ein Konfidenzbereich D_ζ wird aufgestellt und die Distanzfunktion wird in Bezug auf $\boldsymbol{\eta} \in D_\zeta$ für ein fixes $\boldsymbol{x} \in R$ maximiert. Das Maximum, das wir daraus erhalten, ist eine Funktion von \boldsymbol{x} , welche auf R minimiert wird.

Khuri & Conlon (1981) verwenden für die Auswahl von \boldsymbol{x} in R eine zufällige Suchmethode von Price, welche ein globales Optimum finden kann. Auch ein komplexes R stellt für diese Methode kein Problem dar. Um für jedes \boldsymbol{x} , die Metrik ρ zu maximieren, verwenden sie eine Gradienten-basierende Methode, da ρ nach $\boldsymbol{\eta}$ differenzierbar ist, und man damit das Optimum auf dem gegebenen Konfidenzbereich schnell findet. Laut Autoren spart man gegenüber Simplexmethoden oder der oben erwähnten Suchmethode einiges an Rechenzeit.

3.7. Verlustfunktion

Viele Autoren haben Methoden vorgeschlagen, die auf Taguchi's Verlustfunktion basieren. Auch die im vorherigen Abschnitt beschriebene Methode von Khuri und Conlon ist ein Beispiel dafür. Im Folgenden wird die Methode von Vining (1998) betrachtet, die auch mit Hilfe der Verlustfunktion ein Optimum sucht. Die quadratische Verlustfunktion (Squared-Error-Loss-Function) ist im univariaten Fall durch

$$\mathcal{L} = c[\hat{y}(\boldsymbol{x}) - \phi]^2$$

gegeben, wobei $\hat{y}(\boldsymbol{x})$ wieder der geschätzte Wert der Response-Variable ist, den wir für \boldsymbol{x} erhalten, ϕ der optimale Wert der Response-Variable ist und c einen Faktor darstellt, der die Kosten repräsentiert.

Es wird angenommen, dass ϕ bekannt ist. Diese Annahme ist gerechtfertigt, da dies in der Praxis oft der Fall ist. Auch bei der Methode von Derringer und Suich, die in Abschnitt 3.4 beschrieben wurde, haben wir Grenzen festgesetzt, ab welchen wir die Desirability-Funktion auf 1 setzen. Diese Grenzen sind mit ϕ vergleichbar. Das Ziel der Methode ist es jetzt, die optimale Einstellung \boldsymbol{x}_o zu finden, die den erwarteten Verlust $E(\mathcal{L})$ minimiert. Im univariaten Fall ist es äquivalent $E[\hat{y}(\boldsymbol{x}) - \phi]^2$ zu minimieren. Das heißt, dass der Faktor c auf die Minimierung keinen Einfluss hat.

Im multivariaten Fall ergibt sich für die Verlustfunktion

$$\mathcal{L} = [\hat{\boldsymbol{y}}(\boldsymbol{x}) - \boldsymbol{\phi}]' \mathbf{C} [\hat{\boldsymbol{y}}(\boldsymbol{x}) - \boldsymbol{\phi}].$$

Die positiv definite Matrix \mathbf{C} stellt die Kosten dar und $\boldsymbol{\phi}$ ist der Vektor der optimalen Werte der Response-Variablen. Wieder will man die optimale Einstellung \boldsymbol{x}_o

finden, welche den erwarteten Verlust minimiert. Man wird hier sehen, dass \mathbf{C} im multivariaten Fall sehr wohl eine Rolle spielt.

Es wird angenommen, dass die Varianz-Kovarianz Matrix der Response-Variablen $\mathbf{\Sigma}$ bekannt ist. Sollte dies nicht der Fall sein, gibt es die Möglichkeit, diese mit Hilfe der Maximum-Likelihood-Methode zu schätzen. Man erhält für den erwarteten Verlust

$$E(\mathcal{L}) = [E[\hat{\mathbf{y}}(\mathbf{x})] - \boldsymbol{\phi}]' \mathbf{C} [E[\hat{\mathbf{y}}(\mathbf{x})] - \boldsymbol{\phi}] + \text{Spur} [\mathbf{C} \mathbf{\Sigma}_{\hat{\mathbf{y}}}(\mathbf{x})]$$

(Graybill, 1976, S. 139). Ein Schätzer für den erwarteten Verlust ist

$$\hat{E}(\mathcal{L}) = [\hat{\mathbf{y}}(\mathbf{x}) - \boldsymbol{\phi}]' \mathbf{C} [\hat{\mathbf{y}}(\mathbf{x}) - \boldsymbol{\phi}] + \text{Spur} [\mathbf{C} \mathbf{\Sigma}_{\hat{\mathbf{y}}}(\mathbf{x})], \quad (3.17)$$

da $\hat{\mathbf{y}}(\mathbf{x})$ unter der Annahme, dass das gefittete Modell die jeweilige Response-Variable korrekt wiedergibt, ein unverzerrter Schätzer für $E[\hat{\mathbf{y}}(\mathbf{x})]$ ist.

In Gleichung (3.17) kann man erkennen, dass der Schätzer für den erwarteten Verlust aus zwei Termen besteht. Der erste Term bewertet die Abweichung des geschätzten Wertes der Response-Variable vom optimalen Wert der Response-Variable. Der zweite Term bewertet die Qualität der Schätzung. Man sieht, dass in beiden Termen die Matrix \mathbf{C} vorkommt, und diese daher bei der Minimierung des erwarteten Verlusts eine Rolle spielt.

Vorteile dieser Methode sind, dass die Korrelationen zwischen den Response-Variablen durch $\mathbf{\Sigma}_{\hat{\mathbf{y}}}(\mathbf{x})$ und die Kosten durch die Matrix \mathbf{C} in den erwarteten Verlust einfließen und daher beim Auffinden des Optimums berücksichtigt werden. Außerdem werden die Bereiche, in welchen das Modell nicht so gut passt, durch den zweiten Summanden des erwarteten Verlusts bestraft. Ein weiterer Vorteil ist, dass im Falle einer konstanten Matrix \mathbf{C} die Ableitungen der Verlustfunktion sehr leicht berechenbar sind. Es können dann Gradienten-basierende Verfahren zur Optimierung verwendet werden.

Der größte Nachteil der Methode ist, dass die Wahl von \mathbf{C} subjektiv ist. Vining (1998) gibt einige Möglichkeiten an, wie man \mathbf{C} wählen kann.

1. $\mathbf{C} = [\mathbf{\Sigma}_{\hat{\mathbf{y}}}(\mathbf{x})]^{-1}$
2. $\mathbf{C} = \mathbf{\Sigma}^{-1}$
3. $\mathbf{C} = \mathbf{K} \mathbf{\Sigma}^{-1}$

Fall 1 ist äquivalent zur Methode von Khuri und Conlon aus Abschnitt 3.6, wenn man bei deren Methode annimmt, dass der Vektor $\boldsymbol{\phi}$ bekannt ist.

4. Robuste Optimierung

In diesem Kapitel beschäftigen wir uns aufbauend auf Montgomery (2005) und Myers & Montgomery (2002) mit der robusten Optimierung, die in der Praxis sehr oft relevant ist. Bei der robusten Optimierung versucht man die Einstellung der Inputvariablen so zu wählen, dass der Mittelwert der Response-Variable einen bestimmten Zielwert erreicht und gleichzeitig die Variabilität um diesen Zielwert so gering wie möglich gehalten wird.

Ein Vorreiter auf diesem Gebiet war der japanische Mathematiker Genichi Taguchi. Er entwickelte das Robust-Parameter-Design Problem (RPD-Problem) und einige neue Methoden, um die daraus resultierenden Daten zu analysieren. Im Laufe der Zeit fand man einige Schwächen dieses Ansatzes und es wurde versucht, neue Lösungsansätze für dieses Problem zu finden. Unter anderem benutzte man zur Lösung des Problems auch Regressionsflächen Methodik. Das Konzept von Taguchi wurde dabei teilweise übernommen und mit neuen Methoden für die Analyse der Daten verknüpft.

4.1. RPD-Problem

Bei der robusten Optimierung wird das Augenmerk auf einen oder mehrere der folgenden Punkte gelegt.

- Man will das Produkt bzw. den Prozess unempfindlich gegenüber Umwelteinflüsse machen.
- Man will das Produkt bzw. den Prozess gegenüber der Variabilität der Komponenten des Systems unempfindlich machen.
- Man will das Produkt bzw. den Prozess so nah wie möglich an den Zielwert bringen, obwohl man manche Faktoren, wie zum Beispiel Temperatur oder gewisse Faktoren einer chemischen Verbindung nicht ganz genau einstellen kann.
- Man will die wichtigen Leistungs- bzw. Qualitätsmaße des Prozesses bzw. des Produktes so nah wie möglich an den Zielwert bringen und gleichzeitig die Variabilität der Zielgröße um diesen Zielwert so gering wie möglich halten.

Häufig wurden stärkere Komponenten, Komponenten mit strengeren Vorgaben bzw. andere Materialien verwendet um das RPD-Problem zu lösen. Dies führte allerdings dazu, dass Probleme wegen „Overdesigns“ auftraten. Die Produkte wurden sehr teuer bzw. konnten schwer hergestellt werden oder nahmen an Gewicht zu.

Taguchi hatte die Idee, dieses Problem mit Versuchsplanung und statistischen Methoden zu lösen. Dazu unterteilte er die Variablen, die Einfluss auf die Response-Variable haben, in Stör- und Kontrollvariablen. Die Störvariablen verursachen die Variabilität im System und sind in der Realität sehr schwer kontrollierbar. Im Experiment ist dies aber in den meisten Fällen möglich. Aus diesem Grund ist ein wichtiger Schritt zur Lösung des RPD-Problems, dass man diese Störvariablen erkennt, und die Einstellung der Kontrollvariablen findet, welche die Variabilität minimiert, die durch die Störvariablen entsteht.

4.1.1. Produkt-Versuchsplan

Taguchi benutzt in seiner Methode zwei verschiedene Versuchspläne. Ein Versuchsplan, der die Kontrollvariablen und einen, der die Störvariablen beinhaltet. Diese zwei Versuchspläne werden dann miteinander gekreuzt und man erhält einen Produkt- bzw. Crossed-Array Versuchsplan. Dabei wird der Versuchsplan der Kontrollvariablen als inneres Array und der Versuchsplan der Störvariablen als äußeres Array bezeichnet.

Im Crossed-Array Versuchsplan wird jede Einstellung der Kontrollvariablen mit jeder Einstellung der Störvariablen kombiniert. Dadurch wird die Anzahl der Versuche, die man durchführen muss, allerdings sehr groß. Dies ist ein erheblicher Nachteil des Crossed-Array Versuchsplans. Man muss entweder alle Versuche durchführen, was oft sehr teuer ist bzw. sehr viel Zeit in Anspruch nimmt oder man verwendet fraktionierte Versuchspläne, was zu einem Informationsverlust führt. Nehmen wir zum Beispiel für das innere Array einen fraktionierten Versuchsplan, muss man Informationsverlust bei den Interaktionen zwischen Kontroll- und Kontrollvariablen hinnehmen. Abhängig von der Größe des Versuchsplans können wir durch die Aliasbildung sogar falsche Informationen über die Haupteffekte der Kontrollvariablen erhalten. Sehr wichtig ist, dass man aber sehr wohl Informationen über die Interaktionen zwischen Kontroll- und Störvariablen erhält. Diese Interaktionen spielen im RPD-Problem eine wichtige Rolle.

Wir betrachten dazu zwei Beispiele aus Myers & Montgomery (2002). In beiden Beispielen haben wir drei Inputvariablen, wobei A und B Kontrollvariablen sind und C eine Störvariable ist. Die Daten des ersten Beispiels kann man in Tabelle 4.1 ablesen. Wie bereits erwähnt, ist C eine Störvariable und daher im Experiment

Inneres Array		Äußeres Array	
A	B	$C = -1$	$C = 1$
-1	-1	11	15
-1	1	7	8
1	-1	10	26
1	1	10	14

Tabelle 4.1.: Beispiel eines Crossed-Array Versuchsplans

kontrollierbar, in der Realität aber nicht. Betrachtet man jetzt die Interaktionsplots in Abbildung 4.1 zwischen der Störvariablen C und den zwei Kontrollvariablen, sehen wir, dass die Variabilität bei der Einstellung $A = -1$ und $B = 1$ am geringsten ist. Die größte Variabilität tritt hingegen bei der Einstellung $A = 1$ und $B = -1$ auf.

Man unterscheidet zwischen zwei Arten von Effekten. Es gibt die sogenannten Dispersions- und die Lokationseffekte. Die Dispersionswirkungen wirken sich auf die Prozessvarianz, und die Lokationseffekte auf den Prozessmittelwert aus. Man sieht bei den Interaktionsplots in Abbildung 4.1, dass sowohl A als auch B ein Dispersionswirkung ist. Die Varianz zwischen den unterschiedlichen Stufen von C ist bei der Einstellung $A = 1$ viel größer als bei $A = -1$. Weiters kann man auch erkennen, dass beide Faktoren Lokationseffekte sind, da sie Auswirkungen auf den Prozessmittelwert haben. Wollen wir jetzt den Mittelwert des Prozesses maximieren, sieht man sehr gut, dass man bei der robusten Optimierung Kompromisse eingehen muss, da die optimale Einstellung $A = 1$ und $B = -1$ ist. Diese Einstellung maximiert aber, wie vorher festgestellt wurde, die Varianz.

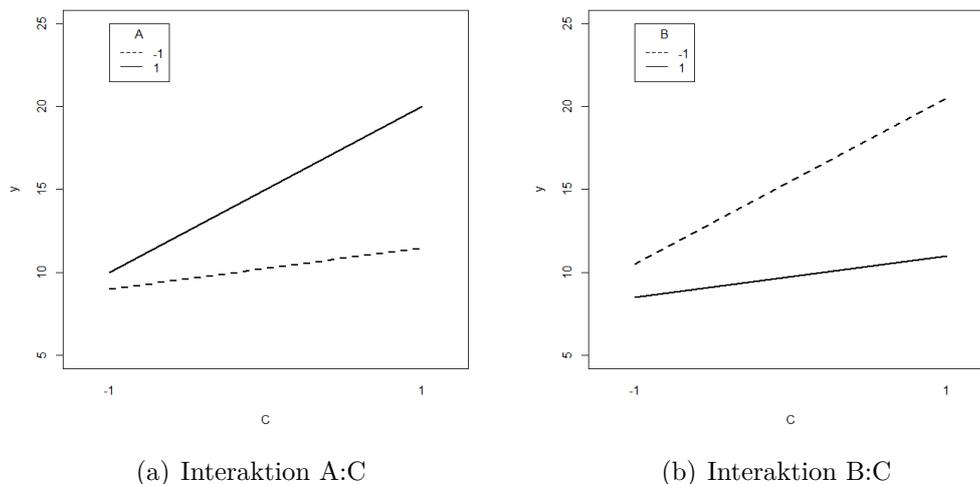
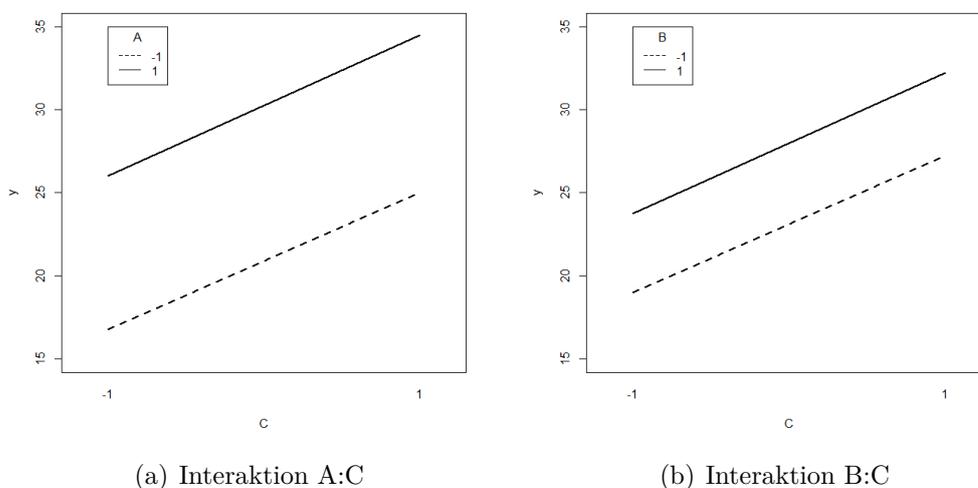


Abbildung 4.1.: Signifikante Kontroll \times Stör Interaktionen

Im Weiteren betrachten wir das zweite Beispiel - gegeben in Tabelle 4.2 - mit den Interaktionsplots in Abbildung 4.2. Man sieht, dass weder die Interaktion $A : C$ noch die Interaktion $B : C$ relevant sind. In beiden Fällen sind die Linien der unterschiedlichen Stufen parallel und daraus schließen wir, dass weder A noch B ein Dispersionswirkung ist. Beide Faktoren sind allerdings Lokationseffekte und man kann erkennen, dass A die größere Auswirkung auf den Mittelwert des Prozesses hat.

Inneres Array		Äußeres Array	
A	B	$C = -1$	$C = 1$
-1	-1	14.5	22.5
-1	1	19.0	27.5
1	-1	23.5	32.0
1	1	28.5	37.0

Tabelle 4.2.: Beispiel eines Crossed-Array Versuchsplans

Abbildung 4.2.: Nicht signifikante Kontroll \times Stör Interaktionen

Wenn wir hier den Mittelwert des Prozesses maximieren wollen, wählen wir die Einstellung $A = 1$ und $B = 1$. Da weder der Faktor A noch der Faktor B Einfluss auf die Varianz hat, liegt hier kein RPD-Problem vor. Das heißt, wir können aus den Interaktionen - natürlich auch Interaktionen höherer Ordnung - zwischen Stör- und Kontrollvariablen die Struktur der Prozessvarianz ablesen.

4.1.2. Analyse des Produkt-Versuchsplans

Taguchi benutzte zur Analyse zwei Statistiken. Zum einen verwendet er den Durchschnitt aller Beobachtungen des inneren Arrays über alle Einstellungen des äußeren Arrays und zum anderen die Statistik Signal-to-noise Ratio (SNR). Diese kombiniert Informationen von Mittelwert und Varianz und ist so definiert, dass der maximale Wert von SNR die Varianz minimiert, die durch die Störvariablen entsteht. Das heißt, man sucht die Einstellung, welche die Response-Variable so nahe wie möglich zum Zielwert bringt und eine maximale SNR hat. Man muss allerdings sehr vorsichtig bei der Anwendung der Statistik SNR sein. Sie vermischt die Lokations- und Dispersionsseffekte und findet nicht immer den Punkt mit der geringsten Variabilität.

Eine weitere Möglichkeit besteht darin, den Mittelwert und die Varianz direkt zu modellieren. Die Werte \bar{y}_i und s^2 werden über alle Beobachtungen berechnet, welche die gleiche Einstellung der Störvariablen haben. Man kann dann die Unterschiede dieser Werte allein auf die unterschiedlichen Einstellungen der Kontrollvariablen zurückführen. Man sucht hier wieder die Einstellung, die den Mittelwert optimiert und die Varianz minimiert. Ein Nachteil dieser Methode ist, dass die Interaktionen zwischen Stör- und Kontrollvariablen nicht direkt betrachtet werden und dadurch diese Beziehung manchmal verdeckt wird. Außerdem ist die Funktion, welche die Beziehung zwischen Response-Variable und Kontrollvariablen beschreibt, in den meisten Fällen nicht linear. Dies erschwert die Modellbildung. Der interessierte Leser findet weitere Details zum Ansatz von Taguchi bei Myers & Montgomery (2002).

4.1.3. Kombierter Versuchsplan

Eine Alternative zum Produkt-Versuchsplan bildet der kombinierte Versuchsplan (Combined Array). Dies ist ein Versuchsplan, der sowohl die Kontroll- als auch die Störvariablen beinhaltet. Man modelliert die Response-Variablen bei diesem Versuchsplan mit Kontroll- und Störvariablen und deren Interaktionen. Betrachtet man ein Modell 1. Ordnung mit zwei Kontrollvariablen x_1 und x_2 und einer Störvariable z_1 erhalten wir

$$Y(\mathbf{x}, z_1) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \gamma_1 z_1 + \delta_{11} x_1 z_1 + \delta_{21} x_2 z_1 + \epsilon. \quad (4.1)$$

Dieses Modell bezeichnet man als Response-Modell.

Wie im Abschnitt 4.1.1 gesehen, erhalten wir nur ein RPD-Problem, wenn Interaktionen zwischen Kontroll- und Störvariablen vorhanden sind. Das heißt, im Modell (4.1) muss zumindest einer der δ -Faktoren ungleich 0 sein. Wir nehmen jetzt an, dass sowohl Kontroll- als auch Störvariablen in der üblichen kodierten Form vorliegen, dass die Störvariablen Zufallsvariablen mit Erwartungswert 0 und Varianz σ_z^2 sind und dass die Kovarianz zwischen den Störvariablen 0 ist. Das Modell für den Mittelwert der Response-Variable erhält man durch Erwartungswertbildung von (4.1). Der Erwartungswert bezüglich beider Zufallsvariablen (z_1 und ϵ) von Y ist

$$E[Y(\mathbf{x}, z_1)] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2.$$

Man sieht hier, dass das Erwartungswertmodell nur aus Kontrollvariablen besteht. Wendet man einen bedingten Varianzoperator auf $Y(\mathbf{x}, z_1)$ an, dann erhält man ein Modell für die Varianz des Prozesses

$$\begin{aligned} \text{Var}[Y(\mathbf{x}, z_1)] &= \text{Var}[\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \gamma_1 z_1 + \delta_{11} x_1 z_1 + \delta_{21} x_2 z_1 + \epsilon] \\ &= \text{Var}[\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + (\gamma_1 + \delta_{11} x_1 + \delta_{21} x_2) z_1 + \epsilon] \\ &= \sigma_z^2 (\gamma_1 + \delta_{11} x_1 + \delta_{21} x_2)^2 + \sigma^2. \end{aligned}$$

Auch hier wird die Varianz bezüglich beiden Zufallsvariablen gebildet. Man sieht, dass die Varianz eine quadratische Funktion der Kontrollvariablen ist und dass das

Varianzmodell auch die δ -Interaktionskoeffizienten enthält. Durch die Interaktionskoeffizienten wird die Response-Variable von den Störvariablen beeinflusst. Da beide Modelle nur die Kontrollvariablen beinhalten, kann man den Zielwert mit Hilfe einer adäquaten Einstellung der Kontrollvariablen erreichen und gleichzeitig die Variabilität mit Hilfe der Störvariablen minimieren.

Hat man einen kombinierten Versuchsplan gegeben, sucht man adäquate Response-Modelle. Die Regressionskoeffizienten in beiden Modellen werden durch ihre Kleinst-Quadrat-Schätzer und σ^2 durch den MSR des Response-Modells ersetzt. Anschließend kann man das Erwartungswert- und das Varianzmodell mit Hilfe von Methoden aus Abschnitt 3 optimieren. Vorteile dieses Ansatzes sind, dass man jetzt einzelne Interaktionen zwischen Stör- und Kontrollvariablen gegen Interaktionen zwischen Kontroll- und Kontrollvariablen tauschen kann und dass der kombinierte Versuchsplan im Allgemeinen effizienter als der Produkt-Versuchsplan ist. Für weitere Informationen zu diesem Thema siehe Myers & Montgomery (2002).

In den folgenden Abschnitten dieses Kapitels werden Methoden vorgestellt, die Prozesse bzw. Produkte mit mehreren Zielgrößen robust optimieren.

4.2. Robustes E-Modell

In diesem Abschnitt betrachten wir eine Methode von Hejazi et al. (2012), welche mehrere Zielgrößen gleichzeitig optimiert, dabei aber auch die Unsicherheit des stochastischen Modells berücksichtigt. Es fließen sowohl die Korrelationen zwischen den Response-Variablen als auch die Korrelationen zwischen den Regressionskoeffizienten β_i mit ein.

Die einzelnen Variablen erhalten je nach Bedeutung, welche der Variable bei der Optimierung zukommen soll, unterschiedliche Gewichte. Diese Gewichte werden am Beginn von einer Expertengruppe festgesetzt. Jeder Experte bewertet, wie wichtig jede einzelne Response-Variable seiner Meinung nach ist. Der Mittelwert und die Varianz dieser unterschiedlichen Bewertungen fließen in die Berechnung mit ein. Man gelangt schließlich mit Hilfe einer Zielfunktion mit Nebenbedingung zu einer robusten optimalen Lösung.

4.2.1. Formulierung des Problems

Man geht vom multivariaten Modell aus Abschnitt 3.1 aus. Zusätzlich wird ein Vektor $\mathbf{w} = (w_1, w_2, \dots, w_l)'$ definiert, der die Gewichte der Response-Variablen enthält. Die Varianz-Kovarianz Matrix dieser Gewichte ist $\Sigma_{\mathbf{w}}$. Seien weiters $\mathbf{l} = (l_1, l_2, \dots, l_k)'$ und $\mathbf{u} = (u_1, u_2, \dots, u_k)'$ die unteren bzw. oberen Grenzen der k Inputvariablen und $\mathbf{t} = (t_1, t_2, \dots, t_l)'$ der Vektor der Zielwerte der Response-Variablen, dann erhalten wir als Optimierungsproblem für mehrere Zielgrößen

$$\min_{\mathbf{x}} \hat{\mathbf{y}}(\mathbf{x}) \tag{4.2}$$

$$\text{u.d.N. } \mathbf{l} < \mathbf{x} < \mathbf{u}.$$

Dies ist ein nichtlineares multivariates Optimierungsproblem und $\mathbf{l} < \mathbf{x} < \mathbf{u}$ bedeutet $l_i < x_i < u_i$ für $i = 1, 2, \dots, k$ (vgl. beispielsweise Miettinen, 1999).

Alternativ kann man dieses multikriterielle Problem auch als Modell mit einer Zielfunktion und Nebenbedingungen definieren. Man erhält das deterministische Problem

$$\min_{\mathbf{x}} F(\mathbf{x}) = \min_{\mathbf{x}} \sum_{i=1}^l w_i(d_i + d'_i) \quad (4.3)$$

$$\text{u.d.N. } \mathbf{l} < \mathbf{x} < \mathbf{u}$$

$$\hat{y}_i(\mathbf{x}) + d_i - d'_i = t_i \quad k = 1, 2, \dots, l,$$

wobei

$$d_i = \frac{1}{2}(|\hat{y}_i(\mathbf{x}) - t_i| + (\hat{y}_i(\mathbf{x}) - t_i))$$

$$d'_i = \frac{1}{2}(|\hat{y}_i(\mathbf{x}) - t_i| - (\hat{y}_i(\mathbf{x}) - t_i)).$$

Wegen $Var(\hat{y}_i(\mathbf{x}) + d_i - d'_i) = 0$ und da $\hat{\beta}$ und $\hat{\Sigma}$ zwei Zufallsmatrizen sind, erhalten wir unter Annahme, dass \mathbf{w} ein unabhängiger Zufallsvektor ist, die stochastische Version des Optimierungsproblems (4.3)

$$\min_{\mathbf{x}} F(\mathbf{x}) = \min_{\mathbf{x}} \sum_{i=1}^l w_i(d_i + d'_i) \quad (4.4)$$

$$\text{u.d.N. } \mathbf{l} < \mathbf{x} < \mathbf{u}$$

$$\hat{y}_i(\mathbf{x}) + d_i - d'_i = t_i \quad k = 1, 2, \dots, l$$

$\mathbf{w} \in \Omega_{\mathbf{w}}$, die Schätzer von $E(\mathbf{w})$ und $Cov(\mathbf{w}) = \Sigma_{\mathbf{w}}$ sind bekannt

$$(\beta, \Sigma) \in \Omega_{\beta, \Sigma} \begin{cases} \hat{\beta} \sim \mathcal{N}_{p \times l}(\beta, (\mathbf{X}'\mathbf{X})^{-1} \otimes \Sigma) \quad \text{und} \\ (n-p)\hat{\Sigma} \sim \mathcal{W}_l(n-p, \Sigma) \end{cases}.$$

Die Bereiche $\Omega_{\mathbf{w}}$ und $\Omega_{\beta, \Sigma}$ sind die Definitionsbereiche der Variablen.

Um das stochastische Problem zu lösen, benutzt man deterministische Ansätze. Man nimmt an, dass die Gewichte und die Regressionskoeffizienten unabhängig sind und dass $\hat{E}(F(\mathbf{x})) \equiv E(\widehat{F}(\mathbf{x}))$ und $\widehat{Var}(F(\mathbf{x})) \equiv Var(\widehat{F}(\mathbf{x}))$, wobei

$$\widehat{E}(F(\mathbf{x})) = \sum_{i=1}^l \widehat{E}(w_i) \widehat{E}(d_i + d'_i) + \underbrace{Cov(w_i, (d_i + d'_i))}_{=0}$$

$$\widehat{Var}(F(\mathbf{x})) = \sum_{i=1}^l \widehat{Var}(w_i(d_i + d'_i)) + 2 \sum_{i=1}^{l-1} \sum_{j=i+1}^l \widehat{Cov}(w_i(d_i + d'_i), w_j(d_j + d'_j)).$$

Hejazi et al. (2012) beschreiben folgende Modelle aus der Literatur. Das nachfolgende Modell (4.5) wird modifiziertes E-Modell genannt und minimiert eine Kombination

aus dem Erwartungswert und der Standardabweichung der Zielfunktion. Das Modell ist durch

$$\min_{\mathbf{x}} r_1 \widehat{E}(F(\mathbf{x})) + r_2 \sqrt{\widehat{Var}(F(\mathbf{x}))} \quad (4.5)$$

$$\text{u.d.N. } \mathbf{l} < \mathbf{x} < \mathbf{u}$$

$$\hat{y}_i(\mathbf{x}) + d_i - d'_i = t_i \quad k = 1, 2, \dots, l$$

gegeben. Die Konstanten r_1 und r_2 sind nicht-negativ, erfüllen im Allgemeinen die Gleichung $r_1 + r_2 = 1$ und geben an wie viel Bedeutung dem Erwartungswert bzw. der Varianz zukommen soll.

Ein weiteres Modell ist das modifizierte V-Modell. Es ist durch

$$\min_{\mathbf{x}} \sqrt{\widehat{Var}(F(\mathbf{x}))} \quad (4.6)$$

$$\text{u.d.N. } \mathbf{l} < \mathbf{x} < \mathbf{u}$$

$$\hat{y}_i(\mathbf{x}) + d_i - d'_i = t_i \quad k = 1, 2, \dots, l$$

$$\widehat{E}(F(\mathbf{x})) \in \Omega_{E(F(\mathbf{x}))} \quad (4.7)$$

definiert. Der Bereich $\Omega_{E(F(\mathbf{x}))}$ ist der gewünschte Bereich für $\widehat{E}(F(\mathbf{x}))$.

Ein weiteres deterministisches Modell, mit welchem man das stochastische Optimierungsproblem lösen kann, ist

$$\min_{\mathbf{x}} \frac{\widehat{E}(F(\mathbf{x}))}{\sqrt{\widehat{Var}(F(\mathbf{x}))}} \quad (4.8)$$

$$\text{u.d.N. } \mathbf{l} < \mathbf{x} < \mathbf{u}$$

$$\hat{y}_i(\mathbf{x}) + d_i - d'_i = t_i \quad k = 1, 2, \dots, l.$$

4.2.2. Robustes E-Modell

Hejazi et al. (2012) schlagen vor das folgende Modell einzusetzen, um das stochastische Optimierungsproblem zu lösen und dabei auch auf Robustheit Rücksicht zu nehmen. Das robuste E-Modell ist durch

$$\min_{\mathbf{x}} \widehat{E}(F(\mathbf{x})) + z_{1-\alpha} \sqrt{\widehat{Var}(F(\mathbf{x}))} \quad (4.9)$$

$$\text{u.d.N. } \mathbf{l} < \mathbf{x} < \mathbf{u}$$

$$\hat{y}_i(\mathbf{x}) + d_i - d'_i = t_i \quad k = 1, 2, \dots, l$$

definiert.

Man sieht, dass die Autoren in (4.9) die obere Grenze des $(1 - \alpha)$ -Konfidenzintervalls vom Erwartungswert der Zielfunktion minimieren wollen. Das heißt, dass diese Methode eine Kombination ist aus dem modifizierten E-Modell und der Minimax Methode, die oft in der robusten Optimierung eingesetzt wird. Der interessierte Leser findet bei Hejazi et al. (2012) ein Beispiel, bei welchem diese Methode mit den davor beschriebenen Methoden verglichen wird.

4.3. Robuste Optimierung mit einem Konfidenzbereich

Die in diesem Abschnitt beschriebene Methode von He et al. (2012) berücksichtigt bei der Optimierung von mehreren Zielgrößen ebenfalls die Unsicherheit des gefitteten Regressionsmodells. Die Robustheit wird bei dieser Methode mit Hilfe eines Konfidenzbereiches miteinbezogen. Nach der Berechnung des Konfidenzbereiches wendet man eine Desirability-Funktion, welche wir in Abschnitt 3.4 eingeführt haben, auf den „Worst-case“ des Konfidenzbereiches an.

4.3.1. Methode

Man definiert sich die Gesamt-Desirability-Funktion für l Response-Variablen mit einem gewichteten geometrischen Mittel

$$D(d_1(y_1), d_2(y_2), \dots, d_l(y_l)) = \left(\prod_{i=1}^l d_i(y_i)^{w_i} \right)^{1/\sum w_i}.$$

Das Ziel ist es, die Einstellung der Inputvariablen zu finden, welche die Funktion

$$\hat{D}(\mathbf{x}) = D(d_1[\hat{y}_1(\mathbf{x})], d_2[\hat{y}_2(\mathbf{x})], \dots, d_l[\hat{y}_l(\mathbf{x})]) = \left(\prod_{i=1}^l d_i[\hat{y}_i(\mathbf{x})]^{w_i} \right)^{1/\sum w_i}$$

maximiert. Die l Response-Variablen werden mit Modellen 2. Ordnung gefittet. Da die Methode von Derringer & Suich (1980) nur dann gute Ergebnisse liefert, wenn die Modellanpassung gut ist, betrachten die Autoren bei dieser Methode anstatt eines Punktschätzers der Response-Variable ein Konfidenzintervall.

Ein $(1 - \alpha)$ -Konfidenzintervall für den wahren Response-Wert $y(\mathbf{x})$ ist gegeben durch

$$[y_U(\mathbf{x}), y_O(\mathbf{x})] = \left[\hat{y}(\mathbf{x}) \mp t_{n-p, 1-\alpha/2} \sqrt{\hat{\sigma}^2 \mathbf{z}'(\mathbf{x})(\mathbf{X}'\mathbf{X})^{-1} \mathbf{z}(\mathbf{x})} \right],$$

wobei hier n wieder die Anzahl der Beobachtungen und p die Anzahl der zu schätzenden Parameter ist. Der Ausdruck unter der Wurzel ist, wie wir bereits in Abschnitt 3.1 gesehen haben, die Varianz der geschätzten Response.

Da wir l Response-Variablen gleichzeitig optimieren wollen, sehen wir uns den gemeinsamen Konfidenzbereich aller l Variablen an. Sei jetzt $[y_{U_i}(\mathbf{x}), y_{O_i}(\mathbf{x})]$ das $(1 - \alpha)^{1/l}$ -Konfidenzintervall der i -ten Response-Variable, so ist der gemeinsame Konfidenzbereich aller l Response-Variablen gegeben durch

$$\Omega(\mathbf{x}) = [y_{U_1}(\mathbf{x}), y_{O_1}(\mathbf{x})] \times [y_{U_2}(\mathbf{x}), y_{O_2}(\mathbf{x})] \times \dots \times [y_{U_l}(\mathbf{x}), y_{O_l}(\mathbf{x})]. \quad (4.10)$$

Die simultane Überdeckungswahrscheinlichkeit ist bei diesem Konfidenzbereich $1 - \alpha$. Wir nehmen an, dass der wahre Response-Wert jeden Punkt $\boldsymbol{\eta} = [\eta_1, \eta_2, \dots, \eta_l]'$ aus dem $(1 - \alpha)$ -Konfidenzbereich annehmen kann und definieren uns die robuste Gesamt-Desirability-Funktion als die Desirability-Funktion des Punktes aus dem Konfidenz-

bereich, der das schlechteste Ergebnis liefert. Das bedeutet

$$\begin{aligned} D_R(\mathbf{x}) &= \min_{\boldsymbol{\eta}} \{D[d_1(\eta_1), d_2(\eta_2), \dots, d_l(\eta_l)] \mid (\eta_1, \eta_2, \dots, \eta_l) \in \Omega(\mathbf{x})\} \\ &= \min_{\boldsymbol{\eta}} \left[\left(\prod_{i=1}^l d_i(\eta_i)^{w_i} \right)^{1/\sum w_i} \mid (\eta_1, \eta_2, \dots, \eta_l) \in \Omega(\mathbf{x}) \right]. \end{aligned}$$

Wenn man die Einstellung findet, welche die Funktion $D_R(\mathbf{x})$ maximiert, hat man das robuste Optimum gefunden.

Die ursprüngliche Desirability-Funktion ist ein Spezialfall dieser Funktion und kann erreicht werden, in dem man α auf 1, also das Konfidenzniveau auf 0 setzt. Bei steigendem Konfidenzniveau ($1 - \alpha$), wächst der Konfidenzbereich und es wird bei der Optimierung mehr Wert auf die Robustheit gelegt. Mit dem Wert α kann man also festlegen, wie stark man die Robustheit gewichtet.

Mit Hilfe von (4.10) kann man die robuste Gesamt-Desirability-Funktion noch ein wenig umschreiben und man erhält

$$D_R(\mathbf{x}) = \left(\prod_{i=1}^l \left[\min_{\eta_i} \{d_i(\eta_i) \mid \eta_i \in [y_{U_i}(\mathbf{x}), y_{O_i}(\mathbf{x})]\} \right]^{w_i} \right)^{1/\sum w_i}.$$

Um die Desirability-Funktionen aller drei Fälle, die auftreten können, angeben zu können, definiert man sich mit

$$d_{Ri}(\mathbf{x}) = \min_{\eta_i} \{d_i(\eta_i) \mid \eta_i \in [y_{U_i}(\mathbf{x}), y_{O_i}(\mathbf{x})]\}$$

die individuelle robuste Desirability-Funktion. Wegen der Definition der Desirability-Funktionen weiß man, dass das Minimum angenommen wird, wenn man η_i entweder auf die obere oder untere Grenze des Konfidenzintervalls setzt. Wir erhalten für den Larger-the-better Fall

$$d_{Ri}(\mathbf{x}) = d_i[y_{U_i}(\mathbf{x})]$$

und für den Smaller-the-better Fall

$$d_{Ri}(\mathbf{x}) = d_i[y_{O_i}(\mathbf{x})].$$

Auch für den Nominal-the-best Fall können wir eine allgemeine Formel angeben. Es muss allerdings gelten, dass die Desirability-Funktion symmetrisch ist. Die individuelle robuste Desirability-Funktion ist gegeben durch

$$d_{Ri}(\mathbf{x}) = \begin{cases} d_i[y_{U_i}(\mathbf{x})] & |y_{U_i}(\mathbf{x}) - c_i| \geq |y_{O_i}(\mathbf{x}) - c_i| \\ d_i[y_{O_i}(\mathbf{x})] & \text{sonst} \end{cases}.$$

Wir erhalten insgesamt

$$D_R(\mathbf{x}) = \left(\prod_{i=1}^l [d_{Ri}(\mathbf{x})]^{w_i} \right)^{1/\sum w_i}$$

und setzen je nach vorliegendem Fall die gewünschte individuelle robuste Desirability-Funktion ein.

4.3.2. Optimierung

He et al. (2012) verwenden für die Optimierung kein direktes Suchverfahren, mit der Begründung, dass man damit, bei Existenz mehrerer Optima, bei einem lokalen Optimum hängen bleiben kann. Man müsste dieses Verfahren mit mehreren Startpunkten durchführen, um das zu vermeiden. Dies kann allerdings zu einem enormen Rechenaufwand führen. Sie verwenden daher einen genetischen Algorithmus, der sich auf Darwin's „Survival of the fittest“-Theorie stützt, um in die Nähe des globalen Optimums zu kommen. Ist man dort angelangt, schlagen sie eine Muster-Suchmethode vor, um das Optimum genau zu bestimmen.

4.4. Robuste Optimierung mit Desirability-Funktionen

Wir betrachten in diesem Abschnitt die Methode von He et al. (2010), die bei der Optimierung mehrerer Zielgrößen als Maß für die Robustheit eine Desirability-Funktion verwendet. Die Optimierung wird ebenfalls mit Hilfe einer Desirability-Funktion durchgeführt. Anschließend werden diese zwei Ansätze zu einer Gesamt-Desirability-Funktion zusammengeführt. Diese Methode berücksichtigt aber die Unsicherheit des Modells nicht und geht davon aus, dass das Modell die Beziehung zwischen Response- und Inputvariablen angemessen wiedergibt.

4.4.1. Robustheitsmaß

Definition 4.4.1. (*Sphärischer Bereich*) Man bezeichnet

$$\sum_{i=1}^k (x_i - C_i)^2 \leq r^2,$$

wobei $0 \leq r \leq \delta$ und üblicherweise $\delta \leq 0.1$, als sphärischen Bereich des Punktes (C_0, C_1, \dots, C_k) .

Das Maximum und das Minimum der geschätzten Response-Variable im sphärischen Bereich erhält man, indem man die Optimierungsprobleme

$$\max \hat{y}_i \text{ oder } \min \hat{y}_i \tag{4.11}$$

$$\text{u.d.N. } \sum_{i=1}^k (x_i - C_i)^2 \leq r^2$$

$$\mathbf{x} \in R$$

löst. R ist auch hier wieder der Bereich, in welchem das Experiment durchgeführt wird. Aus diesem Maximum und Minimum definieren wir uns den geschätzten Bereich der Response-Variable um den Punkt (C_0, C_1, \dots, C_k) .

Definition 4.4.2. (*geschätzter Bereich*) Für einen gegebenen Punkt (C_0, C_1, \dots, C_k) in einem k -dimensionalen Raum, definieren wir uns den geschätzten Bereich der Response-Variable im sphärischen Bereich von (C_0, C_1, \dots, C_k) mit

$$B_i = (\max \hat{y}_i) - (\min \hat{y}_i),$$

wobei die Optima aus (4.11) erhalten werden.

Ist B_i klein genug, bezeichnet man den Punkt (C_0, C_1, \dots, C_k) für die Response-Variable y_i im definierten Bereich als robust. Um eine robuste Desirability-Funktion einzuführen, geht man folgendermaßen vor. Bezeichnet man mit U_i die untere Grenze und mit O_i die obere Grenze des zulässigen Bereichs der Response-Variable y_i , so erhält man sechs verschiedene Fälle, für die Lokation von Maximum und Minimum aus (4.11).

1. $\min \hat{y}_i \leq \max \hat{y}_i \leq U_i \leq O_i$
2. $\min \hat{y}_i \leq U_i \leq \max \hat{y}_i \leq O_i$
3. $\min \hat{y}_i \leq U_i \leq O_i \leq \max \hat{y}_i$
4. $U_i \leq \min \hat{y}_i \leq \max \hat{y}_i \leq O_i$
5. $U_i \leq \min \hat{y}_i \leq O_i \leq \max \hat{y}_i$
6. $U_i \leq O_i \leq \min \hat{y}_i \leq \max \hat{y}_i$

In Abbildung 4.3 sieht man eine grafische Darstellung dieser sechs Fälle.

In den Fällen 1 und 6 setzen wir die Desirability-Funktion sofort auf 0, da hier die Response-Variable außerhalb des zulässigen Bereichs liegt. In den anderen Fällen transformiert man den vorher eingeführten geschätzten Bereich der Response-Variable in einen robusten Desirability-Wert d_{ri} . Ist $B_i = 0$, setzt man $d_{ri} = 1$. Sollte aber $B_i > O_i - U_i$ sein, dann setzen wir d_{ri} auf 0. Man erhält für die i -te individuelle robuste Desirability-Funktion (Nominal the best)

$$d_{ri} = \begin{cases} 0 & B_i \geq O_i - U_i \\ \left(1 - \frac{B_i}{O_i - U_i}\right)^r & 0 < B_i < O_i - U_i \\ 1 & B_i = 0 \end{cases} .$$

Im Larger-the-better Fall setzen wir für die obere Grenze O_i den Zielwert ein. Umgekehrt können wir im Smaller-the-better Fall die untere Grenze U_i durch den Zielwert ersetzen. Der Wert r hat hier die gleiche Bedeutung wie in Abschnitt 3.4.

Die individuellen robusten Desirability-Funktionen werden dann mit Hilfe des geometrischen Mittels zu einer gemeinsamen robusten Desirability-Funktion

$$D_{robust} = (d_{r1}d_{r2} \dots d_{rl})^{1/l} \quad (4.12)$$

zusammengeführt. Das globale Ziel ist es die Funktion D_{robust} zu maximieren.

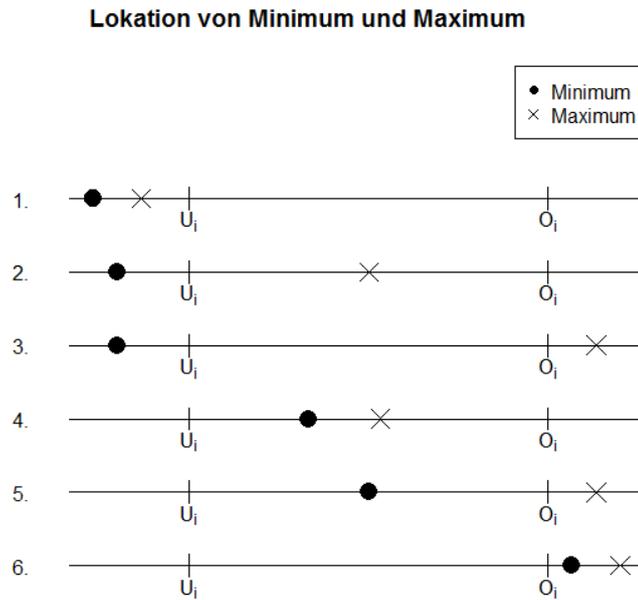


Abbildung 4.3.: Grafische Darstellung der sechs Fälle

4.4.2. Optimierungs-Desirability-Funktion

Die individuellen Optimierungs-Desirability-Funktionen d_{oi} werden wie in Abschnitt 3.4 aufgestellt und mittels geometrischen Mittel zur Optimierungs-Desirability-Funktion

$$D_{opt} = (d_{o1}d_{o2} \dots d_{ol})^{1/l} \quad (4.13)$$

zusammengeführt.

4.4.3. Vorgangsweise

Bei dieser Methode geht man davon aus, dass die Response-Variablen unabhängig und unkorreliert sind. Man stellt sich für jede Response-Variable ein Regressionsmodell auf und berechnet den optimalen Punkt der Funktion (4.13). Man definiert sich die Gesamt-Desirability-Funktion als

$$D_{gesamt} = (D_{opt})^{w_1}(D_{robust})^{w_2}.$$

Es gilt nun die Funktion D_{gesamt} zu maximieren, um den robusten optimalen Punkt zu erhalten. Die Autoren schlagen zur Lösung des Problems das Simplexverfahren vor. Der Anfangspunkt des Verfahrens, dessen Anwendung für diese Methode in A.4 erklärt wird, ist der optimale Punkt der Funktion (4.13).

Die Variablen w_1 und w_2 sind Gewichte, die man je nach Bedeutung, die man der Robustheit bzw. der Optimierung zukommen lassen will, wählen kann. Im Allgemeinen gilt, dass $w_1 + w_2 = 1$. Man kann bei dieser Methode die Gewichtung nicht nur

über die w_1 und w_2 steuern, sondern auch über r , s und t . In der Praxis werden letztere immer auf 1 gesetzt und w_1 und w_2 werden variiert.

5. Anwendung

In diesem Kapitel werden die in Kapitel 4 beschriebenen Methoden auf ein Datenbeispiel angewendet. Es wird dabei untersucht, wie sich bei einem Ätzprozess die Faktoren *Temperatur* (x_1), *Geschwindigkeit* (x_2), *Oszillation* (x_3) und *Druck* (x_4) auf die Zielgrößen $y_1 = C_p$ -Wert und $y_2 = \text{Ätzfaktor}$ auswirken.

Die Ergebnisse des Versuchs findet man in Tabelle 5.1. Das Ziel ist es, die zwei Ziel-

Wiederholung					1	2	3	1	2	3
Nr.	x_1	x_2	x_3	x_4	C_p -Wert			Ätzfaktor		
1	-1	-1	-1	-1	1,59	1,58	1,57	0.64	0.65	0.65
2	1	-1	1	-1	1,33	1,22	1,29	0.51	0.51	0.50
3	-1	1	1	1	1,47	1,47	1,47	1.25	1.26	1.25
4	1	1	-1	1	1,53	1,50	1,69	0.51	0.50	0.54
5	-1	-1	1	1	1,43	1,38	1,30	1.01	1.12	1.21
6	1	-1	-1	1	1,84	1,94	1,79	0.83	0.80	0.75
7	-1	1	-1	-1	1,54	1,59	1,64	0.91	0.84	0.83
8	1	1	1	-1	1,29	1,20	1,23	0.50	0.53	0.50
9	-1	-1	-1	1	1,46	1,35	1,28	0.60	0.59	0.53
10	1	-1	1	1	1,05	0,96	1,03	1.33	1.16	1.33
11	-1	1	1	-1	1,61	1,88	1,60	0.91	0.91	0.97
12	1	1	-1	-1	1,80	2,05	2,24	0.97	0.95	0.95
13	-1	-1	1	-1	1,78	1,80	1,56	0.45	0.42	0.42
14	1	-1	-1	-1	1,59	1,76	1,69	0.44	0.46	0.49
15	-1	1	-1	1	1,73	1,84	1,91	0.87	0.92	0.89
16	1	1	1	1	1,76	1,97	1,89	0.73	0.61	0.59

Tabelle 5.1.: Ergebnisse des Experiments

größen zu maximieren. Da die Faktoren nur auf zwei Stufen untersucht wurden, wird für beide Zielgrößen ein lineares Modell aufgestellt. Wir erhalten folgende Analysen für die zwei Modelle.

```
> cp.lm<-lm(cpvalue~Atemp*Bspeed*Eoscillation*Fpressure, data=etchingneu.dat)
> summary(cp.lm)
```

Call:

```
lm(formula = cpvalue ~ Atemp * Bspeed * Eoscillation * Fpressure,
    data = etchingneu.dat)
```

4.4. Robuste Optimierung mit Desirability-Funktionen

Residuals:

Min	1Q	Median	3Q	Max
-0.230	-0.055	0.000	0.050	0.210

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	1.572292	0.014310	109.874	< 2e-16	***
Atemp	-0.003958	0.014310	-0.277	0.783855	
Bspeed	0.090208	0.014310	6.304	4.52e-07	***
Eoscillation	-0.115208	0.014310	-8.051	3.42e-09	***
Fpressure	-0.028958	0.014310	-2.024	0.051424	.
Atemp:Bspeed	0.020625	0.014310	1.441	0.159213	
Atemp:Eoscillation	-0.101458	0.014310	-7.090	4.83e-08	***
Bspeed:Eoscillation	0.022708	0.014310	1.587	0.122370	
Atemp:Fpressure	0.039792	0.014310	2.781	0.009013	**
Bspeed:Fpressure	0.052292	0.014310	3.654	0.000915	***
Eoscillation:Fpressure	0.003542	0.014310	0.247	0.806105	
Atemp:Bspeed:Eoscillation	0.071458	0.014310	4.994	2.02e-05	***
Atemp:Bspeed:Fpressure	-0.018958	0.014310	-1.325	0.194609	
Atemp:Eoscillation:Fpressure	0.077292	0.014310	5.401	6.18e-06	***
Bspeed:Eoscillation:Fpressure	0.074792	0.014310	5.227	1.03e-05	***
Atemp:Bspeed:Eoscillation:Fpressure	0.116875	0.014310	8.167	2.50e-09	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09914 on 32 degrees of freedom

Multiple R-squared: 0.9127, Adjusted R-squared: 0.8718

F-statistic: 22.31 on 15 and 32 DF, p-value: 8.903e-13

```
> cp.lm<-lm(etCHF~Atemp*Bspeed*Eoscillation*Fpressure, data=etchingneu.dat)
> summary(cp.lm)
```

Call:

```
lm(formula = etCHF ~ Atemp * Bspeed * Eoscillation * Fpressure,
    data = etchingneu.dat)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.113333	-0.017500	-0.003333	0.020000	0.096667

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	0.772708	0.006663	115.963	< 2e-16	***
Atemp	-0.064792	0.006663	-9.723	4.48e-11	***
Bspeed	0.047708	0.006663	7.160	3.97e-08	***
Eoscillation	0.059792	0.006663	8.973	3.00e-10	***
Fpressure	0.109792	0.006663	16.477	< 2e-16	***
Atemp:Bspeed	-0.098958	0.006663	-14.851	6.58e-16	***
Atemp:Eoscillation	-0.034375	0.006663	-5.159	1.25e-05	***
Bspeed:Eoscillation	-0.046042	0.006663	-6.910	8.03e-08	***
Atemp:Fpressure	-0.011042	0.006663	-1.657	0.10728	
Bspeed:Fpressure	-0.103542	0.006663	-15.539	< 2e-16	***
Eoscillation:Fpressure	0.128542	0.006663	19.291	< 2e-16	***
Atemp:Bspeed:Eoscillation	-0.059375	0.006663	-8.911	3.52e-10	***

```

Atemp:Bspeed:Fpressure      -0.071875    0.006663  -10.787  3.44e-12  ***
Atemp:Eoscillation:Fpressure -0.002292    0.006663   -0.344  0.73316
Bspeed:Eoscillation:Fpressure -0.020625    0.006663   -3.095  0.00407  **
Atemp:Bspeed:Eoscillation:Fpressure 0.037708    0.006663    5.659  2.92e-06  ***

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04617 on 32 degrees of freedom
Multiple R-squared: 0.9809, Adjusted R-squared: 0.972
F-statistic: 109.8 on 15 and 32 DF, p-value: < 2.2e-16

Die gefittete Response-Funktion des C_p -Wertes ($\hat{y}_1(\mathbf{x})$) und des Ätzfaktors ($\hat{y}_2(\mathbf{x})$) ist gegeben durch

$$\begin{aligned} \hat{y}_1(\mathbf{x}) = & 1.572 - 0.004x_1 + 0.090x_2 - 0.115x_3 - 0.029x_4 + 0.021x_1x_2 \\ & - 0.101x_1x_3 + 0.040x_1x_4 + 0.023x_2x_3 + 0.052x_2x_4 + 0.004x_3x_4 + 0.071x_1x_2x_3 \\ & - 0.019x_1x_2x_4 + 0.077x_1x_3x_4 + 0.075x_2x_3x_4 + 0.117x_1x_2x_3x_4 \end{aligned} \quad (5.1)$$

und

$$\begin{aligned} \hat{y}_2(\mathbf{x}) = & 0.773 - 0.065x_1 + 0.048x_2 + 0.060x_3 + 0.110x_4 - 0.099x_1x_2 \\ & - 0.034x_1x_3 - 0.011x_1x_4 - 0.046x_2x_3 - 0.104x_2x_4 + 0.129x_3x_4 - 0.059x_1x_2x_3 \\ & - 0.072x_1x_2x_4 - 0.002x_1x_3x_4 - 0.021x_2x_3x_4 + 0.038x_1x_2x_3x_4. \end{aligned} \quad (5.2)$$

Mit Hilfe dieser Funktionen berechnen wir uns die individuellen Optima und können damit herausfinden, auf welchem Bereich unsere Response-Funktionen variieren. Wir sehen in Tabelle 5.2, dass die Maxima der beiden Response-Funktionen nicht bei der gleichen Parametereinstellung angenommen werden. Das heißt, wir werden hier Kompromisse eingehen müssen.

Zielgröße		x_1	x_2	x_3	x_4	Wert
C_p -Wert	Max	1	1	-1	-1	2.03
	Min	1	-1	1	1	1.01
Ätzfaktor	Max	-1	1	1	1	1.25
	Min	1	-1	-1	-1	0.46

Tabelle 5.2.: Individuelle Optima

Wir wollen mit den drei Methoden, die wir kennen gelernt haben, beide Zielgrößen simultan und robust optimieren. Dafür setzen wir uns bestimmte Grenzen fest. Die untere Grenze für den C_p -Wert ist 1.33 und für den Ätzfaktor 0.80. Da in beiden Fällen ein Larger-the-better Fall vorliegt, sind die oberen Grenzen gleich den individuellen Maxima und liegen bei 2.03 und 1.25. Bei den ersten zwei Methoden kann man den Zielgrößen unterschiedliche Gewichte geben. Es werden jeweils die drei Fälle ($w_1 = 0.5, w_2 = 0.5$), ($w_1 = 0.2, w_2 = 0.8$) und ($w_1 = 0.05, w_2 = 0.95$) betrachtet.

5.1. Robustes E-Modell

Die Experten setzen den Zielvektor auf (1.66, 1.00). Wir gehen davon aus, dass in unserem Fall die Gewichte von den Experten fix festgelegt werden und daher keine Varianz besitzen. Man erhält für

$$\hat{\Sigma} = \begin{bmatrix} 0.0098 & -0.0004 \\ -0.0004 & 0.0021 \end{bmatrix}.$$

Weiters ergibt sich für $\widehat{Cov}(\hat{\beta}_1) = 2.05e^{-04}\mathbf{I}_{16}$ und $\widehat{Cov}(\hat{\beta}_2) = 4.44e^{-05}\mathbf{I}_{16}$. Der Schätzer der Kovarianzmatrix der Regressionsflächen ist

$$\widehat{Cov}(\hat{\mathbf{y}}(\mathbf{x})) = (1 + x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1^2x_2^2 + \dots + x_1^2x_2^2x_3^2x_4^2) \begin{bmatrix} 2.05e^{-04} & -9.18e^{-06} \\ -9.18e^{-06} & 4.44e^{-05} \end{bmatrix}.$$

Weiters ist der Schätzer des Erwartungswertes und der Varianz der Zielfunktion durch

$$\hat{E}(F(\mathbf{x})) = w_1|\hat{y}_1(\mathbf{x}) - t_1| + w_2|\hat{y}_2(\mathbf{x}) - t_2|$$

und

$$\widehat{Var}(F(\mathbf{x})) = w_1^2\widehat{Var}(\hat{y}_1(\mathbf{x})) + w_2^2\widehat{Var}(\hat{y}_2(\mathbf{x})) + 2w_1w_2\widehat{Cov}(\hat{y}_1(\mathbf{x}), \hat{y}_2(\mathbf{x}))$$

gegeben.

Wenn wir die Minimierung des robusten E-Modells mit $\alpha = 0.05$ durchführen, erhalten wir die Ergebnisse in Tabelle 5.3.

$w_1 = 0.5, w_2 = 0.5$									
t_1	t_2	x_1	x_2	x_3	x_4	$\hat{y}_1(\mathbf{x})$	$\hat{y}_2(\mathbf{x})$	$\widehat{Var}(\hat{y}_1(\mathbf{x}))$	$\widehat{Var}(\hat{y}_2(\mathbf{x}))$
1.66	1.00	-0.994	1.000	-0.124	0.376	1.66	1.00	$9.44e^{-04}$	$2.05e^{-04}$
1.72	1.00	-0.996	1.000	-0.403	1.000	1.72	1.00	$1.90e^{-03}$	$4.11e^{-04}$
1.66	1.05	-1.000	0.996	-0.076	0.890	1.66	1.05	$1.47e^{-03}$	$3.19e^{-04}$
$w_1 = 0.2, w_2 = 0.8$									
t_1	t_2	x_1	x_2	x_3	x_4	$\hat{y}_1(\mathbf{x})$	$\hat{y}_2(\mathbf{x})$	$\widehat{Var}(\hat{y}_1(\mathbf{x}))$	$\widehat{Var}(\hat{y}_2(\mathbf{x}))$
1.66	1.00	-0.994	1.000	-0.123	0.377	1.66	1.00	$9.44e^{-04}$	$2.05e^{-04}$
1.70	1.00	-1.000	1.000	-0.336	0.801	1.70	1.00	$1.50e^{-03}$	$3.24e^{-04}$
1.66	1.06	-1.000	1.000	-0.065	0.981	1.66	1.06	$1.61e^{-03}$	$3.50e^{-04}$
$w_1 = 0.05, w_2 = 0.95$									
t_1	t_2	x_1	x_2	x_3	x_4	$\hat{y}_1(\mathbf{x})$	$\hat{y}_2(\mathbf{x})$	$\widehat{Var}(\hat{y}_1(\mathbf{x}))$	$\widehat{Var}(\hat{y}_2(\mathbf{x}))$
1.66	1.00	-0.873	1.000	0.164	0.229	1.63	1.00	$7.80e^{-04}$	$1.69e^{-04}$
1.63	1.00	-0.862	0.970	0.191	0.259	1.63	1.00	$7.66e^{-04}$	$1.66e^{-04}$
1.63	1.02	-0.947	0.996	0.189	0.261	1.63	1.02	$8.55e^{-04}$	$1.85e^{-04}$

Tabelle 5.3.: Robustes E-Modell

In allen drei Fällen wird als erstes der Zielvektor (1.66, 1.00) verwendet. Bei der Gewichtung ($w_1 = 0.05, w_2 = 0.95$), letzter Block in Tabelle 5.3, ist es nicht möglich diesen genau zu erreichen. Da der Zielgröße Ätzfaktor so ein großes Gewicht zukommt, erreicht der C_p -Wert hier nur 1.63. Anschließend wird noch getestet, wie weit man den Wert der einen Zielgröße vergrößern kann, um trotzdem mit der anderen Zielgröße noch auf dem ursprünglichen Wert bleiben zu können. Bei der Gewichtung ($w_1 = 0.5, w_2 = 0.5$) sind beispielsweise die Kombinationen (1.72, 1.00) und (1.66, 1.05) möglich.

5.2. Robuste Optimierung mit einem Konfidenzbereich

Die individuellen Optimierungs-Desirability-Funktionen werden mit den vorher genannten unteren und oberen Grenzen gebildet. Der Wert r wird in beiden Fällen auf 1 gesetzt. Die Funktionen sind durch

$$d_1 = \begin{cases} 0 & \hat{y}_1 \leq 1.33 \\ \left[\frac{\hat{y}_1 - 1.33}{2.00 - 1.33} \right]^1 & 1.33 < \hat{y}_1 < 2.00 \\ 1 & \hat{y}_1 \geq 2.00 \end{cases}$$

und

$$d_2 = \begin{cases} 0 & \hat{y}_2 \leq 0.80 \\ \left[\frac{\hat{y}_2 - 0.80}{1.25 - 0.80} \right]^1 & 0.80 < \hat{y}_2 < 1.25 \\ 1 & \hat{y}_2 \geq 1.25 \end{cases}$$

gegeben. Die robusten individuellen Desirability-Funktionen werden bei beiden Zielgrößen (da Larger-the-better) an der unteren Grenze des Konfidenzintervalls minimiert, d.h. der „Worst-case“ des Konfidenzintervalls ist die untere Grenze. Die robuste Gesamt-Desirability-Funktion wird mit Hilfe des Nelder-Mead-Algorithmus maximiert. Um nicht in einem lokalen Maximum hängen zu bleiben, wird das Verfahren mit 625 verschiedenen Startpunkten im zulässigen Bereich durchgeführt.

Tabelle 5.4 zeigt die Ergebnisse der drei verschiedenen Gewichtungen. Die Resultate werden für verschiedene simultane Überdeckungswahrscheinlichkeiten $1 - \alpha$ angegeben. Der Fall $\alpha = 1$ entspricht, wie in Abschnitt 4.3 beschrieben, der optimalen Lösung. Für die robusten Lösungen ($\alpha = 0.1$) sind in Tabelle 5.5 die Konfidenzintervalle angegeben.

Man sieht, dass bei der Gewichtung ($w_1 = 0.5, w_2 = 0.5$) kein Unterschied zwischen optimaler und robuster Lösung ist. Der C_p -Wert nimmt bei dieser Gewichtung sein individuelles Maximum an. Da bei den anderen Einstellungen mehr Gewicht der Zielgröße Ätzfaktor zukommt, nimmt diese bei der optimalen Lösung ihr individuelles Maximum an. Bei der Gewichtung ($w_1 = 0.2, w_2 = 0.8$) können wir die größten Unterschiede zwischen der robusten und optimalen Lösung beobachten. Je mehr Bedeutung wir der Robustheit zukommen lassen, desto größer wird der C_p -Wert. Dabei wird der Ätzfaktor gleichzeitig kleiner.

5.2. Robuste Optimierung mit einem Konfidenzbereich

$w_1 = 0.5, w_2 = 0.5$									
α	x_1	x_2	x_3	x_4	$D_R(\mathbf{x})$	$D_{R1}(\mathbf{x})$	$D_{R2}(\mathbf{x})$	$\hat{y}_1(\mathbf{x})$	$\hat{y}_2(\mathbf{x})$
0.1	1.000	1.000	-1.000	-1.000	0.446	0.872	0.228	2.03	0.96
0.2	1.000	1.000	-1.000	-1.000	0.475	0.902	0.249	2.03	0.96
0.4	1.000	1.000	-1.000	-1.000	0.508	0.939	0.275	2.03	0.96
0.6	1.000	1.000	-1.000	-1.000	0.533	0.967	0.294	2.03	0.96
0.8	1.000	1.000	-1.000	-1.000	0.557	0.994	0.313	2.03	0.96
1.0	1.000	1.000	-1.000	-1.000	0.590	1.000	0.348	2.03	0.96
								>1.66	<1.00
$w_1 = 0.2, w_2 = 0.8$									
α	x_1	x_2	x_3	x_4	$D_R(\mathbf{x})$	$D_{R1}(\mathbf{x})$	$D_{R2}(\mathbf{x})$	$\hat{y}_1(\mathbf{x})$	$\hat{y}_2(\mathbf{x})$
0.1	-1.000	1.000	0.600	1.000	0.558	0.173	0.749	1.54	1.18
0.2	-1.000	1.000	0.661	1.000	0.586	0.179	0.788	1.53	1.19
0.4	-1.000	1.000	0.763	1.000	0.620	0.178	0.847	1.51	1.21
0.6	-1.000	1.000	0.840	1.000	0.648	0.180	0.893	1.50	1.22
0.8	-1.000	1.000	0.914	1.000	0.677	0.183	0.939	1.49	1.24
1.0	-1.000	1.000	0.982	1.000	0.734	0.214	1.000	1.47	1.25
								<1.66	>1.00
$w_1 = 0.05, w_2 = 0.95$									
α	x_1	x_2	x_3	x_4	$D_R(\mathbf{x})$	$D_{R1}(\mathbf{x})$	$D_{R2}(\mathbf{x})$	$\hat{y}_1(\mathbf{x})$	$\hat{y}_2(\mathbf{x})$
0.1	-1.000	1.000	0.953	1.000	0.757	0.053	0.871	1.48	1.24
0.2	-1.000	1.000	1.000	1.000	0.797	0.067	0.909	1.47	1.25
0.4	-1.000	1.000	1.000	1.000	0.837	0.103	0.934	1.47	1.25
0.6	-1.000	1.000	1.000	1.000	0.863	0.131	0.953	1.47	1.25
0.8	-1.000	1.000	1.000	1.000	0.887	0.158	0.972	1.47	1.25
1.0	-1.000	1.000	0.998	0.982	0.926	0.213	1.000	1.47	1.25
								<1.66	>1.00

Tabelle 5.4.: Robuste Optimierung mit Konfidenzbereich

$w_1 = 0.5, w_2 = 0.5$					Zielvektor (1.66,1.00)
Response	U_i	Y_{U_i}	\hat{y}_i	Y_{O_i}	
y_1	1.33	1.914	2.030	2.146	$y_1 > 1.66$
y_2	0.80	0.903	0.957	1.011	überdeckt
$w_1 = 0.2, w_2 = 0.8$					Zielvektor (1.66,1.00)
Response	U_i	Y_{U_i}	\hat{y}_i	Y_{O_i}	
y_1	1.33	1.446	1.541	1.637	nicht überdeckt
y_2	0.80	1.137	1.181	1.226	$y_2 > 1.00$
$w_1 = 0.05, w_2 = 0.95$					Zielvektor (1.66,1.00)
Response	U_i	Y_{U_i}	\hat{y}_i	Y_{O_i}	
y_1	1.33	1.365	1.478	1.591	nicht überdeckt
y_2	0.80	1.192	1.297	1.245	$y_2 > 1.00$

Tabelle 5.5.: Konfidenzbereiche der robusten Lösungen

5.3. Robuste Optimierung mit Desirability-Funktionen

Bei dieser Methode wird mit den gleichen Optimierungs-Desirability-Funktionen, die bei der vorigen Methode verwendet werden, gearbeitet. Weiters werden auch für die Robustheit Desirability-Funktionen gebildet. Die Optimierung wird - anders als bei den Autoren - mit Hilfe des Nelder-Mead-Algorithmus durchgeführt. Dieser ist eine Erweiterung des von den Autoren vorgeschlagenen Simplexverfahrens. Als Startpunkt des Nelder-Mead-Algorithmus wird der optimale Punkt verwendet. Der geschätzte Bereich wird immer mit dem gleichen Radius $r^2 = 0.05$ berechnet.

In Tabelle 5.6 sieht man die Ergebnisse dieser Methode. Die Gewichte w_r und w_o zeigen wie viel Bedeutung die Robustheit bzw. die Optimierung hat.

w_r	w_o	x_1	x_2	x_3	x_4	$D_g(\mathbf{x})$	$D_r(\mathbf{x})$	$D_o(\mathbf{x})$	$\hat{y}_1(\mathbf{x})$	$\hat{y}_2(\mathbf{x})$
1.0	0.0	1.000	1.000	-1.000	-0.767	0.853	0.853	0.476	1.98	0.91
0.9	0.1	1.000	1.000	-1.000	-0.978	0.803	0.832	0.581	2.02	0.95
0.8	0.2	1.000	1.000	-1.000	-1.000	0.774	0.829	0.590	2.03	0.96
0.7	0.3	1.000	1.000	-1.000	-1.000	0.748	0.829	0.590	2.03	0.96
0.6	0.4	1.000	1.000	-1.000	-1.000	0.723	0.829	0.590	2.03	0.96
0.5	0.5	1.000	1.000	-1.000	-1.000	0.699	0.829	0.590	2.03	0.96
0.4	0.6	1.000	1.000	-1.000	-1.000	0.676	0.829	0.590	2.03	0.96
0.3	0.7	1.000	1.000	-1.000	-1.000	0.653	0.829	0.590	2.03	0.96
0.2	0.8	1.000	1.000	-1.000	-1.000	0.631	0.829	0.590	2.03	0.96
0.1	0.9	1.000	1.000	-1.000	-1.000	0.615	0.829	0.590	2.03	0.96
0.0	1.0	1.000	1.000	-1.000	-1.000	0.590	0.829	0.590	2.03	0.96
									>1.66	<1.00

Tabelle 5.6.: Robuste Optimierung mit Desirability-Funktionen

Man sieht, dass es in diesem Fall - vergleichbar mit der vorigen Methode, die mit dem Konfidenzbereich rechnet - nur sehr geringe Unterschiede gibt. Erst bei der Lösung, bei welcher das Gewicht w_r auf 0.9 gesetzt wird, ändern sich die Response-Werte. Einen größeren Sprung haben wir zur Lösung $w_r = 1.0$. Generell wird der Zielwert von 1.00 für den Ätzfaktor nicht erreicht.

Bei dieser Methode ist von den Autoren vorgesehen, dass jeder Zielgröße das gleiche Gewicht zukommt. Sollte man hier den Response-Variablen unterschiedliche Gewichte geben wollen, könnte man die Methode so abändern, dass man die robuste Desirability-Funktion und die Optimierungs-Desirability-Funktion wie bei der vorigen Methode mit einem gewichteten geometrischen Mittel bildet.

6. Prozessfähigkeit

Eine der Zielgrößen in dem betrachteten Datensatz ist der Prozessfähigkeitsindex C_p . In diesem Kapitel wollen wir uns mit der Prozessfähigkeit und den Prozessfähigkeitsindizes beschäftigen. Am Beginn werden die zwei wichtigsten Indizes (C_p , C_{pk}) vorgestellt und ihr Zusammenhang beschrieben. Weitere Indizes und ihre zugrundeliegende Bedeutung werden erklärt. Da man bei den Berechnungen der Prozessfähigkeitsindizes standardmäßig von normalverteilten Daten ausgeht, das aber in der Realität nicht immer zutrifft, werden Möglichkeiten erklärt, wie man bei nicht normalverteilten Daten Prozessfähigkeitsindizes ausrechnen kann.

Im Abschnitt 6.7 werden Daten simuliert, die der Gamma, Lognormal und Weibullverteilung folgen. Wir testen alle beschriebenen Methoden und zeigen, wie groß die Abweichung ist, wenn man in diesen Fällen die Indizes unter der falschen Annahme einer Normalverteilung bestimmt. Schließlich werden noch die unter der Normalverteilung berechneten C_p -Werte des Datensatzes mit C_p -Werten, die mit Methoden für nicht normalverteilte Populationen berechnet werden, verglichen und gegenübergestellt. Das gesamte Kapitel wurde aufbauend auf Kotz & Lovelace (1998) verfasst.

6.1. Normalverteilte und nicht normalverteilte Daten

6.1.1. Normalverteilung und der 6σ -Bereich

Wie bereits vorher erwähnt, geht man bei der Berechnung der Prozessfähigkeitsindizes standardmäßig davon aus, dass die Daten der Normalverteilung folgen. In diesem Fall spielt der 6σ -Bereich eine wichtige Rolle. Wir wollen uns in diesem Abschnitt ansehen, wie viel Prozent der Daten, sollten diese normalverteilt sein, in diesen Bereich fallen.

Die Dichtefunktion einer normalverteilten Zufallsvariable ist

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{(x-\mu)}{\sigma}\right]^2} \quad x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma \in \mathbb{R}_+$$

Es gilt $E[X] = \mu$ und $Var[X] = \sigma^2$. Wir betrachten nun die standardisierte Zufallsvariable

$$Z = \frac{X - \mu}{\sigma},$$

für die $E[Z] = 0$ und $Var[Z] = 1$ gilt. Die Verteilungsfunktion der standardnormalverteilten Zufallsvariable Z ist

$$F_Z(z) = \Phi\left(\frac{x - \mu}{\sigma}\right) = \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{1}{2}t^2} dt.$$

Der Anteil der Beobachtungen, die innerhalb des 6σ -Bereiches liegen, ist daher

$$\begin{aligned} F(\mu + 3\sigma) - F(\mu - 3\sigma) &= \Phi\left[\frac{(\mu + 3\sigma) - \mu}{\sigma}\right] - \Phi\left[\frac{(\mu - 3\sigma) - \mu}{\sigma}\right] \\ &= \Phi(3) - \Phi(-3) \\ &= 0.99865 - 0.00135 \\ &= 0.9973. \end{aligned}$$

Folglich liegen 99.73% der Daten innerhalb des 6σ -Bereiches, wenn die Voraussetzung gegeben ist, dass die Daten aus einer Normalverteilung stammen.

6.1.2. Ansatz für nicht normalverteilte Populationen

Im Allgemeinen betrachten wir bei der Prozessfähigkeitsanalyse drei Punkte: den oberen Tail, den zentralen Punkt und den unteren Tail. Unter der Normalverteilung kann man diese drei Punkte auch mit Hilfe der Quantile beschreiben.

$$\begin{aligned} \Phi^{-1}(0.99865) &= x_{0.99865} = \mu + 3\sigma \\ \Phi^{-1}(0.5) &= x_{0.5} = \mu \\ \Phi^{-1}(0.00135) &= x_{0.00135} = \mu - 3\sigma. \end{aligned} \tag{6.1}$$

Bei nicht normalverteilten Daten entsprechen obige Quantile nicht mehr zwingend diesen drei Punkten. Das Quantil $x_{0.5}$ entspricht im Allgemeinen dem Median. Dem Mittelwert entspricht es nur im normalverteilten Fall bzw. bei anderen symmetrischen Verteilungen, da hier gilt, dass Median und Mittelwert gleich sind. Das heißt, ist eine nicht normalverteilte Population gegeben, wird das Quantil x_p zu einem Quantil x_p^* .

6.2. C_p -Index

In diesem Abschnitt betrachten wir als einfachsten Index den Prozessfähigkeitsindex C_p , der als erstes entwickelt wurde, aber heute eher selten verwendet wird. Am Beginn wird der Index definiert und gezeigt, was dieser Index über den Prozentsatz des fehlerhaften Materials bzw. über die Fähigkeit des Prozesses aussagt. Weiters werden ein Schätzer, ein Konfidenzintervall, Hypothesentests und schließlich Schwächen des C_p -Index aufgezeigt.

Definition 6.2.1. (*C_p -Index*) Der C_p -Index ist durch

$$C_p = \frac{OSG - USG}{6\sigma}$$

definiert, wobei OSG die obere Spezifikationsgrenze, USG die untere Spezifikationsgrenze und σ die Prozessstreuung eines Merkmals X ist.

Um es ganz einfach in Worten zu beschreiben, dividiert man beim C_p -Index den erlaubten Messbereich durch den tatsächlichen Messbereich. Das heißt, je größer der C_p -Wert ist, desto besser arbeitet der Prozess, da im Zähler die Größe des Bereichs, in welchem die Prozessmessungen variieren können und im Nenner die Größe des Bereichs, in welchem die Prozessmessungen tatsächlich variieren, angegeben ist. Vor allem Werte, die kleiner als 1 sind, sind auf keinen Fall akzeptabel, da dies bedeutet, dass der Bereich der Prozessstreuung größer als der Toleranzbereich ist.

Da beim C_p -Index nur zwei Bereiche miteinander verglichen werden, spielt es keine Rolle, wo der Prozess zentriert ist. Die Prozessverteilung könnte sich verschieben: ohne Änderung der Prozessstreuung bleibt der C_p -Index unverändert. Die Voraussetzungen für die Anwendung des C_p -Index sind

- normal-verteilte Daten mit Mittelwert μ und Varianz σ^2
- unabhängige Prozessdaten
- statistisch kontrollierbare Prozessdaten.

Das heißt, man sollte den C_p -Index nur verwenden, wenn die Daten diese drei Bedingungen erfüllen, da dies sonst zu falschen Interpretationen der Prozessfähigkeit führen würde.

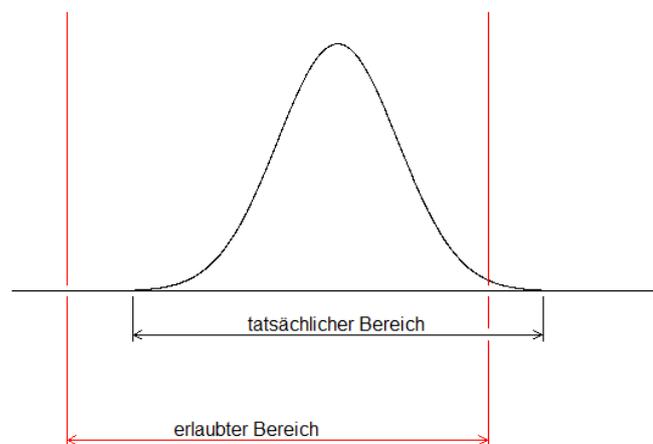


Abbildung 6.1.: Der C_p -Index vergleicht den tatsächlichen mit dem erlaubten Bereich, vgl. Kotz & Lovelace (1998)

6.2.1. Der C_p -Index und der Prozentsatz des fehlerhaften Materials

Obwohl das Inverse des C_p -Index, welches man abgekürzt unter CR (englisch: Capability ratio) kennt, besser mit dem Prozentsatz des außerhalb des Spezifikationsbereiches liegenden Prozessoutputs vergleichbar ist, kann man, wenn man einen C_p -Index

gegeben hat, damit den zugehörigen Prozentsatz des fehlerhaften Materials bestimmen. Im Folgenden werden wir sehen, dass aber auch dies wieder auf der Annahme beruht, dass die Daten normalverteilt sind.

Sei eine Zufallsvariable X normalverteilt, dann ist die Wahrscheinlichkeit des fehlerhaften Materials gegeben mit

$$p = \Phi\left(\frac{USG - \mu}{\sigma}\right) + \left(1 - \Phi\left(\frac{OSG - \mu}{\sigma}\right)\right),$$

wobei Φ die Verteilungsfunktion der Standardnormalverteilung ist. Sollte der Mittelwert μ des Prozesses nicht mit $\frac{(OSG+USG)}{2}$ übereinstimmen, gibt der C_p -Index allerdings nicht den exakten Wert dieses Prozentsatzes wieder, da der C_p -Index den Mittelwert nicht miteinbezieht. Im Allgemeinen erhalten wir also nur eine untere Schranke dieses Prozentsatzes

$$\begin{aligned} p_{C_p} &\geq 1 - \left[\Phi\left(\frac{OSG - \mu}{\sigma}\right) - \Phi\left(\frac{USG - \mu}{\sigma}\right) \right] \\ &\geq 1 - [\Phi(3C_p) - \Phi(-3C_p)] \\ &\geq 2[1 - \Phi(3C_p)]. \end{aligned}$$

Man erhält damit

$$p_{C_p} \geq 2\Phi(-3C_p) \quad \text{für alle } \mu.$$

Ist ein Prozess im Spezifikationsbereich perfekt zentriert, d.h. ist $\mu = \frac{(OSG+USG)}{2}$, dann gilt

$$p_{C_p} = 2\Phi(-3C_p).$$

6.2.2. Wann wird ein Prozess als „fähig“ bezeichnet?

Wann man einen Prozess als fähig bezeichnet, ist meistens eine Vereinbarung zwischen Hersteller eines Produkts und dessen Abnehmer. Es hat sich eingebürgert, dass man den Wert 1.33 oder 1.67 als untere Grenze festlegt. Dies garantiert - sofern der Prozess im Spezifikationsbereich perfekt zentriert ist - dass die Prozessdaten ungefähr 75% oder weniger des Spezifikationsbereiches überdecken. Im Laufe der Zeit sind aber auch andere Grenzen in Betracht gezogen worden, je nachdem wie hoch die Qualitätsansprüche gestellt werden. Der interessierte Leser findet weitere Informationen dazu bei Kotz & Lovelace (1998).

6.2.3. Schätzer für den C_p -Index

Der C_p -Index enthält den unbekannt Parameter σ , der geschätzt werden muss. Da als Schätzer für σ üblicherweise die Standardabweichung S verwendet wird, erhält man als Schätzer für den C_p -Index

$$\hat{C}_p = \frac{OSG - USG}{6S} = \frac{d}{3S} = \frac{\sigma}{S} C_p, \quad (6.2)$$

wobei S und \bar{X} durch

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}}$$

und

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

gegeben sind.

Unter den Bedingungen, die wir am Anfang des Abschnittes 6.2 eingeführt haben, ist S^2 ein unverzerrter Schätzer für σ^2 . Chan et al. (1988b) haben gezeigt, wie \hat{C}_p verteilt ist und dass \hat{C}_p ein verzerrter Schätzer für C_p ist. Der Bias verschwindet aber, wenn die Stichprobengröße n gegen unendlich geht, d.h. \hat{C}_p ist asymptotisch unverzerrt. Weitere Eigenschaften, wie Erwartungswert und Varianz von \hat{C}_p findet man bei Kotz & Lovelace (1998).

6.2.4. Konfidenzintervall für den C_p -Index

Wenn wir annehmen, dass X normalverteilt ist, dann hat S^2 die gleiche Verteilung wie $\sigma^2 \times \frac{\chi_{n-1}^2}{n-1}$. Da S als Schätzer für σ verwendet wird, ist die Verteilung von \hat{C}_p laut Kane (1986) durch

$$\hat{C}_p = \frac{C_p}{\sqrt{\frac{\chi_{n-1}^2}{n-1}}} \quad (6.3)$$

gegeben. Man erhält als konkretes $(1 - \alpha)$ -Konfidenzintervall für σ^2

$$\left(\frac{(n-1)S^2}{\chi_{n-1,1-\alpha/2}^2}, \frac{(n-1)S^2}{\chi_{n-1,\alpha/2}^2} \right).$$

Aus (6.3) und $\hat{C}_p = (OSG - USG)/6S$ ergibt sich das $(1 - \alpha)$ -Konfidenzintervall für C_p mit

$$\begin{aligned} & \left(\frac{OSG - USG}{6S} \frac{\chi_{n-1,\alpha/2}}{\sqrt{n-1}}, \frac{OSG - USG}{6S} \frac{\chi_{n-1,1-\alpha/2}}{\sqrt{n-1}} \right) \\ & \equiv \left(\hat{C}_p \frac{\chi_{n-1,\alpha/2}}{\sqrt{n-1}}, \hat{C}_p \frac{\chi_{n-1,1-\alpha/2}}{\sqrt{n-1}} \right). \end{aligned}$$

6.2.5. Hypothesentests

Wie wir im vorigen Abschnitt gesehen haben, kann man mit Hilfe von Konfidenzintervallen einzelne Schätzer einer Statistik bewerten. Statt einer einzelnen Schätzung des Parameters, kann man sich bei einem Konfidenzintervall mit Wahrscheinlichkeit $(1 - \alpha)$ sicher sein, dass der wahre Parameter vom angegebenen Bereich überdeckt wird. Mit Hypothesentests können wir herausfinden, ob ein hypothetischer Wert eines Parameters, gegeben eine Stichprobe und ein Parameterschätzer dieser Stichprobe, plausibel ist oder nicht.

In der Prozessfähigkeitsanalyse will man mit Hypothesentests herausfinden, wo die geschätzte Fähigkeit bezüglich der wahren Fähigkeit oder einer angenommenen Fähigkeit liegt, bzw. wie groß der Unterschied zwischen der geschätzten und der wahren Fähigkeit ist. Dies erreichen wir, indem wir einen Indexwert schätzen, diesen mit einer unteren Grenze c_0 vergleichen und daraus einen p -Wert berechnen. Der p -Wert gibt dabei an, wie hoch das Risiko ist, dass man fälschlicherweise sagt, dass der Prozess fähig ist und H_0 verwirft, obwohl der Prozess nicht fähig ist. Dies wird auch als Fehler 1. Art bezeichnet. Im Normalfall wählt man p nicht größer als $\alpha = 0.05$.

Wir testen die Hypothese

$$H_0 : C_p \leq c_0 \text{ (der Prozess ist nicht fähig)}$$

$$\text{gegen } H_1 : C_p > c_0 \text{ (der Prozess ist fähig),}$$

wobei man für c_0 das Minimalkriterium für den C_p -Index einsetzt, das im jeweiligem Unternehmen festgesetzt wurde. Im Normalfall liegt diese Grenze entweder bei 1.33 oder bei 1.67. Die Nullhypothese wird verworfen, wenn die Alternativhypothese akzeptiert wird und dies tritt ein, wenn $\hat{C}_p > c$ ist. Wir werden im Folgenden sehen, wie dieser Wert c aussieht.

Da der obige Hypothesentest laut Kane (1986) äquivalent zu

$$H_0 : \sigma \geq \sigma_0$$

$$\text{gegen } H_1 : \sigma < \sigma_0$$

mit $\sigma_0 > 0$ ist, wird H_0 verworfen, wenn $Y = \frac{(n-1)S^2}{\sigma_0^2} < \chi_{n-1,\alpha}^2$. Hierbei ist Y eine χ_{n-1}^2 -Zufallsvariable und $\chi_{n-1,\alpha}^2$ das α -Quantil der χ_{n-1}^2 -Verteilung und daher eine Konstante. Man erhält

$$c = c_0 \sqrt{\frac{n-1}{\chi_{n-1,\alpha}^2}}. \quad (6.4)$$

Aus (6.4) sieht man, dass der Wert von \hat{C}_p größer als der Zielwert des wahren C_p -Index sein muss. Weiters kann man auch erkennen, dass der Unterschied vom Stichprobenumfang n abhängt. Die Güte β (die Wahrscheinlichkeit H_0 zu verwerfen, wenn H_0 falsch ist) dieses Tests ist laut Kane (1986)

$$\beta(C_p) = \mathbb{P}[\hat{C}_p > c | C_p]$$

$$= \mathbb{P}[\chi_{n-1}^2 < (n-1) \frac{C_p^2}{c^2} | C_p].$$

$(1 - \beta)$ wird als Fehler 2. Art bezeichnet.

6.2.6. Schwächen des C_p -Index

Wie man bereits gesehen hat, misst der C_p -Index die Prozessfähigkeit nur in Bezug auf die Prozessstreuung. Die größte Schwäche des C_p -Index ist es also, dass die Lage

des Prozesses in keiner Weise bewertet wird. Damit kann man aber die eigentliche Prozessleistung nicht messen. Aus diesem Grunde wurde ein weiterer Index - der C_{pk} -Index - entwickelt. Wir werden in den folgenden Abschnitten sehen, dass man bei gemeinsamer Verwendung beider Indizes die Prozessfähigkeit gut bewerten kann, da dann sowohl die Prozessstreuung als auch die Prozesslage berücksichtigt werden.

6.3. C_{pk} -Index

In diesem Abschnitt wird ein weiterer Prozessfähigkeitsindex - der C_{pk} -Index - eingeführt. Wir sehen uns an, wie wir mit diesem Index Prozessfähigkeit messen können und leiten einen Schätzer her. Weiters betrachten wir auch hier ein Konfidenzintervall, Hypothesentests und Schwächen des Index.

Definition 6.3.1. (*C_{pk} -Index*) Der C_{pk} -Index ist durch

$$\begin{aligned} C_{pk} &= \min(C_{pu}, C_{pl}) \\ &= \frac{d - |\mu - M|}{3\sigma} \end{aligned}$$

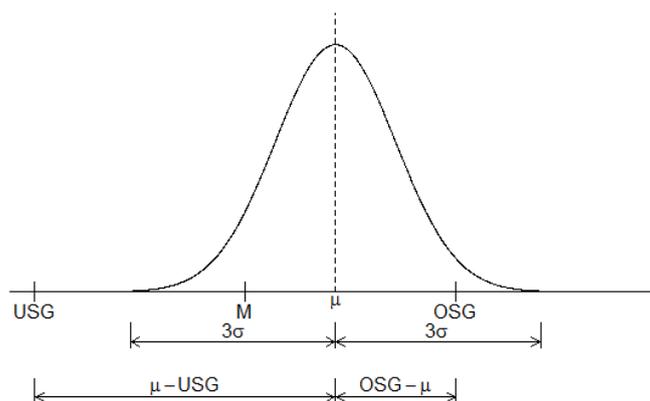
definiert, wobei

$$\begin{aligned} C_{pu} &= \frac{\text{erlaubte obere Streuung}}{\text{tatsächliche obere Streuung}} \\ &= \frac{OSG - \mu}{3\sigma}, \\ C_{pl} &= \frac{\text{erlaubte untere Streuung}}{\text{tatsächliche untere Streuung}} \\ &= \frac{\mu - USG}{3\sigma}, \end{aligned}$$

$d = (OSG - USG)/2$ die Hälfte der Intervalllänge und $M = (OSG + USG)/2$ der Mittelpunkt des Spezifikationsintervalls ist.

Wie vorher erwähnt, wurde der C_{pk} -Index in Japan hauptsächlich entwickelt, um die Schwächen des C_p -Index zu kompensieren. Hersteller, die mit Prozessen arbeiteten, deren Streuung sehr gering war, hatten keinen Anreiz diese Prozesse auf den Zielwert zu zentrieren, da das auf den C_p -Wert keinen Einfluss hat. Wenn sie einen hohen C_p -Wert erreicht hatten, haben sie den Prozess dann so zentriert, dass er näher zu jener Seite des Spezifikationsbereiches war, an welcher man mehr Geld, Rohmaterial, usw. sparen konnte. Dies führte zu einer Verschlechterung der Produkt- bzw. Prozessleistung (vgl. beispielsweise Finley, 1992).

Wenn man die Definition des C_{pk} -Index betrachtet, erkennt man, dass dieser Index die Fähigkeit der schlechteren Hälfte der Daten bewertet. Dies funktioniert so, dass beide Hälften (C_{pu} , C_{pl}) gemessen und anschließend miteinander verglichen werden.

Abbildung 6.2.: C_{pk} -Index, vgl. Kotz & Lovelace (1998)

Das schlechtere Ergebnis wird als Index hergenommen. In Abbildung 6.2 kann man gut sehen, was der C_{pk} -Index misst.

Der C_{pu} -Index vergleicht den oberen 3σ -Bereich der Prozessdaten mit der Distanz des Prozessmittelwertes zur oberen Spezifikationsgrenze. Der C_{pl} -Index wiederum vergleicht den unteren 3σ -Bereich der Prozessdaten mit der Distanz des Prozessmittelwertes zur unteren Spezifikationsgrenze. Der Wert der Hälfte, die das schlechtere Ergebnis ergibt, wird dann als C_{pk} -Index verwendet. Solange sich die Lage der äußeren Tails nicht verändert, kann die Variation innerhalb des Prozesses sehr groß werden, ohne dass sich der C_{pk} -Index verschlechtert. Es ist möglich, dass Prozesse, die Verteilungen mit unterschiedlichen Mittelwerten und Variationen haben, trotzdem den gleichen C_{pk} -Wert haben. Dies ist der Fall, wenn das Verhältnis zwischen dem Tail der „schlechteren Hälfte der Daten“ und der jeweiligen Spezifikationsgrenze bei den jeweiligen Prozessen gleich ist.

Zusammenfassend kann man sagen, dass der C_{pk} -Index die Beziehung des schlechteren Tails zur Spezifikationsgrenze beschreibt. In diesen Index fließt sowohl der Effekt der Prozesslage als auch der Effekt der Prozessstreuung ein.

6.3.1. Messen der Prozessfähigkeit mithilfe des C_{pk} -Index

Heutzutage ist der C_{pk} -Index jener Prozessfähigkeitsindex, der am häufigsten verwendet wird. Im Allgemeinen gelten für diesen Index die gleichen Grenzen wie für den C_p -Index. Ein Wert von 1.33 wird als unterste Grenze betrachtet, ab der ein Prozess als fähig angesehen wird. Ein C_{pk} -Wert von 1.33 bedeutet, dass der Prozess so abläuft, dass zumindest 75% des Spezifikationsbereiches verwendet werden. Zur Erinnerung, ein C_p -Wert von 1.33 bedeutet, dass der Prozess die Fähigkeit hat, innerhalb der Spezifikationsgrenzen zu produzieren und wenn er auf M zentriert ist, nur 75% des Spezifikationsbereiches zu verwenden. Das heißt, dass der C_p -Index eine obere Grenze für den C_{pk} -Index bildet. Gleichheit ist erfüllt, wenn der Prozess auf M zentriert ist.

6.3.2. Schätzer für den C_{pk} -Index

Beim C_{pk} -Index müssen die unbekannt Parameter μ und σ geschätzt werden. Als Schätzer für C_{pk} erhält man daher

$$\hat{C}_{pk} = \frac{d - |\bar{X} - \frac{1}{2}(OSG + USG)|}{3S} = \min(\hat{C}_{pu}, \hat{C}_{pl}),$$

wobei

$$\hat{C}_{pu} = \frac{OSG - \bar{X}}{3S}$$

und

$$\hat{C}_{pl} = \frac{\bar{X} - USG}{3S}.$$

Auch für den C_{pk} -Index muss gelten, dass die Daten normalverteilt, unabhängig und statistisch kontrollierbar sind. Viele Eigenschaften beruhen darauf, dass der zugrundeliegende Prozess normalverteilt ist. Laut Chan et al. (1988b) folgen die Schätzer \hat{C}_{pl} und \hat{C}_{pu} einer nicht zentralen t-Verteilung, wenn der Prozess normalverteilt ist. Weiters haben Kotz & Johnson (1993) gezeigt, dass \hat{C}_{pk} ein verzerrter Schätzer für C_{pk} ist. Der Bias ist positiv, wenn $\mu \neq M$. Sollte $\mu = M$ sein, ist der Bias positiv, wenn $n = 10$. Für größere Werte von n wird er negativ und geht gegen 0, wenn $n \rightarrow \infty$. Weitere Eigenschaften, wie Erwartungswert und Varianz von \hat{C}_{pk} und Informationen zur Verteilung von \hat{C}_{pk} findet man bei Kotz & Johnson (1993) und Kotz & Lovelace (1998).

6.3.3. Konfidenzintervall für den C_{pk} -Index

Die Entwicklung von Konfidenzintervallen für den C_{pk} -Index ist eine sehr schwierige Aufgabe. In der Literatur haben sehr viele Autoren verschiedene approximative Konfidenzintervalle angegeben. Die Konstruktion ist schwierig, da die Verteilung des C_{pk} -Wertes die gemeinsame Verteilung von zwei nicht zentral t-verteiltern Zufallsvariablen umfasst. Bei Kotz & Lovelace (1998) werden einige Konfidenzintervalle beschrieben, wobei keines der Intervalle von den Autoren favorisiert wird.

Wir betrachten hier ein Konfidenzintervall, welches bei Jahn & Braun (2006) hergeleitet wird. Die Autoren gehen davon aus, dass man den C_{pk} -Index auch folgendermaßen umschreiben kann

$$\hat{C}_{pk} = \min\left(\frac{OSG - \bar{X}}{3S}, \frac{\bar{X} - USG}{3S}\right) = \frac{\frac{1}{2}(OSG - USG) - |\bar{X} - M|}{3S}. \quad (6.5)$$

Nun bekommt man mit Hilfe der Konfidenzintervalle für μ und σ^2

$$\bar{X} - \frac{S \cdot t_{n-1, 1-\frac{\alpha}{2}}}{\sqrt{n}} \leq \mu \leq \bar{X} + \frac{S \cdot t_{n-1, 1-\frac{\alpha}{2}}}{\sqrt{n}}$$

$$\frac{(n-1)S^2}{\chi_{n-1, 1-\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{n-1, \frac{\alpha}{2}}}$$

und durch Einsetzen in (6.5) das $(1 - \alpha)$ -Konfidenzintervall für C_{pk}

$$\frac{3\hat{C}_{pk} - a_\alpha}{3b_{u,\alpha}} \leq C_{pk} \leq \frac{3\hat{C}_{pk} + a_\alpha}{3b_{o,\alpha}}$$

mit

$$a_\alpha = \frac{t_{n-1, 1-\frac{\alpha}{2}}}{\sqrt{n}},$$

$$b_{o,\alpha} = \sqrt{\frac{n-1}{\chi_{n-1, 1-\frac{\alpha}{2}}^2}}$$

und

$$b_{u,\alpha} = \sqrt{\frac{n-1}{\chi_{n-1, \frac{\alpha}{2}}^2}}.$$

6.3.4. Hypothesentests

Der Wert des C_{pk} -Index wird geschätzt, da er wie der C_p -Index nicht genau bestimmt werden kann. Da der C_{pk} -Index in der Praxis sehr häufig verwendet wird und täglich weitreichende Entscheidungen aufgrund der Schätzung dieses Wertes getroffen werden, muss man sich über die Eignung dieser Schätzung im Klaren sein. Um gute Entscheidungen zu treffen, ist es daher sinnvoll, den C_{pk} -Index nur gemeinsam mit Konfidenzintervallen und Hypothesentests zu verwenden. Die Theorie für die Hypothesentests mit dem C_{pk} -Index kann aus dem Abschnitt 6.2.5 übernommen werden.

6.3.5. Schwächen des C_{pk} -Index

Auch der C_{pk} -Index weist einige Schwächen auf. Aus diesem Grund warnen mehrere Autoren davor, diesen zu verwenden. Wir werden im Folgenden einige ausgewählte Schwächen anführen. Ein Kritikpunkt ist, dass man an die Daten die Anforderung stellen muss, dass sie normalverteilt, unabhängig und statistisch kontrollierbar sind, um exakte Schätzungen der Prozessfähigkeit zu bekommen. Diese Anforderungen erfüllen jedoch Daten aus der Praxis meistens nicht.

Weiters haben Chou et al. (1990) herausgefunden, dass man einen großen Stichprobenumfang benötigt um akzeptable Genauigkeit der Schätzungen zu erreichen. In der Praxis ist es aber oft schwierig einen großen Stichprobenumfang zu erhalten.

Vorsichtig muss man auch sein, da der C_{pk} -Index eine Kombination aus zwei Indizes ist. Das Problem, das dabei auftritt ist, dass der C_{pu} -Wert die Daten, die nahe der unteren Spezifikationsgrenze liegen, nicht in die Berechnung miteinbezieht. Umgekehrt fließen in den C_{pl} -Wert die Daten, die in der Nähe der oberen Spezifikationsgrenze liegen, nicht mit ein. Wenn sehr große Unterschiede zwischen den zwei Werten auftreten, kann man sagen, dass der eine Wert die eigentliche Fähigkeit des Prozesses überschätzt und der andere Wert diese unterschätzt. Bei Kotz & Lovelace (1998) werden weitere Schwächen beschrieben, die in der Literatur erwähnt werden.

6.4. Der k -Index

Im Folgenden wollen wir jetzt die Beziehung betrachten, die diese Indizes zueinander haben. Der C_p -Index hängt mit dem C_{pk} -Index über den sogenannten k -Index zusammen.

Definition 6.4.1. (*k-Index*) Der k -Index ist durch

$$k = \frac{|M - \mu|}{\frac{1}{2}(OSG - USG)} = \frac{2|\mu - \frac{1}{2}(OSG + USG)|}{OSG - USG}$$

definiert.

In der Definition sehen wir, dass dieser Index die Prozessfähigkeit nur in Bezug auf die Prozesslage misst. Er gibt ein Maß an, welches die Entfernung des Prozesses zum Zielwert bewertet, sofern dieser mit dem Mittelpunkt des Spezifikationsintervalls übereinstimmt. Ist dies nicht der Fall, wird k durch

$$k' = \frac{|T - \mu|}{\min(T - OSG, USG - T)}$$

ersetzt. Der k -Index ist ein skaliertes Maß der Zentriertheit. Ist $k = 0$, dann ist der Prozess genau auf den Zielwert zentriert. Ist $k = 1$, dann ist der Prozess auf einer der Spezifikationsgrenzen zentriert. Das heißt, bei einem k -Wert zwischen 0 und 1 wissen wir, dass der Prozessmittelwert zwischen dem Zielwert und der Spezifikationsgrenze liegt.

Wir betrachten nun den Zusammenhang zwischen C_p , C_{pk} und k .

$$\begin{aligned} C_{pk} &= \frac{\min(OSG - \mu, \mu - USG)}{3\sigma} = \frac{d - |\mu - \frac{1}{2}(OSG + USG)|}{3\sigma} \\ &= \left(1 - \frac{|\mu - \frac{1}{2}(OSG + USG)|}{d}\right) C_p = C_p - \frac{|\mu - \frac{1}{2}(OSG + USG)|}{3\sigma} \\ &= (1 - k)C_p. \end{aligned}$$

6.4.1. Schätzer für den k -Index

Für den Schätzer des k -Index erhalten wir

$$\hat{k} = \frac{|\bar{X} - \frac{1}{2}(OSG + USG)|}{d}.$$

6.5. Weitere Indizes

Da der C_p -Index und der C_{pk} -Index nur den rundherum aufgebauten Spezifikationsbereich, nicht aber den Zielwert in die Berechnungen miteinbeziehen, dieser aber auch

sehr wichtig für die Prozessfähigkeit ist, haben sich im Laufe der Zeit auch weitere Prozessfähigkeitsindizes entwickelt. Die bereits verwendeten Indizes C_p , C_{pk} , C_{pl} , C_{pu} und k sind alle so definiert, dass sie Teile, die innerhalb der Spezifikationsgrenzen liegen, als gut, und Teile, die außerhalb liegen, als schlecht betrachten. Die Reduktion der Variabilität war das primäre Ziel dieser Indizes und die Prozessvariation wurde mittels

$$\sigma^2 = E(X - \mu)^2$$

gemessen, wobei X eine Zufallsvariable ist, die den beobachteten Prozess wiedergibt. Wie man erkennen kann, wurde die Variabilität in Bezug auf den Mittelwert gemessen.

Ein weiteres Maß um die Variabilität zu schätzen ist

$$\tau^2 = E(X - T)^2.$$

Dieses Maß wird bei einer neuen Generation von Prozessfähigkeitsindizes verwendet. Es ist das Ziel, die Variabilität um den Zielwert und nicht die Variabilität um den Mittelwert zu reduzieren. Man arbeitet mit Hilfe der Verlustfunktion

$$L(x) = k(x - T)^2,$$

wobei k eine positive Konstante ist. Ist die Messung genau am Zielwert, ist die Verlustfunktion 0, gibt es eine Abweichung vom Zielwert, ist die Verlustfunktion positiv. Das heißt, dass man mit diesem Ansatz zwischen Prozessen, die am Zielwert liegen und Prozessen, die in der Nähe der Spezifikationsgrenzen liegen, unterscheiden kann.

Ein Vertreter dieser neuen Generation ist der C_{pm} -Index.

Definition 6.5.1. (C_{pm} -Index) Der C_{pm} -Index ist durch

$$\begin{aligned} C_{pm} &= \frac{OSG - USG}{6\tau} \\ &= \frac{d}{3\sqrt{\sigma^2 + (\mu - T)^2}} \\ &= \frac{OSG - USG}{6\sqrt{\sigma^2 + (\mu - T)^2}} \end{aligned}$$

definiert.

Auf diesen folgte der Index C_{pm}^* , da C_{pm} nicht unterscheidet, ob der Prozess δ Einheiten unter dem Zielwert oder δ Einheiten über dem Zielwert zentriert ist.

Definition 6.5.2. (C_{pm}^* -Index) Der C_{pm}^* -Index ist durch

$$C_{pm}^* = \frac{\min(OSG - T, T - USG)}{3\sqrt{\sigma^2 + (\mu - T)^2}}$$

definiert (vgl. Chan et al., 1988a).

Da auch diese Indizes Schwächen aufwiesen, folgte eine weitere Generation von Indizes. Pearn et al. (1992) kombinierten C_{pk} und C_{pm} zu einem Index und es entstand der C_{pmk} -Index.

Definition 6.5.3. (*C_{pmk} -Index*) Der C_{pmk} -Index ist durch

$$C_{pmk} = \frac{\min(OSG - \mu, \mu - USG)}{3\sqrt{\sigma^2 + (\mu - T)^2}}$$

definiert.

Weiters führten Johnson et al. (1994) den C_{pkj} -Index ein. Dieser Index bezieht eine mögliche Asymmetrie der Prozessverteilung mit ein. Der interessierte Leser findet weitere Details zu diesen Indizes und auch weitere Indizes bei Kotz & Lovelace (1998). Obwohl sehr viele neue Indizes entwickelt worden sind, und diese auch gewisse Vorteile gegenüber dem C_p und dem C_{pk} -Index haben, werden diese zwei - unter anderem aufgrund ihrer Einfachheit - heutzutage trotzdem noch am häufigsten verwendet.

6.6. Nicht normalverteilte Populationen

Alle Indizes, die wir bis jetzt betrachtet haben, sind so konstruiert, dass sie nur angewendet werden können, wenn die Daten normalverteilt sind. Alle statistischen Eigenschaften hängen von dieser Voraussetzung ab. Da dies in der Realität sehr oft nicht der Fall ist, muss man mit diesen Prozessfähigkeitsindizes sehr vorsichtig umgehen. Sollten die Daten von der Normalverteilung abweichen, bekommt man ungültige und ungenaue Ergebnisse. Wir wollen uns im Folgenden ansehen, wie man für nicht normalverteilte Populationen Prozessfähigkeitsindizes berechnen kann.

6.6.1. Datentransformation und graphische Methoden

Die einfachste Methode um mit nicht normalverteilten Daten zu arbeiten, ist eine Transformation durchzuführen und dadurch normalverteilte oder annähernd normalverteilte Daten zu erhalten. Auf schiefe Verteilungen wird sehr oft die Wurzeltransformation angewendet um diese symmetrischer zu machen. Anschließend testet man die Verteilung der transformierten Daten auf Normalverteilung. Wenn die transformierten Daten den Test auf Normalverteilung bestehen, kann man mit den transformierten Daten Prozessfähigkeitsindizes schätzen, Konfidenzintervalle konstruieren, Hypothesentests durchführen, usw. Am Ende muss man die Ergebnisse wieder auf die Originalskala zurücktransformieren.

Eine weitere Möglichkeit ist, die Prozessdaten mit einer anderen Verteilung zu modellieren, wie zum Beispiel der Lognormalverteilung oder der Weibullverteilung. Man kann die Indizes der Daten mit Quantilschätzung der jeweiligen Verteilung schätzen. Ob die Methode im jeweiligen Fall passend ist, kann man ganz einfach mit graphischen Methoden oder Goodness of fit Tests überprüfen.

6.6.2. Quantilschätzung für nicht normalverteilte Populationen

Wie wir in Abschnitt 6.1.1 gesehen haben, gilt für normalverteilte Daten, dass 0.135% der Daten unter $\mu - 3\sigma$, 50% der Daten unter μ und 99.867% unter $\mu + 3\sigma$ liegen. Das heißt, dass 99.73% der Daten zwischen den Grenzen $\mu \pm 3\sigma$ liegen. Wenn man das wie in Abschnitt 6.1.2 mit Hilfe von Quantilen betrachtet, kommen wir mit den Gleichungen (6.1) auf

$$\mathbb{P}(Z \leq z_{0.00135}) = 0.00135$$

$$\mathbb{P}(Z \leq z_{0.5}) = 0.5$$

$$\mathbb{P}(Z \leq z_{0.99865}) = 0.99865,$$

mit einer standardnormalverteilten Zufallsvariable Z . Man kann die Prozessfähigkeitsindizes jetzt auch in Quantilschreibweise angeben und man erhält

$$C_{p(q)} = \frac{OSG - USG}{x_{0.99865} - x_{0.00135}} \quad (6.6)$$

$$C_{pk(q)} = \min\left(\frac{OSG - x_{0.5}}{x_{0.99865} - x_{0.5}}, \frac{x_{0.5} - USG}{x_{0.5} - x_{0.00135}}\right). \quad (6.7)$$

Das heißt, wenn wir im Fall nicht normalverteilter Daten eine Verteilung finden, die zum Prozess passt, können wir passende Schätzungen für die drei Quantile bestimmen.

Man kann aufgrund der Schiefe und der Kurtosis der Daten eine passende Verteilung wählen. Hat man keine guten Schätzungen der Schiefe und der Kurtosis, kann man die Daten auch mittels einer Familie von Verteilungen modellieren. Zum Beispiel könnte man dafür die Johnson- oder die Pearson-Verteilungen verwenden. Diese zwei Verteilungsfamilien können Verteilungen mit verschiedensten Formen modellieren. Obwohl man zum Schätzen der Parameter dieser Verteilung ein gewisses statistisches Grundwissen braucht, können auch Nicht-Statistiker mittlerweile mit dieser Methode sehr gut umgehen, da Software entwickelt wurde, die das Fitten mit Hilfe dieser Verteilungen einfacher macht. Die bekannteste Quantil-Transformations Methode wurde von Clements entwickelt. Dieser verwendete Pearson-Kurven um bessere Schätzungen für die drei Quantile zu erhalten. Diese Methode werden wir uns jetzt genauer ansehen.

Clements Methode

Wie bereits erwähnt, benutzte Clements (1989) für die Schätzung der drei benötigten Quantile die Pearson-Kurven. Er schlägt folgende Vorgehensweise vor: Als erstes schätzt man den Mittelwert (\bar{X}), die Standardabweichung (S), die Schiefe (G_1) und die Kurtosis (G_2) der Daten. Mit Hilfe von Schätzungen der Schiefe und der Kurtosis, kann man nun in zwei Tabellen (Kotz & Lovelace, 1998, S. 146ff.) nachsehen, welches $z'_{0.00135}$ und welches $z'_{0.99865}$ (Clements bezeichnet sie als L_p und U_p) verwendet wird. In einer dritten Tabelle (Kotz & Lovelace, 1998, S. 153ff.) kann man für die gegebene

Schiefe und Kurtosis das passende Quantil $z'_{0.5}$ ablesen. Dabei muss man bei positiver Schiefe das Vorzeichen der Tabelle verändern.

Man kann $x_{0.00135}$ mit

$$Q_{0.00135} = \bar{X} - z'_{0.00135} \times S$$

schätzen. Das Quantil $x_{0.5}$ schätzt man mit

$$Q_{0.5} = \bar{X} + z'_{0.5} \times S$$

und $x_{0.99865}$ mit

$$Q_{0.99865} = \bar{X} + z'_{0.99865} \times S.$$

Zum Schluss kann man sich mit Hilfe von (6.6) und (6.7) die Prozessfähigkeitsindizes ausrechnen. Clements definierte die Kurtosis mit $\alpha_4 - 3$ und die Schiefe mit α_3 . Als Schätzer für die Schiefe und die Kurtosis verwendet er

$$G_1 = \frac{M_3}{S^3}$$

$$G_2 = \frac{M_4}{S^4} - 3,$$

wobei M_3 und M_4 durch

$$M_3 = \frac{\sum_{i=1}^n (X_i - \bar{X})^3}{n}$$

$$M_4 = \frac{\sum_{i=1}^n (X_i - \bar{X})^4}{n}$$

gegeben sind.

Ein Vorteil dieser Methode ist, dass sie keine mathematische Transformation erfordert und daher auch für Nicht-Statistiker sehr leicht anwendbar ist. Man muss keine komplizierte Verteilung finden, um damit den Prozess fitten zu können. Die Voraussetzung, dass der Prozess unter statistischer Kontrolle, d.h. stabil ist, muss allerdings auch hier gegeben sein. Nachteile dieser Methode sind, dass die richtige Verteilung der Stichprobenschätzer vernachlässigt wird und dass der Schätzer für die Kurtosis für einen kleinen Stichprobenumfang eine hohe Variabilität aufweist.

Andere Quantil-Transformations-Methoden

Kotz & Lovelace (1998) beschreiben weiters auch die Quantil-Transformations-Methode von Gilchrist und von Chang und Lu. Die Quantil-Transformations-Methode von Gilchrist verwendet an Stelle der Pearson-Kurven die zugrundeliegende Verteilung. Mit Hilfe der Inversen Verteilungsfunktion $x_p = F_X^{-1}(p)$ wird x_p für ein gegebenes p errechnet. Beispielsweise, wenn die Daten einer Exponentialverteilung mit einer Dichtefunktion

$$f_X(x) = \lambda e^{-\lambda x} \quad \text{für } x \geq 0, \lambda > 0$$

und Verteilungsfunktion

$$F_X(x) = 1 - e^{-\lambda x}$$

folgen, verwendet man

$$F_X^{-1}(p) = x_p = -\frac{1}{\lambda} \ln(1 - p)$$

um $x_{0.00135}$, $x_{0.5}$ und $x_{0.99865}$ zu schätzen. Hier ist $\frac{1}{\lambda}$ der Erwartungswert der Exponentialverteilung und $\hat{\lambda} = 1/\bar{X}$ der Schätzer für λ . Sehr wichtig ist bei dieser Methode, dass man einen großen Stichprobenumfang hat um den Prozess gut fitten zu können. Weitere Verteilungen findet man in Tabelle 6.1.

Verteilung	$F_X(x)$	Quantil $x_p = F_X^{-1}(p)$
Normal(μ, σ)	$\Phi\left(\frac{x-\mu}{\sigma}\right)$	$\mu + \sigma\phi^{-1}(p)$
Lognormal(μ, σ)	$\Phi\left(\frac{\ln x - \mu}{\sigma}\right)$	$\exp(\mu + \sigma\phi^{-1}(p))$
Gleichverteilung(a, b)	$\frac{x-a}{b-a}$	$a + p(b - a)$
Weibull(θ, β)	$1 - \exp\left(-\left(\frac{x}{\theta}\right)^\beta\right)$	$\theta \left[\ln\left(\frac{1}{1-p}\right)\right]^{1/\beta}$

Tabelle 6.1.: Inverse Verteilungsfunktionen

Chang und Lu entwickelten eine Methode, bei der man weder statistische Tabellen verwenden, noch Kurtosis und Schiefe berechnen muss. Diese Methode setzt voraus, dass die Stichprobe die Grundgesamtheit perfekt wiedergibt und wird folgendermaßen durchgeführt. Man setzt als erstes die Spezifikationsgrenzen und den Zielwert fest und gibt die beobachteten Daten ein. Anschließend werden die benötigten Quantile der Stichprobe ($Q_{0.00135}$, $Q_{0.5}$ und $Q_{0.99865}$) bestimmt um damit die Prozessfähigkeitssindizes zu berechnen.

6.6.3. Munechika's Annäherung

In diesem Abschnitt betrachten wir die Methode von Munechika, die bei Kotz & Lovelace (1998) beschrieben wird. Bei dieser Methode wird der C_{pk} -Wert für nicht normalverteilte Daten mit Hilfe der Cornish-Fisher-Methode geschätzt. Die Cornish-Fisher-Methode besagt, dass man jede standardisierte Zufallsvariable X (d.h. $\mu = 0$, $\sigma = 1$) mit Hilfe einer standardnormalverteilten Zufallsvariable U durch die Gleichung

$$X = U + B_1(U) + B_2(U) + \dots + B_k(U)$$

ausdrücken kann, wobei

$$B_1 = \frac{1}{6}k_3(U^2 - 1)$$

$$B_2 = \frac{1}{24}k_4(U^3 - 3U) - \frac{1}{36}k_3^2(2U^3 - 5U)$$

und k_i die Kumulanten sind. Die Kumulanten sind durch

$$\begin{aligned} k_1 &= \mu \quad (\text{Mittelwert}) \\ k_2 &= \mu_2 \\ k_3 &= \mu_3 \\ k_4 &= \mu_4 - 3\mu_2^2 \\ k_5 &= \mu_5 - 10\mu_3\mu, \end{aligned}$$

definiert, wobei μ_i das i -te zentrale Moment ist.

In seiner Annäherung führte Munechika die Cornish-Fisher-Methode nur bis zu $B_1(U)$ aus und zeigte, dass für eine χ^2 -verteilte Zufallsvariable X der Fehler, der durch das Weglassen der Restterme entsteht, vernachlässigbar ist, egal welchen Freiheitsgrad die Zufallsvariable hat.

Bei dieser Methode wird der C_{pk} -Wert mit der Funktion g_1 mit

$$g_1(u) = u + \frac{k_3}{6}(u^2 - 1) \quad (6.8)$$

korrigiert. Weiters definiert man

$$C_{p\lambda} = \lambda C_{pk},$$

wobei λ ein Korrekturfaktor ist. Angenommen $3C_{pk} = g(3C_{p\lambda})$, dann ist es möglich, die Gleichung nach λ aufzulösen.

$$\lambda = \frac{g^{-1}(3C_{pk})}{3C_{pk}}, \quad (6.9)$$

wobei g^{-1} die inverse Funktion von g ist.

Aus den Gleichungen (6.8) und (6.9) erhält man

$$\lambda = \frac{\sqrt{k_3^2 + 18C_{pk}(k_3) + 9} - 3}{3C_{pk}(k_3)} \quad k_3 \neq 0. \quad (6.10)$$

Mithilfe dieser Annahmen kann man die Annäherung folgendermaßen durchführen.

- Berechne

$$\begin{aligned} \bar{X} &= \frac{\sum X_i}{n} \\ S &= \sqrt{\frac{\sum (X_i - \bar{X})^2}{n - 1}} \\ \hat{k}_3 &= \frac{\sqrt{n} \sum (X_i - \bar{X})^3}{[\sum (X_i - \bar{X})^2]^{3/2}} \end{aligned}$$

- berechne $\hat{C}_{pk} = (OSG - \bar{X})/3S$
- berechne $\hat{\lambda}$ mit Hilfe von (6.10) unter Verwendung von \hat{C}_{pk} und \hat{k}_3
- berechne $\hat{C}_{p\lambda} = \hat{\lambda} \hat{C}_{pk}$.

Der Vorteil dieser Methode ist, dass wir nur \bar{X} , S und einen Schätzer des dritten Kumulanten benötigen. Es bietet sich natürlich an, zusätzlich auch noch $B_2(u)$ zu addieren. Dies würde allerdings einen Schätzer des vierten Kumulanten benötigen und die Berechnung von λ würde sehr komplex werden. Munechika zeigte, dass sich das nicht lohnt, da sich die Genauigkeit nicht verbessert und der Schätzer des vierten Kumulanten für einen kleinen Stichprobenumfang, wie erwähnt, große Schwankungsbreite hat.

6.6.4. S_{pmk} - Index

Chen & Ding (2001) haben einen neuen Index zur Berechnung der Prozessfähigkeit für nicht normalverteilte Populationen eingeführt. Das Besondere an diesem Index ist, dass er den Prozentsatz des fehlerhaften Materials genau wiedergibt.

Definition 6.6.1. (S_{pmk} -Index) Der S_{pmk} -Index ist gegeben durch

$$S_{pmk} = \frac{\Phi^{-1}\left(\frac{1+F(OSG)-F(USG)}{2}\right)}{3\sqrt{1+\left(\frac{\mu-T}{\sigma}\right)^2}} = \frac{\Phi^{-1}(1-P/2)}{3\sqrt{1+\left(\frac{\mu-T}{\sigma}\right)^2}},$$

wobei $F(x)$ die Verteilungsfunktion der jeweiligen Prozessverteilung ist.

Die Idee, die hinter diesem Prozessfähigkeitsindex steht, beruht auf der Eigenschaft

$$C_{pmk} = C_{pm}(1-k) = \frac{C_p}{\sqrt{1+\left(\frac{\mu-T}{\sigma}\right)^2}}(1-k) = \frac{C_{pk}}{\sqrt{1+\left(\frac{\mu-T}{\sigma}\right)^2}}$$

und der Tatsache, dass C_{pk} unter jeder Verteilung durch $\frac{1}{3}\Phi^{-1}\left(\frac{1+F(OSG)-F(USG)}{2}\right)$ ersetzt werden kann (Chen, 2000). Der Prozentsatz des fehlerhaften Materials p kann mit Hilfe von S_{pmk} genau berechnet werden und ergibt

$$p = 2\left(1 - \Phi\left(3S_{pmk}\sqrt{1+\left(\frac{\mu-T}{\sigma}\right)^2}\right)\right).$$

Weiters gilt unter Normalverteilung und $\mu = M$, dass $S_{pmk} = C_{pmk} = C_{pm}$.

Schätzer

Ein Schätzer für S_{pmk} ist

$$\hat{S}_{pmk} = \frac{\Phi^{-1}\left(\frac{1+\hat{F}(OSG)-\hat{F}(USG)}{2}\right)}{3\sqrt{1+\left(\frac{\bar{X}-T}{S}\right)^2}} = \frac{\Phi^{-1}(1-\hat{P}/2)}{3\sqrt{1+\left(\frac{\bar{X}-T}{S}\right)^2}},$$

wobei $\hat{F}(OSG)$ der Anteil der Stichprobe ist, der kleiner oder gleich OSG ist, $\hat{F}(USG)$ der Anteil der Stichprobe der kleiner USG ist. \bar{X} ist das arithmetische Mittel, S die Standardabweichung und \hat{P} der Anteil des fehlerhaften Materials der Stichprobe.

Beispiel

Wir werden uns jetzt anhand eines Beispiels aus Chen & Ding (2001) \hat{S}_{pmk} berechnen.

Beispiel 1. *Es werden Daten der Firma Tung Pei Industrial Co. Ltd verwendet. Diese Firma stellt Rollenlager her. Der innere Ring ist eine der Komponenten des Rollenlagers. Der Wert d bezeichnet den Durchmesser des inneren Rings. Der Zielwert für diesen liegt bei 60mm. Die obere Spezifikationsgrenze liegt bei 60.004mm und die untere bei 59.981mm. Folgende Daten stehen zur Verfügung ($n = 100$).*

59.984	59.981	59.981	60.003	59.982	60.005	60.004	59.983	59.981	59.980
60.000	59.998	59.982	59.983	59.981	59.982	59.999	60.001	59.982	59.988
59.995	59.998	59.982	59.983	59.981	59.994	60.002	59.988	59.980	59.982
59.982	59.983	59.981	59.986	59.987	60.001	59.982	60.003	60.001	59.984
59.985	59.979	59.987	59.990	59.998	59.984	59.989	59.999	59.985	60.003
60.004	60.001	60.000	59.982	59.981	59.984	59.998	59.983	59.999	59.987
59.991	59.992	59.992	59.983	59.981	59.996	59.997	60.000	60.000	59.991
60.002	60.001	59.990	59.987	59.982	60.006	59.981	59.982	59.984	59.985
60.003	60.004	59.992	59.991	59.986	59.992	59.991	59.981	59.998	59.985
60.001	59.980	59.993	59.984	59.981	59.984	59.988	59.999	60.000	60.001

Mit R wird ein Shapiro-Wilk Test durchgeführt. Es ergeben sich die Werte $W = 0.8834$ und $p = 2.534e^{-07}$. Da der p -Wert sehr klein ist, verwirft der Test die Annahme, dass die Daten normalverteilt sind. Um \hat{S}_{pmk} zu berechnen, benötigen wir folgende Größen: $\bar{x} = 59.9903$, $s = 0.008356$, $\hat{F}(OSG) = 0.98$ und $\hat{F}(USG) = 0.04$. Mit diesen Werten erhalten wir

$$\hat{S}_{pmk} = 0.4092.$$

Dies ist ein sehr kleiner Wert und weist daraufhin, dass der Prozess nicht fähig ist. Der Anteil des fehlerhaften Materials kann genau angegeben werden mit

$$\hat{p} = 2 \left(1 - \Phi \left(3\hat{S}_{pmk} \sqrt{1 + \left(\frac{\bar{x} - T}{s} \right)^2} \right) \right) = 0.06.$$

Es fallen genau sechs Beobachtungen außerhalb des Spezifikationsbereiches.

6.6.5. Konfidenzintervalle für nicht normalverteilte Populationen

Eine Möglichkeit Konfidenzintervalle für nicht normalverteilte Populationen zu berechnen bietet die Bootstrapping Methode. Sundaraiyer (1996) berechnet mit Hilfe von Bootstrapping Konfidenzintervalle für C_{pk} -Werte, die durch die Clements Methode bestimmt worden sind. Als zugrundeliegende Verteilung wird dabei die Inverse-Gauß Verteilung verwendet.

Bootstrapping

Sei x_1, x_2, \dots, x_n eine Stichprobe, die von einer bestimmten Verteilung gezogen worden ist. Bei der Bootstrapping Methode werden wiederholt sogenannte Bootstrap Stichproben $x^{*1}, x^{*2}, \dots, x^{*B}$ mit Größe n aus der ursprünglichen Stichprobe mit Zurücklegen gezogen. Für jede gezogene Stichprobe wird anschließend der Punktschätzer \hat{C}^* berechnet. Die Anzahl B der gezogenen Bootstrap Stichproben sollte dabei ausreichend groß sein. Sundaraiyer (1996) meint, dass man für die Berechnung von Konfidenzintervallen mindestens $B = 1000$ Bootstrap Stichproben ziehen sollte.

Konfidenzintervalle

Mit Hilfe der Bootstrap Stichproben werden jetzt Konfidenzintervalle berechnet. Wir betrachten hier drei verschiedene Konfidenzintervalle, die bei Efron & Tibshirani (1986) beschrieben werden. Das Standard Bootstrap Intervall, das Perzentil Bootstrap Intervall und das bias-korrigierte Bootstrap Intervall werden im Folgenden definiert.

Definition 6.6.2. (Standard Bootstrap Intervall) Sei $\hat{C}^*(i)$ der i -te geordnete Punktschätzer des Prozessfähigkeitsindex C^* , der aus einer der Bootstrap Stichproben berechnet wurde, dann ist

$$\bar{C}^* = \frac{1}{B} \sum_{i=1}^B \hat{C}^*(i)$$

der Mittelwert aller Punktschätzer und

$$S_C^* = \sqrt{\frac{\sum_{i=1}^B (\hat{C}^*(i) - \bar{C}^*)^2}{B - 1}}$$

die Standardabweichung der Punktschätzer. Das $(1 - \alpha)$ -Standard Bootstrap Konfidenzintervall ist gegeben mit

$$\bar{C}^* \mp z_{1-\frac{\alpha}{2}} S_C^*,$$

wobei z_α das α -Quantil der Standardnormalverteilung ist.

Definition 6.6.3. (Perzentil Bootstrap Intervall) Mit den geordneten Punktschätzern $\hat{C}^*(i)$, $i = 1, \dots, B$, kann das $(1 - \alpha)100\%$ -Perzentil Bootstrap Intervall durch

$$\left[\hat{C}^*\left(\frac{\alpha}{2}B\right), \hat{C}^*\left(1 - \frac{\alpha}{2}B\right) \right]$$

definiert werden.

Da die Bootstrap Verteilung möglicherweise auch verzerrt sein könnte, kann man diesen Bias mit nachfolgendem Konfidenzintervall korrigieren.

Definition 6.6.4. (*Bias-korrigiertes Bootstrap Intervall*) Sei

$$p_0 = \mathbb{P}[\hat{C}^* \leq c],$$

wobei c der Punktschätzer C^* der ursprünglichen Stichprobe ist. Weiters sei

$$z_0 = \Phi^{-1}(p_0),$$

$$P_L = \Phi(2z_0 - z_{1-\frac{\alpha}{2}}),$$

$$P_U = \Phi(2z_0 + z_{1-\frac{\alpha}{2}}),$$

wobei $\Phi(\cdot)$ die Verteilungsfunktion der Standardnormalverteilung ist. Dann ist das $(1 - \alpha)$ -Bias-korrigierte Bootstrap Konfidenzintervall durch

$$[\hat{C}^*(P_L B), \hat{C}^*(P_U B)]$$

gegeben.

6.7. Simulierte Zufallsvariablen

In diesem Abschnitt werden Zufallsvariablen aus der Gamma-, Lognormal- und Weibullverteilung simuliert um die Methoden zu testen. Die Daten werden aus Verteilungen mit Erwartungswert $\mu = 100$ und Standardabweichung $\sigma = 5$ simuliert. Wir setzen den Target T gleich 100, die obere Spezifikationsgrenze OSG auf 120 und die untere Spezifikationsgrenze USG auf 80. Insgesamt werden jeweils vier Stichproben simuliert, die den Stichprobenumfang $n = 50, 100, 1000$ und 10000 haben. Für jede Verteilung werden drei verschiedene Parametereinstellungen gewählt, um für jede Verteilung je eine Einstellung mit geringer, mittlerer und hoher Schiefe zu haben. Der Grund dafür ist, dass man die Auswirkung der Schiefe auf die Abweichungen in den Prozessfähigkeitsindizes untersuchen will.

Pro Beispiel gibt es zwei Ergebnistabellen. Im Kopf der ersten Tabelle sind die theoretischen Werte der jeweiligen Verteilung angegeben. In den ersten zwei Spalten findet man folgende Kenngrößen der jeweiligen Stichprobe: der p -Wert des Anderson-Darling-Tests, mit der Nullhypothese, dass die Daten normalverteilt sind, der Mittelwert (\bar{x}), die Standardabweichung (s), die empirische Schiefe ($\hat{\gamma}_e$), die Maximum Likelihood Schätzungen der Verteilungsparameter und die Schiefe ($\hat{\gamma}$), welche die Verteilung mit den Maximum Likelihood Parametern hat.

Weiters sind die Prozessfähigkeitsindizes, die mit fünf verschiedenen Methoden berechnet werden, angegeben. In der ersten Spalte findet man die Werte der Normalverteilungsmethode und in der zweiten Spalte die Werte, die mit Hilfe der Quantile der jeweiligen Verteilung (Gamma, Lognormal und Weibull) berechnet werden. Bei der Methode in der zweiten Spalte werden die Maximum Likelihood Schätzer der Parameterwerte verwendet. In der dritten Spalte sind die Werte der Methode, welche

die empirischen Quantile benutzt, angegeben und die letzten zwei Spalten beinhalten die Ergebnisse der Methode nach Clements und der Annäherung von Munechicka. Zusätzlich wird auch - wenn möglich - der Index S_{pmk} berechnet. Dies ist bei fast symmetrischen Daten aufgrund der Definition von S_{pmk} nicht möglich.

Die zweite Tabelle beinhaltet alle Konfidenzintervalle der jeweiligen Stichprobe. Die Konfidenzintervalle werden in der ersten Spalte unter Annahme der Normalverteilung berechnet. In den anderen drei Spalten sind die Bootstrap Konfidenzintervalle angegeben. Die Berechnung der Prozessfähigkeitsindizes wird für jede Bootstrap Stichprobe wie in der Spalte Gamma, Lognormal und Weibull durchgeführt. Es werden dabei jeweils $B = 1000$ Bootstrap Stichproben gezogen. Die letzte Spalte (Q) der Konfidenzintervall-Tabelle beinhaltet die Länge des Konfidenzintervalls, berechnet unter Annahme der Normalverteilung, dividiert durch die Länge des bias-korrigierten Bootstrap Konfidenzintervalls. In den Spalten, die mit \ddot{U} gekennzeichnet sind, wird angegeben, ob das in der Spalte davorliegende Konfidenzintervall den theoretischen Wert überdeckt.

6.7.1. Gammaverteilung

In den nachfolgenden Tabellen konnte Folgendes beobachtet werden. Die Methode, welche die empirischen Quantile verwendet und die Clements Methode überschätzen die Prozessfähigkeitsindizes für kleinere Stichprobenumfänge. Wird der Stichprobenumfang größer, nähern sich die Werte den Werten der anderen Methoden für nicht normalverteilte Populationen an. Diese zwei Methoden sind bei einem kleinem Stichprobenumfang noch sehr unzuverlässig.

Weiters kann man sehen, dass je größer die Schiefe, desto größer wird der Unterschied zwischen C_{pu} - und C_{pl} -Wert. Der C_{pu} -Wert wird als C_{pk} -Wert hergenommen. Dieser ist sehr klein, und daher kann man erkennen, dass die Methode, die Normalverteilung voraussetzt, zu einer Überschätzung des C_{pk} -Wertes führt. Auch für den C_p -Wert gilt, dass die Werte der Methode, die normalverteilte Daten voraussetzt, mit steigender Schiefe immer mehr von den Methoden für nicht normalverteilte Populationen abweicht. Man kann feststellen, dass der C_p -Wert unter Normalverteilungsannahme bei einem Datensatz mit großer positiver Schiefe den wahren Wert sehr stark überschätzt.

Bei den Konfidenzintervallen des C_p -Wertes sieht man, dass diese bei Beispiel 2 und 3 (Schiefen $\gamma = 1$ und 2) unter Annahme der Normalverteilung viel zu klein sind. Der Quotient, der die Intervalle vergleicht, ist bei diesen Stichproben immer kleiner 1. Die Konfidenzintervalle des C_{pk} -Wertes sind nicht miteinander vergleichbar, da diese durch die Überschätzung des C_{pk} -Wertes bei der Methode, die Normalverteilung voraussetzt, in den meisten Fällen keine gemeinsamen Werte enthalten. In Beispiel 1 - bei geringer Schiefe - passen diese noch gut zusammen. Hier kann man sehen, dass die Bootstrap Konfidenzintervalle immer kleiner sind. Dies ist aber möglicherweise darauf zurückzuführen, dass das Konfidenzintervall von C_{pk} , das Normalverteilung voraussetzt, verzerrt ist.

Insgesamt kann man also sagen, dass die Methoden für nicht normalverteilte Populationen mit steigendem Stichprobenumfang sehr gut zusammenpassen und diese sich

aber - teilweise sehr beträchtlich - von der Methode, die Normalverteilung annimmt, unterscheiden. Die Unterschiede sind bei dem Prozessfähigkeitsindex C_{pk} viel größer.

Gammaverteilung(400,4) - $\mu = 100$, $\sigma = 5$ und $\gamma = 0.1$ theoretische Werte: $C_p = 1.33$, $C_{pk} = 1.27$, $C_{pu} = 1.27$, $C_{pl} = 1.40$, $S_{pmk} = 1.31$							
n = 50			Normal	Gamma	emp.	Clements	Mun.
$p = 0.918$	shape = 360	c_p	1.26	1.26	1.75	1.42	1.25
$\bar{x} = 100.29$	rate = 3.59	c_{pk}	1.24	1.19	1.70	1.40	
$s = 5.29$	$\hat{\gamma} = 0.105$	c_{pu}	1.24	1.19	1.70	1.40	
$\hat{\gamma}_e = -0.012$		c_{pl}	1.28	1.34	1.79	1.45	
n = 100			Normal	Gamma	emp.	Clements	Mun.
$p = 0.324$	shape = 399	c_p	1.33	1.33	1.50	1.31	1.15
$\bar{x} = 100.07$	rate = 3.99	c_{pk}	1.33	1.27	1.41	1.15	
$s = 5.01$	$\hat{\gamma} = 0.100$	c_{pu}	1.33	1.27	1.41	1.15	
$\hat{\gamma}_e = 0.293$		c_{pl}	1.34	1.40	1.59	1.52	
n = 1000			Normal	Gamma	emp.	Clements	Mun.
$p = 0.149$	shape = 358	c_p	1.26	1.26	1.26	1.27	1.19
$\bar{x} = 99.88$	rate = 3.58	c_{pk}	1.26	1.21	1.16	1.22	
$s = 5.28$	$\hat{\gamma} = 0.106$	c_{pu}	1.27	1.21	1.16	1.22	
$\hat{\gamma}_e = 0.123$		c_{pl}	1.26	1.32	1.38	1.32	
n = 10000			Normal	Gamma	emp.	Clements	Mun.
$p = \mathbf{0.001}$	shape = 400	c_p	1.33	1.33	1.32	1.34	1.26
$\bar{x} = 99.99$	rate = 4	c_{pk}	1.33	1.28	1.27	1.28	
$s = 4.99$	$\hat{\gamma} = 0.100$	c_{pu}	1.33	1.28	1.27	1.28	
$\hat{\gamma}_e = 0.097$		c_{pl}	1.33	1.40	1.37	1.40	

Tabelle 6.2.: Gammaverteilung - Beispiel 1

Gammaverteilung(400,4) - $C_p = 1.33$, $C_{pk} = 1.27$								
n		Normal	Ü	Bootstrap	Perzentil	Bias-korr.	Ü	Q
50	C_p	(1.01, 1.51)	✓	(1.06, 1.52)	(1.10, 1.55)	(1.07, 1.51)	✓	1.149
50	C_{pk}	(0.92, 1.60)	✓	(0.98, 1.44)	(1.01, 1.47)	(0.99, 1.43)	✓	1.520
100	C_p	(1.15, 1.52)	✓	(1.14, 1.55)	(1.15, 1.57)	(1.14, 1.55)	✓	0.904
100	C_{pk}	(1.08, 1.59)	✓	(1.06, 1.51)	(1.08, 1.52)	(1.06, 1.50)	✓	1.148
1000	C_p	(1.21, 1.32)	×	(1.21, 1.32)	(1.21, 1.32)	(1.21, 1.32)	×	0.994
1000	C_{pk}	(1.18, 1.33)	✓	(1.15, 1.28)	(1.15, 1.28)	(1.15, 1.27)	✓	1.229
10000	C_p	(1.32, 1.35)	✓	(1.32, 1.35)	(1.32, 1.35)	(1.32, 1.35)	✓	1.016
10000	C_{pk}	(1.31, 1.36)	×	(1.26, 1.30)	(1.26, 1.30)	(1.26, 1.30)	✓	1.287

Tabelle 6.3.: Konfidenzintervalle Gammaverteilung - Beispiel 1

Gammaverteilung(4,1) - $\mu = 100$, $\sigma = 5$ und $\gamma = 1$							
theoretische Werte: $C_p = 1.31$, $C_{pk} = 0.92$, $C_{pu} = 0.92$, $C_{pl} = 2.39$, $S_{pmk} = 1.02$							
n = 50			Normal	Gamma	emp.	Clements	Mun.
$p = \mathbf{0.001}$	shape = 2.304	C_p	1.20	1.11	1.43	1.25	0.80
$\bar{x} = 100.77$	rate = 0.263	C_{pk}	1.15	0.71	0.99	0.78	
$s = 5.57$	$\hat{\gamma} = 1.317$	C_{pu}	1.15	0.71	0.99	0.78	
$\hat{\gamma}_e = 1.337$	$S_{pmk} = 0.77$	C_{pl}	1.24	2.71	2.74	3.43	
n = 100			Normal	Gamma	emp.	Clements	
$p = \mathbf{0.001}$	shape = 2.287	C_p	1.18	1.10	1.53	1.34	0.85
$\bar{x} = 100.81$	rate = 0.260	C_{pk}	1.13	0.71	1.06	0.90	
$s = 5.64$	$\hat{\gamma} = 1.322$	C_{pu}	1.13	0.71	1.06	0.90	
$\hat{\gamma}_e = 0.939$		C_{pl}	1.23	2.71	2.78	2.67	
n = 1000			Normal	Gamma	emp.	Clements	
$p = \mathbf{0.000}$	shape = 2.874	C_p	1.30	1.23	1.32	1.26	0.92
$\bar{x} = 99.97$	rate = 0.320	C_{pk}	1.30	0.83	0.93	0.90	
$s = 5.12$	$\hat{\gamma} = 1.180$	C_{pu}	1.30	0.83	0.93	0.90	
$\hat{\gamma}_e = 1.048$	$S_{pmk} = 0.99$	C_{pl}	1.30	2.57	2.44	2.25	
n = 10000			Normal	Gamma	emp.	Clements	
$p = \mathbf{0.000}$	shape = 4.049	C_p	1.35	1.32	1.35	1.30	0.95
$\bar{x} = 99.98$	rate = 0.406	C_{pk}	1.34	0.93	0.96	0.92	
$s = 4.96$	$\hat{\gamma} = 0.994$	C_{pu}	1.35	0.93	0.96	0.92	
$\hat{\gamma}_e = 0.992$	$S_{pmk} = 1.04$	C_{pl}	1.34	2.40	2.42	2.33	

Tabelle 6.4.: Gammaverteilung - Beispiel 2

Gammaverteilung(4,1) - $C_p = 1.31$, $C_{pk} = 0.92$								
n		Normal	Ü	Bootstrap	Perzentil	Bias-korr.	Ü	Q
50	C_p	(0.96, 1.43)	✓	(0.76, 1.50)	(0.81, 1.55)	(0.82, 1.61)	✓	0.598
50	C_{pk}	(0.85, 1.49)	✓	(0.44, 0.99)	(0.48, 1.05)	(0.51, 1.10)	✓	1.089
100	C_p	(1.02, 1.35)	✓	(0.88, 1.28)	(0.89, 1.29)	(0.93, 1.35)	✓	0.788
100	C_{pk}	(0.92, 1.37)	✓	(0.52, 0.85)	(0.53, 0.86)	(0.57, 0.92)	✓	1.284
1000	C_p	(1.24, 1.36)	✓	(1.15, 1.30)	(1.16, 1.31)	(1.15, 1.30)	×	0.787
1000	C_{pk}	(1.22, 1.38)	×	(0.77, 0.90)	(0.77, 0.90)	(0.77, 0.90)	×	1.259
10000	C_p	(1.33, 1.36)	✓	(1.30, 1.34)	(1.30, 1.34)	(1.30, 1.34)	✓	0.810
10000	C_{pk}	(1.32, 1.37)	×	(0.91, 0.96)	(0.91, 0.95)	(0.91, 0.95)	✓	1.202

Tabelle 6.5.: Konfidenzintervalle Gammaverteilung - Beispiel 2

Gammaverteilung(1,1) - $\mu = 100$, $\sigma = 5$ und $\gamma = 2$							
theoretische Werte: $C_p = 1.21$, $C_{pk} = 0.73$, $C_{pu} = 0.73$, $C_{pl} = 5.34$, $S_{pmk} = 0.90$							
n = 50			Normal	Gamma	emp.	Clements	Mun.
$p = \mathbf{0.000}$	shape = 0.831	c_p	1.13	1.04	1.56	1.34	0.74
$\bar{x} = 100.21$	rate = 0.159	c_{pk}	1.12	0.61	0.97	0.81	
$s = 5.88$	$\hat{\gamma} = 2.193$	c_{pu}	1.12	0.61	0.97	0.81	
$\hat{\gamma}_e = 1.728$	$S_{pmk} = 0.77$	c_{pl}	1.15	5.51	5.50	8.25	
n = 100			Normal	Gamma	emp.	Clements	
$p = \mathbf{0.000}$	shape = 1.049	c_p	1.39	1.28	1.64	1.44	0.87
$\bar{x} = 99.87$	rate = 0.216	c_{pk}	1.38	0.78	1.03	0.89	
$s = 4.79$	$\hat{\gamma} = 1.953$	c_{pu}	1.40	0.78	1.03	0.89	
$\hat{\gamma}_e = 1.649$	$S_{pmk} = 0.86$	c_{pl}	1.38	5.38	5.44	5.48	
n = 1000			Normal	Gamma	emp.	Clements	
$p = \mathbf{0.000}$	shape = 0.975	c_p	1.29	1.15	1.30	1.21	0.78
$\bar{x} = 100.17$	rate = 0.189	c_{pk}	1.27	0.69	0.78	0.73	
$s = 5.19$	$\hat{\gamma} = 2.025$	c_{pu}	1.27	0.69	0.78	0.73	
$\hat{\gamma}_e = 1.937$	$S_{pmk} = 0.86$	c_{pl}	1.30	5.23	5.18	5.43	
n = 10000			Normal	Gamma	emp.	Clements	
$p = \mathbf{0.000}$	shape = 0.994	c_p	1.32	1.20	1.23	1.21	0.79
$\bar{x} = 100.02$	rate = 0.198	c_{pk}	1.32	0.72	0.74	0.73	
$s = 5.06$	$\hat{\gamma} = 2.006$	c_{pu}	1.32	0.72	0.74	0.73	
$\hat{\gamma}_e = 2.007$	$S_{pmk} = 0.89$	c_{pl}	1.32	5.33	5.30	5.60	

Tabelle 6.6.: Gammaverteilung - Beispiel 3

Gammaverteilung(1,1) - $C_p = 1.21$, $C_{pk} = 0.73$								
n		Normal	Ü	Bootstrap	Perzentil	Bias-korr.	Ü	Q
50	C_p	(0.91, 1.36)	✓	(0.67, 1.50)	(0.74, 1.58)	(0.72, 1.49)	✓	0.579
50	C_{pk}	(0.82, 1.46)	×	(0.36, 0.93)	(0.42, 0.98)	(0.41, 0.92)	✓	1.220
100	C_p	(1.20, 1.58)	✓	(1.00, 1.62)	(1.05, 1.66)	(1.04, 1.65)	✓	0.637
100	C_{pk}	(1.13, 1.65)	×	(0.58, 1.01)	(0.62, 1.04)	(0.61, 1.03)	✓	1.222
1000	C_p	(1.23, 1.34)	×	(1.06, 1.25)	(1.06, 1.25)	(1.07, 1.25)	✓	0.604
1000	C_{pk}	(1.20, 1.35)	×	(0.63, 0.75)	(0.63, 0.76)	(0.63, 0.76)	✓	1.200
10000	C_p	(1.30, 1.34)	×	(1.17, 1.23)	(1.17, 1.24)	(1.17, 1.24)	✓	0.579
10000	C_{pk}	(1.29, 1.34)	×	(0.70, 0.74)	(0.70, 0.75)	(0.70, 0.75)	✓	1.128

Tabelle 6.7.: Konfidenzintervalle Gammaverteilung - Beispiel 3

6.7.2. Lognormalverteilung

In den nachfolgenden Tabellen sieht man die Ergebnisse der lognormalverteilten Zufallsvariablen. Man kann beobachten, dass die Aussagen, die wir in Abschnitt 6.7.1 getroffen haben, auch auf die lognormalverteilten Zufallsvariablen zutreffen.

Lognormalverteilung(4.61,0.05) - $\mu = 100$, $\sigma = 5$ und $\gamma = 0.150$ theoretische Werte: $C_p = 1.32$, $C_{pk} = 1.20$, $C_{pu} = 1.20$, $C_{pl} = 1.46$, $S_{pmk} = 1.24$							
n = 50			Normal	Logn.	emp.	Clements	Mun.
$p = 0.514$	meanl = 4.595	c_p	1.35	1.37	1.85	1.73	
$\bar{x} = 99.12$	sdl = 0.049	c_{pk}	1.29	1.34	1.74	1.53	1.22
$s = 4.94$	$\hat{\gamma} = 0.147$	c_{pu}	1.41	1.34	1.74	1.53	
$\hat{\gamma}_e = 0.277$		c_{pl}	1.29	1.40	1.99	2.04	
n = 100			Normal	Logn.	emp.	Clements	Mun.
$p = 0.894$	meanl = 4.618	c_p	1.32	1.32	1.65	1.40	
$\bar{x} = 101.42$	sdl = 0.050	c_{pk}	1.23	1.15	1.45	1.24	1.18
$s = 5.04$	$\hat{\gamma} = 0.149$	c_{pu}	1.23	1.15	1.45	1.24	
$\hat{\gamma}_e = 0.082$		c_{pl}	1.42	1.52	1.85	1.58	
n = 1000			Normal	Logn.	emp.	Clements	Mun.
$p = 0.213$	meanl = 4.607	c_p	1.40	1.39	1.56	1.48	
$\bar{x} = 100.34$	sdl = 0.048	c_{pk}	1.37	1.29	1.43	1.39	1.31
$s = 4.77$	$\hat{\gamma} = 0.143$	c_{pu}	1.37	1.29	1.43	1.39	
$\hat{\gamma}_e = 0.076$		c_{pl}	1.42	1.52	1.71	1.59	
n = 10000			Normal	Logn.	emp.	Clements	Mun.
$p = 0.000$	meanl = 4.609	c_p	1.34	1.33	1.33	1.34	
$\bar{x} = 100.51$	sdl = 0.050	c_{pk}	1.30	1.22	1.24	1.25	1.20
$s = 4.99$	$\hat{\gamma} = 0.149$	c_{pu}	1.30	1.22	1.24	1.25	
$\hat{\gamma}_e = 0.145$	$S_{pmk} = 1.29$	c_{pl}	1.37	1.47	1.43	1.44	

Tabelle 6.8.: Lognormalverteilung - Beispiel 1

Lognormalverteilung(4.61,0.05) - $C_p = 1.32$, $C_{pk} = 1.20$								
n		Normal	Ü	Bootstrap	Perzentil	Bias-korr.	Ü	Q
50	C_p	(1.08, 1.62)	✓	(1.14, 1.67)	(1.19, 1.71)	(1.16, 1.63)	✓	1.146
50	C_{pk}	(0.96, 1.66)	✓	(1.09, 1.63)	(1.12, 1.66)	(1.10, 1.62)	✓	1.338
100	C_p	(1.14, 1.51)	✓	(1.16, 1.53)	(1.17, 1.55)	(1.14, 1.51)	✓	1.014
100	C_{pk}	(1.00, 1.47)	✓	(0.98, 1.35)	(1.00, 1.37)	(0.98, 1.35)	✓	1.303
1000	C_p	(1.34, 1.46)	×	(1.33, 1.46)	(1.33, 1.46)	(1.33, 1.46)	×	0.975
1000	C_{pk}	(1.29, 1.46)	×	(1.22, 1.35)	(1.23, 1.36)	(1.23, 1.36)	×	1.234
10000	C_p	(1.32, 1.35)	✓	(1.32, 1.35)	(1.32, 1.35)	(1.32, 1.35)	✓	1.082
10000	C_{pk}	(1.28, 1.33)	×	(1.20, 1.24)	(1.20, 1.24)	(1.20, 1.24)	✓	1.314

Tabelle 6.9.: Konfidenzintervalle Lognormalverteilung - Beispiel 1

Lognormalverteilung(0,0.21) - $\mu = 100$, $\sigma = 5$ und $\gamma = 0.647$ theoretische Werte: $C_p = 1.29$, $C_{pk} = 1.02$, $C_{pu} = 1.02$, $C_{pl} = 1.81$, $S_{pmk} = 1.08$							
n = 50			Normal	Logn.	emp.	Clements	Mun.
$p = 0.417$	meanl = 3.118	c_p	1.29	1.24	1.74	1.46	1.08
$\bar{x} = 99.66$	sdl = 0.222	c_{pk}	1.26	0.98	1.53	1.23	
$s = 5.18$	$\hat{\gamma} = 0.685$	c_{pu}	1.31	0.98	1.53	1.23	
$\hat{\gamma}_e = 0.443$		c_{pl}	1.26	1.74	2.06	1.82	
n = 100			Normal	Logn.	emp.	Clements	Mun.
$p = 0.261$	meanl = 3.160	c_p	1.46	1.41	1.95	1.87	1.22
$\bar{x} = 100.49$	sdl = 0.191	c_{pk}	1.42	1.10	1.68	1.54	
$s = 4.57$	$\hat{\gamma} = 0.585$	c_{pu}	1.42	1.10	1.68	1.54	
$\hat{\gamma}_e = 0.302$		c_{pl}	1.50	1.95	2.34	2.36	
n = 1000			Normal	Logn.	emp.	Clements	Mun.
$p = \mathbf{0.000}$	meanl = 3.135	c_p	1.28	1.25	1.18	1.21	0.97
$\bar{x} = 100.05$	sdl = 0.217	c_{pk}	1.28	0.97	0.89	0.94	
$s = 5.20$	$\hat{\gamma} = 0.669$	c_{pu}	1.28	0.97	0.89	0.94	
$\hat{\gamma}_e = 0.740$	$S_{pmk} = 0.99$	c_{pl}	1.29	1.77	1.83	1.73	
n = 10000			Normal	Logn.	emp.	Clements	Mun.
$p = \mathbf{0.000}$	meanl = 3.136	c_p	1.32	1.28	1.28	1.25	1.01
$\bar{x} = 100.03$	sdl = 0.212	c_{pk}	1.32	1.00	1.02	0.97	
$s = 5.05$	$\hat{\gamma} = 0.652$	c_{pu}	1.32	1.00	1.02	0.97	
$\hat{\gamma}_e = 0.675$	$S_{pmk} = 1.08$	c_{pl}	1.32	1.80	1.76	1.78	

Tabelle 6.10.: Lognormalverteilung - Beispiel 2

Lognormalverteilung(0,0.21) - $C_p = 1.29$, $C_{pk} = 1.02$								
n		Normal	Ü	Bootstrap	Perzentil	Bias-korr.	Ü	Q
50	C_p	(1.03, 1.54)	✓	(0.99, 1.54)	(1.03, 1.56)	(1.02, 1.54)	✓	0.988
50	C_{pk}	(0.94, 1.63)	✓	(0.73, 1.28)	(0.77, 1.31)	(0.75, 1.28)	✓	1.318
100	C_p	(1.26, 1.66)	✓	(1.23, 1.61)	(1.23, 1.64)	(1.21, 1.60)	✓	1.043
100	C_{pk}	(1.17, 1.70)	×	(0.92, 1.31)	(0.92, 1.32)	(0.90, 1.30)	✓	1.323
1000	C_p	(1.23, 1.34)	✓	(1.18, 1.32)	(1.19, 1.32)	(1.18, 1.32)	✓	0.841
1000	C_{pk}	(1.20, 1.36)	×	(0.91, 1.04)	(0.91, 1.04)	(0.91, 1.04)	✓	1.159
10000	C_p	(1.30, 1.34)	×	(1.26, 1.30)	(1.26, 1.30)	(1.26, 1.30)	✓	0.890
10000	C_{pk}	(1.29, 1.34)	×	(0.98, 1.02)	(0.98, 1.02)	(0.98, 1.02)	✓	1.234

Tabelle 6.11.: Konfidenzintervalle Lognormalverteilung - Beispiel 2

Lognormalverteilung(0,0.55) - $\mu = 100$, $\sigma = 5$ und $\gamma = 1.993$ theoretische Werte: $C_p = 1.11$, $C_{pk} = 0.70$, $C_{pu} = 0.70$, $C_{pl} = 3.23$, $S_{pmk} = 0.91$							
n = 50			Normal	Logn.	emp.	Clements	Mun.
$p = \mathbf{0.002}$	meanl = 2.037	c_p	1.41	1.11	2.13	1.97	1.00
$\bar{x} = 100.31$	sdl = 0.530	c_{pk}	1.39	0.69	1.50	1.31	
$s = 4.74$	$\hat{\gamma} = 1.894$	c_{pu}	1.39	0.69	1.50	1.31	
$\hat{\gamma}_e = 0.885$		c_{pl}	1.43	3.14	3.89	4.48	
n = 100			Normal	Logn.	emp.	Clements	Mun.
$p = \mathbf{0.000}$	meanl = 2.013	c_p	1.17	0.92	1.51	1.29	0.78
$\bar{x} = 100.44$	sdl = 0.594	c_{pk}	1.14	0.57	1.00	0.80	
$s = 5.70$	$\hat{\gamma} = 2.228$	c_{pu}	1.14	0.57	1.00	0.80	
$\hat{\gamma}_e = 1.441$	$S_{pmk} = 0.86$	c_{pl}	1.20	3.05	3.53	4.28	
n = 1000			Normal	Logn.	emp.	Clements	Mun.
$p = \mathbf{0.000}$	meanl = 2.020	c_p	1.35	1.10	1.28	1.26	0.85
$\bar{x} = 100.21$	sdl = 0.538	c_{pk}	1.33	0.69	0.82	0.79	
$s = 4.95$	$\hat{\gamma} = 1.934$	c_{pu}	1.33	0.69	0.82	0.79	
$\hat{\gamma}_e = 1.599$	$S_{pmk} = 0.92$	c_{pl}	1.36	3.15	3.31	3.71	
n = 10000			Normal	Logn.	emp.	Clements	Mun.
$p = \mathbf{0.000}$	meanl = 1.981	c_p	1.35	1.12	1.09	1.15	0.82
$\bar{x} = 99.91$	sdl = 0.544	c_{pk}	1.35	0.71	0.69	0.73	
$s = 4.93$	$\hat{\gamma} = 1.962$	c_{pu}	1.36	0.71	0.69	0.73	
$\hat{\gamma}_e = 1.939$	$S_{pmk} = 0.91$	c_{pl}	1.35	3.22	3.20	3.37	

Tabelle 6.12.: Lognormalverteilung - Beispiel 3

Lognormalverteilung(0,0.55) - $C_p = 1.11$, $C_{pk} = 0.70$								
n		Normal	Ü	Bootstrap	Perzentil	Bias-korr.	Ü	Q
50	C_p	(1.13, 1.69)	×	(0.78, 1.50)	(0.83, 1.55)	(0.81, 1.50)	✓	0.800
50	C_{pk}	(1.04, 1.77)	×	(0.45, 1.00)	(0.49, 1.04)	(0.47, 1.01)	✓	1.372
100	C_p	(1.01, 1.33)	✓	(0.69, 1.22)	(0.73, 1.24)	(0.69, 1.21)	✓	0.617
100	C_{pk}	(0.93, 1.38)	×	(0.39, 0.79)	(0.43, 0.81)	(0.41, 0.78)	✓	1.202
1000	C_p	(1.29, 1.41)	×	(1.01, 1.19)	(1.02, 1.19)	(1.02, 1.19)	✓	0.693
1000	C_{pk}	(1.26, 1.41)	×	(0.63, 0.76)	(0.63, 0.76)	(0.63, 0.76)	✓	1.227
10000	C_p	(1.33, 1.37)	×	(1.09, 1.15)	(1.09, 1.15)	(1.09, 1.15)	✓	0.640
10000	C_{pk}	(1.32, 1.37)	×	(0.69, 0.74)	(0.69, 0.74)	(0.69, 0.74)	✓	1.125

Tabelle 6.13.: Konfidenzintervalle Lognormalverteilung - Beispiel 3

6.7.3. Weibullverteilung

Zum Schluss sehen wir uns noch die Ergebnisse der weibullverteilten Zufallsvariablen an. Auch hier bestätigen sich die Aussagen aus Abschnitt 6.7.1.

Weibullverteilung(4,1) - $\mu = 100$, $\sigma = 5$ und $\gamma = -0.087$ theoretische Werte: $C_p = 1.44$, $C_{pk} = 1.42$, $C_{pu} = 1.46$, $C_{pl} = 1.42$, $S_{pmk} = 1.62$							
n = 50			Normal	Weibull	emp.	Clements	Mun.
$p = 0.166$	shape = 3.600	C_p	1.15	1.28	1.79	1.77	
$\bar{x} = 100.32$	scale = 20.384	C_{pk}	1.13	1.22	1.68	1.64	1.11
$s = 5.78$	$\hat{\gamma} = 0.000$	C_{pu}	1.13	1.22	1.68	1.64	
$\hat{\gamma}_e = 0.050$		C_{pl}	1.17	1.35	1.92	1.93	
n = 100			Normal	Weibull	emp.	Clements	Mun.
$p = 0.577$	shape = 4.103	C_p	1.33	1.43	1.62	1.41	
$\bar{x} = 100.31$	scale = 20.165	C_{pk}	1.31	1.42	1.58	1.36	1.37
$s = 5.02$	$\hat{\gamma} = -0.107$	C_{pu}	1.31	1.45	1.67	1.46	
$\hat{\gamma}_e = -0.070$		C_{pl}	1.35	1.42	1.58	1.36	
n = 1000			Normal	Weibull	emp.	Clements	Mun.
$p = 0.876$	shape = 3.900	C_p	1.31	1.40	1.30	1.31	
$\bar{x} = 100.00$	scale = 19.876	C_{pk}	1.31	1.39	1.20	1.31	1.28
$s = 5.10$	$\hat{\gamma} = -0.067$	C_{pu}	1.31	1.41	1.20	1.31	
$\hat{\gamma}_e = 0.030$		C_{pl}	1.31	1.39	1.40	1.31	
n = 10000			Normal	Weibull	emp.	Clements	Mun.
$p = \mathbf{0.000}$	shape = 4.025	C_p	1.33	1.43	1.45	1.41	
$\bar{x} = 99.99$	scale = 19.842	C_{pk}	1.33	1.41	1.40	1.34	1.42
$s = 5.02$	$\hat{\gamma} = -0.092$	C_{pu}	1.33	1.46	1.50	1.48	
$\hat{\gamma}_e = -0.100$		C_{pl}	1.33	1.41	1.40	1.34	

Tabelle 6.14.: Weibullverteilung - Beispiel 1

Weibullverteilung(4,1) - $C_p = 1.44$, $C_{pk} = 1.42$								
n		Normal	Ü	Bootstrap	Perzentil	Bias-korr.	Ü	Q
50	C_p	(0.93, 1.38)	×	(1.14, 1.47)	(1.15, 1.48)	(1.13, 1.45)	✓	1.421
50	C_{pk}	(0.83, 1.47)	✓	(1.03, 1.46)	(1.03, 1.45)	(0.99, 1.41)	×	1.519
100	C_p	(1.14, 1.51)	✓	(1.29, 1.60)	(1.31, 1.62)	(1.30, 1.59)	✓	1.240
100	C_{pk}	(1.07, 1.56)	✓	(1.28, 1.54)	(1.25, 1.52)	(1.25, 1.52)	✓	1.835
1000	C_p	(1.25, 1.36)	×	(1.35, 1.45)	(1.35, 1.45)	(1.35, 1.45)	✓	1.093
1000	C_{pk}	(1.23, 1.39)	×	(1.34, 1.43)	(1.33, 1.42)	(1.33, 1.42)	✓	1.662
10000	C_p	(1.31, 1.35)	×	(1.42, 1.45)	(1.42, 1.45)	(1.42, 1.46)	✓	1.168
10000	C_{pk}	(1.30, 1.35)	×	(1.40, 1.42)	(1.40, 1.42)	(1.40, 1.42)	✓	2.977

Tabelle 6.15.: Konfidenzintervalle Weibullverteilung - Beispiel 1

6.7. Simulierte Zufallsvariablen

Weibullverteilung(2,1) - $\mu = 100$, $\sigma = 5$ und $\gamma = 0.631$ theoretische Werte: $C_p = 1.46$, $C_{pk} = 1.10$, $C_{pu} = 1.10$, $C_{pl} = 2.26$, $S_{pmk} = 1.15$							
n = 50			Normal	Weibull	emp.	Clements	Mun.
$p = 0.605$	shape = 2.000	c_p	1.30	1.41	1.97	1.63	1.19
$\bar{x} = 99.94$	scale = 11.171	c_{pk}	1.29	1.07	1.76	1.46	
$s = 5.15$	$\hat{\gamma} = 0.631$	c_{pu}	1.30	1.07	1.76	1.46	
$\hat{\gamma}_e = 0.172$		c_{pl}	1.29	2.17	2.24	1.84	
n = 100			Normal	Weibull	emp.	Clements	Mun.
$p = 0.118$	shape = 2.162	c_p	1.39	1.54	1.83	1.53	1.11
$\bar{x} = 99.76$	scale = 11.030	c_{pk}	1.37	1.21	1.49	1.24	
$s = 4.80$	$\hat{\gamma} = 0.530$	c_{pu}	1.41	1.21	1.49	1.24	
$\hat{\gamma}_e = 0.524$		c_{pl}	1.37	2.20	2.42	2.05	
n = 1000			Normal	Weibull	emp.	Clements	Mun.
$p = 0.000$	shape = 2.033	c_p	1.30	1.44	1.50	1.41	1.02
$\bar{x} = 99.89$	scale = 11.172	c_{pk}	1.29	1.09	1.13	1.10	
$s = 5.12$	$\hat{\gamma} = 0.609$	c_{pu}	1.31	1.09	1.13	1.10	
$\hat{\gamma}_e = 0.625$		c_{pl}	1.29	2.17	2.30	2.02	
n = 10000			Normal	Weibull	emp.	Clements	Mun.
$p = 0.000$	shape = 2.115	c_p	1.33	1.47	1.42	1.36	1.02
$\bar{x} = 100.04$	scale = 11.353	c_{pk}	1.33	1.12	1.05	1.07	
$s = 5.02$	$\hat{\gamma} = 0.558$	c_{pu}	1.33	1.12	1.05	1.07	
$\hat{\gamma}_e = 0.672$	$S_{pmk} = 1.12$	c_{pl}	1.33	2.16	2.27	1.94	

Tabelle 6.16.: Weibullverteilung - Beispiel 2

Weibullverteilung(2,1) - $C_p = 1.46$, $C_{pk} = 1.10$								
n		Normal	Ü	Bootstrap	Perzentil	Bias-korr.	Ü	Q
50	C_p	(1.04, 1.55)	✓	(1.14, 1.74)	(1.17, 1.76)	(1.15, 1.73)	✓	0.879
50	C_{pk}	(0.96, 1.66)	✓	(0.78, 1.43)	(0.82, 1.46)	(0.81, 1.44)	✓	1.118
100	C_p	(1.20, 1.58)	✓	(1.33, 1.78)	(1.35, 1.80)	(1.34, 1.79)	✓	0.864
100	C_{pk}	(1.12, 1.64)	×	(0.97, 1.49)	(1.01, 1.51)	(0.99, 1.50)	✓	1.012
1000	C_p	(1.24, 1.36)	×	(1.37, 1.51)	(1.37, 1.51)	(1.37, 1.51)	✓	0.812
1000	C_{pk}	(1.22, 1.37)	×	(1.02, 1.17)	(1.02, 1.17)	(1.02, 1.17)	✓	1.020
10000	C_p	(1.31, 1.35)	×	(1.45, 1.49)	(1.45, 1.49)	(1.45, 1.49)	✓	0.855
10000	C_{pk}	(1.30, 1.35)	×	(1.10, 1.15)	(1.10, 1.15)	(1.10, 1.15)	✓	1.040

Tabelle 6.17.: Konfidenzintervalle Weibullverteilung - Beispiel 2

Weibullverteilung(1,1) - $\mu = 100$, $\sigma = 5$ und $\gamma = 2$							
theoretische Werte: $C_p = 1.21$, $C_{pk} = 0.73$, $C_{pu} = 0.73$, $C_{pl} = 5.34$, $S_{pmk} = 0.90$							
n = 50			Normal	Weibull	emp.	Clements	Mun.
$p = \mathbf{0.002}$	shape = 1.047	c_p	1.57	1.32	2.48	2.23	1.11
$\bar{x} = 99.91$	scale = 4.993	c_{pk}	1.56	0.80	1.72	1.54	
$s = 4.24$	$\hat{\gamma} = 1.867$	c_{pu}	1.58	0.80	1.72	1.54	
$\hat{\gamma}_e = 0.837$		c_{pl}	1.56	5.28	4.71	4.46	
n = 100			Normal	Weibull	emp.	Clements	Mun.
$p = \mathbf{0.000}$	shape = 1.023	c_p	1.36	1.31	1.69	1.49	0.85
$\bar{x} = 99.77$	scale = 4.818	c_{pk}	1.35	0.80	1.05	0.90	
$s = 4.89$	$\hat{\gamma} = 1.933$	c_{pu}	1.38	0.80	1.05	0.90	
$\hat{\gamma}_e = 1.727$		c_{pl}	1.35	5.47	6.12	7.40	
n = 1000			Normal	Weibull	emp.	Clements	Mun.
$p = \mathbf{0.000}$	shape = 0.999	c_p	1.23	1.13	1.19	1.11	0.75
$\bar{x} = 100.35$	scale = 5.347	c_{pk}	1.21	0.67	0.71	0.66	
$s = 5.40$	$\hat{\gamma} = 2.002$	c_{pu}	1.21	0.67	0.71	0.66	
$\hat{\gamma}_e = 2.006$	$S_{pmk} = 0.90$	c_{pl}	1.26	5.06	5.08	4.73	
n = 10000			Normal	Weibull	emp.	Clements	Mun.
$p = \mathbf{0.000}$	shape = 1.001	c_p	1.36	1.22	1.24	1.28	0.82
$\bar{x} = 99.96$	scale = 4.959	c_{pk}	1.36	0.74	0.75	0.77	
$s = 4.91$	$\hat{\gamma} = 1.997$	c_{pu}	1.36	0.74	0.75	0.77	
$\hat{\gamma}_e = 1.895$	$S_{pmk} = 0.92$	c_{pl}	1.36	5.37	5.34	5.70	

Tabelle 6.18.: Weibullverteilung - Beispiel 3

Weibullverteilung(1,1) - $C_p = 1.21$, $C_{pk} = 0.73$								
n		Normal	Ü	Bootstrap	Perzentil	Bias-korr.	Ü	Q
50	C_p	(1.26, 1.88)	×	(0.95, 1.79)	(1.01, 1.85)	(0.98, 1.73)	✓	0.828
50	C_{pk}	(1.18, 1.99)	×	(0.54, 1.14)	(0.59, 1.18)	(0.58, 1.09)	✓	1.571
100	C_p	(1.17, 1.55)	✓	(0.97, 1.71)	(1.00, 1.74)	(0.98, 1.70)	✓	0.532
100	C_{pk}	(1.10, 1.61)	×	(0.56, 1.08)	(0.59, 1.10)	(0.58, 1.07)	✓	1.026
1000	C_p	(1.18, 1.29)	✓	(1.03, 1.24)	(1.03, 1.25)	(1.03, 1.25)	✓	0.490
1000	C_{pk}	(1.14, 1.29)	×	(0.60, 0.75)	(0.61, 0.76)	(0.60, 0.75)	✓	0.986
10000	C_p	(1.34, 1.38)	×	(1.19, 1.26)	(1.18, 1.26)	(1.18, 1.26)	✓	0.487
10000	C_{pk}	(1.33, 1.38)	×	(0.71, 0.76)	(0.71, 0.76)	(0.71, 0.76)	✓	0.971

Tabelle 6.19.: Konfidenzintervalle Weibullverteilung - Beispiel 3

6.8. Datenbeispiel

In diesem Abschnitt wird die C_p -Wert Berechnung des Datensatzes aus Abschnitt 5 genauer betrachtet. Von 16 Versuchen wurden jeweils drei Wiederholungen ausgewertet. Jede Wiederholung hatte 324 Top- und 324 Bottom-Werte. Von den Top-Werten und von den Bottom-Werte wurde ein C_p -Wert berechnet. Von diesen zwei Werten wurde der Mittelwert gebildet und dieser als C_p -Wert der jeweiligen Wiederholung verwendet. Insgesamt wurden $3 \times 16 = 48$ C_p -Werte ermittelt. Die C_p -Werte wurden immer unter Annahme der Normalverteilung berechnet, obwohl diese nicht gegeben war. Wir wollen in diesem Abschnitt untersuchen, welche Abweichungen in der C_p -Wert Berechnung wegen dieser falschen Annahme entstanden sind.

Zu jedem Datensatz wird eine Verteilung gesucht, welche den Daten gut angepasst ist, wobei die Parameter dieser Verteilung mittels Maximum Likelihood Schätzung berechnet werden. Der C_p -Wert wird dann mit Hilfe der Quantile dieser Verteilung und über die empirischen Quantile berechnet. Beide Methoden werden mit der Berechnung, bei der man Normalverteilung angenommen hat, verglichen.

6.8.1. Fitten der Verteilung

Das Fitten der Verteilung wird mit Hilfe des Packages `fitdistrplus` durchgeführt. Wir betrachten dies jetzt anhand der 1. Wiederholung des vierten Versuchs genauer. In Tabelle 6.20 sieht man die Kenngrößen dieser zwei Stichproben. Da es bei den Top-Werten - wie wir später sehen werden - einen Ausreißer gibt, beinhaltet die zweite Spalte die Kenngrößen des Datensatzes ohne diesen Ausreißer. Aus der Tabelle erkennt man, dass der Shapiro-Wilk Test bei beiden Stichproben die Annahme, dass die Daten normalverteilt sind, verwirft.

	Top	Top (o. A.)	Bottom
\bar{x}	101.20	101.12	135.01
s	3.61	3.30	5.53
γ_e	0.297	-1.156	0.440
$kurt_e$	9.982	2.549	0.125
p	1.02e-14	1.18e-10	3.33e-03

Tabelle 6.20.: Wichtige Kenngrößen

Als erstes betrachten wir die Bottom Werte der 1. Wiederholung. Anhand von Tabelle 6.20 und des Histogramms (Abbildung 6.3) kann man erkennen, dass der Datensatz leicht rechtsschief ist. Mit der Funktion `descdist` kann man sich einen Skewness-Kurtosis Plot (Abbildung 6.4) ausgeben lassen. Dieser zeigt mit Hilfe der quadrierten Schiefe und der Kurtosis an, mit welcher Verteilung man den Datensatz am besten fitten kann. Die Gamma und Lognormalverteilung stellen in diesem Plot eine Linie dar. Die Weibullverteilung ist auf diesem Plot nicht eingezeichnet, dennoch können Beobachtungen, die sehr nahe bei der Gamma und Lognormalverteilungslinie liegen, meistens sehr gut mit der Weibullverteilung gefittet werden. Dies könnte

möglicherweise auf die Bottom-Werte zutreffen. Die gelben Punkte stellen Bootstrap Stichproben dar, die aus diesem Datensatz gezogen werden.

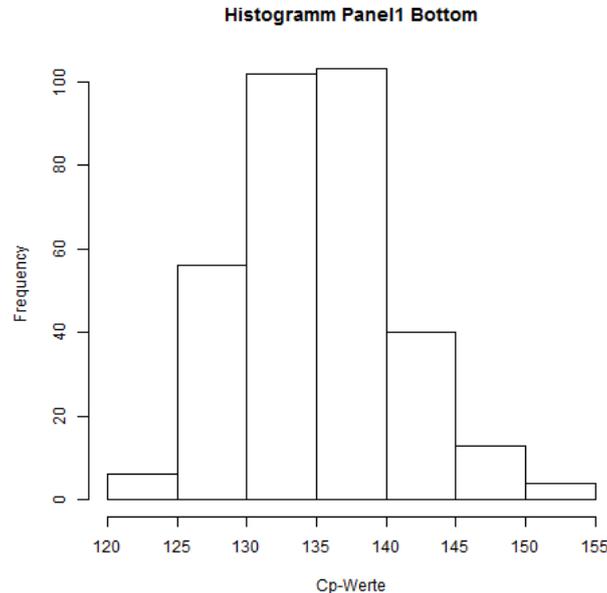


Abbildung 6.3.: Histogramm - Bottom

Um genauer zu untersuchen, welche Verteilung den Datensatz am besten fitten würde, vergleichen wir die Weibull-, Gamma-, Lognormal- und Gumbelverteilung anhand von vier Plots (`graphcomp`) und Goodness of Fit Tests (`gofstat`).

In Abbildung 6.5 ist die empirische Verteilungsfunktion gegen die theoretischen Verteilungsfunktionen der vier Verteilungen aufgezeichnet. In Abbildung 6.6 sieht man ein Histogramm der Daten mit den Dichtefunktionen der vier Verteilungen. Anhand dieser zwei Plots kann man sagen, dass die Weibullverteilung am besten zum Fitten der Daten geeignet ist. Weiters betrachten wir noch einen PP-Plot (Abbildung 6.7) und einen QQ-Plot (Abbildung 6.8) der vier Verteilungen. Auch hier sehen wir, dass die Weibullverteilung am besten zu dem Datensatz passt. Dies bestätigen auch die p -Werte der Goodness of Fit Tests in Tabelle 6.21. Der Cramer von Mises und der Anderson Darling Test verwerfen die Annahme, dass die Daten einer Weibullverteilung folgen, bei einem Signifikanzniveau von $\alpha = 0.05$ nicht.

	Weibull	Gamma	Lognormal	Gumbel
CvM	0.63	0.32	0.02	0.32
Ad	0.60	0.32	0.01	0.28

Tabelle 6.21.: Goodness of Fit Tests - Bottom

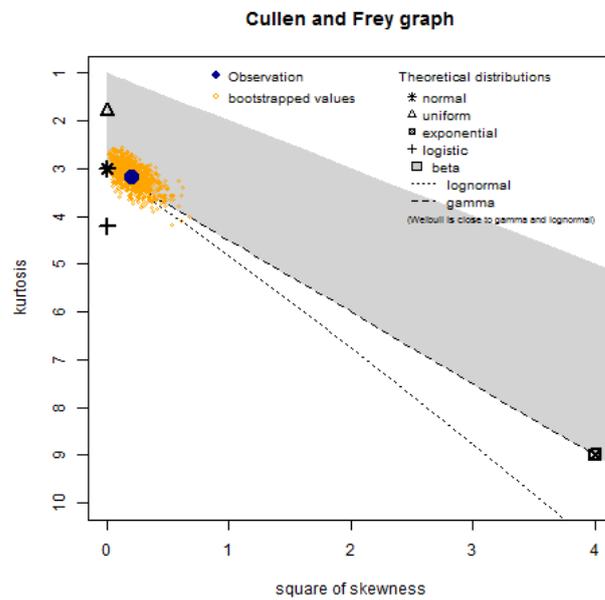


Abbildung 6.4.: Skewness-Kurtosis Plot - Bottom

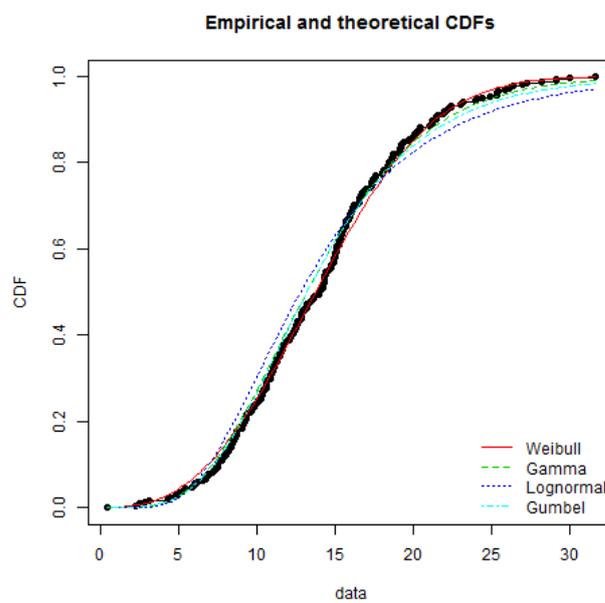


Abbildung 6.5.: Verteilungsfunktion - Bottom

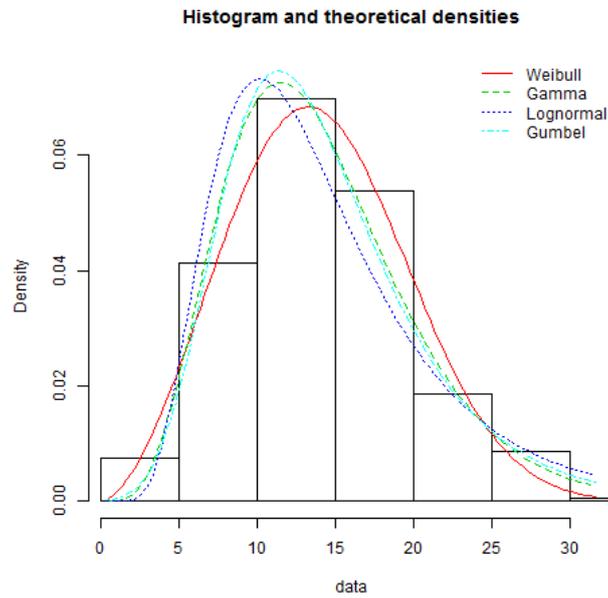


Abbildung 6.6.: Dichtefunktion - Bottom

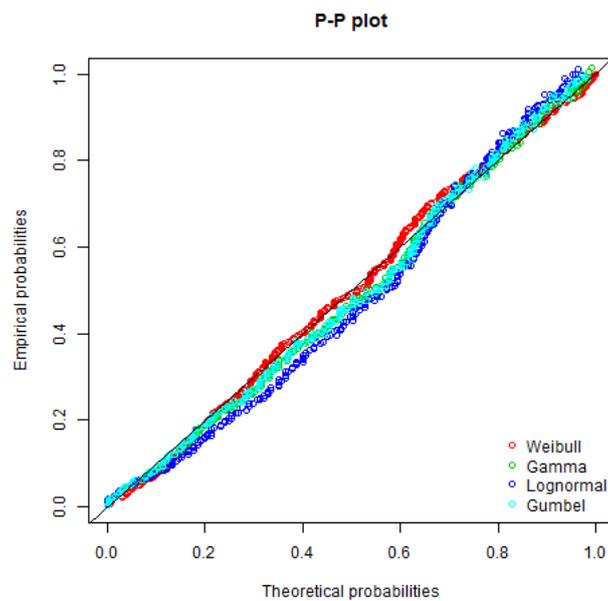


Abbildung 6.7.: PP-Plot - Bottom

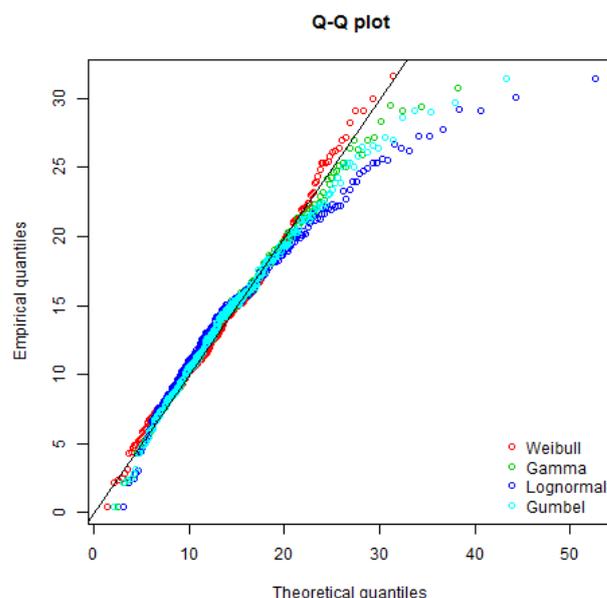


Abbildung 6.8.: QQ-Plot - Bottom

In Abbildung 6.9 kann man alle vier Plots der gefitteten Weibullverteilung sehen. Mit Hilfe der Quantile dieser gefitteten Verteilung wird der C_p -Wert in Tabelle 6.22 in der Spalte Weibull berechnet. Weiters sind die C_p -Werte, die unter Annahme der Normalverteilung und mit den empirischen Quantilen berechnet werden, in der Tabelle angegeben. Man sieht, dass der C_p -Wert, der unter Annahme der Normalverteilung berechnet wird, den C_p -Wert unterschätzt.

	Normal	Weibull	empirisch
c_p	1.21	1.32	1.35

Tabelle 6.22.: Vergleich der Prozessfähigkeitsindizes - Bottom

Nun betrachten wir die Top Werte dieser Wiederholung. Im Histogramm 6.10 können wir erkennen, dass ein Ausreißer, der mehr als 6 Standardabweichungen vom Mittelwert entfernt ist, im Datensatz vorhanden ist. Dieser Ausreißer wird aus dem Datensatz entfernt.

Man kann im Histogramm (Abbildung 6.11) und in Tabelle 6.20 erkennen, dass der Datensatz ohne den Ausreißer linksschief ist. Um linksschiefe Verteilungen zu fitten, werden folgende Möglichkeiten in Betracht gezogen. Man könnte die Daten $m - X$ betrachten, wobei m ein vorher festgelegter Maximalwert ist, welchen die Daten nicht überschreiten können. Diese Daten sind rechtsschief. In manchen Fällen - wie auch in diesem - ist es allerdings nicht so einfach eine solche Obergrenze m festzulegen. Möglich ist dies zum Beispiel bei einem Datensatz, der das Alter von

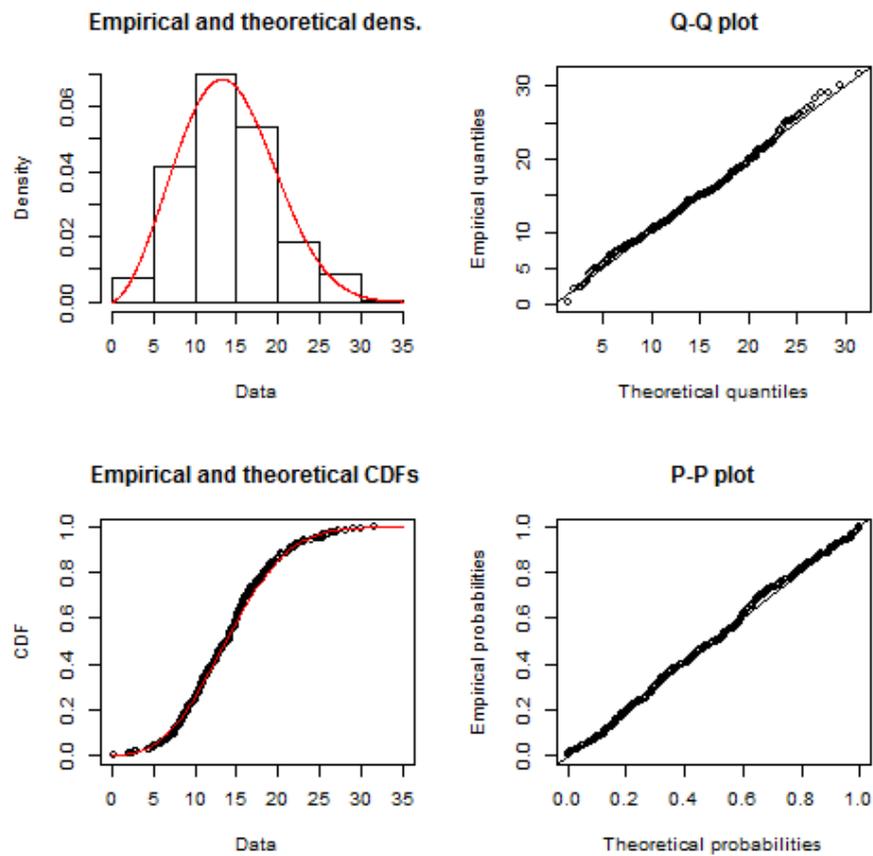


Abbildung 6.9.: Plot der gefitteten Weibullverteilung

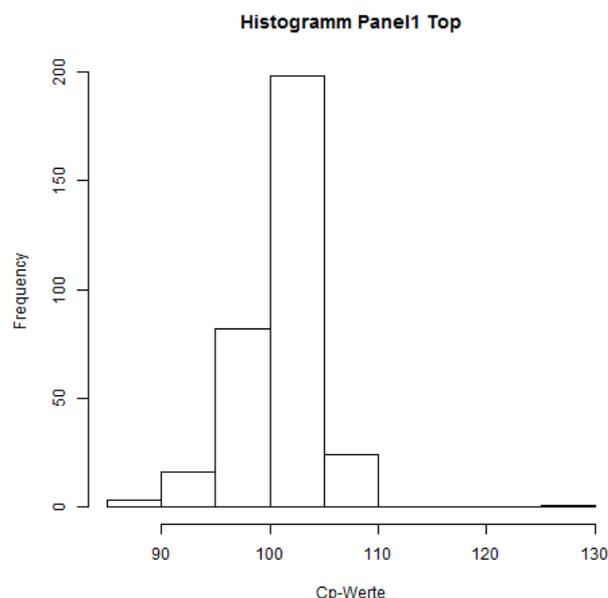


Abbildung 6.10.: Histogramm - Top

Personen beschreibt.

Eine weitere Möglichkeit eine linksschiefe Verteilung zu fitten ist den Datensatz zu quadrieren und X^2 mit der Normalverteilung zu fitten. Außerdem wollen wir versuchen X mit Hilfe der Gompertzverteilung und der Weibullverteilung zu fitten.

Da die erste Möglichkeit, wie bereits erwähnt, bei diesem Datensatz nicht anwendbar ist, vergleichen wir im Folgenden die anderen zwei Methoden. In Abbildung 6.13 sehen wir, dass die Normalverteilung die Daten X^2 sehr gut fitten würde. Die Goodness of Fit Tests in Tabelle 6.23 bestätigen dies. Der Cramer von Mises und der Anderson Darling Test verwerfen die Annahme, dass X^2 normalverteilt ist, bei einem Signifikanzniveau von $\alpha = 0.05$ nicht.

	Weibull	Gompertz	Normal (X^2)
CvM	0.08	0.35	0.83
Ad	0.05	0.38	0.82

Tabelle 6.23.: Goodness of Fit Tests - Top

Im Weiteren betrachten wir die vier Plots (Abbildung 6.14 - 6.17), in welchen die Weibullverteilung mit der Gompertzverteilung verglichen wird. Wir können beobachten, dass die Gompertzverteilung für diesen Datensatz besser passt. Die Goodness of Fit Tests in Tabelle 6.23 verwerfen die Annahme der Gompertzverteilung bei einem Signifikanzniveau von $\alpha = 0.05$ nicht. Bei diesem Datensatz fällt die Entscheidung also auf die Gompertzverteilung (Abbildung 6.18).

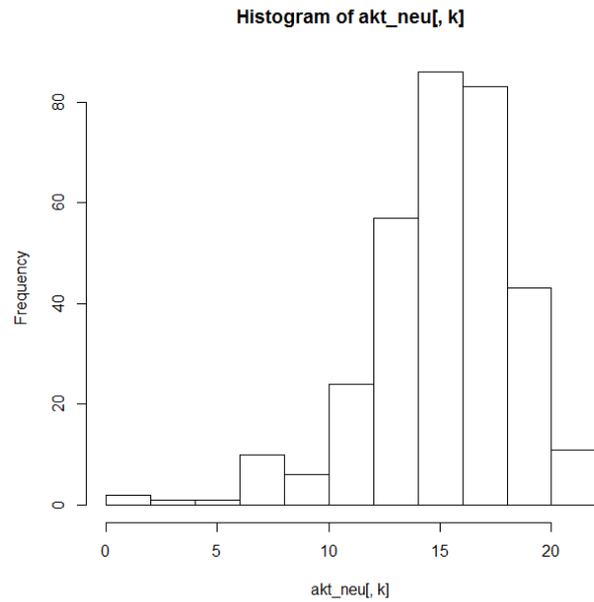


Abbildung 6.11.: Histogramm - Top (o.A.)

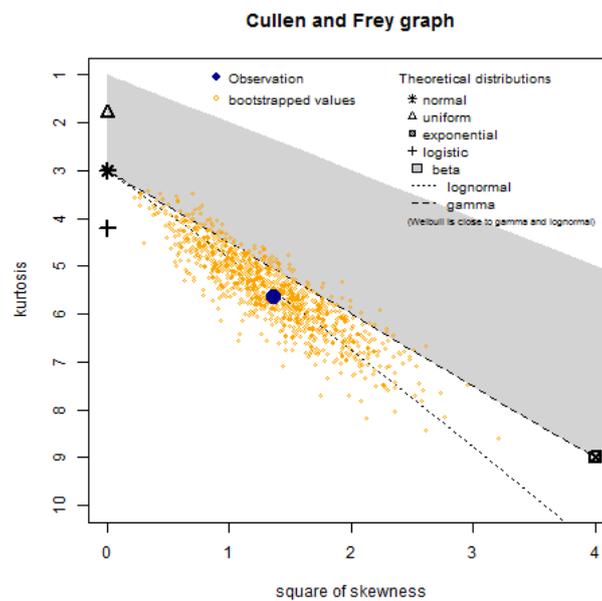


Abbildung 6.12.: Skewness-Kurtosis Plot - Top

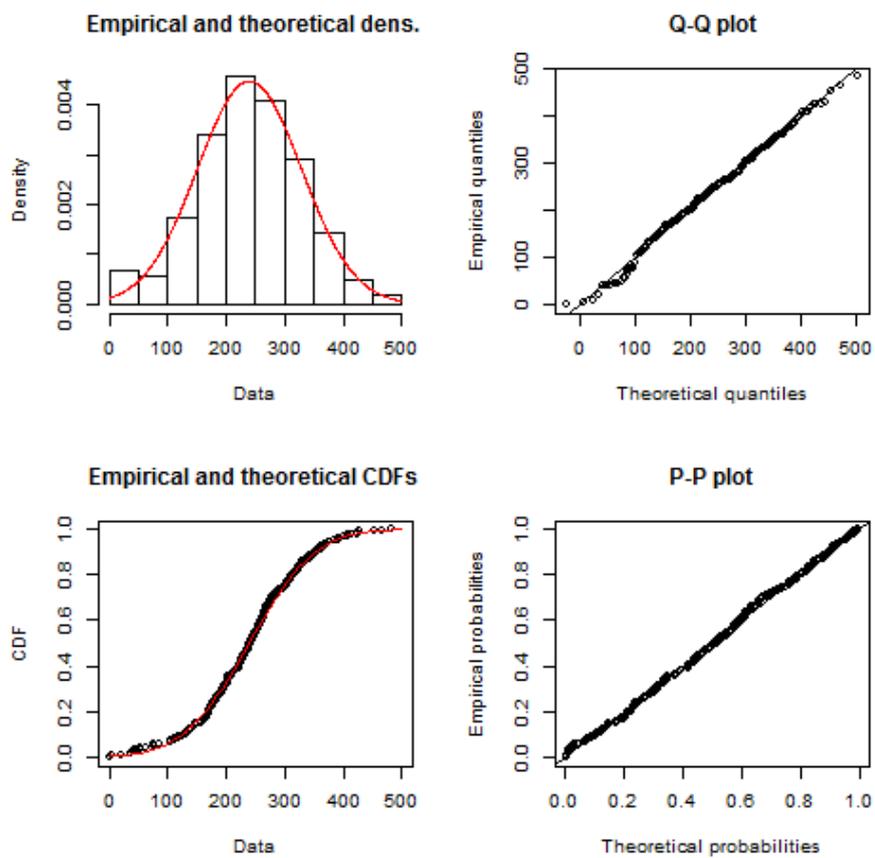


Abbildung 6.13.: Plot der gefitteten Normalverteilung für X^2

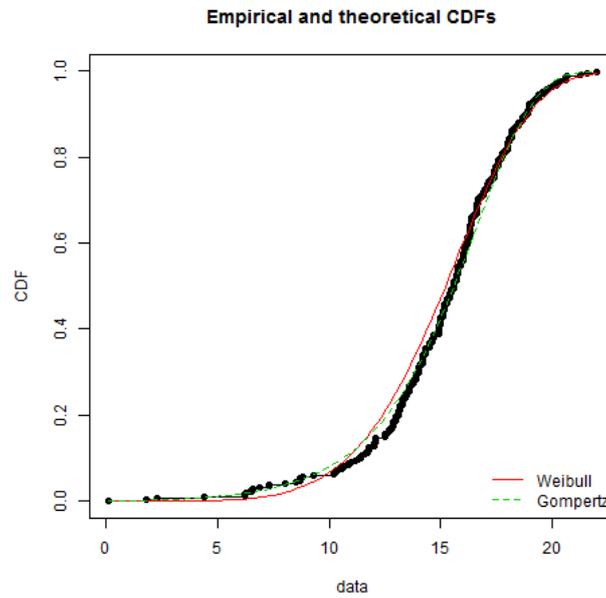


Abbildung 6.14.: Verteilungsfunktion - Top

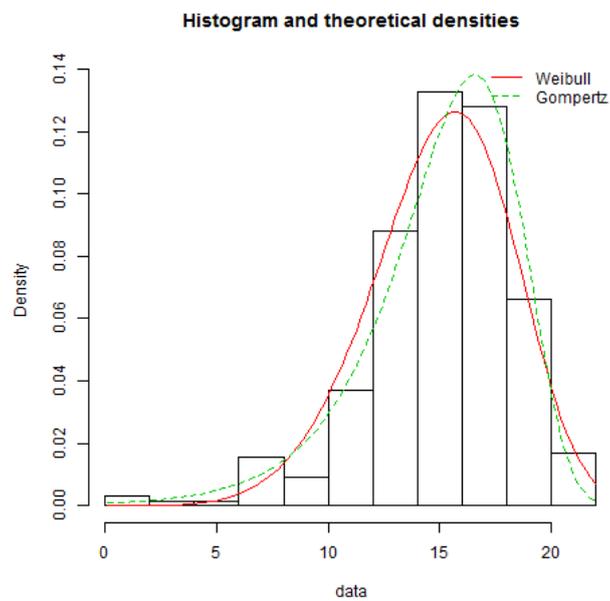


Abbildung 6.15.: Dichtefunktion - Top

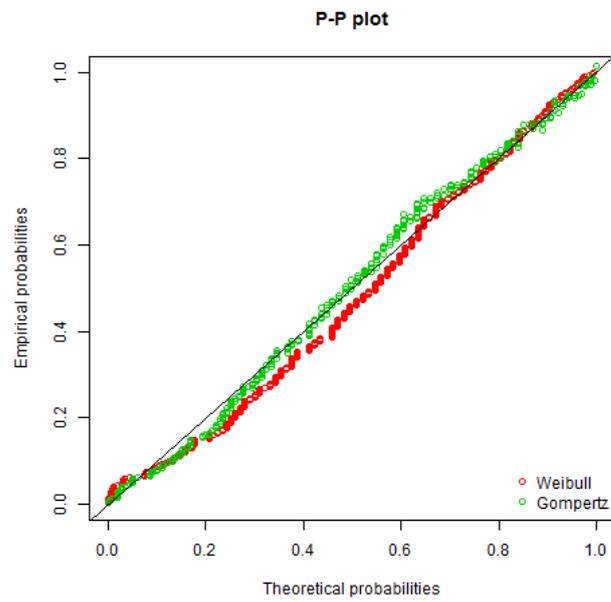


Abbildung 6.16.: PP-Plot - Top

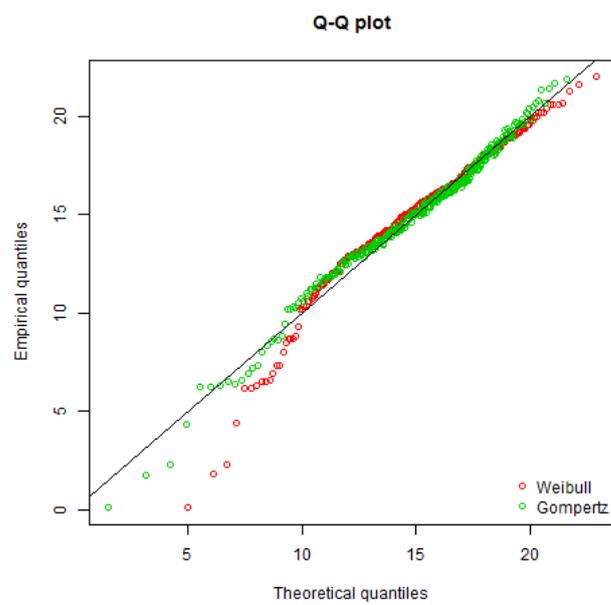


Abbildung 6.17.: QQ-Plot - Top

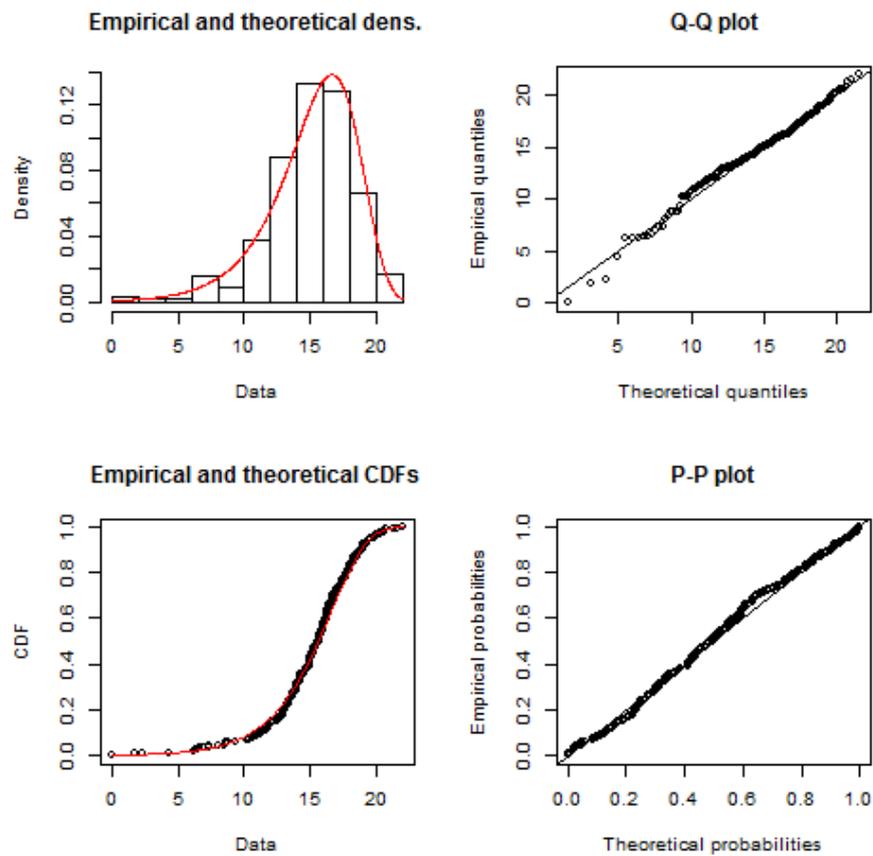


Abbildung 6.18.: Plot der gefitteten Gompertzverteilung

Das heißt, dass wir bei diesem Datensatz die Möglichkeit haben, die Daten X^2 mit der Normalverteilung oder X mit der Gompertzverteilung zu fitten. Die Entscheidung fiel auf die Gompertzverteilung, da man bei der anderen Methode den C_p -Wert von X^2 berechnet und dieser dann nicht direkt mit den anderen C_p -Werten vergleichbar ist. In Tabelle 6.24 sieht man den Vergleich der drei C_p -Werte. Wir erkennen, dass der C_p -Wert, der unter Normalverteilung berechnet wird, den C_p -Wert in diesem Fall überschätzt.

	Normal	Gompertz	emp
c_p	2.02	1.98	1.91

Tabelle 6.24.: Vergleich der Prozessfähigkeitsindizes - Top

6.8.2. Ergebnisse

Wir betrachten das Gesamtergebnis der 1. Wiederholung des 4. Versuchs. Die Werte in der Tabelle wurden auf zwei Kommastellen gerundet. In der Spalte Unterschied ist jeweils der Unterschied zwischen der sich in der davorliegenden Spalte befindlichen Berechnung zur Berechnung unter Annahme der Normalverteilung angegeben. Man sieht, dass, obwohl in diesem Fall die individuellen Unterschiede groß sind, sich diese nach der Mittelwertbildung verringern. Bei positiven Vorzeichen überschätzt der C_p -Wert, berechnet unter der Normalverteilung den C_p -Wert, berechnet aus der gefitteten Verteilung. Umgekehrt gilt für negatives Vorzeichen, dass der Normalverteilungswert zu einer Unterschätzung führt.

Versuch4	Normal	alt. Vert.	Unterschied	emp	Unterschied
Wd. 1 T	2.02	1.98	0.05	1.91	0.12
Wd. 1 B	1.21	1.32	-0.11	1.35	-0.14
MW	1.61	1.65	-0.03	1.63	-0.01

Tabelle 6.25.: Vergleich der C_p -Werte für Versuch 4 Wd. 1

Auf diese Weise wurden alle weiteren Wiederholungen der 16 Versuche untersucht. Für jeden Datensatz konnte eine passende Verteilung gewählt werden. Mit dieser wurde der C_p -Wert berechnet. Es ergaben sich teilweise beträchtliche Unterschiede zum Normalverteilungsansatz. Dabei fiel auch hier auf, dass die Unterschiede bei größerer Schiefe größer wurden. Nur in einigen wenigen Fällen wurden diese, wie im vorigen Beispiel, durch die Mittelwertbildung verringert.

Zusammenfassend kann man also feststellen, dass die C_p -Wert Berechnung unter der Normalverteilungsannahme großteils zu verfälschten Ergebnissen (Über- und Unterschätzung) führten.

A. Anhang

A.1. Kanonische Reduktion von Matrizen

Für diesen Abschnitt vergleiche Rao (1973).

Satz A.1.1. *Seien \mathbf{B} und \mathbf{C} reelle, symmetrische $k \times k$ Matrizen. Die Matrix \mathbf{C} ist außerdem positiv definit. Dann gibt es eine reguläre Matrix \mathbf{S} , sodass $\mathbf{B} = \mathbf{S}^{-1'} \mathbf{\Lambda} \mathbf{S}^{-1}$ und $\mathbf{C} = \mathbf{S}^{-1'} \mathbf{S}^{-1}$, wobei $\mathbf{\Lambda}$ eine Diagonalmatrix ist.*

Beweis: Seien $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$ die Lösungen von $|\mathbf{B} - \lambda \mathbf{C}| = 0$. Es folgt, dass das i -te Diagonalelement von $\mathbf{\Lambda}$ die Lösung λ_i ist und der i -te Spaltenvektor von \mathbf{S} die Gleichung $\mathbf{B} \mathbf{S}_i = \lambda \mathbf{C} \mathbf{S}_i$ erfüllt.

Eine positiv definite Matrix lässt sich durch eine invertierbare Matrix und ihrer Transponierten darstellen (vgl. zum Beispiel Rao, 1973). Es gilt daher für die Matrix $\mathbf{C} = \mathbf{T}' \mathbf{T}$. Daraus folgt

$$0 = |\mathbf{B} - \lambda \mathbf{C}| = |\mathbf{B} - \lambda \mathbf{T}' \mathbf{T}| \Leftrightarrow |\mathbf{T}^{-1'} \mathbf{B} \mathbf{T}^{-1} - \lambda \mathbf{I}| = 0.$$

Man setzt $\mathbf{A} := \mathbf{T}^{-1'} \mathbf{B} \mathbf{T}^{-1}$. Auch \mathbf{A} ist eine symmetrische Matrix und daher gilt wegen Satz A.1.2 für eine orthogonale Matrix \mathbf{P} , dass

$$\mathbf{P}' \mathbf{A} \mathbf{P} = \mathbf{\Lambda}.$$

Aus der Definition von \mathbf{A} erhält man

$$\mathbf{P}' \mathbf{T}^{-1'} \mathbf{B} \mathbf{T}^{-1} \mathbf{P} = \mathbf{\Lambda}.$$

Setzt man $\mathbf{S} = \mathbf{T}^{-1} \mathbf{P}$ kommt man auf

$$\mathbf{S}' \mathbf{B} \mathbf{S} = \mathbf{\Lambda}$$

und

$$\begin{aligned} \mathbf{S}' \mathbf{C} \mathbf{S} &= \mathbf{P}' \mathbf{T}^{-1'} \mathbf{C} \mathbf{T} \mathbf{P} \\ &= \mathbf{P}' \underbrace{\mathbf{T}^{-1'} \mathbf{T}'}_{\mathbf{I}} \underbrace{\mathbf{T} \mathbf{T}^{-1}}_{\mathbf{I}} \mathbf{P} \\ &= \mathbf{P}' \mathbf{P} \\ &\stackrel{\mathbf{P} \text{ orth.}}{=} \mathbf{I}. \end{aligned}$$

Satz A.1.2. Sei \mathbf{A} eine reelle, symmetrische $k \times k$ Matrix. Dann gibt es eine orthogonale Matrix \mathbf{P} und eine Diagonalmatrix $\mathbf{\Lambda}$, für welche

$$\mathbf{P}'\mathbf{A}\mathbf{P} = \mathbf{\Lambda} \quad \text{oder} \quad \mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}'$$

gilt.

Beweis: Siehe Rao (1973).

A.2. Die Varianz von λ^0

Die folgenden Überlegungen wurden aufbauend auf Khuri & Conlon (1981) verfasst. Sei y_{ij} der exakte i -te Response-Wert des j -ten Versuchs und y_{ij}^0 der gerundete Wert von y_{ij} . Die $n \times l$ Matrix \mathbf{Y}^0 kann durch die Gleichung

$$\mathbf{Y}^0 = \mathbf{Y} + \mathbf{Z} \tag{A.1}$$

angegeben werden, wobei \mathbf{Z} eine $n \times l$ Matrix ist, welche die Rundungsfehler enthält.

Aus der Annahme, dass die Rundungsfehler gleichverteilt $U(-\delta, \delta)$ sind, erhalten wir

$$E(z_{ij}) = \frac{\delta - (-\delta)}{2} = 0$$

und

$$\text{Var}(z_{ij}) = \frac{1}{2}(\delta + \delta)^2 = \frac{\delta^2}{3}.$$

Wenn man Gleichung (A.1) in Gleichung (3.11) einsetzt, ergibt es

$$\mathbf{D} = \underbrace{\mathbf{Y}^{0'}(\mathbf{I}_n - \mathbf{1}_n\mathbf{1}'_n/n)}_{=\mathbf{D}^0} - \mathbf{Z}'(\mathbf{I}_n - \mathbf{1}_n\mathbf{1}'_n/n).$$

Im Weiteren erhalten wir

$$\mathbf{D}^0\mathbf{D}^{0'} = \mathbf{D}\mathbf{D}' + \mathbf{L},$$

wobei

$$\mathbf{L} = \mathbf{D}(\mathbf{I}_n - \mathbf{1}_n\mathbf{1}'_n/n)\mathbf{Z} + \mathbf{Z}'(\mathbf{I}_n - \mathbf{1}_n\mathbf{1}'_n/n)\mathbf{D}' + \underbrace{\mathbf{Z}'(\mathbf{I}_n - \mathbf{1}_n\mathbf{1}'_n/n)\mathbf{Z}}_{=\mathbf{S}}.$$

Sei $\lambda = 0$ der Eigenwert von $\mathbf{D}\mathbf{D}'$, dann gilt, dass für den dazugehörigen Eigenwert λ^0 von $\mathbf{D}^0\mathbf{D}^{0'}$ und \mathbf{a} mit Einheitslänge

$$\lambda^0 \sim \mathbf{a}'\mathbf{L}\mathbf{a} \tag{A.2}$$

ist, wenn $\delta \rightarrow 0$ (Wilkinson, 1969, S. 172ff.). Das heißt, dass λ^0 von der gleichen Größenordnung wie der Term auf der rechten Seite ist. Wir wissen, dass $\mathbf{a}'\mathbf{D}\mathbf{D}' = \mathbf{0}'$ und folglich auch $\mathbf{a}'\mathbf{D} = \mathbf{0}'$. Daraus erhalten wir für Gleichung (A.2), dass

$$\lambda^0 \sim \mathbf{a}'\mathbf{S}\mathbf{a}, \tag{A.3}$$

da die ersten zwei Terme der Matrix \mathbf{L} , wenn man \mathbf{a}' linksseitig multipliziert, jeweils eine Nullmatrix ergeben.

Für den quadrierten Erwartungswert folgt

$$E(\lambda^0)^2 \sim E(\mathbf{a}'\mathbf{S}\mathbf{a})^2$$

für $\delta \rightarrow 0$. Weiters erhalten wir für δ klein genug

$$\text{Var}(\lambda^0) \sim E(\mathbf{a}'\mathbf{S}\mathbf{a})^2 - (n-1)^2\sigma_{rf}^4. \quad (\text{A.4})$$

Daraus ergibt sich mit Hilfe des Verschiebungssatzes die Gleichung (3.12).

Da die Matrix $\mathbf{Z}'\mathbf{Z} - \mathbf{S}$ positiv semidefinit ist, wissen wir, dass

$$E(\mathbf{a}'\mathbf{S}\mathbf{a})^2 \leq E(\mathbf{a}'\mathbf{Z}'\mathbf{Z}\mathbf{a})^2$$

ist. Weiters gilt, dass $\mathbf{a}'\mathbf{Z}'\mathbf{Z}\mathbf{a} \leq \text{Spur}(\mathbf{Z}'\mathbf{Z})$ (Lancaster, 1969, S. 109). Mit der Gleichung (A.4) erhalten wir für die Varianz des Eigenwerts approximativ

$$\text{Var}(\lambda^0) \leq E\left(\sum_{i=1}^l \sum_{j=1}^n z_{ij}^2\right)^2 - (n-1)\sigma_{rf}^4.$$

Um die obere Grenze zu berechnen, schreiben wir

$$E\left(\sum_{i=1}^l \sum_{j=1}^n z_{ij}^2\right)^2 = \sum_{i=1}^l \sum_{j=1}^n E(z_{ij}^4) + 2 \sum_I E(z_{ij}^2)E(z_{km}^2),$$

wobei $I = I_1 \cup I_2$ mit

$$I_1 = \{(i, j, k, m) : 1 \leq i < k \leq l\}$$

$$I_2 = \{(i, j, k, m) : 1 \leq j < m \leq n\}.$$

Aus dem Verschiebungssatz erhalten wir

$$E(z_{ij}^2) = \text{Var}(z_{ij}) + E(z_{ij})^2 = \sigma_{rf}^2$$

für $i = 1, \dots, l, j = 1, \dots, n$. Mit

$$E(z_{ij}^4) = \int_{-\delta}^{\delta} x^4 \frac{1}{\delta + \delta} dx = \frac{\delta^4}{5} = \frac{9\sigma_{rf}^4}{5}$$

für $i = 1, \dots, l, j = 1, \dots, n$ bekommen wir

$$E\left(\sum_{i=1}^l \sum_{j=1}^n z_{ij}^2\right)^2 = \frac{9nl}{5}\sigma_{rf}^4 + nl(nl-1)\sigma_{rf}^4.$$

Zusammen ergibt das für δ ausreichend klein, dass

$$\text{Var}(\lambda^0) \leq \underbrace{[9nl/5 + nl(nl-1) - (n-1)^2]}_q \sigma_{rf}^4.$$

A.3. Entfernen von linear abhängigen Response-Variablen

Khuri & Conlon (1981) schlagen zum Entfernen der Response-Variablen, die lineare Funktionen einer anderen Response-Variable sind, folgende Methode vor. Angenommen $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$ sind die m zu dem Eigenwert 0 von $\mathbf{D}\mathbf{D}'$ zugehörigen linear unabhängigen Eigenvektoren. Jedes \mathbf{a}_i ist ein l -dimensionaler Vektor. Sei $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m]$ eine $m \times l$ Matrix und \mathbf{D} die $l \times n$ Matrix aus (3.11), dann gilt $\mathbf{D}\mathbf{D}'\mathbf{A}' = \mathbf{0}$ und folglich $\mathbf{A}\mathbf{D}\mathbf{D}'\mathbf{A}' = \mathbf{0}$. Dies können wir auch schreiben als $\mathbf{A}\mathbf{D}(\mathbf{A}\mathbf{D})' = \mathbf{0}$, daher muss gelten, dass $\mathbf{A}\mathbf{D} = \mathbf{0}$. Das heißt, dass

$$\mathbf{A}[\mathbf{y}_1 - \bar{y}_1 \mathbf{1}_n, \dots, \mathbf{y}_l - \bar{y}_l \mathbf{1}_n]' = \mathbf{0}. \quad (\text{A.5})$$

Da auch hier Rundungsfehler auftreten, schreiben wir die Gleichung folgendermaßen an

$$\mathbf{A}[\mathbf{y}_1 - \bar{y}_1 \mathbf{1}_n, \dots, \mathbf{y}_l - \bar{y}_l \mathbf{1}_n]' = \mathbf{E}, \quad (\text{A.6})$$

wobei \mathbf{E} eine $m \times n$ Matrix ist, deren Elemente genau 0 wären, wenn keine Rundungsfehler auftreten. O.B.d.A. kann man die Matrix \mathbf{A} partitionieren und man erhält $\mathbf{A} = [\mathbf{A}_1 : \mathbf{A}_2]$, wobei \mathbf{A}_1 eine $m \times m$ Matrix mit vollem Rang ist. Wir partitionieren \mathbf{D} auf die gleiche Weise und erhalten wegen (A.6)

$$\mathbf{D}_1 = -\mathbf{A}_1^{-1} \mathbf{A}_2 \mathbf{D}_2 + \mathbf{A}_1^{-1} \mathbf{E}.$$

Sei $\|\mathbf{A}\|$ die euklidische Norm der Matrix \mathbf{A} , d.h. $\|\mathbf{A}\| = \sqrt{\text{Spur}(\mathbf{A}\mathbf{A}')}$. Um m Response-Variablen entfernen zu können, wählt man jetzt \mathbf{A}_1 so, dass $\|\mathbf{A}_1^{-1}\|$ das Minimum in der Klasse aller regulären $m \times m$ Untermatrizen von \mathbf{A} ist. Durch diese Wahl von \mathbf{A}_1 wird der Effekt des Rundungsfehlers minimiert. Dieser hat fälschlicherweise dazu geführt, dass auch Response-Variablen die linear abhängig waren, die Gleichung (A.5) nicht genau erfüllten.

A.4. Simplex-Suchverfahren

Das Simplex-Suchverfahren ist ein direktes Suchverfahren, da es zur Optimierung des gegebenen Problems keine Information der Ableitung benötigt. Wir betrachten in diesem Abschnitt wie He et al. (2010) dieses Verfahren in deren robusten Optimierungsmethode verwenden. Weitere Details zum Simplexverfahren findet der interessierte Leser bei Spendley et al. (1962).

Das Simplexverfahren wird mit einem zulässigen Punkt, einem Punkt im Bereich, in welchem das Experiment durchgeführt wird, gestartet. In unserem Fall ist dieser Anfangspunkt der Punkt $\mathbf{V}_0 = (x_{1o}, x_{2o}, \dots, x_{ko})'$, den wir durch die Maximierung der Optimierungs-Desirability-Funktion erhalten haben. Aus diesem Punkt konstruiert man sich ein reguläres Simplex. Sei jetzt $c \in \mathbb{R}^+$ die Seitenlänge des Simplex, für welche üblicherweise $0 < c < 1$ gilt, dann ist das Startsimplex S_0 mit Seitenlänge c durch folgende Design-Matrix gegeben.

$$\mathbf{S}_0 = \begin{bmatrix} x_{1o} & x_{2o} & x_{3o} & \cdots & x_{ko} \\ x_{1o} + c\alpha & x_{2o} + c\beta & x_{3o} + c\beta & \cdots & x_{ko}c\beta \\ x_{1o} + c\beta & x_{2o} + c\alpha & x_{3o} + c\beta & \cdots & x_{ko}c\beta \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1o} + c\beta & x_{2o} + c\beta & x_{3o} + c\beta & \cdots & x_{ko}c\alpha \end{bmatrix},$$

wobei

$$\alpha = \frac{1}{k\sqrt{2}}((k-1) + \sqrt{k+1})$$

und

$$\beta = \frac{1}{k\sqrt{2}}(\sqrt{k+1} - 1).$$

Man sieht hier, dass die erste Zeile der Matrix die k Koordinaten des Startpunktes \mathbf{V}_0 enthält. Im Weiteren enthält die $(i+1)$ -te Zeile der Matrix die Koordinaten des Punktes \mathbf{V}_i . Die Punkte $\mathbf{V}_0, \mathbf{V}_1, \dots, \mathbf{V}_k$ ergeben das reguläre Startsimplex S_0 . Nun wird für jeden Eckpunkt der Wert von D_{gesamt} ausgerechnet und anschließend alle Werte miteinander verglichen. Der Punkt \mathbf{V}_i , der den minimalen Wert von D_{gesamt} hat, wird entfernt und man rechnet mit dem benachbarten Simplex weiter. Das benachbarte Simplex erhält man, indem man anstelle des entfernten Punktes, den Spiegelpunkt \mathbf{V}_{k+1} dieses entfernten Punktes in der Hyperebene der verbleibenden Punkte, hinzufügt. Das heißt, das neue Simplex S_1 enthält k Eckpunkte des alten Simplex, nämlich $\mathbf{V}_0, \mathbf{V}_1, \dots, \mathbf{V}_{i-1}, \mathbf{V}_{i+1}, \dots, \mathbf{V}_k$ und den neuen Punkt \mathbf{V}_{k+1} , der durch folgende Gleichung

$$\mathbf{V}_{k+1} = \frac{2}{k}(\mathbf{V}_0 + \mathbf{V}_1 + \cdots + \mathbf{V}_{i-1} + \mathbf{V}_{i+1} + \cdots + \mathbf{V}_k) - \mathbf{V}_i$$

gegeben ist.

Auch in diesem Simplex sucht man den Punkt mit dem minimalen D_{gesamt} um diesen durch einen neuen Punkt zu ersetzen. Diese Schritte werden jetzt immer wieder durchgeführt bis das Abbruchkriterium erfüllt ist. Während der Anwendung des Simplexverfahrens, werden folgende Regeln befolgt.

1. Ist ein Eckpunkt, des Simplex außerhalb des zulässigen Bereichs, wird die Seitenlänge c halbiert.
2. Wenn das Simplex S_i wieder zum Simplex S_{i-1} übergehen würde, konstruiert man sich S_{i+1} , indem man die Mittelpunkte der Kanten, die zum Eckpunkt mit dem größten Wert von D_{gesamt} in S_{i-1} und damit auch in S_i führen, in Simplex S_{i-1} verwendet. Die Seitenlänge wird in diesem Fall halbiert.
3. Sollte ein neuer Punkt \mathbf{V}_{k+i} außerhalb des zulässigen Bereichs sein, wird die Seitenlänge halbiert.
4. Wenn die Seitenlänge c halbiert wird, wird der Wert von r^2 in (4.11) auch halbiert.

Durch Anwenden dieser Regeln wird das Simplex immer kleiner. Den robusten optimalen Punkt erhalten wir, wenn das Simplex klein genug ist, z.B. wenn $c < 0.01$. Der robuste optimale Punkt \mathbf{x}^* ist der Eckpunkt des letzten Simplex, der den maximalen Wert von D_{gesamt} hat.

Bemerkungen

- Je kleiner der zulässige Bereich ist, desto kleiner ist die Seitenlänge c , mit welcher wir beginnen.
- Die Wahl von c und r beeinflusst die robuste optimale Lösung kaum.
- Wenn es mehrere getrennte zulässige Bereiche gibt, muss man das Simplexverfahren für jeden Bereich durchführen. Man erhält mehrere robuste optimale Punkte, und kann daraus den Punkt mit dem maximalen D_{gesamt} -Wert - den robusten optimalen Punkt - berechnen.
- Laut He et al. (2010) kann mit diesem Algorithmus die Komplexität und die Rechenzeit in ihrer Methode verringert werden.

A.5. R-Routinen

In diesem Abschnitt werden einige R-Routinen und R-Funktionen angeführt, die in dieser Arbeit verwendet wurden. Es werden nur Details der Routinen erklärt, die tatsächlich benutzt wurden. Weitere Informationen zu den R-Routinen findet man online unter <http://cran.r-project.org/>.

A.5.1. Kapitel 5

Robustes E-Modell

Erwartungswert der Zielfunktion:

```
erwartungswert_funktion<-function(x){  
  erw_w[1,1] * abs(y_dach_1(x) - t_vec[1]) + erw_w[2,1] * abs(y_dach_2(x) -  
  t_vec[2])  
}
```

Varianz der Zielfunktion:

```
varianz_funktion<-function(x){  
  (erw_w[1,1])^2 * vari_y_dach(x)[1,1] +  
  (erw_w[2,1])^2 * vari_y_dach(x)[2,2] + 2 * erw_w[1,1] * erw_w[2,1] *  
  vari_y_dach(x)[1,2]  
}
```

Funktion, die minimiert wird:

```
optimierungsf<-function(x){  
  erwartungswert_funktion(x) + qnorm(1-alpha)*sqrt(varianz_funktion(x))  
}
```

Robuste Optimierung mit einem Konfidenzbereich

Optimierungs-Desirability-Funktion für den C_p -Wert:

```
docp<-function(l, t, x){
  if (y_dach_1(x)>t)
  {1}
  else if(y_dach_1(x)<l)
  {0}
  else (y_dach_1(x)-l)/(t- l)
}
```

Optimierungs-Desirability-Funktion für den Ätzfaktor:

```
doef<-function(l, t, x){
  if (y_dach_2(x)>t)
  {1}
  else if(y_dach_2(x)<l)
  {0}
  else (y_dach_2(x)-l)/(t- l)
}
```

Optimierungs-Desirability-Funktion:

```
dopt<-function(x){
  out<-(docp(l[1],target[1],x))^(w1)*(doef(l[2],target[2],x))^(w2)
  if(any(abs(x) > 1.0))
  {
    out <- 0
  }
  out
}
```

Konfidenzintervall für die gefittete Response-Funktion des C_p -Wertes:

```
ki_cp<-function(x){
  ki_c<-vector()
  ki_c[1]<-y_dach_1(x)- qt(1-alpha_2/2,48-16)* sqrt(sigma_dach[1,1]* multi(x))
  ki_c[2]<-y_dach_1(x)+ qt(1-alpha_2/2,48-16)* sqrt(sigma_dach[1,1] * multi(x))
  return(ki_c)
}
```

Konfidenzintervall für die gefittete Response-Funktion des Ätzfaktors:

```
ki_ef<-function(x){
  ki_e<-vector()
  ki_e[1]<-y_dach_2(x)- qt(1-alpha_2/2,48-16)* sqrt(sigma_dach[2,2]* multi(x))
  ki_e[2]<-y_dach_2(x)+ qt(1-alpha_2/2,48-16)* sqrt(sigma_dach[2,2] * multi(x))
  return(ki_e)
}
```

Robuste Desirability-Funktion für den C_p -Wert:

```
drcp<-function(l, t, x){
  if (ki_cp(x)[1]>t)
  {1}
  else if(ki_cp(x)[1]<l)
  {0}
  else (ki_cp(x)[1]-l)/(t- l)
}
```

Robuste Desirability-Funktion für den Ätzfaktor:

```
dref<-function(l, t, x){
  if (ki_ef(x)[1]>t)
  {1}
  else if(ki_ef(x)[1]<1)
  {0}
  else (ki_ef(x)[1]-1)/(t- 1)
}
```

Robuste Gesamt-Desirability-Funktion:

```
overallrobust<-function(x){
  out<-drcp(l[1],target[1],x)^w1 * dref(l[2],target[2],x)^w2
  if(any(abs(x) > 1.0))
  {
    out <- 0
  }
  out
}
```

Startpunkte für den Nelder-Mead Algorithmus:

```
anfangsp <- expand.grid(temp = seq(-1, 1, length = 5),
  speed = seq(-1, 1, length = 5), oscillation = seq(-1, 1, length = 5),
  press= seq(-1, 1, length = 5))
```

Robuste Optimierung mit Desirability-Funktionen

Sphärischer Bereich des C_p -Wertes:

```
range_cp<-function(x, y, r, methode = "min"){
  if(((x[1]-y[1])^2 + (x[2]-y[2])^2 + (x[3]-y[3])^2+ (x[4]-y[4])^2) <= r^2)
  {
    y_dach_1(x)
  }
  else if(methode == "max")
  {0}
  else 300
}
```

Sphärischer Bereich des Ätzfaktors:

```
range_ef<-function(x, y, r, methode = "min"){
  if(((x[1]-y[1])^2 + (x[2]-y[2])^2 + (x[3]-y[3])^2+ (x[4]-y[4])^2) <= r^2)
  {
    y_dach_2(x)
  }
  else if(methode == "max")
  {0}
  else 300
}
```

Desirability-Funktion:

```
dr<-function(l,t,x){
  if(x>=t-1)
  {0}
  else if(x==0)
  {1}
  else (1-x/(t-1))
}
```

Robustheits-Desirability-Funktion:

```
robustdes<-function(x, radius){
  wert1<-constrOptim(x, range_cp,y=x,r=sqrt(radius), methode = "max",
  NULL, ui=a,ci=b, control = list(fnscale = -1))$value-
  constrOptim(x, range_cp,y=x,r=sqrt(radius), methode = "min",
  NULL, ui=a,ci=b)$value

  wert2<-constrOptim(x, range_ef,y=x,r=sqrt(radius), methode = "max",
  NULL, ui=a,ci=b, control = list(fnscale = -1))$value-
  constrOptim(x, range_ef,y=x,r=sqrt(radius), methode = "min",
  NULL, ui=a,ci=b)$value

  out<-((dr(l[1],target[1],wert1))^gewicht_1 * (dr(l[2],target[2],wert2))^gewicht_2)
  if(any(abs(x) > 1.0))
  {
    out <- 0
  }
  out
}
```

Gesamt-Desirability-Funktion:

```
gesamt_des<-function(x){
  out<-((dopt(x))^g1)* (robustdes(x, 0.05))^g2
}
```

Desirability-Funktionen

Verwendet man für die Gesamt-Desirability-Funktionen keine gewichteten geometrischen Mittel kann man mit dem Package `Desirability` arbeiten. In den Funktionen `dMax`, `dMin` und `dTarget` sind die Desirability-Funktionen implementiert. Mit der Funktion `dOverall` kann man die Gesamt-Desirability-Funktion bilden.

Optimierung

Für die Optimierung wurden die Funktionen `optim` und `constrOptim` verwendet.

Funktion `optim`:

```
optim(par, fn, gr = NULL, ...,
      method = c("Nelder-Mead", "BFGS", "CG", "L-BFGS-B", "SANN",
                 "Brent"),
      lower = -Inf, upper = Inf,
      control = list(), hessian = FALSE)
```

Man kann bei dieser Funktion zwischen verschiedenen Optimierungsverfahren wählen. Defaultmäßig wird mit dem Nelder-Mead Algorithmus optimiert. Diese Funktion minimiert die Zielfunktion. Wenn man die Zielfunktion maximieren will, setzt man `control = list(fnscale = -1)`.

Funktion `constrOptim`:

```
constrOptim(theta, f, grad, ui, ci, mu = 1e-04, control = list(),
            method = if(is.null(grad)) "Nelder-Mead" else "BFGS",
            outer.iterations = 100, outer.eps = 1e-05, ...,
            hessian = FALSE)
```

Diese Funktion wird verwendet, wenn man bei der Optimierung Nebenbedingungen hat. Der zulässige Bereich wird mit den Befehlen `ui` und `ci` eingeschränkt. Es gilt `ui` `%%` `theta - ci >= 0`.

A.5.2. Kapitel 6

Simulation

Die Simulation der Zufallsvariablen wurde mit den Funktionen `rgamma`, `rlnorm` und `rweibull` durchgeführt.

Prozessfähigkeitsindizes

Funktion `pcr` im Package `qualityTools`:

```
pcr(x, distribution = "normal", lsl, usl, target, main, xlim, ylim, ...)
```

Diese Funktion berechnet C_p , C_{pk} , C_{pl} und C_{pu} für eine gegebene Stichprobe und Verteilung. Es wird ein Histogramm der Stichprobe mit der Dichtefunktion der gewählten Verteilung und ein QQ-Plot ausgegeben. Die Parameterschätzung wird mit der Funktion `fitdistr` des Packages `MASS` durchgeführt. Die Funktion gibt auch den p -Wert des Anderson Darling Tests aus.

Funktion `cp` im Package `qualityTools`:

```
cp(x, distribution = "normal", lsl, usl, target, main, xlim, ylim, ...)
```

Diese Funktion ist mit der Funktion `pcr` vergleichbar.

Fitten der Verteilung

Funktion `descdist` im Package `fitdistrplus`:

```
descdist(data, discrete = FALSE, boot = NULL, method = "unbiased",
graph = TRUE, obs.col = "darkblue", obs.pch = 16, boot.col = "orange")
```

Diese Funktion gibt einen Skewness-Kurtosis Plot aus. Es wird die quadrierte Schiefe gegen die Kurtosis gezeichnet. Der Plot zeigt an, mit welcher Verteilung die Stichprobe gefittet werden kann. Die Gamma- und Lognormalverteilung bilden eine Linie, die

Beta Verteilung bildet einen ganzen Bereich und die Normalverteilung einen Punkt. Wenn man `boot = n` setzt, werden n Bootstrap Stichproben aus der Stichprobe gezogen und in den Plot gezeichnet.

Funktion `fitdist` im Package `fitdistrplus`:

```
fitdist(data, distr, method = c("mle", "mme", "qme", "mge"),
        start=NULL, ...)
```

Es wird für die Stichprobe eine univariate Verteilung gefittet. Man kann zwischen Maximierung der Likelihood (`mle`), Momenten-Anpassung (`mme`), Quantils-Anpassung (`qme`) und Maximierung der Anpassungsgüte (`mge`) wählen.

Funktion `cdfcomp` im Package `fitdistrplus`:

```
cdfcomp(ft, xlim, ylim, ...)
```

Diese Funktion wird auf ein `fitdist`-Objekt oder auf eine Liste von `fitdist`-Objekten angewendet. Es wird die empirische Verteilungsfunktion gegen jede gefittete Verteilungsfunktion gezeichnet.

Funktion `denscomp` im Package `fitdistrplus`:

```
denscomp(ft, xlim, ylim, probability = TRUE,...)
```

Diese Funktion wird auf ein `fitdist`-Objekt oder auf eine Liste von `fitdist`-Objekten angewendet. Es wird ein Histogramm gegen jede gefittete Dichtefunktion gezeichnet.

Funktion `qqcomp` im Package `fitdistrplus`:

```
qqcomp(ft, xlim, ylim, xlogscale = FALSE, ylogscale = FALSE, ...)
```

Diese Funktion wird auf ein `fitdist`-Objekt oder auf eine Liste von `fitdist`-Objekten angewendet. Es werden die theoretischen Quantile jeder gefitteten Verteilung gegen die empirischen Quantile gezeichnet.

Funktion `ppcomp` im Package `fitdistrplus`:

```
ppcomp(ft, xlim, ylim, ...)
```

Diese Funktion wird auf ein `fitdist`-Objekt oder auf eine Liste von `fitdist`-Objekten angewendet. Es werden die theoretischen Wahrscheinlichkeiten der gefitteten Verteilung gegen die empirischen Wahrscheinlichkeiten gezeichnet.

Die vier oben beschriebenen Plots können auch gemeinsam für ein `fitdist`-Objekt oder eine Liste von `fitdist`-Objekten ausgegeben werden, indem man die Funktion `plot(ft)` aufruft.

Funktion `gofstat` im Package `fitdistrplus`:

```
gofstat(f, chisqbreaks, meancount, discrete, fitnames=NULL)
```

Diese Funktion wird auf ein `fitdist`-Objekt oder auf eine Liste von `fitdist`-Objekten angewendet. Es werden Goodness-of-Fit Statistiken für jede gefittete Verteilung berechnet.

Literaturverzeichnis

- Box, G. E. P., Hunter, W. G., Mac Gregor, J. F. & Erjavec, J. (1973). Some problems associated with the analysis of multiresponse data. *Technometrics*, 15 (1), 33-51.
- Chan, L., Cheng, S. & Spriring, F. (1988a). A new measure of process capability: C_{pm} . *Journal of Quality Technology*, 20 (3), 162-173.
- Chan, L., Cheng, S. & Spriring, F. (1988b). The robustness of the process capability index to departures from normality. *Statistical Theory and Data Analysis II (Proc. Second Pacific Area Statistical Conference) Tokyo*, 223-239.
- Chen, J. (2000). Reevaluating the process capability indices for non-normal distributions. *International Journal of Production Research*, 38 (6), 1311-1324.
- Chen, J. & Ding, C. (2001). A new process capability index for non-normal distributions. *International Journal of Quality and Reliability Management*, 18, 762-770.
- Chou, Y. M., Owen, D. B. & Borrego, S. A. (1990). Lower confidence limits on process capability indices. *Journal of Quality Technology*, 22 (3), 223-229.
- Clements, J. A. (1989, September). Process capability calculations for non-normal distributions. *Quality Progress*, 95-100.
- Derringer, G. & Suich, R. (1980). Simultaneous optimization of several response variables. *Journal of Quality Technology*, 12 (4), 214-219.
- Efron, B. & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1 (1), 54-77.
- Finley, J. (1992). What is capability? Or what is C_p and C_{pk} ? *ASQC Quality Congress Transactions, Nashville*, 186-191.
- Graybill, F. (1976). *Theory and application of the linear model*. Boston: Duxbury Press.
- He, Z., Wang, J., Oh, J. & Park, S. H. (2010). Robust optimization for multiple responses using response surface methodology. *Applied Stochastic Models in Business and Industry*, 26 (2), 157-171.
- He, Z., Zhu, P.-F. & Park, S.-H. (2012). A robust desirability function method for multi-response surface optimization considering model uncertainty. *European Journal of Operational Research*, 221, 241-247.

- Hejazi, T. H., Bashiri, M., Diaz-Garcia, J. A. & Noghondarian, K. (2012). Optimization of probabilistic multiple response surfaces. *Applied Mathematical Modelling*, *36*, 1275-1285.
- Jahn, W. & Braun, L. (2006). *Praxisleitfaden Qualität: Prozessoptimierung mit multivariater Statistik in 150 Beispielen*. München: Hanser.
- Johnson, N. L., Kotz, S. & Pearn, W. L. (1994). Flexible process capability indices. *Pakistan Journal of Statistic*, *10* (1), 23-31.
- Kane, V. E. (1986). Process capability indices. *Journal of Quality Technology*, *18* (1), 41-52.
- Khuri, A. I. & Conlon, M. (1981). Simultaneous optimization of multiple responses represented by polynomial regression functions. *Technometrics*, *23* (4), 363-375.
- Kotz, S. & Johnson, N. L. (1993). *Process capability indices*. London: Chapman and Hall.
- Kotz, S. & Lovelace, C. R. (1998). *Process capability indices in theory and practice*. London: Arnold.
- Lancaster, P. (1969). *Theory of matrices*. New York: Academic Press.
- Miettinen, K. (1999). *Non linear multiobjective optimization*. Boston: Kluwer Academic Publisher.
- Montgomery, D. (2005). *Design and analysis of experiments* (6th Aufl.). New York: John Wiley & Sons.
- Myers, R. H. & Carter, W. H. (1973). Response surface techniques for dual response systems. *Technometrics*, *15* (2), 301-317.
- Myers, R. H. & Montgomery, D. C. (2002). *Response surface methodology: Process and product in optimization using designed experiments* (2nd Aufl.). New York: John Wiley & Sons.
- Pearn, W. L., Kotz, S. & Johnson, N. L. (1992). Distributional and inferential properties of process capability indices. *Journal of Quality Technology*, *24* (4), 216-231.
- Rao, C. R. (1973). *Linear statistical inference and its applications*. New York: John Wiley & Sons.
- Spendley, W., Hext, G. & Himsworth, F. (1962). Sequential application of simplex designs in optimisation and evolutionary operation. *Technometrics*, *4*, 441-461.
- Sundaraiyer, V. H. (1996). Estimation of a process capability index for inverse gaussian distribution. *Communications in Statistics - Theory and Methods*, *25* (10), 2381-2393.

Vining, G. G. (1998). A compromise approach to multiresponse optimization. *Journal of Quality technology*, 30 (4), 309-313.

Wilkinson, J. H. (1969). *Rundungsfehler*. Berlin: Springer.