

Dipl.-Ing. Verena FEIRER

Generalisierte Binomialverteilungen zur statistischen Modellierung von unbedruckten Stellen auf Papier

DISSERTATION

zur Erlangung des akademischen Grades einer Doktorin der technischen
Wissenschaften

Doktoratsstudium der Technischen Wissenschaften im Rahmen der
Doktoratsschule „Mathematik und Wissenschaftliches Rechnen“



Betreuer:
Ao.Univ.-Prof. Dipl.-Ing. Dr.techn. Herwig FRIEDL

Institut für Statistik

Graz, im Jänner 2012

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am (Unterschrift)

Zusammenfassung

Praktische Motivation dieser Arbeit ist ein Problem aus der Papier- und Druckindustrie: das Auftreten unbedruckter Stellen im Flexodruck und dessen Zusammenhänge mit den Eigenschaften des verwendeten Papiers. Daher wurden generalisierte lineare Modelle betrachtet, welche das Auftreten unbedruckter Beobachtungen anhand von Topografie und Formation des Papiers erklären sollen.

Die vorliegenden Daten variieren jedoch je nach Druckqualität stärker oder auch schwächer, als von einem logistischen Modell unter Binomialverteilungsannahme beschrieben werden kann. Daher beschäftigt sich diese Arbeit mit zwei Verallgemeinerungen der Binomialverteilung, welche durch einen zusätzlichen Parameter größere Freiheiten aufweisen und Variabilitäten größer respektive kleiner jener der klassischen Binomialverteilung ermöglichen.

Für diese beiden Verteilungen, die multiplikative (Altham, 1978) sowie die doppelte (Efron, 1986) Binomialverteilung, wird eine Maximum Likelihood Schätzung entwickelt, welche eine simultane Schätzung der Punktschätzer eines linearen Prädiktors sowie jene des zusätzlichen Parameters der beiden Verteilungen ermöglicht. Monte-Carlo Simulationen untersuchen die Möglichkeiten und Grenzen der Parameterschätzungen unter den beiden verallgemeinerten Binomialverteilungsannahmen. Die Arbeit schließt mit Modellschätzungen unter multiplikativer sowie doppelter Binomialverteilung der Daten aus dem genannten Problem der Papier- und Druckindustrie.

Abstract

Practical motivation of this work is a problem from the paper and print industries: the occurrence of unprinted areas in flexo print and its connection to the properties of the employed paper. Generalized linear models were used to explain the occurrence of unprinted areas based on the paper's formation and topography.

But depending on the quality of the print, the present data vary much more or much less than a logistic model would be able to describe. That is why this work focuses on two generalizations of the classic binomial distribution that due to a second parameter gain more flexibility and so are able to model variabilities larger or smaller than the one of a classic binomial distribution.

For both of these distributional families, the multiplicative (Altham, 1978) and the double (Efron, 1986) binomial distribution, a maximum likelihood estimation will be developed. It will allow the simultaneous estimation of both the point estimators of the linear predictor and the additional parameter of the two distributions. Capabilities and limitations of the parameter estimation under the two distributions are scrutinized with the help of Monte-Carlo simulations. The work is topped off with model estimations of the printability dataset from the paper and print industries under the assumption of multiplicative or double binomial distribution of the response variable.

Inhaltsverzeichnis

1. Einleitung	1
2. Generalisierte Lineare Modelle	5
2.1. Das lineare Regressionsmodell	6
2.2. Das generalisierte lineare Modell	7
2.2.1. Exponentialfamilien	8
2.2.2. Likelihood- und log-Likelihoodfunktionen	17
2.2.3. Maximum Likelihood Schätzung	18
2.2.4. Tests auf Güte der Modellanpassung	20
2.2.5. Das logistische Regressionsmodell	24
2.2.6. Über- und Unterdispersion im logistischen Modell	27
3. Die multiplikative Binomialverteilung	31
3.1. Motivation	31
3.2. Herleitung	32
3.3. Eigenschaften	38
3.4. Die multiplikative Binomialverteilung in der logistischen Regression	45
3.4.1. Schätzung von ω	47
4. Die doppelte Binomialverteilung	49
4.1. Definition	49
4.2. Eigenschaften	50
4.3. Die doppelte Binomialverteilung in der logistischen Regression	56
5. Simulationen und Code	59
5.1. Diskussion der Optimierungsroutine	60
5.1.1. Abriss zum Newton- und Quasi-Newton-Verfahren	60
5.1.2. Approximation der Hessematrix mittels BFGS-Verfahren	61
5.2. Beschreibung der verwendeten R-Funktionen	62
5.2.1. Abbruchkriterien der Optimierungsroutine	62
5.2.2. Zufallszahlengeneratoren	63
5.3. Simulationen basierend auf multiplikativ binomialverteilten Zufallsvariablen	64
5.3.1. Vergleiche von Erwartungswert und Varianz	64

5.3.2.	Simulationsergebnisse der Parameterschätzung mittels <code>optim()</code>	65
5.4.	Simulationen basierend auf doppelt binomialverteilten Zufallsvariablen	79
5.4.1.	Vergleiche von Erwartungswert und Varianz	79
5.4.2.	Simulationsergebnisse der Parameterschätzung mittels <code>optim()</code>	80
5.5.	R-Code zur Modellierung unter verallgemeinerter Binomialverteilungsannahme	89
6.	Modellierung der Farbübertragung	97
6.1.	Der Datensatz	98
6.1.1.	Sackpapiere	98
6.1.2.	Flexodruck	98
6.1.3.	Formation	100
6.1.4.	Topografie und gefilterte Topografie	100
6.1.5.	Papiereigenschaften der vermessenen Proben	105
6.1.6.	Druckqualität	108
6.2.	Binomiale Modelle	110
6.2.1.	Modellschätzung für einen Ausschnitt von Probe 201 (A)	111
6.2.2.	Vergleich der Modellierungsergebnisse der zwölf Proben	121
6.3.	Verallgemeinerte Binomialmodelle	124
6.3.1.	Modellschätzung für einen Ausschnitt von Probe 201 (A)	124
6.3.2.	Modellierungsergebnisse der zwölf Proben	132
6.3.3.	Andere Blockgrößen	134
7.	Zusammenfassung und Ausblick	139
7.1.	Zusammenfassung der praxisbezogenen Resultate	139
7.2.	Zusammenfassung des theoretischen Teils	140
7.3.	Ausblick	141
A.	R-Funktionen	143
A.1.	Log-Likelihoodfunktionen	143
A.2.	Modellierungsfunktionen	147
A.3.	Normierungsterme	150
A.4.	Funktionen für den Simulationsteil	153

Abbildungsverzeichnis

2.1.	Verschiedene Linkfunktionen des logistischen Modells	25
2.2.	Beispiel einer logistischen Regression	27
3.1.	Ausprägungsmöglichkeiten dreier zweistufiger Faktoren	32
3.2.	Wahrscheinlichkeitsfunktion der multiplikativen Binomialverteilung bei $\pi = 0,5$	37
3.3.	Wahrscheinlichkeitsfunktion der multiplikativen Binomialverteilung bei $\pi = 0,8$	37
3.4.	Erwartungswert der multiplikativen Binomialverteilung in π und ω	43
3.5.	Standardabweichung der multiplikativen Binomialverteilung in π und ω	43
3.6.	Standardabweichung im Erwartungswert und ω	43
3.7.	Varianzen der klassischen und der multiplikativen Binomialverteilung	44
4.1.	Wahrscheinlichkeitsfunktion der doppelten Binomialverteilung bei $\pi = 0,5$	51
4.2.	Wahrscheinlichkeitsfunktion der doppelten Binomialverteilung bei $\pi = 0,8$	51
4.3.	Erwartungswert der doppelten Binomialverteilung in π und ϕ	54
4.4.	Standardabweichung der doppelten Binomialverteilung in π und ϕ	54
4.5.	Standardabweichung der doppelten Binomialverteilung im Erwartungswert und ϕ	54
4.6.	Varianzen der klassischen und der doppelten Binomialverteilung	55
5.1.	Ergebnisse der Parameterschätzer $\hat{\beta}_0$ und $\hat{\omega}$ aus der Monte-Carlo Simulation	66
5.2.	Empirische Momente bei multiplikativ binomialverteilten Zufallsvariablen	67
5.3.	Resultierende Punktschätzer $\hat{\beta}_0$ für $\omega < 1$ und $\omega = 1$	68
5.4.	Resultierende Punktschätzer $\hat{\omega}$ für $\omega < 1$ und $\omega = 1$	68
5.5.	Boxplot der Iterationen sowie der $-2 \cdot l(\hat{\beta}, \hat{\omega} \mathbf{y})$ bei $\omega < 1$ und $\omega = 1$	68
5.6.	Resultierende Punktschätzer $\hat{\beta}_0$ für $\omega > 1$ und $\omega = 1$	71
5.7.	Resultierende Punktschätzer $\hat{\omega}$ für $\omega > 1$ und $\omega = 1$	71
5.8.	Boxplot der Iterationen sowie der $-2 \cdot l(\hat{\beta}, \hat{\omega} \mathbf{y})$ bei $\omega > 1$ und $\omega = 1$	71
5.9.	Erwartungswerte und Standardfehler unter $\boldsymbol{\eta} = X\boldsymbol{\beta}$ und ω	72
5.10.	Resultierende Punktschätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ für $\omega < 1$	73
5.11.	Resultierende Punktschätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ für $\omega = 1$	73
5.12.	Resultierende Punktschätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ für $\omega > 1$	73
5.13.	Resultierende Punktschätzer $\hat{\omega}$ bei $\omega < 1$	74
5.14.	Resultierende Punktschätzer $\hat{\omega}$ bei $\omega = 1$	74
5.15.	Resultierende Punktschätzer $\hat{\omega}$ bei $\omega > 1$	74

5.16. Erwartungswerte und Standardfehler unter $\boldsymbol{\eta} = X\boldsymbol{\beta}$ und ω	75
5.17. Resultierende Punktschätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ für $\omega < 1$	76
5.18. Resultierende Punktschätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ für $\omega = 1$	76
5.19. Resultierende Punktschätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ für $\omega > 1$	76
5.20. Resultierende Punktschätzer $\hat{\omega}$ bei $\omega < 1$	77
5.21. Resultierende Punktschätzer $\hat{\omega}$ bei $\omega = 1$	77
5.22. Resultierende Punktschätzer $\hat{\omega}$ bei $\omega > 1$	77
5.23. Boxplot der Iterationen sowie der $-2 \cdot l(\hat{\boldsymbol{\beta}}, \hat{\omega} \mathbf{y})$ bei verschiedenen ω	78
5.24. Empirische Momente bei doppelt binomialverteilten Zufallsvariablen	81
5.25. Resultierende Punktschätzer $\hat{\beta}_0$ für $\phi < 1$ und $\phi = 1$	82
5.26. Resultierende Punktschätzer $\hat{\phi}$ für $\phi < 1$ und $\phi = 1$	82
5.27. Boxplot der Iterationen sowie der $-2 \cdot l(\hat{\boldsymbol{\beta}}, \hat{\phi} \mathbf{y})$ bei $\phi < 1$ und $\phi = 1$	82
5.28. Resultierende Punktschätzer $\hat{\beta}_0$ für $\phi > 1$ und $\phi = 1$	84
5.29. Resultierende Punktschätzer $\hat{\phi}$ für $\phi > 1$ und $\phi = 1$	84
5.30. Boxplot der Iterationen sowie der $-2 \cdot l(\hat{\boldsymbol{\beta}}, \hat{\phi} \mathbf{y})$ bei $\phi > 1$ und $\phi = 1$	84
5.31. Erwartungswerte und Standardfehler unter $\boldsymbol{\eta} = X\boldsymbol{\beta}$ und ϕ	85
5.32. Resultierende Punktschätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ für $\phi < 1$	86
5.33. Resultierende Punktschätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ für $\phi = 1$	86
5.34. Resultierende Punktschätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ für $\phi > 1$	86
5.35. Resultierende Punktschätzer $\hat{\phi}$ bei $\phi < 1$	87
5.36. Resultierende Punktschätzer $\hat{\phi}$ bei $\phi = 1$	87
5.37. Resultierende Punktschätzer $\hat{\phi}$ bei $\phi > 1$	87
5.38. Bei einem Aufruf von <code>gbplot()</code> erstellte Plotserie.	94
5.39. Vergleich geschätzter Momente unter verschiedenen Verteilungsannahmen.	95
6.1. Farbaufnahme und Binärbild einer Probenfläche	97
6.2. Skizze der Funktionsweise des Flexodrucks	99
6.3. Im Flexodruck typische Quetschränder	99
6.4. Formation eines Papiers	101
6.5. Gemessene Rohtopografie eines Papiers	101
6.6. Gefilterte Topografie eines Papiers	101
6.7. Beispiel für ein Bild im Ortsraum und im Frequenzraum	103
6.8. Beispiel für einen Bandpassfilterkern im Frequenzraum.	103
6.9. Gefiltertes Bild im Frequenzraum und im Ortsraum	103
6.10. Farbaufnahme von Papier 201 (A)	106
6.11. Dichotomisierte Aufnahme der Farbübertragung	107
6.12. Druckqualität gegen Rauigkeit	109
6.13. Ausschnitt aus Probe 201 (A)	110
6.14. Binärisierte Farbübertragung des Ausschnitts von Probe 201 (A)	110
6.15. Gefilterte Topografie des Ausschnitts von Probe 201 (A)	111
6.16. Formation des Ausschnitts von Probe 201 (A)	111
6.17. Geschätzte Wahrscheinlichkeiten des Ausschnitts von Probe 201 (A)	113
6.18. Sensitivität und Relevanz einer Probenfläche	115
6.19. Diskretisierte Modellprognose	115
6.20. Überlagerung von Realität und Modellprognose	117
6.21. Überlagerung von Realität und Prognose in einem Modell mit Formation	117

6.22. Überlagerung von Realität und Prognose in einem Modell mit Topografie	117
6.23. Druckqualität, geplottet gegen die Modelldevianz	122
6.24. AIC Werte der zwölf Binomialmodelle	122
6.25. Geschätzte Wahrscheinlichkeiten im Fall der klassischen Binomialverteilung . .	130
6.26. Geschätzte Wahrscheinlichkeiten im Fall der multiplikativen Binomialverteilung	130
6.27. Geschätzte Wahrscheinlichkeiten im Fall der doppelten Binomialverteilung . . .	130
6.28. Vergleich der resultierenden Erwartungswerte dreier Modelle	131
6.29. Vergleich der resultierenden Standardabweichungen dreier Modelle	131
6.30. AIC Werte verschiedener Modelle	133
6.31. AIC-Werte der drei Modelle aller zwölf Proben.	133
6.32. Zusätzlicher Parameter ω gegen die Druckqualität	135
6.33. Zusätzlicher Parameter ϕ gegen die Druckqualität	135
6.34. Druckqualität gegen AIC-Werte im Fall $n = 81$	136

Tabellenverzeichnis

2.1.	Verschiedene Linkfunktionen des binomialen Regressionsmodells	25
2.2.	Datensatz für ein Beispiel logistischer Regression	26
3.1.	Darstellung der Datensituation der multiplikativen Binomialverteilung	31
5.1.	Unterschiede der empirischen und theoretischen Erwartungswerte	64
5.2.	Unterschiede der empirischen und theoretischen Varianzen	65
5.3.	Entwicklung der Punktschätzer in den einzelnen Iterationsschritten	69
5.4.	Iterationsanzahlen bei Simulationsserien mit Modell $X\beta$ und verschiedene ω . .	75
5.5.	Wertebereiche der ersten beiden Momente	78
5.6.	Unterschiede der empirischen und theoretischen Erwartungswerte	80
5.7.	Unterschiede der empirischen und theoretischen Varianzen	80
5.8.	Wertebereiche der ersten beiden Momente	88
6.1.	Kennwerte der untersuchten Papiere	105
6.2.	Druckqualitäten dq der untersuchten Papiere.	108
6.3.	Vier mögliche Fälle im Vergleich von Realität und Modellprognose	114
6.4.	Sensitivitäten, Devianzen und AICs verschiedener Modelle	118
6.5.	Sensitivitäts- und Devianzwerte verschiedener Modelle	119
6.6.	Kennwerte der Modellierungsergebnisse der einzelnen Papiere	121
6.7.	Punktschätzer mit Standardfehlern der Wechselwirkungsmodelle	123

Funktionenverzeichnis

5.1.	Definition der log-Likelihoodfunktion	62
5.2.	Definition der negativen log-Likelihoodfunktion	62
5.3.	Wahrscheinlichkeitsfunktion, Verteilungsfunktion, Quantilsfunktion und Zufallszahlengenerator der verallgemeinerten Binomialverteilungen.	63
5.4.	Definition der Funktion gbmod()	90
5.5.	Modellzusammenfassung aus gbmod()	91
5.6.	Output der Funktion gbsummary()	92
6.1.	Ergebnis des Binomialmodells für den Ausschnitt aus 201 (A)	112
6.2.	Ergebnis des Binomialmodells des Ausschnitts aus 201 (A)	120
6.3.	Modellzusammenfassung eines Binomialmodells aus gbsummary()	125
6.4.	Modellzusammenfassung unter Annahme der multiplikativen Binomialverteilung	127
6.5.	Modellzusammenfassung unter Annahme der doppelten Binomialverteilung . . .	128

Einleitung

Seit 2007 läuft am Institut für Papier-, Zellstoff- und Fasertechnik der Technischen Universität Graz unter dem Titel „Modellierung der Farbübertragung aus lokalen Papiereigenschaften“ ein Projekt, welches die Zusammenhänge zwischen Druckqualität und Papier untersucht. Die Erforschung solcher Abhängigkeiten ist von Interesse, da mangelhafte Druckqualität zu einem immer präsenteren Thema wird und die kontinuierlich steigenden Anforderungen an das Druckbild zu einer wachsenden Zahl an Reklamationen führen¹. Aus diesem Grund wurde das Projekt der Modellierung der Farbübertragung initiiert und von der österreichischen Forschungsförderungsgesellschaft (FFG) sowie den Projektpartnern Feinpapierfabrik Franz Feuerstein (Traun, OÖ), Mondi Packaging (Frantschach, Ktn), Myllykoski (Ettringen, D), Norske Skog (Bruck, Stmk), Omya (Oftringen, CH), Sappi (Gratkorn, Stmk), SCA (Laakirchen, OÖ), Stora Enso (Mönchengladbach, D) und Voith Paper (Heidenheim, D) unterstützt und ermöglicht.

Einer der Forschungsschwerpunkte des Projekts beschäftigt sich mit der fehlerhaften Farbübertragung im Flexodruck und der Zielsetzung, eine „Modellbildung und Ursachenfindung für schlechtes Druckbild“ durchzuführen. Flexodruck ist hauptsächlich im Verpackungsdruck zu finden, wobei die Verpackung eines Gutes heutzutage nicht mehr ausschließlich für Schutz und Verpackung gedacht ist. Sie wird inzwischen auch zum Informationsträger und Werbemittel und trägt daher entscheidend zur Kaufentscheidung eines Kunden bei. Aufgrund dessen ist auch in der Sparte des Verpackungsdrucks eine gute Bedruckbarkeit des Mediums von großer Bedeutung.

In Zusammenarbeit mit dem Institut für Statistik der Technischen Universität Graz wurde daher ein logistisches Modell erarbeitet, welches eine Modellierung der erfolgreichen beziehungsweise fehlgeschlagenen Farbübertragung anhand mehrerer Papiereigenschaften ermöglicht. Untersuchungen zum Zusammenhang zwischen einzelnen Papiereigenschaften und der lokalen Dichte der Druckfarbe am Papier wurden unter anderem bereits von Gregersen, Jonsen und Helle (1995) oder Barros, Fahlcrantz und Johansson (2005) durchgeführt. Statistische Modellierungsansätze jedoch, bei denen mehrere erklärende Größen zur Beschreibung der Aus-

¹siehe den Branchenreport über die Druckindustrie der Europäischen Union (Ernst & Young, 2007) oder jenen des Deutschen Sparkassen Verlags (Jankowski, 2010)

prägung der Zielvariable beitragen, wurden im Kontext der erfolgreichen oder fehlgeschlagenen Farbübertragung im Druck noch nicht betrachtet. Daher wurde bereits im ersten Forschungsjahr des Projekts die Modellierung der Farbübertragung anhand logistischer Regressionsmodelle begonnen und war Hauptbestandteil der dieser Arbeit vorangegangenen Diplomarbeit (Feirer, 2008). Um in weiterer Folge auf diese logistischen Modelle aufbauen zu können, widmet sich Kapitel 2 der vorliegenden Arbeit einer Einführung in die verwendeten Modellierungstechniken.

Wie sich im Zuge der Arbeiten gezeigt hat, kann die Anzahl erfolgreicher beziehungsweise fehlgeschlagener Farbübertragungen jedoch nicht von Vorhinein als binomialverteilt angenommen werden. Statt dessen variieren die Daten im Fall eines guten Druckbildes mitunter schwächer, als von einer Binomialverteilung angenommen. Auf einem besonders schlechten Druckbild hingegen ist die Variabilität der Daten deutlich größer, als von der Binomialverteilung erlaubt wäre. Daher liegt die Zielsetzung der vorliegenden Arbeit darin, ein Modell zur Beschreibung von Häufigkeiten anhand mehrerer Prädiktoren zu finden, welches sowohl bei größerer als auch bei kleinerer Variabilität der Daten als jener der Binomialverteilung angemessen funktioniert.

In der statistischen Literatur sind Datensätze mit einer größeren Variabilität als jener der Binomialverteilung sowie Lösungsansätze zur besseren Modellierung derselben durchaus zu finden. Datensätze mit kleinerer Variabilität jedoch sind selten. Und eine gemeinsame Modellierungsmöglichkeit von Datensätzen aus beiden Situationen ist nicht vorhanden. Aus diesem Grund widmet sich die vorliegende Arbeit zwei in der Literatur vorgestellten Verteilungen, mit denen sich sowohl Über- als auch Unterdispersion eines Datensatzes beschreiben lassen. Es sind dies die multiplikative Verallgemeinerung der Binomialverteilung wie von Altham (1978) vorgestellt und die doppelte Binomialverteilung von Efron (1986). Beide Verteilungen sind theoretisch in der Lage, auf Variabilitäten größer oder kleiner jener der Binomialverteilung zu reagieren und sollen daher zur stochastischen Modellierung der Farbübertragung herangezogen werden.

Kapitel 3 ist der multiplikativen Verallgemeinerung der Binomialverteilung gewidmet und beginnt mit einer Vorstellung der Arbeiten von Altham. Danach wird gezeigt, dass die multiplikative Verallgemeinerung ein Mitglied der Exponentialfamilie ist. In weiterer Folge wendet sich das Kapitel der Entwicklung von geschlossenen Darstellungen für Erwartungswert und Varianz der Verteilung zu. Weiters wird die Verteilung in die Theorie der generalisierten linearen Modelle eingebettet. Innerhalb dieses Modellierungskontextes werden ihre log-Likelihoodfunktion sowie ihre partiellen Ableitungen erster und zweiter Ordnung hergeleitet.

In Kapitel 4 wird ähnlich vorgegangen, um die doppelte Binomialverteilung einzuführen. Nach einem allgemeinen Einstieg für doppelte Exponentialfamilien wird die doppelte Binomialverteilung vorgestellt und festgehalten, dass auch diese Verteilung Mitglied der Exponentialfamilie ist. Es folgt eine Herleitung von geschlossenen Darstellungen für Erwartungswert und Varianz der Verteilung sowie ein Vergleich dieser beiden Momente mit jenen der klassischen Binomialverteilung. Den Abschluss des Kapitels macht abermals die Einbettung der vorgestellten Verteilungsfamilie in den Kontext der generalisierten linearen Modelle. Aus der log-Likelihoodfunktion der Verteilung werden die beiden ersten partiellen Ableitungen gebildet, welche in weiterer Folge für Parameterschätzungen unter doppelter Binomialverteilungsannahme verwendet werden sollen.

Einer Vorstellung dieser beiden Verteilungen in den Kapiteln 3 und 4 folgt in Kapitel 5 eine Beschreibung von R-Funktionen, welche eine Modellschätzung unter diesen Verteilungsannahmen ermöglichen. Dafür wird auf die in den vorangegangenen Kapiteln vorgestellten log-Likelihoodfunktionen und die ersten beiden partiellen Ableitungen der beiden Verteilungen zurückgegriffen. Mit diesen R-Funktionen zur Modellierung verallgemeinert binomialverteilter

Responsevariablen wurden ausführliche Simulationen durchgeführt, um ihre Funktionalität zu untersuchen sowie das Verhalten der beiden Verteilungen im Zuge einer Modellschätzung zu verstehen. Ausführliche Darlegungen der Ergebnisse dieser durchgeführten Monte-Carlo Simulationen machen den Großteil der in Kapitel 5 diskutierten Ergebnisse aus.

Nach der Konstruktion einer logistischen Modellschätzung für Responsevariablen, deren Varianzstruktur nicht jener der Binomialverteilung folgt, kehrt Kapitel 6 wieder zum vorliegenden Praxisproblem der fehlerhaften Farbübertragung im Flexodruck zurück. Dieses Kapitel versucht, dem Leser einen Überblick über die betrachtete Situation zu geben. Dafür werden einige wichtige Begriffe aus der Papierherstellung und -bedruckung knapp erläutert und ihre Zusammenhänge beschrieben. Aufbauend auf dieses Wissen werden Modelle besprochen, die mit einer der beiden genannten Verteilungsannahmen die binäre Größe der Farbübertragung zu modellieren versuchen. Neben der Diskussion der eigentlichen Modellierungsergebnisse werden zahlreiche weitere Erkenntnisse angesprochen. So werden die Einflüsse der einzelnen Prädiktoren auf das Druckbild besprochen oder Vergleiche der vorliegenden Papiere vorgenommen. Weiters werden Werte wie Rauigkeit, Luftdurchlässigkeit oder Dicke der Proben sowie gewisse fertigungstechnische Größen berücksichtigt.

Den Abschluss macht eine Zusammenfassung der sowohl aus statistischer als auch aus papiertechnischer Sicht wichtigsten Resultate in Kapitel 7. Der Anhang enthält die Beschreibung aller im Zuge dieser Arbeit entstandenen R-Funktionen.

Generalisierte Lineare Modelle

Dieser Teil beinhaltet den grundlegenden statistischen Aspekt der Arbeit. Dem mathematisch versierten Leser mögen Teile dieses Kapitels vertraut sein. In Anbetracht der Tatsache, dass es sich bei dieser Arbeit jedoch um ein in Kooperation mit der Verfahrenstechnik entstandenes Werk handelt, wird in der nun folgenden Beschreibung der verwendeten Theorie bewusst derart vorgegangen, um auch einem der Statistik fremden Leser einen Zugang zu diesem Thema zu ermöglichen.

Um in weiterer Folge von logistischen Modellen sprechen zu können, werden hier unterschiedliche statistische Modelle vorgestellt. Deren einheitliches Ziel ist, eine Zielgröße anhand mehrerer erklärenden Variablen möglichst gut zu beschreiben.

Das einfachste dieser Modelle ist ein lineares Regressionsmodell, welches in Abschnitt 2.1 aus Gründen der Vollständigkeit kurz erwähnt werden soll. Auf die in Abschnitt 6 beschriebenen Daten aus der Praxis kann ein lineares Regressionsmodell jedoch nicht angewandt werden, da diese den Anforderungen eines linearen Modells nicht genügen. Statt dessen muss ein logistisches Modell aus der Klasse der generalisierten linearen Modelle verwendet werden. Daher liegt der Schwerpunkt des Kapitels 2 auf der Betrachtung der erstmals von Nelder und Wedderburn (1972) beschriebenen generalisierten linearen Modelle. Besondere Aufmerksamkeit wird dabei dem logistischen Modell in Abschnitt 2.2.5 geschenkt, welches zur Modellierung binomialverteilter Zielgrößen verwendet wird.

Die bereits erwähnten Daten aus Kapitel 6 haben jedoch in weiterer Folge die besondere Eigenschaft, eine Variabilität größer oder kleiner jener der Binomialverteilung aufzuweisen. In diesen Fällen spricht man von einer vorliegenden Über- oder Unterdispersion, auf die im letzten Abschnitt des Kapitels noch eingegangen wird.

Dieses Vorliegen von Über- und Unterdispersion ist auch der Grund für die zentrale Zielsetzung dieser Arbeit, eine Erweiterung des logistischen Modells zu finden, sodass auch für Daten mit kleinerer oder größerer Varianz als jener der Binomialverteilung ein angemessenes Modell geschätzt werden kann. Zwei Verallgemeinerungen, welche diesen Herausforderungen genügen, werden in den Kapiteln 3 und 4 vorgestellt und ausführlich diskutiert.

Festgehalten soll an dieser Stelle noch werden, dass in dieser Arbeit sowohl für eine Zufallsvariable als auch für ihre Realisation y geschrieben wird, ohne mittels Groß- und Kleinschreibung

zwischen den beiden zu unterscheiden.

2.1 Das lineare Regressionsmodell

Ziel der linearen Regressionsanalyse ist es, einen Zusammenhang zwischen einer Zielgröße oder Responsevariablen y und einer oder mehreren erklärenden Größen x_1, \dots, x_{p-1} zu beschreiben. Der gesuchte Zusammenhang zwischen der Response und den Prädiktoren lässt sich notieren als

$$E(y) = \beta_0 + x_1\beta_1 + \dots + x_{p-1}\beta_{p-1}$$

in den unbekanntem Parametern $\beta_0, \dots, \beta_{p-1}$. Diese unbekanntem Parameter sollen in weiterer Folge geschätzt werden, indem für festgelegte Werte der Tupel $(1, x_{i1}, \dots, x_{i,p-1})$ mit $i = 1, \dots, N$ die Realisationen y_i erhoben werden. Es ergeben sich also N Gleichungen der Art

$$\mu_i = \beta_0 + x_{i1}\beta_1 + \dots + x_{i,p-1}\beta_{p-1} \quad (2.1)$$

für die Beschreibung des Zusammenhanges zwischen dem Erwartungswert μ_i von y_i und den erklärenden Größen. Der Parameter β_0 wird in Anlehnung an die klassische Geradengleichung $y = kx + d$ gerne Intercept genannt und beschreibt den konstanten Teil der Geraden. Im Fall eines Modells, welches nur eine erklärende Variable verwendet, also $p = 2$ gilt, vereinfacht sich das Modell zu $\mu = \beta_0 + x_1\beta_1$. Der Parameter β_1 beschreibt die Steigung der Geraden und somit auch die Wichtigkeit der Einflussgröße x_1 . Ebenso können auch im komplexeren Fall mit mehreren erklärenden Größen die Parameter $\beta_1, \dots, \beta_{p-1}$ als Steigungen einer Ebene (im Fall $p = 3$) oder Hyperebene (wenn $p > 3$) interpretiert werden.

Alternativ wird Modell (2.1) auch häufig in Matrixnotation als

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$$

geschrieben, indem die Einzelbeobachtungen y_i zusammengefasst werden zum $N \times 1$ -Vektor $\mathbf{y} = (y_1, \dots, y_N)^t$ und ihre Erwartungswerte μ_i zu $\boldsymbol{\mu} = (\mu_1, \dots, \mu_N)^t \in \mathbb{R}^{N \times 1}$. Die aus den Zeilenvektoren bestehende Matrix $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N)^t \in \mathbb{R}^{N \times p}$ enthält die erklärenden Größen $\mathbf{x}_i = (1, x_{i1}, \dots, x_{i,p-1})$ einer jeden Beobachtung i , wobei die „1“ Einträge der ersten Spalte zum Interceptparameter β_0 gehören. In beiden Notationsvarianten der Modellgleichung enthält die rechte Seite die unbekanntem Parameter $\boldsymbol{\beta} = (\beta_0, \dots, \beta_{p-1})^t \in \mathbb{R}^{p \times 1}$, für die im Zuge der Modellschätzung Punktschätzer $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \dots, \hat{\beta}_{p-1})^t$ ermittelt werden.

Aus einem erfolgreich geschätzten Modell lassen sich dann unterschiedliche Schlussfolgerungen ziehen. So existieren Tests, mit denen sich die Relevanz der einzelnen erklärenden Größen ermitteln lässt und es ermöglichen, Prädiktoren zu identifizieren, welche keinen Einfluss auf die Ausprägungen der y_i haben. Weiters ist es möglich, verschachtelte Modelle (*nested models*) miteinander zu vergleichen und Aussagen darüber zu tätigen, welches der betrachteten Modelle die Daten besser zu beschreiben vermag.

All diese Aussagen sind jedoch nur korrekt, wenn gewisse Modellannahmen nicht verletzt werden. Das bedeutet:

- Die y_i müssen unabhängig und normalverteilt (*independent and ident*, kurz iid) sein mit Erwartungswert μ_i und Varianz σ^2 , also $y_i \stackrel{iid}{\sim} N(\mu_i, \sigma^2)$.

- Außerdem müssen alle y_i dieselbe Varianz σ^2 aufweisen. (Diese Anforderung ist als sogenannte Homoskedastizität bekannt.)

Weitaus genauere Einführungen in die lineare Regressionsanalyse als dieser kurze Überblick sind unter anderem in Jørgensen (1993) oder Rawlings, Pantula und Dickey (1998) zu finden.

Nun existieren jedoch Datensätze, welche die oben genannten Anforderungen nicht erfüllen können. Ist die Annahme der Homoskedastizität verletzt, so kann unter Umständen eine Transformation der Zielvariablen bereits Abhilfe schaffen. Handelt es sich bei der Zielvariablen jedoch nicht um eine stetige, sondern eine binäre Größe oder eine Zählvariable, kann auch eine Transformation die oben geforderten Annahmen nicht mehr erreichen. In solchen Fällen bietet sich die Verwendung eines generalisierten linearen Modells an, welches die strengen Voraussetzungen der linearen Regressionsanalyse etwas lockert und in die Lage versetzt, auch nicht-normalverteilte Zielvariablen zu modellieren.

2.2 Das generalisierte lineare Modell

In einem generalisierten linearen Modell wird von der Responsevariablen nur noch verlangt, dass ihre Verteilung ein Mitglied der einparametrischen, linearen Exponentialfamilie in kanonischer Form ist. Die im linearen Regressionsmodell betrachtete Normalverteilung ist, sofern der Varianzparameter σ^2 als fest angenommen werden kann, ein Mitglied dieser Familie, ebenso aber auch noch etliche andere gängige Verteilungsfunktionen. Genaueres zu Exponentialfamilien folgt in Abschnitt 2.2.1.

Ist nun die Dichte- oder Wahrscheinlichkeitsfunktion der Zufallsvariablen y ein Mitglied dieser Exponentialfamilie, lässt sich das Verhalten der Responsevariablen mittels eines generalisierten linearen Modells beschreiben. Der Zusammenhang des linearen Prädiktors $\mathbf{x}^t \boldsymbol{\beta}$, basierend auf den $p-1$ erklärenden Größen x_1, \dots, x_{p-1} , und des Erwartungswerts μ muss im Gegensatz zum linearen Regressionsmodell hier jedoch nicht mehr streng linearer Natur sein. Statt dessen kann eine Funktion des Erwartungswerts als Linearkombination der erklärenden Größen modelliert werden. Notiert wird solch ein Modell üblicherweise in der Form

$$g(\mu) = \mathbf{x}^t \boldsymbol{\beta}, \quad (2.2)$$

wobei sich für die Funktion $g(\cdot)$ der Name Linkfunktion eingebürgert hat.

Im Vergleich zum linearen Regressionsmodell sind die wichtigsten Verallgemeinerungen somit:

- Die Responsevariable y kann aus einer anderen Verteilung als der Normalverteilung stammen und muss nicht mehr stetiger Natur sein. Die Verteilung muss jedoch Mitglied der Exponentialfamilie sein, die in Abschnitt 2.2.1 noch näher beschrieben wird.
- Eine Funktion des Erwartungswertes μ soll durch den linearen Prädiktor $\mathbf{x}^t \boldsymbol{\beta}$ beschrieben werden.

Eine weitere Neuerung, die hier nicht sofort ersichtlich ist, aber in späterer Folge von Bedeutung sein wird, soll hier ebenfalls festgehalten werden:

- Die Varianz kann in Abhängigkeit des Erwartungswerts beschrieben werden, anstatt wie im linearen Modell konstant im Erwartungswert zu sein.

2.2.1 Exponentialfamilien

Da bereits mehrmals erwähnt wurde, dass die Dichte- oder Verteilungsfunktion der Responsevariablen y im Fall eines generalisierten linearen Modells ein Mitglied der Exponentialfamilie sein muss, soll diese Familie nun vorgestellt werden. Definition und grundlegende Eigenschaften dieser Familie sind in der Literatur an vielen Stellen zu finden, zum Beispiel bei Dobson und Barnett (2008), Molenberghs und Verbeke (2005) oder Aitkin, Francis und Hinde (2005), um nur einige zu nennen. Die nun folgende allgemeine Einführung der Exponentialfamilie jedoch stammt aus Casella und Berger (2002).

Allgemeine Exponentialfamilien

Kann die Dichte- oder Wahrscheinlichkeitsfunktion einer Zufallsvariablen y für einen Parametervektor $\boldsymbol{\theta}$ geschrieben werden in der Form

$$f(y|\boldsymbol{\theta}) = h(y) c(\boldsymbol{\theta}) \exp\left(\sum_{k=1}^K w_k(\boldsymbol{\theta}) t_k(y)\right), \quad (2.3)$$

so ist $f(y|\boldsymbol{\theta})$ ein Mitglied der K -parametrischen Exponentialfamilie. Dabei müssen die Funktionen $h(y)$, $c(\boldsymbol{\theta})$, $w_1(\boldsymbol{\theta}), \dots, w_K(\boldsymbol{\theta})$ und $t_1(y), \dots, t_K(y)$ bekannt sein. Außerdem müssen $h(y)$ und $c(\boldsymbol{\theta})$ positiv sein.

Üblicherweise wird eine Darstellung gewählt, in der K minimal ist. In der Praxis ist in den meisten Fällen $K = 1$, seltener gilt $K = 2$ und fast nie ist $K \geq 3$. Dabei kann K nie größer sein als die Anzahl der Parameter in $\boldsymbol{\theta}$. Im Fall $K < d$ für die Dimension d von $\boldsymbol{\theta} \in \mathbb{R}^{d \times 1}$ wird von einer gekrümmten und bei $K = d$ von einer vollen Exponentialfamilie gesprochen. Erwähnt sei außerdem noch, dass eine gefundene Darstellung einer Dichte- oder Wahrscheinlichkeitsfunktion als Exponentialfamilie nicht eindeutig ist und es oft mehrere äquivalente Möglichkeiten gibt, eine Funktion in der geforderten Form (2.3) zu notieren.

Unter anderem handelt es sich bei der Poissonverteilung, der standardisierten Binomialverteilung, der Normalverteilung oder der Gammaverteilung um Mitglieder der Exponentialfamilie.

Beispiel: Die Dichtefunktion der Normalverteilung lässt sich umformen zu

$$\begin{aligned} f(y|\mu, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y^2 - 2y\mu + \mu^2}{2\sigma^2}\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \exp\left(\frac{\mu}{\sigma^2} y - \frac{1}{2\sigma^2} y^2\right) \end{aligned} \quad (2.4)$$

für $y \in \mathbb{R}$, $\mu \in \mathbb{R}$ und $\sigma > 0$. Um zur Darstellung der in (2.3) geforderten Form zu gelangen, wird für $\boldsymbol{\theta} = (\mu, \sigma^2)$ die Parametrisierung

$$w_1(\boldsymbol{\theta}) = \frac{\mu}{\sigma^2} \quad \text{und} \quad w_2(\boldsymbol{\theta}) = -\frac{1}{2\sigma^2}$$

gewählt. In weiterer Folge ergibt sich somit $t_1(y) = y$ sowie $t_2(y) = y^2$. In diesem Fall ist die Funktion $h(y) = 1$ und $c(\boldsymbol{\theta})$ enthält die restlichen Terme aus (2.4),

$$c(\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\mu^2}{2\sigma^2}\right).$$

Die Normalverteilung ist somit ein Mitglied der zweiparametrischen Exponentialfamilie. \lrcorner

Beispiel: Auch die Dichtefunktion der Gammaverteilung kann auf die in (2.3) geforderte Form gebracht werden. Durch umschreiben der Dichtefunktion

$$\begin{aligned} f(y|\alpha, \beta) &= \frac{1}{\Gamma(\alpha)\beta^\alpha} y^{\alpha-1} e^{-\frac{y}{\beta}} \\ &= \frac{1}{y} \frac{1}{\Gamma(\alpha)\beta^\alpha} \exp\left(\alpha \log y - \frac{y}{\beta}\right) \end{aligned}$$

für $0 \leq y < \infty$ und positive α und β ergibt sich eine Darstellung, bei der die Funktion in die geforderten Komponenten

$$\begin{aligned} h(y) &= \frac{1}{y} & \text{und} & & c(\alpha, \beta) &= \frac{1}{\Gamma(\alpha)\beta^\alpha}, \\ t_1(y) &= \log(y) & \text{mit} & & w_1(\alpha, \beta) &= \alpha & \text{und} \\ t_2(y) &= y & \text{mit} & & w_2(\alpha, \beta) &= -\frac{1}{\beta} \end{aligned}$$

zerfällt. Die Gammaverteilung ist daher für unbekannte α und β ein Mitglied der Exponentialfamilie in zwei Parametern. \lrcorner

Exponentialfamilien sind von besonderem Interesse, da ihre Mitglieder einige spezielle Eigenschaften aufweisen. So halten Casella und Berger (2002) fest, dass für Erwartungswert und Varianz einer Zufallsvariablen y beziehungsweise eine Funktion $t_k(y)$ derselben aus der Exponentialfamilie

$$\mathbb{E}\left(\sum_{k=1}^K \frac{\partial w_k(\boldsymbol{\theta})}{\partial \theta_r} t_k(y)\right) = -\frac{\partial}{\partial \theta_r} \log c(\boldsymbol{\theta}) \quad (2.5)$$

sowie

$$\text{Var}\left(\sum_{k=1}^K \frac{\partial w_k(\boldsymbol{\theta})}{\partial \theta_r} t_k(y)\right) = -\frac{\partial^2}{\partial \theta_r^2} \log c(\boldsymbol{\theta}) - \mathbb{E}\left(\sum_{k=1}^K \frac{\partial^2 w_k(\boldsymbol{\theta})}{\partial \theta_r^2} t_k(y)\right) \quad (2.6)$$

gilt.

Exponentialfamilien in kanonischer Form

Kann eine Exponentialfamilie aus Darstellung (2.3) derart reparametrisiert werden, dass

$$f(y|\boldsymbol{\nu}) = h(y) c^*(\boldsymbol{\nu}) \exp\left(\sum_{k=1}^K \nu_k t_k(y)\right) \quad (2.7)$$

gilt, so ist von einer K -parametrischen Exponentialfamilie in kanonischer Form die Rede. Dabei handelt es sich bei den Funktionen $h(y)$ und $t_k(y)$ für $k = 1, \dots, K$ um dieselben Funktionen wie

in der originalen Parametrisierung (2.3). Der Parameter $\boldsymbol{\nu}$ wird dabei als kanonischer Parameter bezeichnet.

In weiterer Folge lässt sich der natürliche Parameterraum der Familie,

$$\mathcal{H} = \left\{ \boldsymbol{\nu} \mid \int_{-\infty}^{\infty} h(y) \exp\left(\sum_{k=1}^K \nu_k t_k(y)\right) dy < \infty \right\},$$

definieren. Casella und Berger (2002) halten fest, dass diese Menge konvex ist.

Beispiel: So lässt sich die Dichtefunktion der Gammaverteilung,

$$f(y|\alpha, \beta) = \frac{1}{y} \frac{1}{\Gamma(\alpha)\beta^\alpha} \exp\left(\alpha \log y - \frac{y}{\beta}\right),$$

derart reparametrisieren, dass sie einer Exponentialfamilie in kanonischer Form entspricht. Dafür sei

$$\nu_1 = \alpha \quad \text{und} \quad \nu_2 = -\frac{1}{\beta}.$$

Dann folgt für die Dichtefunktion

$$f(y|\alpha, \beta) = \frac{1}{y} \frac{-\nu_2^{\nu_1}}{\Gamma(\nu_1)} \exp(\nu_1 \log y + \nu_2 y)$$

die geforderte Form mit

$$c^*(\boldsymbol{\nu}) = \frac{-\nu_2^{\nu_1}}{\Gamma(\nu_1)}.$$

┘

Für Exponentialfamilien in kanonischer Form lassen sich für Erwartungswert und Varianz spezielle Formeln zur Berechnung angeben. Ausgehend von Darstellung (2.7) im kanonischen Parameter $\boldsymbol{\nu}$ lässt sich zeigen, dass für den Erwartungswert

$$E(t_k(y)) = -\frac{\partial}{\partial \nu_k} \log c^*(\boldsymbol{\nu}) \tag{2.8}$$

und für die Varianz

$$\text{Var}(t_k(y)) = -\frac{\partial^2}{\partial \nu_k^2} \log c^*(\boldsymbol{\nu}) \tag{2.9}$$

gilt. Diese Formeln sind zum Beispiel bei Lindsey (1996, S. 30) oder Witting (1985, Satz 1.164) angegeben oder lassen sich aus den oben genannten, aus Casella und Berger (2002, Theorem 3.4.2) entnommenen Formeln (2.5) und (2.6), herleiten.

Beispiel: Um Erwartungswert und Varianz der Normalverteilung für unbekanntes μ und σ^2 über (2.8) und (2.9) zu berechnen, wird die kanonische Parametrisierung dieser Exponentialfamilie benötigt. Diese ergibt sich, indem in der Darstellung als Exponentialfamilie,

$$f(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \exp\left(\frac{\mu}{\sigma^2} y - \frac{1}{2\sigma^2} y^2\right)$$

für die Reparametrisierung die Terme

$$\nu_1 = \frac{\mu}{\sigma^2} \quad \text{sowie} \quad \nu_2 = -\frac{1}{2\sigma^2}$$

gewählt werden. Damit läßt sich (2.4) derart umschreiben, dass die Exponentialfamilie in kanonischer Form

$$f(y|\boldsymbol{\nu}) = \sqrt{-\frac{\nu_2}{\pi}} \exp\left(\frac{\nu_1^2}{4\nu_2}\right) \exp\left(\nu_1 y + \nu_2 y^2\right)$$

vorliegt. Für die Funktion $c^*(\boldsymbol{\nu})$ rein in den Parametern und frei von y ergibt sich

$$c^*(\boldsymbol{\nu}) = \sqrt{-\frac{\nu_2}{\pi}} \exp\left(\frac{\nu_1^2}{4\nu_2}\right).$$

Daraus läßt sich nun der Erwartungswert von y über (2.8) berechnen und es gilt

$$E(y) = -\frac{\partial}{\partial \nu_1} \log\left(\sqrt{-\frac{\nu_2}{\pi}} \exp\left(\frac{\nu_1^2}{4\nu_2}\right)\right) = -\frac{\partial}{\partial \nu_1} \frac{\nu_1^2}{4\nu_2} = -\frac{2\nu_1}{4\nu_2} = -\frac{\frac{\mu}{\sigma^2}}{-2\frac{1}{2\sigma^2}} = \mu.$$

Ebenso läßt sich die Varianz errechnen, indem die zweite Ableitung von $-\log c^*(\boldsymbol{\nu})$,

$$\text{Var}(y) = -\frac{\partial^2}{\partial \nu_1^2} \log\left(\sqrt{-\frac{\nu_2}{\pi}} \exp\left(\frac{\nu_1^2}{4\nu_2}\right)\right) = -\frac{2}{4\nu_2} = \sigma^2,$$

betrachtet wird und es folgen die für die Normalverteilung bekannten Momente. \square

Lineare Exponentialfamilien

Für $K = 1$ läßt sich ein Mitglied der Exponentialfamilie in etwas anderer Schreibweise auch darstellen als

$$f(y|\theta) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right) \quad (2.10)$$

für bekannte Funktionen $a(\phi)$, $b(\theta)$ und $c(y, \phi)$, wobei $a(\phi) > 0$ gelten muss. Der Parameter ϕ wird in diesem Fall als bekannter Wert aufgefasst. Außerdem vereinfacht sich die Funktion $a(\phi)$ meistens zu $a \cdot \phi$ oder gar a für eine Konstante a . Eine Exponentialfamilie in dieser Darstellungsform wird als einparametrig, lineare Exponentialfamilie im kanonischen Parameter θ bezeichnet.

Beispiel: Die Wahrscheinlichkeitsfunktion der Poissonverteilung im Parameter λ läßt sich auf die in (2.10) erwähnte Form bringen, indem sie notiert wird als

$$\begin{aligned} f(y|\lambda) &= \exp(-\lambda) \frac{\lambda^y}{y!} = \exp(y \log(\lambda) - \lambda - \log y!) \\ &= \exp(y\theta - \exp(\theta) - \log y!) \quad \text{für } y = 0, 1, 2, \dots \end{aligned}$$

Diese Darstellung entspricht der geforderten Form, wobei $\theta = \log(\lambda)$ der kanonische Parameter ist. Somit folgt wegen $\lambda = \exp(\theta)$

$$b(\theta) = \exp(\theta).$$

Im Fall der Poissonverteilung gilt $a(\phi) = 1$ und in weiterer Folge $c(y) = -\log y!$. Daher ist die Poissonverteilung ein Mitglied der einparametrischen Exponentialfamilie im kanonischen Parameter $\log(\lambda)$. \lrcorner

Beispiel: Für festes σ^2 ist auch die Normalverteilung ein Mitglied der einparametrischen, linearen Exponentialfamilie. Dafür wird die Dichtefunktion zu

$$\begin{aligned} f(y|\mu, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{y^2 - 2y\mu + \mu^2}{2\sigma^2}\right) \\ &= \exp\left(\frac{y\mu - 1/2\mu^2}{\sigma^2} + \left(-\frac{1}{2}\log(2\pi\sigma^2) - \frac{y^2}{2\sigma^2}\right)\right). \end{aligned} \quad (2.11)$$

umgeformt. Wird nun $a(\phi) = \sigma^2$ als fest angenommen, so läßt sich für $\theta = \mu$ zeigen, dass die Normalverteilungsdichte ein Mitglied der einparametrischen Exponentialfamilie ist. Mit

$$a(\phi) = \sigma^2, \quad b(\theta) = \frac{\theta^2}{2} \quad \text{und} \quad c(y, \phi) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{y^2}{2\sigma^2}$$

ergibt sich für (2.11) die in (2.10) geforderte Darstellung. Somit ist auch die Normalverteilung in dieser Form ein Mitglied der einparametrischen Exponentialfamilie in kanonischer Form. \lrcorner

Beispiel: Ein letztes Beispiel soll nun noch zeigen, dass die Binomialverteilung für festes n ebenfalls ein Mitglied der Exponentialfamilie ist. Ihre Wahrscheinlichkeitsfunktion

$$P(y) = \binom{n}{y} \pi^y (1-\pi)^{n-y}$$

kann umgeformt werden zu

$$P(y) = \exp\left(y \log\left(\frac{\pi}{1-\pi}\right) + n \log(1-\pi) + \log\binom{n}{y}\right)$$

für absolute Häufigkeiten $y = 0, 1, 2, \dots, n$. Mit der Parametrisierung

$$\theta = \log\left(\frac{\pi}{1-\pi}\right) \quad \text{beziehungsweise} \quad \pi = \frac{\exp(\theta)}{1 + \exp(\theta)} \quad (2.12)$$

ergibt sich dann die in (2.10) geforderte Darstellung für bekanntes n , $a(\phi) = 1$ und $b(\theta) = n \log(1 + \exp(\theta))$. Dieser Zusammenhang folgt aus (2.12) und

$$-n \log(1-\pi) = -n \log\left(1 - \frac{\exp(\theta)}{1 + \exp(\theta)}\right) = -n \log(1 + \exp(\theta)).$$

Der Logarithmus des Binomialkoeffizienten hängt nur von y , nicht aber von θ ab und kann somit wie gefordert als $c(y, \phi)$ geschrieben werden.

Da es bei der Binomialverteilung später jedoch von Interesse sein wird, relative Häufigkeiten zu modellieren, soll hier auch die Wahrscheinlichkeitsfunktion einer standardisierten Binomialverteilung betrachtet werden, bei der nicht y selber binomialverteilt ist, sondern $ny \sim \text{Bin}(n, \pi)$ gilt. Für die relativen Häufigkeiten y gilt $y \in \{0, 1/n, 2/n, \dots, 1\}$ und die Wahrscheinlichkeitsfunktion läßt sich in der für eine Exponentialfamilie geforderten Darstellung schreiben als

$$\begin{aligned} P(ny) &= \binom{n}{ny} \pi^{ny} (1 - \pi)^{n - ny} \\ &= \exp \left(ny \log \left(\frac{\pi}{1 - \pi} \right) + n \log(1 - \pi) + \log \binom{n}{ny} \right) \\ &= \exp \left(\frac{y \log \left(\frac{\pi}{1 - \pi} \right) + \log(1 - \pi)}{1/n} + \log \binom{n}{ny} \right). \end{aligned}$$

In dieser Schreibweise gilt für θ ebenfalls wieder

$$\theta = \log \left(\frac{\pi}{1 - \pi} \right),$$

jedoch ist nun $a(\phi) = a \cdot \phi = 1/n \cdot 1$ und

$$b(\theta) = -\log(1 - \pi) = \log(1 + \exp(\theta))$$

für die beiden Funktionen $a(\phi)$ und $b(\theta)$. ┘

Wie bereits erwähnt ist die Exponentialfamilie deshalb so interessant, da sie einige besondere Eigenschaften besitzt. So gelten für die ersten beiden Momente einer wie in (2.10) definierten einparametrischen, linearen Exponentialfamilie

$$E(y) = b'(\theta) \tag{2.13}$$

und

$$\text{Var}(y) = a(\phi) b''(\theta). \tag{2.14}$$

Diese Resultate basieren auf den besonderen Eigenschaften der Ableitungen der logarithmierten Dichte- oder Wahrscheinlichkeitsfunktionen (d.h. der Score Funktionen, vgl. Abschnitt 2.2.2) von Mitgliedern der Exponentialfamilie, für die

$$E \left(\frac{\partial}{\partial \theta} \log f(y|\theta) \right) = 0 \tag{2.15}$$

und

$$\text{Var} \left(\frac{\partial}{\partial \theta} \log f(y|\theta) \right) = E \left(\left(\frac{\partial}{\partial \theta} \log f(y|\theta) \right)^2 \right) = E \left(-\frac{\partial^2}{\partial \theta^2} \log f(y|\theta) \right) \tag{2.16}$$

gilt. Ausgehend von der Tatsache, dass für die Dichte

$$\int_{-\infty}^{\infty} f(y|\theta) dy = 1$$

gelten muss sowie unter Verwendung der Kettenregel für die Ableitung der logarithmierten Dichtefunktion nach θ ,

$$\frac{\partial}{\partial \theta} \log f(y|\theta) = \frac{1}{f(y|\theta)} \frac{\partial}{\partial \theta} f(y|\theta),$$

lassen sich die beiden Aussagen (2.15) und (2.16) zeigen. So ergibt sich (2.15) über

$$\begin{aligned} E\left(\frac{\partial}{\partial \theta} \log f(y|\theta)\right) &= E\left(\frac{1}{f(y|\theta)} \frac{\partial}{\partial \theta} f(y|\theta)\right) \\ &= \int_{-\infty}^{\infty} \frac{\partial}{\partial \theta} f(y|\theta) dy \stackrel{(*)}{=} \frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} f(y|\theta) dy = \frac{\partial}{\partial \theta} 1 = 0. \end{aligned}$$

Der mit (*) markierte Schritt benötigt dabei die Vertauschbarkeit von Integration und Differentiation, welche zum Beispiel bei Casella und Berger (2002, Korollar 2.4.4) zu finden ist. Aus (2.15) und aus der Definition der Exponentialfamilie in (2.10) folgt somit

$$\begin{aligned} 0 &= E\left(\frac{\partial}{\partial \theta} \log f(y|\theta)\right) = E\left(\frac{\partial}{\partial \theta} \left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right)\right) = E\left(\frac{y - b'(\theta)}{a(\phi)}\right) \\ &= \frac{1}{a(\phi)} E(y - b'(\theta)), \end{aligned}$$

woraus sich in weiterer Folge $E(y) = b'(\theta)$ von (2.13) ergibt. Ebenso folgt daraus auch sofort die erste Gleichheit aus (2.16). Die zweite ergibt sich aus dem Zusammenhang

$$\begin{aligned} E\left(-\frac{\partial^2}{\partial \theta^2} \log f(y|\theta)\right) &= E\left(\frac{1}{f(y|\theta)^2} \left(\frac{\partial}{\partial \theta} f(y|\theta)\right)^2 - \frac{1}{f(y|\theta)} \frac{\partial^2}{\partial \theta^2} f(y|\theta)\right) \\ &= \int_{-\infty}^{\infty} \frac{1}{f(y|\theta)} \left(\frac{\partial}{\partial \theta} f(y|\theta)\right)^2 dy - \int_{-\infty}^{\infty} \frac{\partial^2}{\partial \theta^2} f(y|\theta) dy \\ &= \int_{-\infty}^{\infty} f(y|\theta) \left(\frac{\partial}{\partial \theta} \log f(y|\theta)\right)^2 dy - \frac{\partial^2}{\partial \theta^2} \int_{-\infty}^{\infty} f(y|\theta) dy \\ &= E\left(\left(\frac{\partial}{\partial \theta} \log f(y|\theta)\right)^2\right) - \frac{\partial^2}{\partial \theta^2} \cdot 1 = E\left(\left(\frac{\partial}{\partial \theta} \log f(y|\theta)\right)^2\right). \end{aligned}$$

Somit folgt für die Varianz einer einparametrischen Exponentialfamilie über

$$\begin{aligned} 0 &= E\left(\frac{\partial^2}{\partial \theta^2} \log f(y|\theta)\right) + E\left(\left(\frac{\partial}{\partial \theta} \log f(y|\theta)\right)^2\right) \\ &= E\left(\frac{\partial^2}{\partial \theta^2} \left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right)\right) + E\left(\left[\frac{\partial}{\partial \theta} \left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right)\right]^2\right) \\ &= -\frac{b''(\theta)}{a(\phi)} + \frac{1}{a^2(\phi)} E(y - b'(\theta))^2 = -\frac{b''(\theta)}{a(\phi)} + \frac{1}{a^2(\phi)} \text{Var}(y) \end{aligned}$$

der Zusammenhang $-a(\phi)b''(\theta) + \text{Var}(y) = 0$. Daher gilt für die Varianz einer Zufallsvariablen einer Exponentialfamilie wie in (2.14) angegeben $\text{Var}(y) = a(\phi)b''(\theta)$.

Festgehalten sei an dieser Stelle noch der Zusammenhang zwischen der sogenannten kumulantenerzeugenden Funktion $K(s)$ und der Funktion $b(\theta)$. Die kumulantenerzeugende Funktion ergibt sich aus der momenterzeugenden Funktion $M_y(s) = \text{E}(e^{sy})$ via

$$K(s) = \log M_y(s) = \log \text{E}(e^{sy}). \quad (2.17)$$

Von Interesse ist diese Funktion, da sie in Relation mit den Momenten der Verteilung steht. Um das k -te Moment aus der kumulantenerzeugenden Funktion zu errechnen, wird die k -te Kumulante κ_k benötigt. Diese ergibt sich aus der k -ten Ableitung von $K(s)$ nach s in $s = 0$,

$$\kappa_k = K^{(k)}(s)|_{s=0}.$$

Nun läßt sich zeigen, dass

$$\begin{aligned} \kappa_1 &= \text{E}(y) \\ \kappa_2 &= \text{E}(y - \text{E}(y))^2 \\ \kappa_3 &= \text{E}(y - \text{E}(y))^3 \\ \kappa_4 &= \text{E}(y - \text{E}(y))^4 - 3 (\text{E}(y - \text{E}(y))^2)^2 \\ \kappa_5 &= \text{E}(y - \text{E}(y))^5 - 10 \text{E}(y - \text{E}(y))^2 \text{E}(y - \text{E}(y))^3 \end{aligned}$$

gilt. Der Erwartungswert einer einparametrischen Exponentialfamilie ist somit gleich der ersten Ableitung der kumulantenerzeugenden Funktion in $s = 0$. Formal notiert bedeutet dies

$$\text{E}(y) = \kappa_1 = K'(0) = \frac{M'(0)}{M(0)}$$

für die momenterzeugende Funktion $M(s)$ von y in $s = 0$. Aufgrund dieses Zusammenhanges gilt für einparametrische Exponentialfamilien, dass $b(\theta)$ der kumulantenerzeugenden Funktion κ_1 der Verteilung entspricht. Auch die Varianz läßt sich über die kumulantenerzeugende Funktion errechnen, indem

$$\text{Var}(y) = \kappa_2 = K''(s)|_{s=0} = \frac{M''(0)}{M(0)} - \left(\frac{M'(0)}{M(0)} \right)^2$$

betrachtet wird.

Eine genauere Ausführung der Zusammenhänge zwischen den Momenten einer Verteilung und ihrer momenterzeugenden Funktion ist zum Beispiel bei Davison (2003, Kapitel 2.4) nachzulesen.

Mit der üblichen Schreibweise von $b'(\theta) = \text{E}(y) =: \mu$ läßt sich die Varianz somit als Produkt einer Funktion des Erwartungswerts μ und der von μ unabhängigen Funktion $a(\phi)$ schreiben als $\text{Var}(y) = a(\phi)V(\mu)$. Bei der Funktion $V(\mu) := b''(\theta)$ handelt es sich um die sogenannte Varianzfunktion, mithilfe derer sich die Varianz einer Verteilung als Funktion im Erwartungswert beschreiben läßt.

Somit wird es im generalisierten linearen Modell auch möglich, Daten zu beschreiben, deren Varianz nicht konstant im Erwartungswert ist.

Beispiel: Die ersten beiden Kumulanten der Poissonverteilung, die ja wie zuvor gezeigt ein Mitglied der Exponentialfamilie mit $\theta = \log(\lambda)$, $b(\theta) = \exp(\theta)$ und $a(\phi) = 1$ ist, lassen sich also errechnen als

$$E(y) = b'(\theta) = \frac{\partial}{\partial \theta} \exp(\theta) = \exp(\theta) = \lambda$$

und

$$\text{Var}(y) = a(\phi)b''(\theta) = 1 \cdot \frac{\partial^2}{\partial \theta^2} \exp(\theta) = \exp(\theta) = \lambda.$$

Auch alle höheren Momente der Poissonverteilung besitzen diese Eigenschaft, da für alle Kumulanten der Verteilung

$$\kappa_k = \lambda \quad \forall k = 1, 2, \dots$$

gilt. ┘

Beispiel: Für die Normalverteilung mit Parametrisierung als Exponentialfamilie in nur einem Parameter (siehe Darstellung in (2.11)) ergeben sich unter Verwendung von $a(\phi) = \sigma^2$, $b(\theta) = \theta^2/2$ und $\theta = \mu$ für Erwartungswert und Varianz

$$E(y) = b'(\theta) = \frac{\partial}{\partial \theta} \frac{\theta^2}{2} = \theta = \mu$$

und

$$\text{Var}(y) = a(\phi)b''(\theta) = \sigma^2 \cdot \frac{\partial^2}{\partial \theta^2} \frac{\theta^2}{2} = \sigma^2.$$

┘

Beispiel: In der Binomialverteilung mit $y \sim \text{Bin}(n, \pi)$ ergibt sich auf dieselbe Art und Weise aus $a(\phi) = 1$,

$$\theta = \log\left(\frac{\pi}{1-\pi}\right) \quad \text{beziehungsweise} \quad \pi = \frac{\exp(\theta)}{1 + \exp(\theta)}$$

und $b(\theta) = n \log(1 + \exp(\theta))$ für das erste Moment

$$E(y) = \frac{\partial}{\partial \theta} n \log(1 + \exp(\theta)) = n \frac{\exp(\theta)}{1 + \exp(\theta)} = n\pi,$$

das bekannte Ergebnis für den Erwartungswert einer binomialverteilten Zufallsvariablen. Ebenso läßt sich für die Varianz beobachten, dass

$$\text{Var}(y) = 1 \cdot \frac{\partial^2}{\partial \theta^2} n \log(1 + \exp(\theta)) = n \frac{\exp(\theta)}{(1 + \exp(\theta))^2} = n\pi(1 - \pi)$$

gilt. Natürlich halten diese Aussagen auch für die standardisierte Binomialverteilung mit der Parametrisierung $b(\theta) = \log \pi/(1-\pi)$ und $a(\phi) = 1/n$, und es ergeben sich die Momente $E(y) = \pi$ und $\text{Var}(y) = 1/n\pi(1 - \pi)$. ┘

2.2.2 Likelihood- und log-Likelihoodfunktionen

Für eine Stichprobe von N unabhängigen Zufallsvariablen y_1, \dots, y_N mit Dichte oder Wahrscheinlichkeitsfunktion $f(y_i|\theta_i)$ für $i = 1, \dots, N$ läßt sich die gemeinsame Dichte oder Wahrscheinlichkeitsfunktion dieser Zufallsvariablen schreiben als Produkt

$$f(y_1, \dots, y_N|\theta_1, \dots, \theta_N) = \prod_{i=1}^N f(y_i|\theta_i) \quad (2.18)$$

ihrer jeweiligen Dichte- beziehungsweise Wahrscheinlichkeitsfunktionen. Die Funktion der gemeinsamen Dichte oder Wahrscheinlichkeitsfunktion (2.18) für feste, bekannte Werte y_1, \dots, y_N und unbekannte Parameter $\theta_1, \dots, \theta_N$ wird als Likelihoodfunktion bezeichnet und als

$$L(\theta_1, \dots, \theta_N|y_1, \dots, y_N) = \prod_{i=1}^N f(y_i|\theta_i) \quad (2.19)$$

notiert. Definition (2.19) ist somit eine Funktion im Parametervektor $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)^t$, der in weiterer Folge geschätzt werden soll. Da das Produkt aus (2.18) schnell unhandlich werden kann, wird aus Gründen der Einfachheit anstelle der Likelihoodfunktion oft auch die logarithmierte Likelihoodfunktion

$$l(\boldsymbol{\theta}|\mathbf{y}) = \log L(\boldsymbol{\theta}|\mathbf{y}) = \log \prod_{i=1}^N f(y_i|\theta_i) = \sum_{i=1}^N \log f(y_i|\theta_i)$$

betrachtet.

Beispiel: Die Likelihoodfunktion von N unabhängig und identisch Poisson-verteilten Zufallsvariablen $y_i \stackrel{iid}{\sim} \text{Pois}(\lambda)$ läßt sich notieren als

$$L(\lambda|\mathbf{y}) = \prod_{i=1}^N \frac{e^{-\lambda} \lambda^{y_i}}{y_i!}.$$

Die log-Likelihoodfunktion derselben ist durch

$$l(\lambda|\mathbf{y}) = \sum_{i=1}^N \log \left(\frac{e^{-\lambda} \lambda^{y_i}}{y_i!} \right) = \sum_{i=1}^N (-\lambda + y_i \log \lambda - \log(y_i!))$$

gegeben. ┘

Beispiel: Für N unabhängig binomialverteilte Zufallsvariablen $y_i \stackrel{ind}{\sim} \text{Bin}(n_i, \pi_i)$ ergeben sich die Likelihood- und log-Likelihoodfunktion als

$$L(\boldsymbol{\pi}|\mathbf{y}) = \prod_{i=1}^N \binom{n_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{n_i - y_i}$$

und

$$l(\boldsymbol{\pi}|\mathbf{y}) = \sum_{i=1}^N \left(\log \binom{n_i}{y_i} + y_i \log(\pi_i) + (n_i - y_i) \log(1 - \pi_i) \right)$$

für Erfolgswahrscheinlichkeiten $\boldsymbol{\pi} = (\pi_1, \dots, \pi_N)^t$. Ausgehend von der Darstellung als Mitglied der Exponentialfamilie gilt für die log-Likelihoodfunktion

$$l(\boldsymbol{\pi}|\mathbf{y}) = \sum_{i=1}^N \left(y_i \log \frac{\pi_i}{1 - \pi_i} - n_i \log \frac{1}{1 - \pi_i} + \log \binom{n_i}{y_i} \right).$$

┘

Allgemein haben sich die Begriffe Scorefunktion für die erste Ableitung der log-Likelihoodfunktion nach dem Parametervektor $\boldsymbol{\theta}$,

$$s(\boldsymbol{\theta}) := \frac{\partial}{\partial \boldsymbol{\theta}} l(\boldsymbol{\theta}|\mathbf{y}),$$

und Informationsmatrix für den negativen Erwartungswert der Hessematrix der log-Likelihoodfunktion,

$$\mathcal{I}(\boldsymbol{\theta}) := -\mathbb{E} \left(\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^t} l(\boldsymbol{\theta}|\mathbf{y}) \right), \quad (2.20)$$

eingebürgert.

Ausgehend von den Likelihoodfunktionen läßt sich nun die Idee der Maximum Likelihood Schätzung erklären.

2.2.3 Maximum Likelihood Schätzung

Die Maximum Likelihood Schätzung ist eine der gängigsten Methoden zur Konstruktion von Punktschätzern. Das Prinzip beruht darauf, die Dichte- oder Wahrscheinlichkeitsfunktion der Stichprobe zu maximieren. Der Punktschätzer $\hat{\boldsymbol{\theta}}$ wird also derart gewählt, dass für die Likelihoodfunktion

$$L(\hat{\boldsymbol{\theta}}|\mathbf{y}) \geq L(\boldsymbol{\theta}|\mathbf{y}) \quad \forall \boldsymbol{\theta} \in \Theta$$

für alle $\boldsymbol{\theta}$ aus dem Parameterraum Θ der zulässigen Werte von $\boldsymbol{\theta}$ gilt. Für eine vorliegende Stichprobe y_1, \dots, y_N wird also jener Wert $\hat{\boldsymbol{\theta}}$ gesucht, der die Likelihoodfunktion und somit auch die Wahrscheinlichkeit für das Auftreten dieser Stichprobe maximiert. Für den gefundenen Punktschätzer $\hat{\boldsymbol{\theta}}$ ist das Auftreten der Stichprobe somit maximal plausibel (*likely*).

Der Wert für $\hat{\boldsymbol{\theta}}$ wird üblicherweise durch die Suche eines Maximums der log-Likelihoodfunktion über Quasi-Newton Verfahren beziehungsweise, falls die Funktion differenzierbar ist, durch suchen einer Nullstelle der Scorefunktion gefunden. Wichtig ist dabei, dass das Maximum im Inneren des Parameterraums auftritt. Eventuell wird dabei für das Auffinden des Maximums ein iteratives Verfahren wie jenes von Newton-Raphson, das Fisher-Scoring Verfahren oder der EM-Algorithmus benötigt.

Um die Suche nach dem Maximum zu vereinfachen, kann aufgrund der Monotonie des Logarithmus auch das Maximum der log-Likelihoodfunktion $\log L(\boldsymbol{\theta}|\mathbf{y}) = l(\boldsymbol{\theta}|\mathbf{y})$ berechnet werden.

Beispiel: Am Beispiel der Poissonverteilung von $y_i \stackrel{iid}{\sim} \text{Pois}(\lambda)$ für $i = 1, \dots, N$ sieht die Maximum Likelihood Schätzung wie folgt aus: Zu Beginn wird die log-Likelihoodfunktion

$$l(\lambda|\mathbf{y}) = \sum_{i=1}^N (-\lambda + y_i \log \lambda - \log(y_i!))$$

nach λ differenziert. Durch Null setzen und auflösen nach λ ergibt sich somit aus

$$\frac{\partial}{\partial \lambda} l(\lambda|\mathbf{y}) = -N + \sum_{i=1}^N \frac{y_i}{\lambda} \stackrel{!}{=} 0.$$

der Punktschätzer für den Parameter λ der Poissonverteilung als

$$\hat{\lambda} = \frac{1}{N} \sum_{i=1}^N y_i = \bar{y}.$$

Dass dieser Wert die Likelihoodfunktion maximiert, ergibt sich aus der zweiten Ableitung

$$\frac{\partial^2}{\partial \lambda^2} l(\lambda|\mathbf{y}) = -\frac{1}{\lambda^2} \sum_{i=1}^N y_i,$$

die für alle $\lambda > 0$ negativ ist, da $y_i \in \{0, 1, 2, \dots\}$. Somit maximiert der Schätzer $\hat{\lambda} = \bar{y}$ die Likelihoodfunktion im Parameter λ . ┘

Beispiel: Für N Zufallsvariablen $y_i \stackrel{iid}{\sim} \text{Bin}(n_i, \pi_i)$ liefert die Maximum Likelihood Schätzung ausgehend von

$$l(\boldsymbol{\pi}|\mathbf{y}) = \sum_{i=1}^N \left(\log \binom{n_i}{y_i} + y_i \log(\pi_i) + (n_i - y_i) \log(1 - \pi_i) \right)$$

$$\frac{\partial}{\partial \pi_i} l(\boldsymbol{\pi}|\mathbf{y}) = \frac{y_i}{\pi_i} - \frac{n_i - y_i}{1 - \pi_i} \stackrel{!}{=} 0 \quad i = 1, \dots, N.$$

In weiterer Folge ergibt sich für den Schätzer einer einzelnen Erfolgswahrscheinlichkeit $\hat{\pi}_i$

$$\frac{y_i}{\hat{\pi}_i} = \frac{n_i - y_i}{1 - \hat{\pi}_i}$$

$$\hat{\pi}_i = \frac{y_i}{n_i}$$

für $i = 1, \dots, N$. ┘

Der Maximum Likelihood Schätzer besitzt einige ausgezeichnete Eigenschaften, welche die Maximum Likelihood Schätzung zu einem beliebigen Werkzeug macht. So garantiert die Konstruktionsmethode, dass der Schätzer $\hat{\boldsymbol{\theta}}$ aus dem Bereich des Parameterraums Θ stammt. Der

Punktschätzer $\hat{\theta}$ ist weiters zumindest asymptotisch erwartungstreu, erfüllt also für große Stichproben $E(\hat{\theta}) = \theta$. Außerdem ist der Schätzer konsistent, konvergiert also für wachsenden Stichprobenumfang N in Wahrscheinlichkeit gegen den wahren Wert θ . Weiters läßt sich zeigen, dass der Maximum Likelihood Schätzer asymptotisch normalverteilt ist, also $\hat{\theta} \stackrel{as}{\sim} N(\theta, \mathcal{I}^{-1})$ erfüllt und die unter (2.20) definierte Informationsmatrix als Grenzvarianzkovarianz besitzt. Somit ist der Schätzer auch asymptotisch effizient. Die Varianzkovarianzmatrix von $\hat{\theta}$ erreicht also asymptotisch die Cramér-Rao Schranke und $\hat{\theta}$ besitzt daher die minimal mögliche Varianz, die ein Schätzer $\hat{\theta}$ überhaupt erreichen kann.

Für die Parameterschätzung mittels Maximum Likelihoodfunktion im Fall von generalisierten linearen Modellen existieren Verfahren, welche speziell auf die Eigenschaften einer einparametrischen, linearen Exponentialfamilie zugeschnitten sind. In den bereits genannten Werken über generalisierte lineare Modelle wird beschrieben, warum im Fall dieser Modelle iterative Methoden für die Parameterschätzung verwendet werden müssen. Ausgangspunkt für die iterativen Optimierungsverfahren ist wie in vielen anderen Fällen auch ein Ansatz ähnlich dem Newton-Verfahren. Ausgehend davon läßt sich für generalisierte lineare Modelle das sogenannte Newton-Raphson-Verfahren, auch bekannt als iterativ gewichtete kleinste Quadrate Methode, oder die Fisher-Scoring Methode entwickeln.

In Anbetracht der Tatsache, dass weder die Newton-Raphson noch die Fisher-Scoring Technik in dieser Arbeit für die Parameterschätzung verwendet werden, werden diese hier auch nicht weiter diskutiert. Bei Bedarf sind sie zum Beispiel in Dobson und Barnett (2008) oder Lindsey (1996) nachzulesen.

Auf jeden Fall können mit der Maximum Likelihood Methode — sofern sie existieren — Schätzer für den Parameter β ermittelt werden. Weiters gilt, dass dieses $\hat{\beta}$ zumindest asymptotisch erwartungstreu ist und wie oben erwähnt asymptotisch auch

$$\text{Var}(\hat{\beta}) = \mathcal{I}^{-1}$$

erfüllt. Daraus ergibt sich wiederum asymptotisch

$$(\hat{\beta} - \beta)^t \mathcal{I}(\hat{\beta})(\hat{\beta} - \beta) \sim \chi^2(p)$$

beziehungsweise $\hat{\beta} \sim N(\beta, \mathcal{I}^{-1})$.

2.2.4 Tests auf Güte der Modellanpassung

Im Fall des linearen Regressionsmodells kann aufgrund der Annahme der Normalverteilung aus der Quadratsummenzerlegung das Bestimmtheitsmaß R^2 errechnet werden. Dieses kann für Aussagen über die Güte der Anpassung des Modells herangezogen werden. Im Fall des generalisierten linearen Modells ist diese Zerlegung jedoch nicht mehr zulässig und es müssen andere Möglichkeiten in Betracht gezogen werden, um einen Test für die Güte eines Modells zu finden.

Eine Möglichkeit ist dabei, das aktuelle Modell mit einem Modell zu vergleichen, welches den Daten uneingeschränkte Freiheit gestattet, also einem Modell mit maximaler Anzahl an Parametern. Das bedeutet, jeder Beobachtung y_i wird ein eigenes μ_i gestattet. Dieses oft als gesättigtes oder volles Modell bezeichnete Modell ist dabei ebenfalls wieder ein generalisiertes lineares Modell mit derselben Linkfunktion wie das aktuell betrachtete Modell, diesmal jedoch

in N Parametern, wohingegen das aktuelle Modell mit nur p Parametern auskommt. Die beiden Modelle lassen sich als die Hypothesen

$$H_0 : \mu_i = g^{-1}(\eta_i) \quad \text{für } i = 1, \dots, N$$

sowie $H_A : \mu_1, \dots, \mu_N$ beliebig,

formulieren, wobei η_i den linearen Prädiktor $\mathbf{x}_i^t \boldsymbol{\beta}$ bezeichne. Um nun zu testen, ob die Daten durch das unter H_0 definierte Modell angemessen beschrieben können, kann ein Likelihood-Quotiententest verwendet werden. Dafür sei $\hat{\boldsymbol{\mu}}$ der Maximum Likelihood Schätzer des Erwartungswerts unter H_0 . Unter der Alternativhypothese H_A bekommt jedes y_i seinen eigenen Schätzer, also gilt in diesem Fall $\hat{\mu}_i = y_i$ für $i = 1, \dots, N$. Dann läßt sich anhand des Verhältnisses der beiden maximierten Likelihoodfunktionen,

$$\lambda = \frac{L(\hat{\boldsymbol{\mu}}|\mathbf{y})}{L(\mathbf{y}|\mathbf{y})},$$

auf die Anpassungsgüte des betrachteten Modells schließen.

Aufgrund der Tatsachen, dass für die beiden unter H_0 und H_A maximierten Likelihoodfunktionen $L(\boldsymbol{\theta}|\mathbf{y}) > 0$ und $L(\hat{\boldsymbol{\theta}}|\mathbf{y}) \leq L(\mathbf{y}|\mathbf{y})$ gilt, lassen sich aus dem Quotienten Aussagen über die Anpassungsgüte des vorliegenden Modells folgern. So ist aufgrund der genannten Aussagen generell $0 < \lambda \leq 1$. Ist nun das aktuelle Modell jenem aus der Alternativhypothese nur geringfügig überlegen, so sind die Werte der beiden Likelihoodfunktionen ähnlich und λ wird nur unerheblich kleiner eins sein. Ein Wert deutlich kleiner eins deutet hingegen darauf hin, dass das aktuelle Modell gegenüber dem Modell aus der Alternativhypothese sehr wohl eine entscheidende Datenreduktion erlaubt.

In der Praxis wird statt dieser Statistik jedoch eher

$$-2 \log \lambda = 2 (l(\mathbf{y}|\mathbf{y}) - l(\hat{\boldsymbol{\mu}}|\mathbf{y})) = 1/\phi D(\hat{\boldsymbol{\mu}}|\mathbf{y}) \quad (2.21)$$

betrachtet, die doppelte Differenz der log-Likelihoodfunktionen des vollen sowie des aktuellen Modells. Da $l(\hat{\boldsymbol{\mu}}|\mathbf{y}) \leq l(\mathbf{y}|\mathbf{y})$ gilt, ist diese Differenz von log-Likelihoodfunktionen nicht-negativ.

Es läßt sich zeigen, dass die Verteilung dieser Statistik zumindest unter gewissen Regularitätsbedingungen asymptotisch durch jene der χ^2 -Verteilung angenähert werden kann. Der Freiheitsgrad der Verteilung entspricht dabei genau $N - p$, der durch das aktuelle Modell erfahrenen Variablenreduktion im Vergleich zum vollen Modell. Für diese Teststatistik haben Nelder und Wedderburn (1972) in ihrer Arbeit den Begriff der skalierten Devianz geprägt.

Aufgrund der approximativen Nähe zur χ^2 -Verteilung sollte die Devianz unter den oben erwähnten Regularitätsbedingungen bei einem guten Modell in etwa der Anzahl der Freiheitsgrade entsprechen. Deutliche Abweichungen von diesem Wert weisen auf ein unangemessenes Modell hin, welches unter Umständen noch verbessert werden kann. Eine genauere Diskussion dieser Fälle wird in Abschnitt 2.2.6 begonnen.

Wie bereits erwähnt wird die mit $1/\phi$ multiplizierte Devianz als skalierte Devianz bezeichnet. Außerdem hat sich für den Ausdruck $-2l(\hat{\boldsymbol{\theta}}|\mathbf{y})$ der Begriff Disparität eingebürgert. Unter Berücksichtigung aller Konstanten einer Likelihoodfunktion lassen sich mit Hilfe der Disparität laut Aitkin et al. (2005) selbst nicht genestete Modelle miteinander vergleichen.

Beispiel: Für eine Stichprobe von N unabhängig Poisson-verteilten Beobachtungen y_i mit Parametern λ_i gilt für die log-Likelihoodfunktion

$$l(\boldsymbol{\lambda}|\mathbf{y}) = \sum_{i=1}^N (-\lambda_i + y_i \log \lambda_i - \log y_i!).$$

Für die Punktschätzer $\hat{\boldsymbol{\lambda}}$ eines Modells $\log \boldsymbol{\lambda} = \mathbf{X}\boldsymbol{\beta}$ in p Parametern β_j folgt für die log-Likelihoodfunktion weiters

$$l(\hat{\boldsymbol{\lambda}}|\mathbf{y}) = \sum_{i=1}^N (-\hat{\lambda}_i + y_i \log \hat{\lambda}_i - \log y_i!).$$

In einem vollen Modell, in dem jedes y_i durch einen eigenen Parameter beschrieben wird, gilt für die Maximum Likelihood Schätzer $\hat{\lambda}_i = y_i$. Daher ergibt sich für die log-Likelihoodfunktion in diesem Fall

$$l(\mathbf{y}|\mathbf{y}) = \sum_{i=1}^N (-y_i + y_i \log y_i - \log y_i!).$$

Aus diesen beiden log-Likelihoodfunktionen erhält man

$$\begin{aligned} D(\hat{\boldsymbol{\lambda}}|\mathbf{y}) &= 2 \left[\sum_{i=1}^N (-y_i + y_i \log y_i - \log y_i!) - \sum_{i=1}^N (-\hat{\lambda}_i + y_i \log \hat{\lambda}_i - \log y_i!) \right] \\ &= 2 \left[\sum_{i=1}^N y_i \log \frac{y_i}{\hat{\lambda}_i} - \sum_{i=1}^N (y_i - \hat{\lambda}_i) \right], \end{aligned}$$

die Devianz im Fall eines loglinearen Poissonmodells. Da zumindest für Modelle mit Intercept jedoch weiters

$$\sum_{i=1}^N y_i = \sum_{i=1}^N \hat{\lambda}_i$$

gilt, reduziert sich die Devianz dann zu

$$D(\hat{\boldsymbol{\lambda}}|\mathbf{y}) = 2 \sum_{i=1}^N y_i \log \frac{y_i}{\hat{\lambda}_i}.$$

┘

Beispiel: Im Fall einer Normalverteilung mit $y_i \stackrel{ind}{\sim} N(\mu_i, \sigma^2)$ gilt für die dazugehörige log-Likelihoodfunktion

$$l(\boldsymbol{\mu}, \sigma^2|\mathbf{y}) = -\frac{N}{2} \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - \mu_i)^2,$$

woraus sich für die Devianz folgender Ausdruck ergibt

$$D(\hat{\boldsymbol{\mu}}, \sigma^2 | \mathbf{y}) = 2[l(\mathbf{y} | \mathbf{y}) - l(\hat{\boldsymbol{\mu}} | \mathbf{y})] = \frac{1}{\sigma^2} \sum_{i=1}^N (y_i - \hat{\mu}_i)^2.$$

In diesem Fall entspricht die Devianz somit genau der mit $1/\sigma^2$ skalierten Fehlerquadratsumme $SSE(\hat{\boldsymbol{\mu}})$, was exakt einer χ^2 -Verteilung entspricht. \lrcorner

Beispiel: Wie die vorangegangenen Abschnitte soll auch dieser mit einem Beispiel für die Binomialverteilung abschließen, diesmal jedoch für den standardisierten Fall mit $y_i | n_i \stackrel{\text{ind}}{\sim} \text{Bin}(n_i, \pi_i)$. In Abschnitt 2.2.2 wurde die log-Likelihoodfunktion der Binomialverteilung als

$$l(\boldsymbol{\pi} | \mathbf{y}) = \sum_{i=1}^N \left(n_i y_i \log \frac{\pi_i}{1 - \pi_i} + n_i \log (1 - \pi_i) + \log \binom{n_i}{n_i y_i} \right)$$

definiert. Zur Berechnung der Devianz werden nun die log-Likelihood Funktionen des vollen Modells mit N Parametern sowie eines Modells mit $p < N$ Parametern betrachtet. Die Differenz $D(\hat{\boldsymbol{\pi}} | \mathbf{y}) = 2(l(\mathbf{y} | \mathbf{y}) - l(\hat{\boldsymbol{\pi}} | \mathbf{y}))$ der beiden log-Likelihoodfunktionen liefert dann

$$D(\hat{\boldsymbol{\pi}} | \mathbf{y}) = 2 \left[\sum_{i=1}^N \left(n_i y_i \log \frac{y_i}{1 - y_i} + n_i \log (1 - y_i) + \log \binom{n_i}{n_i y_i} \right) - \sum_{i=1}^N \left(n_i y_i \log \frac{\hat{\pi}_i}{1 - \hat{\pi}_i} + n_i \log (1 - \hat{\pi}_i) + \log \binom{n_i}{n_i y_i} \right) \right],$$

woraus sich in weiterer Folge durch Vereinfachen

$$D(\hat{\boldsymbol{\pi}} | \mathbf{y}) = 2 \sum_{i=1}^N \left(n_i y_i \log \frac{y_i}{\hat{\pi}_i} + n_i (1 - y_i) \log \frac{1 - y_i}{1 - \hat{\pi}_i} \right)$$

für die Devianz eines Modells mit standardisiert binomialverteilten Daten ergibt. Diese Devianz ist selbst für $y_i \in \{0, 1\}$ definiert, wobei dann $-n_i \log(1 - \hat{\pi}_i)$ im Fall $y_i = 0$ beziehungsweise $-n_i \log \hat{\pi}_i$ für $y_i = 1$ in die Summe eingeht. \lrcorner

Als Alternative zur Devianz kann auch die Pearson χ^2 Statistik X^2 verwendet werden. Diese ist definiert als die Summe über die Quadrate der Pearson Residuen,

$$X^2 = \sum \frac{(o - e)^2}{e},$$

wobei o die beobachteten und e die erwarteten Häufigkeiten bezeichne. Es läßt sich zeigen, dass diese Statistik asymptotisch äquivalent zur Devianz ist. Daraus folgt, dass auch die X^2 Teststatistik unter gewissen Regularitätsbedingungen $X^2 \sim \chi_{N-p}^2$ erfüllt. Vor- und Nachteile der beiden Statistiken X^2 und D werden zum Beispiel bei Dobson und Barnett (2008) besprochen.

Weiters soll noch das auf der log-Likelihoodfunktion basierende Akaike Informationskriterium (AIC) genannt werden, welches ebenfalls zum Vergleich von Modellen verwendet werden kann. Dieses von Akaike (1973) vorgestellte Informationskriterium ist definiert als

$$\text{AIC} = -2 \cdot \left(l(\hat{\boldsymbol{\theta}}|\mathbf{y}) - p \right), \quad (2.22)$$

wobei p die Anzahl der Modellparameter sei. Es handelt sich also um eine durch die doppelte Anzahl der Parameter bestrafte Form der Disparität. Diese Statistik eignet sich besonders gut dafür, genestete Modelle miteinander zu vergleichen.

2.2.5 Das logistische Regressionsmodell

Sollen binomialverteilte Daten mittels eines generalisierten linearen Modells beschrieben werden, so spricht man gerne von einem logistischen Modell. Es soll hier besonders hervorgehoben werden, da die Daten aus dem praktischen Teil der Arbeit (siehe Kapitel 6) mit einem derartigen Modell beschrieben werden.

Ein logistisches Regressionsmodell kommt zum Einsatz, wenn die Verteilung der Responsevariablen einer Binomialverteilung folgt. Wie bereits in Abschnitt 2.2.1 gezeigt, ist die klassische Binomialverteilung für festen Gruppenumfang n ein Mitglied der Exponentialfamilie. So ist die Wahrscheinlichkeitsfunktion N binomialverteilter Zufallsvariablen y_i mit Parametern π_i für $i = 1, \dots, N$ und konstanter Gruppengröße n von der Form

$$f(\mathbf{y}|\boldsymbol{\pi}) = \sum_{i=1}^N \left(\binom{n}{y_i} \pi_i^{y_i} (1 - \pi_i)^{n-y_i} \right)$$

und die gemeinsame log-Likelihoodfunktion der standardisierten Häufigkeiten y_i (also $y_i \in \{0, 1/n, 2/n, \dots, 1\}$ mit $ny_i \sim \text{Bin}(n, \pi_i)$) läßt sich notieren als

$$l(\boldsymbol{\pi}|\mathbf{y}) = \sum_{i=1}^N \left(\frac{y_i \log \frac{\pi_i}{1-\pi_i} + \log(1 - \pi_i)}{\frac{1}{n}} + \log \binom{n}{ny_i} \right).$$

Wie bereits beschrieben gilt für den Erwartungswert und die Varianz von y_i in diesem Fall

$$\text{E}(y_i) = \frac{\exp(\theta_i)}{1 + \exp(\theta_i)} = \pi_i$$

und

$$\text{Var}(y_i) = \frac{1}{n} \frac{\exp(\theta_i)}{(1 + \exp(\theta_i))^2} = \frac{1}{n} \pi_i (1 - \pi_i).$$

Linkfunktionen

Das Modell soll nun die Erfolgswahrscheinlichkeit π_i als Funktion im linearen Prädiktor beschreiben.

Hier wird auch deutlich, warum ein klassisches lineares Regressionsmodell nicht angebracht wäre. Der lineare Prädiktor $\mathbf{x}^t \boldsymbol{\beta}$ kann auf den ganzen reellen Zahlen realisieren, π_i hingegen ist eine Wahrscheinlichkeit und darf somit nur zwischen 0 und 1 liegen. Daher wird hier eine Funktion des Erwartungswerts und nicht der Erwartungswert direkt modelliert. Durch geschickte

Wahl dieser Funktion, der sogenannten Linkfunktion, lässt sich der Wertebereich von π_i auf das Einheitsintervall einschränken.

Genau genommen wird eine Funktion benötigt, die monoton und differenzierbar ist. Außerdem soll sie im Fall des logistischen Modells die reellen Zahlen auf das Einheitsintervall abbilden, also $g : \mathbb{R} \rightarrow [0, 1]$ erfüllen. Diese Anforderung wird prinzipiell von jeder stetigen Verteilungsfunktion erfüllt. Die weitaus gängigste Wahl ist jedoch jene der kanonischen Linkfunktion, welche $g(\mu_i) = \theta_i$ erfüllt. Wegen $\mu_i = b'(\theta_i)$ für den Erwartungswert μ_i folgt für die Linkfunktion somit die Äquivalenz $g(\mu_i) = (b'(\mu_i))^{-1}$. Im Fall des logistischen Regressionsmodells erfüllt die logit-Funktion

$$\theta_i = \log\left(\frac{\pi_i}{1 - \pi_i}\right) = \eta_i = \mathbf{x}_i^t \boldsymbol{\beta} \quad (2.23)$$

diese Anforderung. Theoretisch können auch andere Funktionen, welche die Voraussetzungen erfüllen, in Betracht gezogen werden. So zeigt Tabelle 2.1 zwei weitere Linkfunktionen, die im Fall des logistischen Regressionsmodells üblich sind. Abbildung 2.1 jedoch lässt erkennen, wie geringfügig die Unterschiede zwischen den einzelnen Linkfunktionen sind. Daher wird für die in dieser Arbeit diskutierten Modelle das Augenmerk auf den kanonischen Link gelegt und keine anderen Linkfunktionen in Betracht gezogen.

Linkfunktion	Formel
logit	$\exp(\eta)/(1+\exp(\eta))$
probit	$\Phi(\eta)$
cloglog	$1 - \exp(-\exp(\eta))$

Tabelle 2.1.: Verschiedene Linkfunktionen des binomialen Regressionsmodells.

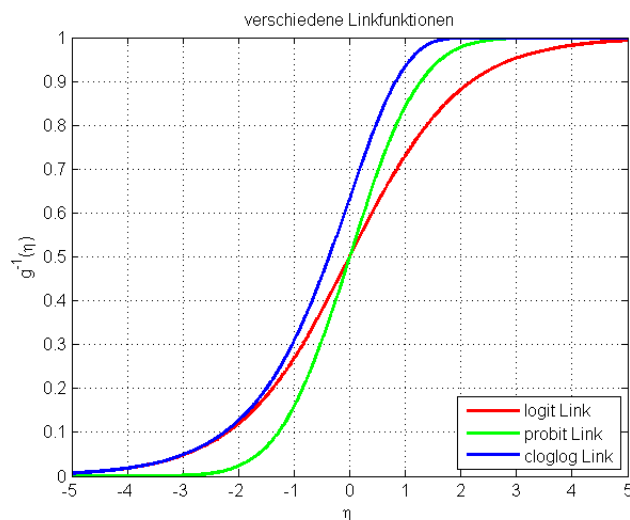


Abbildung 2.1.: Visualisierung der in Tabelle 2.1 aufgelisteten Linkfunktionen.

Ein Beispiel

Zur Abrundung der Theorie der generalisierten linearen Modelle soll an dieser Stelle ein Beispiel eines logistischen Regressionsmodells betrachtet werden. Die konkrete Anwendung auf die bereits erwähnten Datensätze aus der Papier- und Druckindustrie wird in Kapitel 6 besprochen.

Beispiel: Für dieses Beispiel wird ein Datensatz aus Dobson und Barnett (2008) herangezogen, welcher die Sterblichkeit von Käfern beschreibt, die unterschiedlichen Dosierungen eines Gifts ausgesetzt sind. Die genauen Daten sind in Tabelle 2.2 aufgelistet. Wie aus dieser Tabelle abzulesen ist, variieren die Größen n_i der $N = 8$ Gruppen auf den unterschiedlichen Stufen der Dosierung. Die Anzahl y_i an verstorbenen Tieren innerhalb einer Gruppe wird als binomialverteilt in den Parametern n_i und π_i angenommen.

$\log_2(\text{Dosis})$	Tiere	Verstorben
1,6907	59	6
1,7242	60	13
1,7552	62	18
1,7842	56	28
1,8113	63	52
1,8369	59	53
1,8610	62	61
1,8839	60	60

Tabelle 2.2.: Datensatz für logistische Regression aus Dobson und Barnett (2008).

Um nun die Wahrscheinlichkeit für das Verenden eines Tieres in Abhängigkeit der Dosis d_i des Gifts zu beschreiben, wird ein logistisches Regressionsmodell verwendet. Das Modell

$$\eta_i = \beta_0 + \beta_1 \cdot d_i$$

geht über den logit-Link

$$\log \frac{\pi_i}{1 - \pi_i} = \eta_i$$

in die Parameterschätzung ein. Die Parameter β_0 und β_1 werden also mithilfe der log-Likelihoodfunktion

$$l(\boldsymbol{\beta}|\mathbf{y}) = \sum_{i=1}^8 \left(y_i (\beta_0 + \beta_1 d_i) - n_i \log (1 + \exp (\beta_0 + \beta_1 d_i)) + \log \binom{n_i}{y_i} \right)$$

geschätzt und es ergeben sich für die Punktschätzer die Werte (und Standardfehler)

$$\hat{\beta}_0 = -60,72 \quad (se = 5,18)$$

und $\hat{\beta}_1 = 34,27 \quad (se = 2,91).$

Das Modell besitzt eine Devianz von 11,23 bei 6 Freiheitsgraden. Die Devianz des Nullmodells liegt hingegen bei 284. Das gefundene Modell ist also durchaus in der Lage,

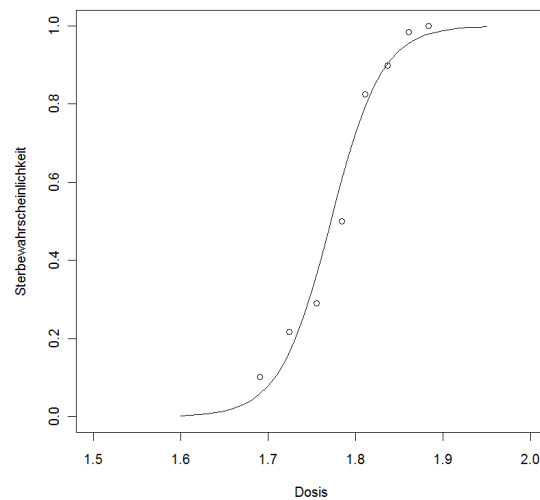


Abbildung 2.2.: Visualisierung des Zusammenhangs zwischen der Sterbewahrscheinlichkeit von Versuchstieren und der Dosierung des Gifts.

den vorliegenden Zusammenhang zwischen der Dosierung des Gifts und der Sterbewahrscheinlichkeit zu beschreiben.

Die Höhe der Dosierung spielt in diesem Beispiel also eine entscheidende Rolle, wenn es um die Überlebens- beziehungsweise Sterberate der Tiere geht. Ein Zusammenhang, der auch in Abbildung 2.2 zur Geltung kommt. Diese zeigt den Zusammenhang zwischen der Dosis und der Sterbewahrscheinlichkeit, wobei die beobachteten relativen Häufigkeiten der Werte aus Tabelle 2.2 ebenso eingezeichnet sind wie die resultierende Modellschätzung (durchgezogene Linie).

┘

2.2.6 Über- und Unterdispersion im logistischen Modell

Im Fall eines logistischen Regressionsmodells wird der Erwartungswert μ (bzw. π) in Abhängigkeit des linearen Prädiktors beschrieben. Für die Variabilität der vorliegenden Daten wird angenommen, dass für die absoluten Häufigkeiten

$$\text{Var}(y) = n\pi(1 - \pi)$$

gilt. Diese Annahme wird von tatsächlich binomialverteilten Zufallsvariablen erfüllt, kann in der Praxis jedoch unangemessen sein. Mitunter kann es vorkommen, dass die Variabilität der Responsevariablen eines Datensatzes deutlich kleiner oder größer ist als die vom Modell angenommene. In diesen Fällen spricht man von Unter- beziehungsweise Überdispersion.

Hinweise auf das Vorliegen von Unter- oder Überdispersion kann zum Beispiel die Devianz eines Modells liefern, welche unter Regularitätsbedingungen zumindest asymptotisch χ^2 -verteilt ist. Daher kann eine Devianz deutlich größer der Anzahl der Freiheitsgrade ein Hinweis darauf sein, dass die Variabilität in den Daten größer ist als vom Modell angenommen.

Überdispersion

Von Überdispersion wird gesprochen, wenn die Variabilität in den vorliegenden Daten größer ist als die vom Modell angenommene Varianz der Binomialverteilung. Dieser Fall ist in der Realität nicht selten zu finden und kann verschiedene Ursachen haben. Zu den wichtigsten gehören laut Hardy (2002)

1. das Fehlen einer oder mehrerer wichtiger erklärender Größen;
2. das Vorliegen von Ausreißern;
3. eine unangemessene Wahl der Linkfunktion oder
4. ein kleiner Datensatz, bei dem ein Vergleich mit der χ^2 -Verteilung nicht angemessen ist.

Können diese Fälle jedoch ausgeschlossen werden, schreibt Hardy weiters, so sind am wahrscheinlichsten Abhängigkeiten zwischen den einzelnen Beobachtungen Schuld an der vorliegenden Überdispersion.

Eine einfache Möglichkeit, auf eine vorliegende Überdispersion einzugehen, wäre, anstelle der Varianz aus der klassischen Binomialverteilung ein Skalierungsfaktor ψ einzuführen und statt $n\pi(1 - \pi)$ mit einer Varianz der Art

$$\text{Var}(y) = \psi \cdot n\pi(1 - \pi)$$

zu arbeiten. Diese Vorgehensweise führt zur Theorie der Quasi-Likelihood Schätzung, welche erstmals von Wedderburn (1974) beschrieben wurde und zum Beispiel von Annis (2007) oder Engel und te Brake (1993) zur Überwindung der Überdispersion verwendet wird.

Alternativ kann auch eine andere Verteilung aus der Exponentialfamilie zur Beschreibung der Daten verwendet werden. So zogen in der Vergangenheit viele Autoren die negative Binomialverteilung oder die Beta-Binomialverteilung als Alternativen zur Binomialverteilung heran, wenn es sich um einen Datensatz mit Überdispersion handelte. Siehe dazu Lindsey und Altham (1998) und ihre Untersuchungen am Datensatz von Kindern aus sächsischen Großfamilien im 19. Jahrhundert von Geissler oder die Arbeiten von Aitkin und Clayton (1980) sowie Kupper und Haseman (1978).

Kann davon ausgegangen werden, dass das Fehlen wichtiger erklärender Größen für die Überdispersion verantwortlich ist, bieten sich weiters sogenannte Random Effect Modelle an, um das vorliegende Problem der Überdispersion zu überwinden. Diese Modelle werden unter anderem in Aitkin et al. (2005), Kapitel 8, ausführlich beschrieben. Anwendungsfälle sind zum Beispiel Klaas, Enevoldsen, Ersbøll und Tölle (2005), Krackow und Tkadlec (2001) oder Atwill, Mohammed, Scarlett und McCulloch (1995), deren Untersuchungen sich dabei von der Betrachtung von Risikofaktoren bei Milchkühen über Geschlechterverhältnisse bei Geschwistern bis zu Virusinfektionen bei Hühnern und Welpen spannen, dabei jedoch alle aus dem veterinären beziehungsweise humanmedizinischen Bereich stammen.

Wie aus den Quellen aus der Literatur zu erkennen ist, tritt Überdispersion vor allem in biologischen Untersuchungen auf. Ein häufiger Grund dafür sind Verwandtschaftsverhältnisse zwischen den untersuchten Individuen, wenn zum Beispiel Tiere oder Würfe teils verschiedener Muttertiere betrachtet werden. In solchen Fällen muss klarerweise davon ausgegangen werden, dass Geschwister gewisse Gemeinsamkeiten aufweisen und auf diese Verwandtschaftsbeziehungen in einer Modellierung nicht verzichtet werden sollte.

Ein weiteres Beispiel für Überdispersion im Vergleich zur Poissonverteilung ist in dem Buch über Parasitologie von Cox (1993) zu finden, der die Anzahl von Parasiten in Wirten betrachtet. Dabei tritt Überdispersion häufig auf, da Parasiten üblicherweise innerhalb der Wirtpopulation in Gruppen auftreten. So beherbergen viele Wirte meist wenig Parasiten und nur wenige Wirte sind von vielen Parasiten befallen.

In der Regel kann Überdispersion als extrem große Heterogenität einer Population in Hinsicht auf das betrachtete Merkmal verstanden werden. Eine weitere Möglichkeit, Überdispersion zu beschreiben ist, sie als Resultat eines ansteckenden Prozesses zu sehen.

Unterdispersion

Unterdispersion, also Variabilität deutlich kleiner als im Fall der klassischen Binomialverteilung, tritt deutlich seltener auf als ihr Gegenteil. Viele Arbeiten beschränken sich daher rein auf die Diskussion der Überdispersion und lassen die Unterdispersion beiseite.

Einzelne Beispiele lassen sich jedoch abermals in der Biologie finden. So schreibt Hardy (2002), dass die Ausprägung des Geschlechts bei Hautflüglern (Wespen, Ameisen, Bienen,...), Spinnmilben oder Rädertierchen häufig zu einer Variabilität kleiner jener der Binomialverteilung führt. Der Grund dafür liegt in den Fortpflanzungsverhalten dieser haplodiploiden¹ Lebewesen, welche auf extrem ungleichmäßige Geschlechterverhältnisse führen.

Auf das Beispiel der Verteilung von Parasiten in Wirten aus dem bereits erwähnten Werk von Cox (1993) zurückkommend kann Unterdispersion durch eine Immunisierung der Wirte gegen den Parasiten oder eine höhere Sterberate der Parasiten erklärt werden.

Unterdispersion kann als abstoßender Effekt interpretiert werden, bei dem eine positive Ausprägung weitere positive Fälle deutlich seltener werden läßt. Ein solcher Fall wird in King (1989) erwähnt und konnte 2009/10 auch in Österreich beobachtet werden, als nach Bekanntwerden der Wiederkandidatur des Bundespräsidenten für das höchste Amt im Staat Politiker anderer Parteien ihre Kandidaturen zurückzogen.

Generell tritt Unterdispersion in Fällen auf, in denen große Homogenität vorherrscht. So wie Überdispersion als Resultate eines ansteckenden Effekts beschrieben werden kann, läßt sich Unterdispersion als abstoßender Effekt ansehen.

Die Proben des in Kapitel 6 zur Betrachtung kommenden Datensatzes weisen jedoch sowohl Über- als auch Unterdispersion auf. Daher werden für eine angemessene Modellierung dieser Situationen in den kommenden beiden Kapiteln zwei Verteilungsfamilien vorgestellt, welche sowohl mit Über- als auch Unterdispersion im Vergleich zur klassischen Binomialverteilung umgehen können.

¹**haplodiploid:** Eines der beiden Geschlechter trägt den einfachen Chromosomensatz (*haploid*), das andere Geschlecht den doppelten (*diploid*).

Die multiplikative Binomialverteilung

Um Häufigkeiten mit größerer oder kleinerer Variabilität als jener der Binomialverteilung besser beschreiben zu können, hat Altham (1978) in ihrer Arbeit zwei Verallgemeinerungen der Binomialverteilung vorgeschlagen. Eine der beiden, die multiplikative Verallgemeinerung der Binomialverteilung, ist ein Mitglied der in Abschnitt 2.2.1 vorgestellten, zweiparametrischen Exponentialfamilie in kanonischer Form.

Dieses Kapitel wird zu Beginn die Motivation und Konstruktion dieser Verteilung erläutern, bevor in weiterer Folge ihr Erwartungswert und ihre Varianz besprochen werden. Den Abschluss dieses Kapitels macht eine Darstellung der multiplikativen Binomialverteilung im Kontext der generalisierten linearen Modelle inklusive ersten und zweiten partiellen Ableitungen der korrespondierenden log-Likelihoodfunktion.

3.1 Motivation

Altham (1978) motiviert ihre Arbeit anhand eines toxikologischen Experiments, bei dem Jungtiere aus mehreren Würfen einer giftigen Substanz ausgesetzt werden. Die Dosierung des Gifts variiert dabei über die betrachteten Würfe. Wie in vielen Untersuchungen dieser Art ist nun von Interesse, wieviele Tiere eines Wurfs eine bestimmte Dosis überleben können. Die Responsevariable y_i gibt somit für jede Dosierung der untersuchten Substanz die Anzahl der am Ende der Untersuchungen lebenden Tiere innerhalb einer Gruppe an. Die hier beschriebene Datensituation ist zum besseren Verständnis in Tabelle 3.1 skizziert.

Faktorstufe	1	...	N
Gruppenumfang	n_1	...	n_N
Anzahl Erfolge	y_1	...	y_N

Tabelle 3.1.: Darstellung der im Text beschriebenen Datensituation.

In vielen Fällen kann bei diesen Responsevariablen y_i von einer Binomialverteilung ausge-

gangen werden. Das besondere in diesem speziellen Fall ist jedoch, dass alle Tiere innerhalb einer Faktorstufe aus demselben Wurf stammen. Daher führt Altham für die Beobachtungen innerhalb einer Gruppe Interaktionen zwischen den Einzelereignissen ein, um die besagten Verwandtschaftsverhältnisse berücksichtigen zu können.

Der Großteil der zitierten Arbeit betrachtet die resultierenden Verteilungen, wenn einerseits die additive und andererseits die multiplikative Definition für Interaktionen verwendet wird. Aus der additiven Definition für Interaktionen ergibt sich eine Verteilung, die kein Mitglied der Exponentialfamilie ist und daher an dieser Stelle nicht genauer betrachtet werden soll. Statt dessen wird hier auf die Konstruktion der multiplikativen Verallgemeinerung der Binomialverteilung eingegangen. Die Ausführungen folgen dabei den Arbeiten von Altham (1978) sowie Lovison (1998), der eine alternative Herleitung für dieselbe Verteilung skizziert.

3.2 Herleitung

Ausgangspunkt für die Betrachtung dieser Verallgemeinerung sind n identisch verteilte, binäre Beobachtungen z_k . Im Beispiel von vorhin wären dies die Versuchsausgänge der n Tiere eines Wurfs. Unter Annahme von paarweiser Unabhängigkeit der einzelnen z_k ist deren Summe $y = \sum_{k=1}^n z_k$ binomialverteilt mit Parametern n und $\pi = P(z_k = 1)$. In der oben beschriebenen Datensituation wäre y die Anzahl an überlebenden Tieren eines Wurfs. Des weiteren tritt nun anstelle der üblichen Annahme der Unabhängigkeit der Beobachtungen die Annahme von paarweisen Abhängigkeiten der Einzelbeobachtungen. Damit sollen die Verwandtschaftsverhältnisse der Versuchstiere berücksichtigt werden.

Altham betrachtet in ihrer Herleitung zu Beginn den einfachen Fall mit $n = 3$ Gruppenmitgliedern. Für den multivariat Bernoulli-verteilten Vektor $\mathbf{z} = (z_1, z_2, z_3)$ sind hier 2^3 unterschiedliche Realisationen möglich. Diese Fälle sind in Abbildung 3.1 als die möglichen Realisationen dreier zweistufiger Faktoren A , B und C skizziert.

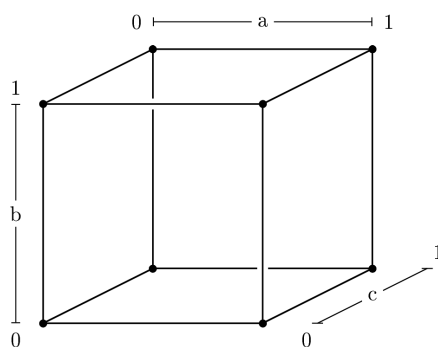


Abbildung 3.1.: Skizze der verschiedenen Ausprägungsmöglichkeiten dreier zweistufiger Faktoren.

Nun sollen die zweifachen Wechselwirkungen dieser drei Zufallsgrößen beschrieben werden. Darroch (1974) betrachtet dafür Zufallsvariablen mit Auftrittswahrscheinlichkeiten p_a , p_b und p_c und gemeinsamer Wahrscheinlichkeit $p_{abc} = P(\mathbf{z} = (a, b, c))$ mit $\sum_a \sum_b \sum_c p_{abc} = 1$. Für

diese Zufallsvariablen notiert er eine Darstellung ohne multiplikative dreifache Wechselwirkung als

$$H_m: p_{abc} = r_{bc} \cdot s_{ac} \cdot t_{ab} \quad (3.1)$$

für alle a, b und c sowie geeignet definierte $\{r_{bc}\}$, $\{s_{ac}\}$ und $\{t_{ab}\}$. Im hier betrachteten Fall seien die Variablen $a, b, c \in \{0, 1\}$, generell jedoch könnten diese Variablen auch mehrere Ausprägungsmöglichkeiten besitzen. In weiterer Folge ist es nun naheliegend, für die Einzelbeobachtungen innerhalb einer Gruppe idente Verteilungsannahmen zu treffen. Daher lässt sich Hypothese (3.1) anhand von $r_{..} = s_{..} = t_{..} =: \psi_{..}$ vereinfachen zu

$$H'_m: p_{abc} = \psi_{bc} \cdot \psi_{ac} \cdot \psi_{ab}. \quad (3.2)$$

Wie bereits erwähnt sind a, b und c entweder 0 oder 1, wobei nun weiters noch $\psi_{01} = \psi_{10}$ gelten soll. Da es sich hier um Wahrscheinlichkeiten handelt, folgt aus

$$\begin{aligned} 1 &\stackrel{!}{=} \sum_a \sum_b \sum_c p_{abc} = \sum_a \sum_b \sum_c \psi_{bc} \psi_{ac} \psi_{ab} \\ &= \sum_a \sum_b (\psi_{b0} \psi_{a0} \psi_{ab} + \psi_{b1} \psi_{a1} \psi_{ab}) \\ &= \sum_a (\psi_{00} \psi_{a0} \psi_{a0} + \psi_{01} \psi_{a1} \psi_{a0} + \psi_{10} \psi_{a0} \psi_{a1} + \psi_{11} \psi_{a1} \psi_{a1}) \\ &= \psi_{00} \psi_{00} \psi_{00} + \psi_{01} \psi_{01} \psi_{00} + \psi_{10} \psi_{00} \psi_{01} + \psi_{11} \psi_{01} \psi_{01} \\ &\quad + \psi_{00} \psi_{10} \psi_{10} + \psi_{01} \psi_{11} \psi_{10} + \psi_{10} \psi_{10} \psi_{11} + \psi_{11} \psi_{11} \psi_{11} \\ &= (\psi_{00})^3 + 3 \psi_{00} (\psi_{10})^2 + 3 (\psi_{10})^2 \psi_{11} + (\psi_{11})^3. \end{aligned}$$

Dies sind alle möglichen Fälle von 0-1 Kombinationen der drei zweistufigen Faktoren. Der erste Term, $\psi_{00} \psi_{00} \psi_{00}$, bezeichne die Wahrscheinlichkeit, dass alle drei Faktoren auf der Stufe 0 liegen. Der letzte Term entspricht $\psi_{11} \psi_{11} \psi_{11} = P(\mathbf{z} = (1, 1, 1))$. Die gemischten Terme zählen die $\binom{3}{1}$ Fälle, in denen zwei Erfolge und einen Misserfolg vorliegen sowie die $\binom{3}{2}$ Situationen, in denen ein Erfolg und zwei Misserfolge beobachtet werden.

Aus dieser Betrachtung folgt nun unmittelbar, dass es nur zwei frei wählbare Parameter $\psi_{..}$ geben kann. Die gemeinsame Verteilung von z_1, z_2 und z_3 ist also symmetrisch und durch nur zwei frei wählbare Parameter definiert.

Für den allgemeineren Fall mit nicht drei sondern n Einzelereignissen z_k mit

$$z_k = \begin{cases} 1 & \text{bei Erfolg} \\ 0 & \text{bei Misserfolg} \end{cases} \quad \text{für } k = 1, \dots, n,$$

notiert Altham ausgehend von Hypothese (3.2) die gemeinsame Wahrscheinlichkeit dieser n Beobachtungen mit nur zweifachen und keinen höheren Wechselwirkungen als Produkt

$$P(\mathbf{z} = (z_1, \dots, z_n)) = C \prod_{1 \leq u < v \leq n} \psi_{z_u z_v} \quad (3.3)$$

für eine Normierungskonstante C . Auch hier sind die Einzelereignissen z_k entweder 0 oder 1 und es gelten dieselben Resultate wie bei $n = 3$. In weiterer Folge wird nun die Wahrscheinlichkeit $P(y)$ für $y = \sum_{k=1}^n z_k$ betrachtet. Aus (3.3) folgernd lässt sich $P(y)$ als

$$P(y) = C \binom{n}{y} (\psi_{00})^{1/2(n-y)(n-y-1)} (\psi_{10})^{(y-n)y} (\psi_{11})^{1/2y(y-1)}$$

schreiben. Durch eine Reparametrisierung mit

$$\omega = \frac{\psi_{10}}{\sqrt{\psi_{00}\psi_{11}}} \quad \text{und} \quad \pi = \frac{(\psi_{11})^{(n-1)/2}}{(\psi_{00})^{(n-1)/2} + (\psi_{11})^{(n-1)/2}}$$

ergibt sich in weiterer Folge

$$P(y) = C \binom{n}{y} \pi^y (1 - \pi)^{n-y} \omega^{(n-y)y}$$

für die Wahrscheinlichkeit, bei n abhängigen Einzelereignissen y Erfolge beobachten zu können. Der Normierungsterm C muss dabei so gewählt werden, dass $\sum_{y=0}^n P(y) = 1$ erfüllt wird.

Lovison (1998) verwendet in seiner alternativen Herleitung der multiplikativen Binomialverteilung anstelle der Definition „keiner multiplikativen Dreifachinteraktionen“ aus Hypothese (3.1) eine von Cox (1972) vorgeschlagene Formulierung für Interaktionen von Zufallsvariablen. Dafür formt Cox die Zufallsvariablen z_k aus $\{0, 1\}$ für $k = 1, \dots, n$ zu neuen Variablen

$$w_k := 2z_k - 1$$

um. Die Zufallsvariablen w_k sind ebenfalls wieder binär, für ihre möglichen Realisationen gilt jedoch

$$w_k = \begin{cases} 1 & \text{bei Erfolg} \\ -1 & \text{bei Misserfolg} \end{cases} \quad \text{für } k = 1, \dots, n.$$

Nun behauptet Cox, dass die gemeinsame Wahrscheinlichkeit $P(\mathbf{z})$ als

$$P(\mathbf{z}) = \exp\left(\sum_{k=1}^n \lambda_k w_k + \sum_{k_1 < k_2} \lambda_{k_1 k_2} w_{k_1} w_{k_2} + \sum_{k_1 < k_2 < k_3} \lambda_{k_1 k_2 k_3} w_{k_1} w_{k_2} w_{k_3} + \dots + \lambda_{k_1 \dots k_n} w_{k_1} w_{k_2} \dots w_{k_n} - \Lambda\right) \quad (3.4)$$

geschrieben werden kann. Der Ausdruck $\exp(-\Lambda)$ bezeichne einen Normierungsterm der derart gewählt wird, dass sich die einzelnen Wahrscheinlichkeiten zu eins summieren. Die λ_{\cdot} innerhalb der einzelnen Summen seien gewisse Interaktionsparameter. Die erste Summe in (3.4) enthält die Randwahrscheinlichkeiten $P(z_k)$ bzw. $P(w_k)$, die zweite Summe die zweifachen Wechselwirkungen der einzelnen Variablen, die dritte die dreifachen Interaktionen und so weiter bis zur letzten Summe, welche die n -fache Wechselwirkung aller n vorliegenden Einzelbetrachtungen beschreibt.

In dieser Arbeit sollen nun nur Zweifachinteraktionen berücksichtigt werden, für die höheren Wechselwirkungen gilt somit $\lambda_{k_1 k_2 k_3} = \dots = \lambda_{k_1 \dots k_n} = 0$. Außerdem gelte für die Interaktionsparameter $\lambda_k := \gamma \forall k$ und $\lambda_{k_1 k_2} := \delta \forall k_1, k_2$. Daher läßt sich Modell (3.4) vereinfachen zu

$$P(\mathbf{z}) = \exp(-\Lambda) \exp\left(\gamma \sum_{k=1}^n w_k + \delta \sum_{k_1 < k_2} w_{k_1} w_{k_2}\right). \quad (3.5)$$

Kombinatorische Überlegungen erlauben es nun, die Summen aus (3.5) explizit zu berechnen. So läßt sich leicht einsehen, dass die Anzahl aller möglicher Paarungen (z_i, z_j) aus $\mathbf{z} = (z_1, \dots, z_n)$

gleich $\binom{n}{2} = 1/2 n(n-1)$ für $i \neq j$ ist. Außerdem können nur drei unterschiedliche Fälle unterschieden werden, nämlich $(0,0)$, wo sowohl z_i als auch z_j Misserfolge sind; $(1,1)$, in dem beide Ereignisse erfolgreich sind und der Fall $(0,1)$ beziehungsweise $(1,0)$, also Paare mit einem Misserfolg und einem Erfolg. Sein nun $y = \sum_{i=1}^n z_i$, dann gibt es

$$\begin{aligned} & \frac{y(y-1)}{2} \quad \text{Paare vom Typ } (1,1) \\ & \frac{(n-y)(n-y+1)}{2} \quad \text{Paare vom Typ } (0,0) \\ & \text{und } (n-y)y \quad \text{Paare vom Typ } (0,1) \text{ bzw. } (1,0). \end{aligned}$$

Daraus ergibt sich für die Summe über die w_k

$$\sum_{k=1}^n w_k = \#(\text{Erfolge}) - \#(\text{Misserfolge}) = y - (n-y),$$

da die w_k genau so definiert sind, dass sie bei Erfolg 1 und bei Misserfolg -1 sind. Ebenso gilt für ein Produkt $w_k w_h$, dass es 1 wird wenn es sich bei den Einzelereignissen um zwei Erfolge oder zwei Misserfolge handelt und das Produkt wird -1 , wenn es aus einem Einzelerfolg und einem Misserfolg entsteht. Daher läßt sich für die zweite Summe aus (3.5) zeigen, dass

$$\begin{aligned} \sum_{k < h} w_k w_h &= \#(\text{gleiche Paare}) - \#(\text{gemischte Paare}) \\ &= (\#(\text{alle Paare}) - \#(\text{gemischte Paare})) - \#(\text{gemischte Paare}) \\ &= \left(\frac{n(n-1)}{2} - (n-y)y \right) - (n-y)y \end{aligned}$$

gilt und es ergibt sich

$$\begin{aligned} P(\mathbf{z}) &= \exp(-\Lambda) \exp\left(\gamma \sum_{k=1}^n w_k + \delta \sum_{k < h} w_k w_h\right) \\ &= \exp(-\Lambda) \exp\left(\gamma(y - (n-y)) + \delta\left(\frac{n(n-1)}{2} - 2(n-y)y\right)\right) \\ &= \exp(-\Lambda) \exp\left(\gamma y - \gamma(n-y) - 2\delta(n-y)y\right) \end{aligned} \quad (3.6)$$

für einen Normierungsterm $\exp(-\Lambda)$. Durch Erweitern des Ausdrucks (3.6) mit $(\exp(\gamma) + \exp(-\gamma))^{-n}$ und unter Verwendung von

$$\frac{\exp(\gamma)}{\exp(\gamma) + \exp(-\gamma)} = \frac{\exp(2\gamma)}{1 + \exp(2\gamma)}$$

wird aus obiger Gleichung

$$\begin{aligned} P(\mathbf{z}) &= \exp(-\Lambda) \frac{\exp(\gamma)^y \exp(-\gamma)^{n-y} \exp(2\delta)^{(n-y)y}}{(\exp(\gamma) + \exp(-\gamma))^n} \\ &= \exp(-\Lambda) \left(\frac{\exp(\gamma)}{\exp(\gamma) + \exp(-\gamma)} \right)^y \left(\frac{\exp(-\gamma)}{\exp(\gamma) + \exp(-\gamma)} \right)^{n-y} \exp(-2\delta)^{(n-y)y} \\ &= \exp(-\Lambda) \left(\frac{\exp(\gamma)}{\exp(\gamma) + \exp(-\gamma)} \right)^y \left(1 - \frac{\exp(\gamma)}{\exp(\gamma) + \exp(-\gamma)} \right)^{n-y} \exp(-2\delta)^{(n-y)y}. \end{aligned}$$

Nun läßt sich durch geeignetes Substituieren von

$$\pi := \frac{\exp(\gamma)}{\exp(\gamma) + \exp(-\gamma)} \quad \text{und} \quad \omega := \exp(-2\delta)$$

die Darstellung

$$P(\mathbf{z}) = \exp(-\Lambda) \pi^y (1 - \pi)^{n-y} \omega^{(n-y)y} \quad (3.7)$$

erhalten. Da sich die einzelnen Wahrscheinlichkeiten $P(\mathbf{z})$ zu eins summieren sollen, ergibt sich für den Normierungsterm

$$\exp(\Lambda) = \sum_{j=0}^n \binom{n}{j} \pi^j (1 - \pi)^{n-j} \omega^{(n-j)j}.$$

Für die Verteilung von $y = \sum_{k=1}^n z_k$ folgt daher

$$P(y) = \frac{\binom{n}{y} \pi^y (1 - \pi)^{n-y} \omega^{(n-y)y}}{\sum_{j=0}^n \binom{n}{j} \pi^j (1 - \pi)^{n-j} \omega^{(n-j)j}}. \quad (3.8)$$

Dabei läßt sich in dieser auf der Darstellung aus (3.4) basierenden Konstruktionsweise eine sinnvolle Interpretation des Parameters ω angeben. Diese stammt aus der Betrachtung des bedingten Kreuzproduktverhältnisses (*odds ratio*) zweier Einzelereignisse gegeben alle anderen Beobachtungen,

$$\text{CPR}(z_k, z_h | \text{Rest}) = \frac{P(z_k = 0, z_h = 0 | \text{Rest}) P(z_k = 1, z_h = 1 | \text{Rest})}{P(z_k = 0, z_h = 1 | \text{Rest}) P(z_k = 1, z_h = 0 | \text{Rest})} = \frac{\exp(2\delta)}{\exp(-2\delta)}.$$

Dieses Verhältnis ist genau dann 1, wenn die beiden Zufallsvariablen unabhängig sind. Das ist jedoch genau dann der Fall, wenn für den Interaktionsparameter δ aus (3.5) $\delta = 0$ gilt.

Für ω folgt somit, dass es als ein Maß für den Grad der Interaktionen zwischen den beiden Zufallsvariablen z_k und z_h aufgefasst werden kann, da

$$\omega = \frac{1}{\sqrt{\text{CPR}(z_k, z_h | \text{Rest})}}$$

gilt. Sind die beiden Zufallsvariablen z_k und z_h unabhängig, so ist ihr Kreuzproduktverhältnis genau 1 und es folgt $\omega = 1$. Für unabhängige Zufallsvariablen vereinfacht sich die multiplikative Binomialverteilung somit zur klassischen Binomialverteilung.

Als Abrundung dieser Einführung visualisieren Abbildungen 3.2 und 3.3 die Wahrscheinlichkeitsfunktion für $\pi = \{0, 5; 0, 8\}$ und $\omega \in [0, 9; 1, 1]$ bei $n = 36$. Für $\omega = 1$ wird die multiplikative Verallgemeinerung zur klassischen Binomialverteilung mit $P(y) = \binom{n}{y} \pi^y (1 - \pi)^{n-y}$. Sobald ω etwas größer respektive kleiner 1 ist, verändert sich das Verhalten dieser Verteilung drastisch. Für Werte von $\omega < 1$ werden die Ränder $y = 0$ und $y = n$ immer wahrscheinlicher und Werte im Zentrum haben Wahrscheinlichkeiten (fast) 0. Für $\omega > 1$ ist genau das Gegenteil der Fall. Die Fälle $y = 0$ und $y = n$ werden extrem unwahrscheinlich, jene im Zentrum hingegen werden immer wahrscheinlicher.

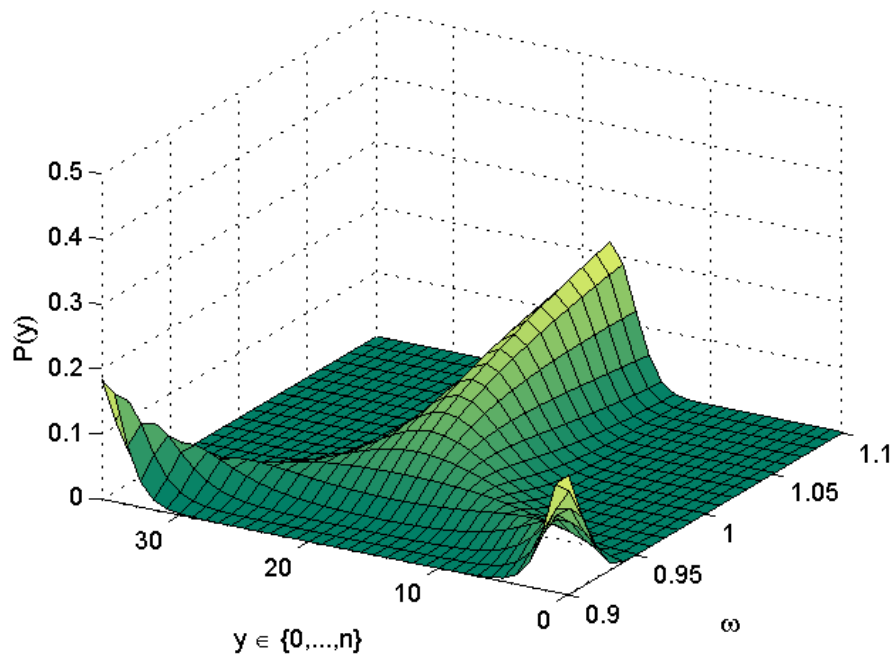


Abbildung 3.2.: Wahrscheinlichkeitsfunktion der multiplikativen Binomialverteilung bei $\pi = 0,5$ über verschiedene Werte von ω und $n = 36$.

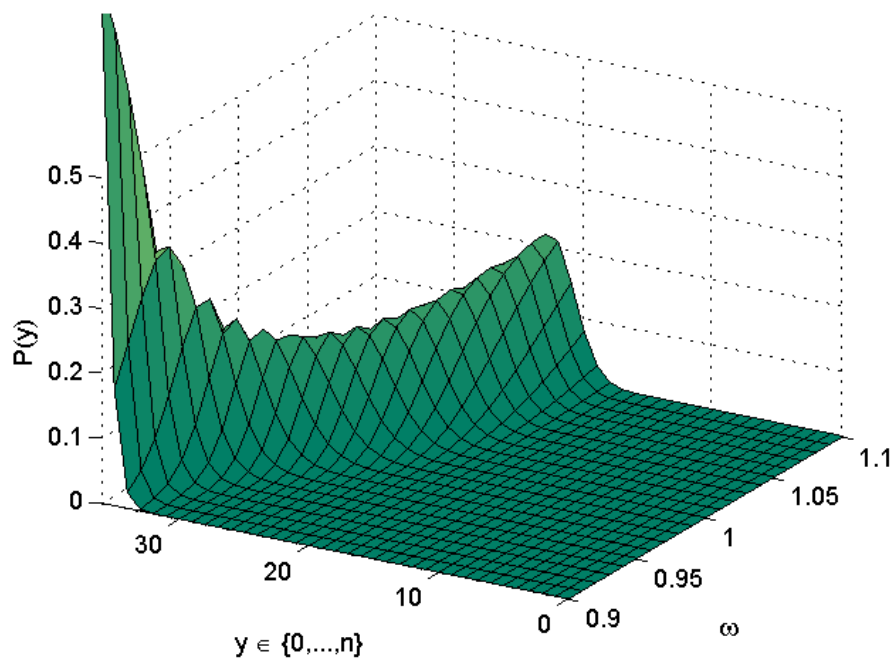


Abbildung 3.3.: Wahrscheinlichkeitsfunktion der multiplikativen Binomialverteilung bei $\pi = 0,8$ über verschiedene Werte von ω und $n = 36$.

3.3 Eigenschaften

Es lässt sich zeigen, dass die multiplikative Verallgemeinerung der Binomialverteilung ein Mitglied der zweiparametrischen Exponentialfamilie ist. Dafür muss die in (3.8) verwendete Darstellung umgeschrieben werden zu

$$\begin{aligned} P(y) &= \frac{\binom{n}{y} \left(\frac{\pi}{1-\pi}\right)^y (1-\pi)^n \omega^{(n-y)y}}{\sum_{j=0}^n \binom{n}{j} \left(\frac{\pi}{1-\pi}\right)^j (1-\pi)^n \omega^{(n-j)j}} \\ &= \binom{n}{y} \frac{1}{\sum_{j=0}^n \binom{n}{j} \left(\frac{\pi}{1-\pi}\right)^j \omega^{(n-j)j}} \exp\left(y \log \frac{\pi}{1-\pi} + (n-y)y \log \omega\right). \end{aligned} \quad (3.9)$$

Diese Schreibweise entspricht der in (2.3) geforderten mit

$$w_1(\boldsymbol{\theta}) = \log \frac{\pi}{1-\pi} \quad \text{und} \quad w_2(\boldsymbol{\theta}) = \log \omega,$$

wobei $\boldsymbol{\theta} = (\pi, \omega)$ gilt. Des Weiteren sind

$$h(y) = \binom{n}{y}, \quad t_1(y) = y \quad \text{und} \quad t_2(y) = (n-y)y.$$

Der restliche Ausdruck ist von y unabhängig und kann zur Funktion $c(\boldsymbol{\theta})$ zusammengefasst werden.

Wie in den Gleichungen (2.5) und (2.6) festgehalten, gibt es für K -parametrische Exponentialfamilien Formeln zur Berechnung von Erwartungswert und Varianz. Dafür muss jedoch die Exponentialfamilie in ihrem natürlichen Parameterraum vorliegen. Für die multiplikative Binomialverteilung erhält man diese Darstellung, indem man die Wahrscheinlichkeitsfunktion mit

$$\nu_1 = \log \frac{\pi}{1-\pi} \quad \text{und} \quad \nu_2 = \log \omega$$

reparametrisiert. Dadurch ergibt sich

$$P(y|\boldsymbol{\nu}) = \binom{n}{y} \frac{1}{\sum_{j=0}^n \binom{n}{j} e^{j\nu_1} e^{(n-j)j\nu_2}} e^{y\nu_1} e^{(n-y)y\nu_2} \quad (3.10)$$

und es lassen sich Erwartungswert und Varianz von y aus den ersten beiden Ableitungen von

$$-\log(c^*(\boldsymbol{\nu})) = \log \sum_{j=0}^n \binom{n}{j} e^{j\nu_1} e^{(n-j)j\nu_2}$$

nach ν_1 ermitteln, da für die zu ν_1 gehörende suffiziente Statistik $t_1(y) = y$ gilt. Für den Erwartungswert folgt daher aus (3.10)

$$\begin{aligned} E(y) &= -\frac{\partial}{\partial \nu_1} \log(c^*(\boldsymbol{\nu})) = \frac{\partial}{\partial \nu_1} \log \sum_{j=0}^n \binom{n}{j} e^{j\nu_1} e^{(n-j)j\nu_2} \\ &= \frac{\sum_{j=0}^n j \binom{n}{j} e^{j\nu_1} e^{(n-j)j\nu_2}}{\sum_{j=0}^n \binom{n}{j} e^{j\nu_1} e^{(n-j)j\nu_2}}. \end{aligned}$$

Der Einfachheit halber soll nun $d_j^{(\nu)} := \binom{n}{j} e^{j\nu_1} e^{(n-j)\nu_2}$ gelten und es folgt

$$E(y) = \frac{\sum_{j=0}^n j d_j^{(\nu)}}{\sum_{j=0}^n d_j^{(\nu)}}. \quad (3.11)$$

Diese Darstellung des Erwartungswerts wurde auch von Seeber (1997) verwendet. Ebenso ergibt sich für die Varianz mit (2.6)

$$\text{Var}(y) = \frac{\sum_{j=0}^n j^2 d_j^{(\nu)} \sum_{j=0}^n d_j^{(\nu)} - \left\{ \sum_{j=0}^n j d_j^{(\nu)} \right\}^2}{\left\{ \sum_{j=0}^n d_j^{(\nu)} \right\}^2}. \quad (3.12)$$

Im Gegensatz zu Altham (1978) gibt erst Lovison (1998) in seiner Arbeit geschlossene Darstellungen für die beiden ersten Momente an. Er verwendet die Wahrscheinlichkeiten $P(\mathbf{z})$ aus (3.7), um in einem ersten Schritt Erwartungswert und Varianz von z_k zu errechnen. Dafür verwendet Lovison die Tatsache, dass sich die Summe zur Berechnung des Erwartungswerts von z_k , $E(z_k) = 0 \cdot P(z_k = 0) + 1 \cdot P(z_k = 1)$, hier zu $E(z_k) = P(z_k = 1)$ vereinfacht und erhält

$$\begin{aligned} E(z_k) &= P(z_k = 1) = \sum_{z_1=0}^1 \cdots \sum_{z_n=0}^1 P(z_1, \dots, z_k = 1, \dots, z_n) \\ &= \frac{\sum_{y=1}^n \binom{n-1}{y-1} \pi^y (1-\pi)^{n-y} \omega^{(n-y)y}}{\sum_{y=0}^n \binom{n}{y} \pi^y (1-\pi)^{n-y} \omega^{(n-y)y}} = \pi \frac{\sum_{y=1}^n \binom{n-1}{y-1} \pi^{y-1} (1-\pi)^{n-y} \omega^{(n-y)y}}{\sum_{y=0}^n \binom{n}{y} \pi^y (1-\pi)^{n-y} \omega^{(n-y)y}}. \end{aligned}$$

Für diese beiden Summen schreibt Lovison (1998) nun

$$\xi_{n-a}(\pi, \omega) := \sum_{j=0}^{n-a} \binom{n-a}{j} \pi^j (1-\pi)^{n-a-j} \omega^{(n-a-j)(a+j)}$$

und erhält die kompakte Darstellung

$$E(z_k) = P(z_k = 1) = \pi \frac{\xi_{n-1}(\pi, \omega)}{\xi_n(\pi, \omega)}.$$

Daraus läßt sich nun der Erwartungswert von $y = \sum_{k=1}^n z_k$ als

$$E(y) = n\pi \frac{\xi_{n-1}(\pi, \omega)}{\xi_n(\pi, \omega)} \quad (3.13)$$

ermitteln. Es läßt sich zeigen, dass diese Darstellung des Erwartungswerts von y jener unter (3.11) entspricht. Dafür wird der Zusammenhang

$$\binom{n}{j} = \frac{n}{j} \binom{n-1}{j-1}$$

benötigt, mit dessen Hilfe dann

$$\begin{aligned}
 \frac{\sum_{j=0}^n j d_j^{(\pi, \omega)}}{\sum_{j=0}^n d_j^{(\pi, \omega)}} &= \frac{\sum_{j=0}^n j \binom{n}{j} \pi^j (1-\pi)^{n-j} \omega^{(n-j)j}}{\sum_{j=0}^n \binom{n}{j} \pi^j (1-\pi)^{n-j} \omega^{(n-j)j}} \\
 &= \frac{\sum_{j=0}^n j \frac{n}{j} \binom{n-1}{j-1} \pi^j (1-\pi)^{(n-j)} \omega^{(n-j)j}}{\sum_{j=0}^n \binom{n}{j} \pi^j (1-\pi)^{n-j} \omega^{(n-j)j}} \\
 &= n\pi \frac{\sum_{j=0}^n \binom{n-1}{j-1} \pi^{j-1} (1-\pi)^{(n-j)} \omega^{(n-j)j}}{\sum_{j=0}^n \binom{n}{j} \pi^j (1-\pi)^{n-j} \omega^{(n-j)j}} \\
 &= n\pi \frac{\sum_{j=0}^{n-1} \binom{n-1}{j} \pi^j (1-\pi)^{(n-j-1)} \omega^{(n-j-1)(j+1)}}{\sum_{j=0}^n \binom{n}{j} \pi^j (1-\pi)^{n-j} \omega^{(n-j)j}} \\
 &= n\pi \frac{\xi_{n-1}(\pi, \omega)}{\xi_n(\pi, \omega)}
 \end{aligned}$$

folgt.

Für die Varianz von z_k gilt bekannter Weise

$$\text{Var}(z_k) = \text{E}(z_k^2) - \text{E}(z_k)^2 = \text{P}(z_k = 1) - \text{P}(z_k = 1)^2.$$

Für die Varianz von y muss die Summe der Varianzen der Einzelereignisse z_k sowie der Kovarianzen zwischen zwei Ereignissen z_k und z_h ,

$$\text{Var}(y) = \sum_{k=1}^n \text{Var}(z_k) + \sum_{k=1}^n \sum_{h \neq k}^n \text{Cov}(z_k, z_h)$$

betrachtet werden, wobei für die Kovarianz die Zerlegung

$$\text{Cov}(z_k, z_h) = \text{E}(z_k z_h) - \text{E}(z_k)\text{E}(z_h) = \text{P}(z_k = 1, z_h = 1) - \text{P}(z_k = 1)^2$$

gilt. Daher berechnet Lovison

$$\begin{aligned}
 \text{E}(z_k z_h) &= \text{P}(z_k = 1, z_h = 1) = \sum_{z_1=0}^1 \cdots \sum_{z_n=0}^1 \text{P}(z_1, \dots, z_k = 1, \dots, z_h = 1, \dots, z_n) \\
 &= \frac{\sum_{y=2}^n \binom{n-2}{y-2} \pi^y (1-\pi)^{n-y} \omega^{(n-y)y}}{\sum_{y=0}^n \binom{n}{y} \pi^y (1-\pi)^{n-y} \omega^{(n-y)y}} = \pi^2 \frac{\xi_{n-2}(\pi, \omega)}{\xi_n(\pi, \omega)}
 \end{aligned}$$

und kann die Varianz von y errechnen aus

$$\begin{aligned}
 \text{Var}(y) &= \sum_{k=1}^n \text{Var}(z_k) + \sum_{k=1}^n \sum_{h \neq k}^n \text{Cov}(z_k, z_h) \\
 &= n(\text{P}(z_1 = 1) - \text{P}(z_1 = 1)^2) + n(n-1)(\text{P}(z_1 = 1, z_2 = 1) - \text{P}(z_1 = 1)^2) \\
 &= n\text{P}(z_1 = 1) + n(n-1)\text{P}(z_1 = 1, z_2 = 1) - n^2\text{P}(z_1 = 1)^2 \\
 &= n\pi \frac{\xi_{n-1}(\pi, \omega)}{\xi_n(\pi, \omega)} + n(n-1)\pi^2 \frac{\xi_{n-2}(\pi, \omega)}{\xi_n(\pi, \omega)} - n^2\pi^2 \left(\frac{\xi_{n-1}(\pi, \omega)}{\xi_n(\pi, \omega)} \right)^2 \\
 &= n\pi \left[\frac{\xi_{n-1}(\pi, \omega)}{\xi_n(\pi, \omega)} - \pi \frac{n \xi_{n-1}^2(\pi, \omega) - (n-1) \xi_{n-2}(\pi, \omega) \xi_n(\pi, \omega)}{\xi_n^2(\pi, \omega)} \right]. \tag{3.14}
 \end{aligned}$$

Durch geschicktes Umformen lässt sich zeigen, dass auch die beiden Darstellungen (3.12) und (3.14) für die Varianz von y äquivalent sind. Dafür sollen zuerst die Funktionen $\xi_{n-a}(\pi, \omega)$ für $a = 0, 1, 2$ betrachtet werden. Für $a = 0$ gilt

$$\xi_n(\pi, \omega) = \sum_{j=0}^n \binom{n}{j} \pi^j (1-\pi)^{n-j} \omega^{(n-j)j} = \sum_{j=0}^n d_j^{(\pi, \omega)}, \quad (3.15)$$

für $a = 1$ lässt sich $\xi_{n-1}(\pi, \omega)$ umschreiben zu

$$\begin{aligned} \xi_{n-1}(\pi, \omega) &= \sum_{j=0}^{n-1} \binom{n-1}{j} \pi^j (1-\pi)^{n-j-1} \omega^{(n-j-1)(j+1)} \\ &= \sum_{j=1}^n \binom{n-1}{j-1} \pi^{j-1} (1-\pi)^{n-j} \omega^{(n-j)j} \\ &= \sum_{j=1}^n \frac{j}{n} \binom{n}{j} \frac{1}{\pi} \pi^j (1-\pi)^{n-j} \omega^{(n-j)j} \\ &= \frac{1}{n\pi} \sum_{j=1}^n j \binom{n}{j} \pi^j (1-\pi)^{n-j} \omega^{(n-j)j} = \frac{1}{n\pi} \sum_{j=0}^n d_j^{(\pi, \omega)}. \end{aligned}$$

Es gilt also

$$\xi_{n-1}(\pi, \omega) = \frac{1}{n\pi} \sum_{j=0}^n d_j^{(\pi, \omega)}. \quad (3.16)$$

Dieser Zusammenhang wurde auch bereits beim Erwartungswert verwendet. Natürlich lässt sich auch für $\xi_{n-2}(\pi, \omega)$ zeigen, dass

$$\xi_{n-2}(\pi, \omega) = \frac{1}{n(n-1)\pi^2} \left[\sum_{j=0}^n j^2 d_j^{(\pi, \omega)} - \sum_{j=0}^n j d_j^{(\pi, \omega)} \right] \quad (3.17)$$

gilt. Mit diesen Umformungen lässt sich nun zeigen, dass die beiden Darstellungen für die Varianz multiplikativ binomialverteilter Zufallsvariablen äquivalent sind. Ausgehend von der bei Lovison (1998) vorgestellten Notation folgt mithilfe der Zusammenhänge aus (3.15) bis (3.17)

$$\begin{aligned} \text{Var}(y) &= n\pi \left[\frac{\xi_{n-1}(\pi, \omega)}{\xi_n(\pi, \omega)} - \pi \frac{n \xi_{n-1}^2(\pi, \omega) - (n-1) \xi_{n-2}(\pi, \omega) \xi_n(\pi, \omega)}{\xi_n^2(\pi, \omega)} \right] \\ &= n\pi \frac{\frac{1}{n\pi} \sum_j j d_j}{\sum_j d_j} - n\pi^2 \frac{n \left\{ \frac{1}{n\pi} \sum_j j d_j \right\}^2 - (n-1) \frac{1}{n(n-1)\pi^2} \left[\sum_j j^2 d_j - \sum_j j d_j \right] \sum_j d_j}{\left\{ \sum_j d_j \right\}^2} \\ &= \frac{\sum_j j d_j}{\sum_j d_j} - \frac{(n\pi)^2 \frac{1}{(n\pi)^2} \left\{ \sum_j j d_j \right\}^2 - n(n-1)\pi^2 \frac{1}{n(n-1)\pi^2} \left[\sum_j j^2 d_j - \sum_j j d_j \right] \sum_j d_j}{\left\{ \sum_j d_j \right\}^2} \end{aligned}$$

$$\begin{aligned}
 &= \frac{\sum_j j d_j \sum_j d_j - \left\{ \sum_j j d_j \right\}^2 + \sum_j j^2 d_j \sum_j d_j - \sum_j j d_j \sum_j d_j}{\left\{ \sum_j d_j \right\}^2} \\
 &= \frac{\sum_{j=0}^n j^2 d_j^{(\pi, \omega)} \sum_{j=0}^n d_j^{(\pi, \omega)} - \left\{ \sum_{j=0}^n j d_j^{(\pi, \omega)} \right\}^2}{\left\{ \sum_{j=0}^n d_j^{(\pi, \omega)} \right\}^2}.
 \end{aligned}$$

Die beiden Notationen für die Varianz der multiplikativen Binomialverteilung stimmen somit überein.

Abbildung 3.4 zeigt Visualisierungen des Erwartungswerts, einmal als Fläche in Abhängigkeit von π und ω und einmal im Aufriss für $\pi \in [0; 1]$ für verschiedene Werte von ω . Abbildungen 3.5 und 3.6 zeigen die Standardabweichungen als Fläche in Abhängigkeit von π und ω beziehungsweise dem Erwartungswert und ω und im Aufriss für verschiedene Werte von ω . Es ist deutlich zu erkennen, dass die Standardabweichung bereits für kleine Abweichungen von ω von 1 extreme Werte annimmt. So ist für ein $\omega > 1$ die Standardabweichung über weite Bereiche von π fast konstant und nur an den Rändern, also um $\pi = 0$ und $\pi = 1$, verändert sich die Standardabweichung sichtbar. Im Vergleich zur Standardabweichung der klassischen Binomialverteilung (schwarze Linie bei $\omega = 1$) ist diese Standardabweichung deutlich breiter und entspricht einer Überdispersion. Bei $\omega = 1$ entspricht die Standardabweichung jener der Binomialverteilung. Für Werte von $\omega < 1$ zeigt sich ein starker Ausschlag bei $\pi = 1/2$. Für Werte von π weiter weg vom Zentrum ist die Standardabweichung hingegen fast konstant 0. Für $\omega < 1$ ist die Standardabweichung also deutlich kleiner als jene der klassischen Binomialverteilung und es liegt Unterdispersion vor.

Festgehalten sei an dieser Stelle nochmals, dass für sämtliche Abbildungen eine feste Gruppengröße von $n = 36$ angenommen wurde.

Wie sich die Varianz der multiplikativen Binomialverteilung im Gegensatz zur klassischen Binomialverteilung verhält, zeigt Abbildung 3.7. Aufgrund der bereits angesprochenen extrem großen Werte der Varianz wurde für die z-Achse des Plots eine logarithmische Skalierung verwendet. Die Abbildung lässt erkennen, dass für $\omega < 1$ im Zentrum des Einheitsintervalls die multiplikative Binomialverteilung Überdispersion aufweist, an den Rändern jedoch die Varianz der klassischen Binomialverteilung deutlich größer ist. Für $\omega > 1$ verhält es sich umgekehrt: im Zentrum des Einheitsintervalls ist die Varianz multiplikativ binomialverteilter Zufallsvariablen kleiner als jene der klassischen Binomialverteilung; an den Rändern des Einheitsintervalls verhält es sich umgekehrt. Eine scharfe Trennung zwischen Über- und Unterdispersion allein anhand des Wertes von ω ist also nicht möglich.

Zu bemerken ist noch, dass sich für Gruppengrößen von $n = 1$ die Wahrscheinlichkeitsfunktion der multiplikativen Verallgemeinerung soweit vereinfacht, dass sie nur noch der Wahrscheinlichkeitsfunktion der Bernoulli-Verteilung mit $P(y) = \pi^y (1 - \pi)^{1-y}$ entspricht und der zweite Parameter ω aus der Verteilung verschwunden ist. Daher kann mittels dieser Verteilung nur für Gruppengrößen mit $n \geq 2$ Über- und Unterdispersion beschrieben werden.

Um zu prüfen, ob die Summe zweier multiplikativ binomialverteilter Zufallsvariablen abermals wieder multiplikativ binomialverteilt ist, wird die momenterzeugende Funktion $M_y(s) = E(e^{sy})$ der Verteilung benötigt. Wie bereits in Kapitel 2.2.1 erwähnt, lässt sich diese Funktion aus der kanonischen Darstellung der Exponentialfamilie erhalten. Ausgehend von der Darstellung

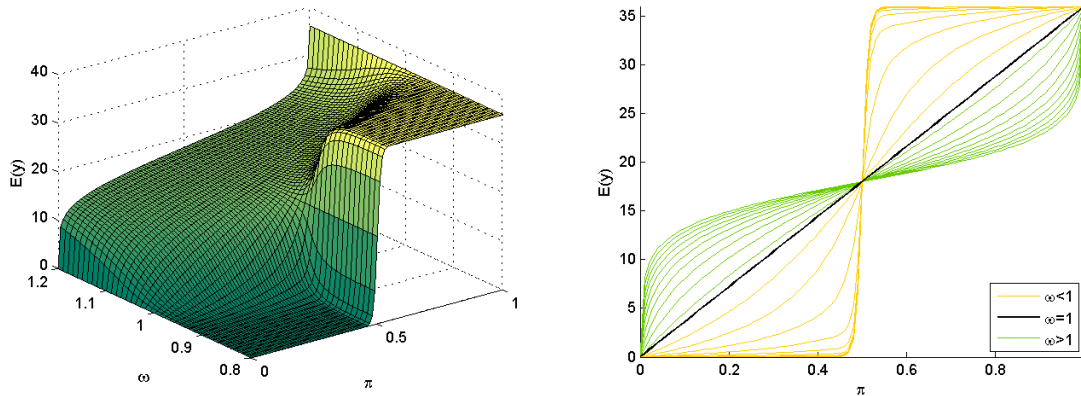


Abbildung 3.4.: Visualisierung des Erwartungswerts in Abhängigkeit von π und ω .

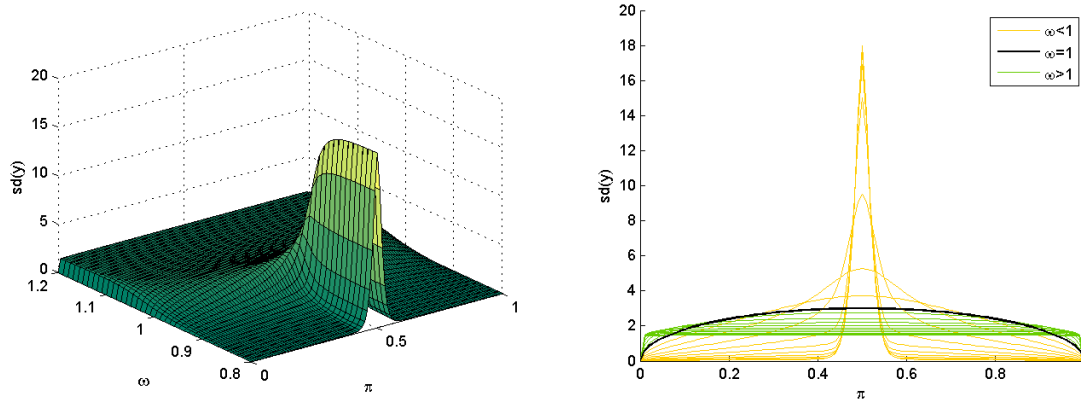


Abbildung 3.5.: Visualisierung der Standardabweichung in Abhängigkeit von π und ω .

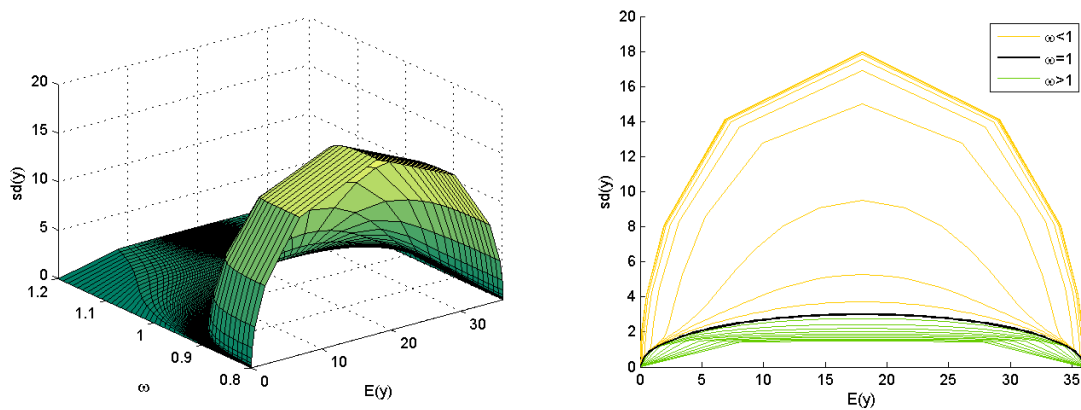


Abbildung 3.6.: Visualisierung der Standardabweichung in Abhängigkeit vom Erwartungswert und ω .

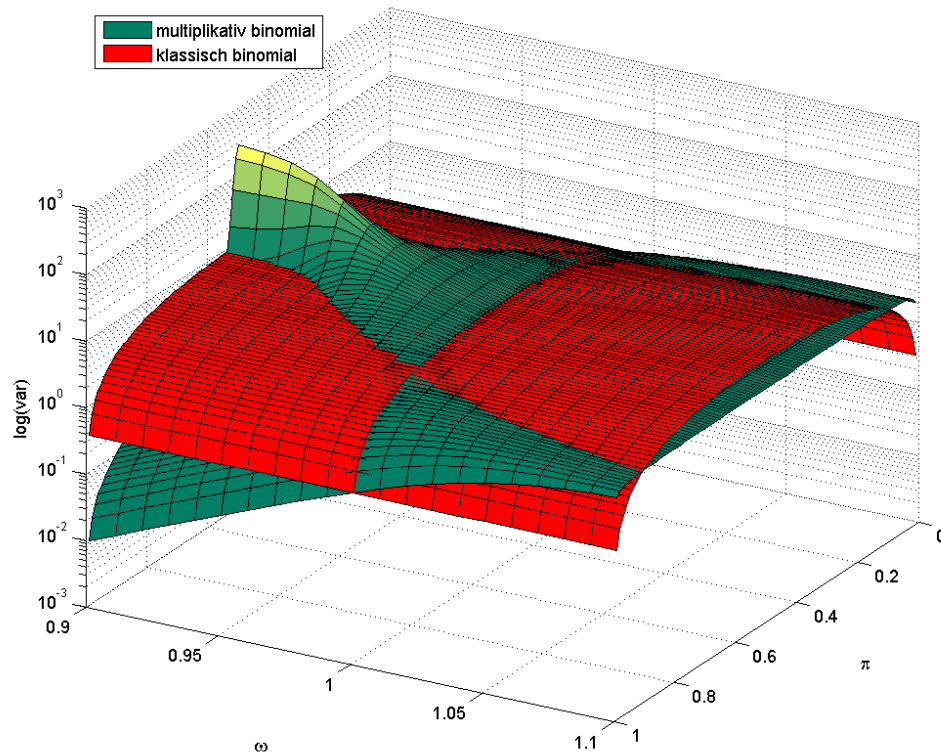


Abbildung 3.7.: Gegenüberstellung der Varianzen unter klassischer sowie unter multiplikativer Binomialverteilung für verschiedene Werte von ω und $p \in [0, 1]$ bei $n = 36$.

in (3.10) folgt aus

$$\int f(y|\boldsymbol{\nu}) dy = \int h(y) c^*(\nu_1, \nu_2) e^{y\nu_1 + (n-y)y\nu_2} dy = 1,$$

dass für den in y konstanten Term $c(\nu_1, \nu_2)$

$$\int h(y) e^{y\nu_1 + (n-y)y\nu_2} dy = \frac{1}{c^*(\nu_1, \nu_2)}$$

gilt. Daraus folgt für die bezüglich ν_1 lineare Exponentialfamilie, dass die momentenerzeugende Funktion aus

$$\begin{aligned} M_y(s) &= E(e^{sy}) = \int e^{sy} h(y) c^*(\boldsymbol{\nu}) e^{y\nu_1 + (n-y)y\nu_2} dy \\ &= c^*(\boldsymbol{\nu}) \int h(y) e^{y(\nu_1+s) + (n-y)y\nu_2} dy \\ &= \frac{c^*(\nu_1, \nu_2)}{c^*(\nu_1 + s, \nu_2)} \end{aligned}$$

berechnet werden kann. Somit ist die durch

$$M_y(s) = \frac{\sum_{j=0}^{n_i} \binom{n_i}{j} e^{j\nu_1} e^{(n_i-j)\nu_2} e^{js}}{\sum_{j=0}^{n_i} \binom{n_i}{j} e^{j\nu_1} e^{(n_i-j)\nu_2}}$$

gegebene Funktion die momentenerzeugende Funktion der multiplikativen Binomialverteilung.

Die momentenerzeugende Funktion der Summe zweier unabhängig und ident multiplikativ binomialverteilter Zufallsvariablen erfüllt

$$\begin{aligned} M_{y_1+y_2}(s) &= E(e^{s(y_1+y_2)}) = E(e^{sy_1})E(e^{sy_2}) = M_{y_1}(s)M_{y_2}(s) \\ &= \left(\frac{\sum_{j=0}^{n_i} \binom{n_i}{j} e^{j\nu_1} e^{(n_i-j)\nu_2} e^{js}}{\sum_{j=0}^{n_i} \binom{n_i}{j} e^{j\nu_1} e^{(n_i-j)\nu_2}} \right)^2. \end{aligned}$$

Diese neue Funktion entspricht jedoch keiner momentenerzeugenden Funktion der multiplikativen Binomialverteilung mehr. Daher ist diese Verallgemeinerung der Binomialverteilung nicht abgeschlossen bezüglich ihrer Faltung.

Dieses Resultat ist auch einleuchtend, wenn abermals auf die Motivation der Verteilung zurückgegriffen wird. So wird diese multiplikative Verallgemeinerung der Binomialverteilung eingeführt, um paarweise Abhängigkeiten zwischen den Individuen einer Gruppe zu berücksichtigen. Diese paarweisen Abhängigkeitsstrukturen aller Einzelbeobachtungen können bei einer Faltung zweier derartiger Zufallsvariablen jedoch nicht realisiert werden. Daher entspricht die Summe zweier multiplikativ binomialverteilter Zufallsvariablen keiner multiplikativen Binomialverteilung mehr. Dies hat direkte Auswirkungen auf die Wahl der Gruppengröße n und wird in Kapitel 6 nochmals aufgenommen.

3.4 Die multiplikative Binomialverteilung in der logistischen Regression

Um mit der multiplikativen Verallgemeinerung der Binomialverteilung ein generalisiertes lineares Modell schätzen zu können, wird die Erfolgswahrscheinlichkeit π_i einer jeden Beobachtung in (3.8) durch ein Funktion im linearen Prädiktor $\eta_i = \mathbf{x}_i^t \boldsymbol{\beta}$ beschrieben. Wie bereits im linearen Regressionsmodell sei hier $\mathbf{X} \in \mathbb{R}^{N \times p}$ die Designmatrix von N Beobachtungen, wobei jede Zeile die $p - 1$ erklärenden Größen einer Beobachtung und den Intercept-Eintrag enthält. Außerdem sei $\boldsymbol{\beta} \in \mathbb{R}^p$ wie gewohnt der Spaltenvektor der p unbekannt Parameter $\beta_0, \dots, \beta_{p-1}$.

Um die Erfolgswahrscheinlichkeit π_i mit dem linearen Prädiktor η_i zu verknüpfen, wird bei der logistischen Regression der zur Binomialverteilung kanonische Link verwendet. Somit gilt für das Modell

$$\log \frac{\pi_i}{1 - \pi_i} = \eta_i = \mathbf{x}_i^t \boldsymbol{\beta} \quad \text{beziehungsweise} \quad \pi_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} = \frac{\exp(\mathbf{x}_i^t \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i^t \boldsymbol{\beta})}.$$

Ausgehend von der log-Likelihoodfunktion der multiplikativen Binomialverteilung bei N Be-

obachtungen y_i in den Parametern n_i , π_i und einem globalen ω für alle Gruppen,

$$l(\boldsymbol{\pi}, \omega | \mathbf{y}) = \sum_{i=1}^N \left[\log \binom{n_i}{y_i} - \log \sum_{j=0}^{n_i} \binom{n_i}{j} \left(\frac{\pi_i}{1 - \pi_i} \right)^j \omega^{(n_i-j)j} + y_i \log \frac{\pi_i}{1 - \pi_i} + (n_i - y_i) y_i \log \omega \right],$$

läßt sich die log-Likelihoodfunktion im linearen Prädiktor $\eta_i = \mathbf{x}_i^t \boldsymbol{\beta}$ somit schreiben als

$$l(\boldsymbol{\beta}, \alpha | \mathbf{y}) = \sum_{i=1}^N \left[\log \binom{n_i}{y_i} - \log \sum_{j=0}^{n_i} \binom{n_i}{j} e^{j \mathbf{x}_i^t \boldsymbol{\beta}} e^{(n_i-j)j\alpha} + y_i \mathbf{x}_i^t \boldsymbol{\beta} + (n_i - y_i) y_i \alpha \right]. \quad (3.18)$$

Für den Parameter ω wurde hier die log-lineare Reparametrisierung

$$\omega = \exp(\alpha)$$

verwendet, welche $\omega > 0$ garantiert. Bei einer Modellschätzung soll nun zum einen der Parametervektor $\boldsymbol{\beta}$ geschätzt werden, zum anderen soll auch für α ein Punktschätzer ermittelt werden.

Die partiellen Ableitungen erster Ordnung der log-Likelihoodfunktion lassen sich ermitteln als

$$\frac{\partial}{\partial \beta_k} l(\boldsymbol{\beta}, \alpha | \mathbf{y}) = \sum_{i=1}^N x_{ik} \left[y_i - \frac{\sum_{j=0}^{n_i} j \binom{n_i}{j} e^{j \mathbf{x}_i^t \boldsymbol{\beta}} e^{(n_i-j)j\alpha}}{\sum_{j=0}^{n_i} \binom{n_i}{j} e^{j \mathbf{x}_i^t \boldsymbol{\beta}} e^{(n_i-j)j\alpha}} \right]$$

für $k = 0, \dots, p-1$ und

$$\frac{\partial}{\partial \alpha} l(\boldsymbol{\beta}, \alpha | \mathbf{y}) = \sum_{i=1}^N \left[(n_i - y_i) y_i - \frac{\sum_{j=0}^{n_i} (n_i - j) j \binom{n_i}{j} e^{j \mathbf{x}_i^t \boldsymbol{\beta}} e^{(n_i-j)j\alpha}}{\sum_{j=0}^{n_i} \binom{n_i}{j} e^{j \mathbf{x}_i^t \boldsymbol{\beta}} e^{(n_i-j)j\alpha}} \right].$$

Übersichtlicher gestaltet sich die Diskussion, wenn anstelle der ausgeschriebenen Summanden des Normierungsterms die Vereinfachung $d_j^{(\boldsymbol{\beta}, \alpha)}$ geschrieben wird. Dann lassen sich die partiellen Ableitungen erster Ordnung vereinfachen zu

$$\frac{\partial}{\partial \beta_k} l(\boldsymbol{\beta}, \alpha | \mathbf{y}) = \sum_{i=1}^N x_{ik} \left[y_i - \frac{\sum_{j=0}^{n_i} j d_j^{(\boldsymbol{\beta}, \alpha)}}{\sum_{j=0}^{n_i} d_j^{(\boldsymbol{\beta}, \alpha)}} \right]$$

für $k = 0, \dots, p-1$ und

$$\frac{\partial}{\partial \alpha} l(\boldsymbol{\beta}, \alpha | \mathbf{y}) = \sum_{i=1}^N \left[(n_i - y_i) y_i - \frac{\sum_{j=0}^{n_i} (n_i - j) j d_j^{(\boldsymbol{\beta}, \alpha)}}{\sum_{j=0}^{n_i} d_j^{(\boldsymbol{\beta}, \alpha)}} \right].$$

Für die partiellen Ableitungen zweiter Ordnung gilt in weiterer Folge

$$\frac{\partial^2 l}{\partial \beta_k \partial \beta_h} l(\boldsymbol{\beta}, \alpha | \mathbf{y}) = - \sum_{i=1}^N x_{ik} x_{ih} \frac{\sum_{j=0}^{n_i} j^2 d_j^{(\boldsymbol{\beta}, \alpha)} \sum_{j=0}^{n_i} d_j^{(\boldsymbol{\beta}, \alpha)} - \left\{ \sum_{j=0}^{n_i} j d_j^{(\boldsymbol{\beta}, \alpha)} \right\}^2}{\left\{ \sum_{j=0}^{n_i} d_j^{(\boldsymbol{\beta}, \alpha)} \right\}^2}$$

für $k, h = 0, \dots, p-1$,

$$\frac{\partial^2 l}{\partial \beta_k \partial \alpha} l(\boldsymbol{\beta}, \alpha | \mathbf{y}) = - \sum_{i=1}^N x_{ik} \frac{\sum_{j=0}^{n_i} (n_i - j) j^2 d_j^{(\boldsymbol{\beta}, \alpha)} \sum_{j=0}^{n_i} d_j^{(\boldsymbol{\beta}, \alpha)} - \sum_{j=0}^{n_i} j d_j^{(\boldsymbol{\beta}, \alpha)} \sum_{j=0}^{n_i} (n_i - j) j d_j^{(\boldsymbol{\beta}, \alpha)}}{\left\{ \sum_{j=0}^{n_i} d_j^{(\boldsymbol{\beta}, \alpha)} \right\}^2}$$

für $k = 0, \dots, p-1$ und

$$\frac{\partial^2 l}{\partial \alpha^2} l(\boldsymbol{\beta}, \alpha | \mathbf{y}) = - \sum_{i=1}^N \frac{\sum_{j=0}^{n_i} [(n_i - j) j]^2 d_j^{(\boldsymbol{\beta}, \alpha)} \sum_{j=0}^{n_i} d_j^{(\boldsymbol{\beta}, \alpha)} - \left\{ \sum_{j=0}^{n_i} (n_i - j) j d_j^{(\boldsymbol{\beta}, \alpha)} \right\}^2}{\left\{ \sum_{j=0}^{n_i} d_j^{(\boldsymbol{\beta}, \alpha)} \right\}^2}.$$

Diese partiellen Ableitungen werden in weiterer Folge bei der Parameterschätzung eine wichtige Rolle spielen.

3.4.1 Schätzung von ω

Außerdem muss an dieser Stelle noch darauf hingewiesen werden, dass die Modellschätzung in der hier beschriebenen Parametrisierung Punktschätzer für $\boldsymbol{\beta}$ und α liefert. Um aus $\hat{\alpha}$ den Schätzer für ω zu errechnen, ist nur die Relation $\hat{\omega} = \exp(\hat{\alpha})$ vonnöten. Um jedoch die Standardfehler für $\hat{\omega}$ zu erhalten, muss die Delta Methode verwendet werden, die unter anderem in Agresti (2002) vorgestellt wird. Ausgehend von der Tatsache, dass für $\hat{\alpha}$ als Maximum Likelihood Schätzer

$$\sqrt{n}(\hat{\alpha} - \alpha) \xrightarrow{d} N(0, \sigma_\alpha^2)$$

gilt, läßt sich die Verteilung von $\hat{\omega}$ als Funktion in $\hat{\alpha}$ ermitteln. Die Funktion $\omega = \exp(\alpha)$ ist stetig differenzierbar und es läßt sich mittels Taylor-Entwicklungen zeigen, dass mit $g(\alpha) = \exp(\alpha)$ in etwa

$$\sqrt{n}(g(\hat{\alpha}) - g(\alpha)) \approx \sqrt{n}(\hat{\alpha} - \alpha) g'(\alpha)$$

gilt. Daher folgt für die Verteilung von $\omega = \exp(\alpha)$

$$\sqrt{n}(\hat{\omega} - \omega) \xrightarrow{d} N\left(0, \sigma_\alpha^2 [g'(\alpha)]^2\right)$$

beziehungsweise

$$\sqrt{n}(\hat{\omega} - \omega) \xrightarrow{d} N\left(0, \sigma_\alpha^2 \exp(2\alpha)\right).$$

Bevor nun eine Modellschätzung mit multiplikativ binomialverteilten Zufallsvariablen in Angriff genommen wird, soll noch eine zweite Verteilung betrachtet werden. Diese sogenannte doppelte Binomialverteilung erlaubt ebenso wie die multiplikative Verallgemeinerung der Binomialverteilung, Daten mit größerer respektive kleinerer Variabilität als jener der Binomialverteilung zu beschreiben. Daher wird sie im nun folgenden Abschnitt näher betrachtet.

Die doppelte Binomialverteilung

Dieser Abschnitt ist einer weiteren Verteilung gewidmet, welche ebenfalls anhand eines zweiten Parameters sowohl auf Über- als auch auf Unterdispersion im Vergleich zur Binomialverteilung zu reagieren vermag. Die Verteilung wurde von Efron (1986) - basierend auf seinen früheren Arbeiten - als doppelte Exponentialfamilie vorgestellt. Die Bezeichnung doppelte Exponentialfamilie wurde dabei gewählt, da die Verteilung für jeden der beiden Parameter die Eigenschaften der Exponentialfamilie erfüllt.

Nach einer Definition der doppelten Binomialverteilung aus der doppelten Exponentialfamilie bringt dieses Kapitel eine ausführliche Herleitung von Erwartungswert und Varianz der Verteilung. Abschließend wird die Verteilung in den Kontext der generalisierten linearen Modelle eingebettet und wie bei der multiplikativen Binomialverteilung die ersten beiden partiellen Ableitungen aufgestellt.

4.1 Definition

Definiert wird die doppelte Exponentialfamilie für eine Zufallsvariable y als

$$f(y|\zeta, \phi) = c(\zeta, \phi) \sqrt{\phi} f(y|\zeta)^\phi f(y|\hat{\zeta}_s)^{1-\phi}, \quad (\phi > 0) \quad (4.1)$$

in den Parametern ζ und nicht negativem ϕ , wobei $\hat{\zeta}_s$ für den Punktschätzer ζ eines gesättigten Modells steht. Bei der Funktion $c(\zeta, \phi)$ handelt es sich um eine Normierungsfunktion.

Seit der Veröffentlichung der Arbeit 1986 haben verschiedene Autoren die vorgeschlagene Familie aufgenommen, so unter anderem Dey, Gelfand und Peng (1997). Gelfand und Dalal (1990) betonen, dass die doppelte Exponentialfamilie eng verwandt ist mit den bei Jørgensen (1987) behandelten Exponential Dispersion Models. Ebenso wird zum Beispiel in Lee und Nelder (2000) die Nähe zu den Extended Quasi Likelihood Models von Nelder und Pregibon (1987) hervorgehoben.

In dieser Arbeit soll nun die doppelte Binomialverteilung von weiterem Interesse sein. Diese wird in der bereits genannten Arbeit von Efron nicht explizit vorgestellt, ist jedoch in anderen

Werken zu finden. So verwenden sie zum Beispiel Lindsey und Altham (1998) in ihrer Arbeit über das Geschlechterverhältnis bei Geburtsraten oder Lindsey (1999) in seiner allgemeineren Literaturarbeit zum Thema Generalisierte Lineare Modelle.

Aus (4.1) ergibt sich durch Einsetzen der Wahrscheinlichkeitsfunktion der Binomialverteilung

$$\begin{aligned} f(y|\pi, \phi) &= c(\pi, \phi) \sqrt{\phi} \left(\binom{n}{y} \pi^y (1-\pi)^{n-y} \right)^\phi \left(\binom{n}{y} \left(\frac{y}{n}\right)^y \left(1 - \frac{y}{n}\right)^{n-y} \right)^{1-\phi} \\ &= c(\pi, \phi) \sqrt{\phi} \binom{n}{y} \pi^{y\phi} (1-\pi)^{(n-y)\phi} \left(\frac{n^n}{y^y (n-y)^{n-y}} \right)^{\phi-1} \end{aligned}$$

die Wahrscheinlichkeitsfunktion der doppelten Binomialverteilung in den Parametern π und ϕ und dem Normierungsterm $c(\pi, \phi)$.

4.2 Eigenschaften

Die in dieser Arbeit betrachtete doppelte Binomialverteilung lässt sich notieren als

$$P(y) = \frac{\binom{n}{y} \left(\frac{n^n}{y^y (n-y)^{n-y}} \right)^{\phi-1} \pi^{y\phi} (1-\pi)^{(n-y)\phi}}{\sum_{j=0}^n \binom{n}{j} \left(\frac{n^n}{j^j (n-j)^{n-j}} \right)^{\phi-1} \pi^{j\phi} (1-\pi)^{(n-j)\phi}} \quad (4.2)$$

für $y = 0, 1, \dots, n$ beziehungsweise vereinfacht als

$$P(y) = \frac{\binom{n}{y} (y^y (n-y)^{n-y})^{1-\phi} \left(\frac{\pi}{1-\pi} \right)^{y\phi}}{\sum_{j=0}^n \binom{n}{j} (j^j (n-j)^{n-j})^{1-\phi} \left(\frac{\pi}{1-\pi} \right)^{j\phi}} \quad (4.3)$$

in den beiden Parametern π und ϕ . Mit π wird wie üblich die Erfolgswahrscheinlichkeit eines Einzelereignisses beschrieben, der Parameter ϕ erlaubt nun eine Modellierung von Über- und Unterdispersion. So ist im Fall von Unterdispersion $0 < \phi < 1$, im Fall von Überdispersion gilt $\phi > 1$. Für $\phi = 1$ ergibt sich abermals die klassische Binomialverteilung.

Dass die doppelte Binomialverteilung wie bereits in der Einleitung dieses Kapitels erwähnt ein Mitglied der zweiparametrischen Exponentialfamilie ist lässt sich einsehen, wenn die Wahrscheinlichkeitsfunktion aus (4.2) umgeformt wird zu

$$\begin{aligned} P(y) &= \binom{n}{y} y^y (n-y)^{n-y} \cdot \frac{1}{\sum_{j=0}^n \binom{n}{j} (j^j (n-j)^{n-j})^{1-\phi} \left(\frac{\pi}{1-\pi} \right)^{j\phi}} \\ &\quad \exp \left(-[y \log y + (n-y) \log(n-y)] \phi + y\phi \log \frac{\pi}{1-\pi} \right). \end{aligned}$$

Diese Darstellung entspricht der in (2.3) geforderten Form mit

$$\begin{aligned} h(y) &= \binom{n}{y} y^y (n-y)^{n-y} \quad \text{für } y = 0, 1, \dots, n, \\ t_1(y) &= -[y \log y + (n-y) \log(n-y)] \quad \text{mit } w_1(\boldsymbol{\theta}) = \phi, \\ t_2(y) &= y \quad \text{mit } w_2(\boldsymbol{\theta}) = \phi \log \frac{\pi}{1-\pi}, \end{aligned}$$

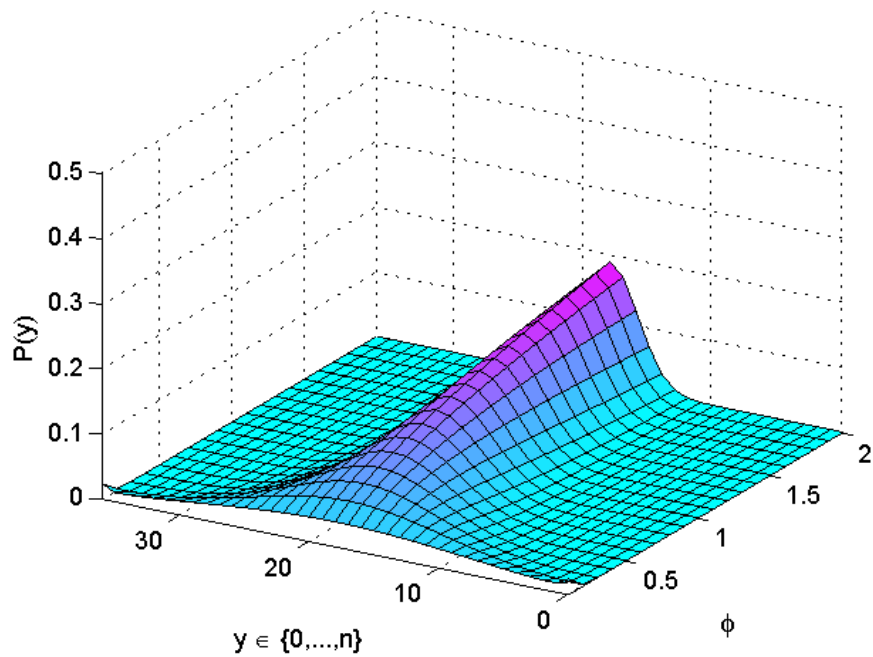


Abbildung 4.1.: Wahrscheinlichkeitsfunktion der doppelten Binomialverteilung bei $\pi = 0,5$ über verschiedene Werten von ϕ und $n = 36$.

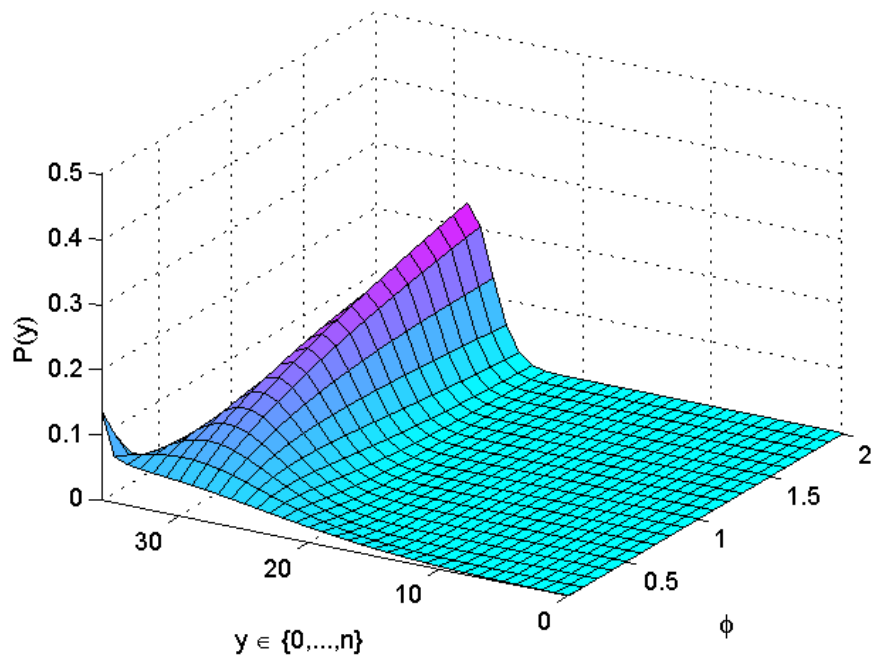


Abbildung 4.2.: Wahrscheinlichkeitsfunktion der doppelten Binomialverteilung bei $\pi = 0,8$ über verschiedene Werten von ϕ und $n = 36$.

und

$$c(\theta) = \frac{1}{\sum_{j=0}^n \binom{n}{j} (j^j (n-j)^{n-j})^{1-\phi} \left(\frac{\pi}{1-\pi}\right)^{j\phi}}$$

für $\theta = (\pi, \phi)$.

Um das Verhalten dieser doppelten Binomialverteilung besser verstehen zu können, zeigen die beiden Grafiken in den Abbildungen 4.1 und 4.2 die Wahrscheinlichkeitsfunktionen für $\pi = 0,5$ und $\pi = 0,8$. Die Wahrscheinlichkeitsfunktion der klassischen Binomialverteilung ist bei $\phi = 1$ zu finden. Im Gegensatz zur multiplikativen Verallgemeinerung der Binomialverteilung verändert sich die doppelte Binomialverteilung für Werte von ϕ etwas größer oder kleiner als 1 deutlich gemächlicher.

Im Gegensatz zur multiplikativen Binomialverteilung zerfällt hier der Ausdruck innerhalb der Exponentialfunktion nicht in Terme rein in θ_1 oder θ_2 . Statt dessen treten hier gemischte Summanden auf.

Um die Formeln (2.5) und (2.6) für Erwartungswert und Varianz verwenden zu können, muss die Darstellung als Exponentialfamilie etwas umständlich reparametrisiert werden, um eine Darstellung in kanonischer Form zu erhalten. Dafür seien

$$\nu_1 = \phi \quad \text{und} \quad \nu_2 = \phi \log \frac{\pi}{1-\pi}$$

die kanonischen Parameter der Verteilung und es ergibt sich für die Wahrscheinlichkeitsfunktion als zweiparametrische Exponentialfamilie in kanonischer Form

$$P(y) = \binom{n}{y} y^y (n-y)^{n-y} \cdot \frac{1}{\sum_{j=0}^n \binom{n}{j} (j^j (n-j)^{n-j})^{1-\nu_1} e^{j\nu_2}} \cdot \exp\left(-[y \log y + (n-y) \log(n-y)]\nu_1 + y\nu_2\right).$$

Die vorliegende Exponentialfamilie ist linear bezüglich ν_2 . Ausgehend von der in y konstanten Funktion $c^*(\nu)$ aus der Darstellung der Exponentialfamilie in kanonischer Form lassen sich nun Erwartungswert und Varianz errechnen, indem die Ableitungen von

$$-\log(c^*(\nu)) = -\log\left(\frac{1}{\sum_{j=0}^n \binom{n}{j} (j^j (n-j)^{n-j})^{1-\nu_1} e^{j\nu_2}}\right)$$

betrachtet werden. Der Erwartungswert ergibt sich somit aus

$$E(y) = -\frac{\partial}{\partial \nu_2} \log(c^*(\nu)) = \frac{\partial}{\partial \nu_2} \log \sum_{j=0}^n \binom{n}{j} (j^j (n-j)^{n-j})^{1-\nu_1} e^{j\nu_2} = \frac{\sum_{j=0}^n j d_j^{(\nu)}}{\sum_{j=0}^n d_j^{(\nu)}},$$

wobei zur besseren Lesbarkeit die einzelnen Summanden im Parameter ν durch $d_j^{(\nu)}$ ersetzt wurden, also

$$d_j^{(\nu)} := \binom{n}{j} (j^j (n-j)^{n-j})^{1-\nu_1} e^{j\nu_2}$$

gilt. Durch Rücksubstituieren folgt für den Erwartungswert einer doppelt binomialverteilten Zufallsvariablen

$$E(y) = \frac{\sum_{j=0}^n j d_j^{(\pi, \phi)}}{\sum_{j=0}^n d_j^{(\pi, \phi)}} \quad (4.4)$$

für

$$d_j^{(\pi, \phi)} = \binom{n}{j} (j^j (n-j)^{n-j})^{1-\phi} \left(\frac{\pi}{1-\pi} \right)^{j\phi}.$$

Ebenso läßt sich aus (2.6) eine geschlossene Darstellung der Varianz ermitteln, indem die zweite partielle Ableitung von $-\log c^*(\boldsymbol{\nu})$ nach ν_2 gebildet wird. Für die Varianz der doppelten Binomialverteilung ergibt sich also

$$\text{Var}(y) = \frac{\sum_{j=0}^n j^2 d_j^{(\pi, \phi)} \sum_{j=0}^n d_j^{(\pi, \phi)} - \left\{ \sum_{j=0}^n j d_j^{(\pi, \phi)} \right\}^2}{\left\{ \sum_{j=0}^n d_j^{(\pi, \phi)} \right\}^2}. \quad (4.5)$$

Abermals sollen Visualisierungen von Erwartungswert und Varianz helfen, die Zusammenhänge zwischen den einzelnen Parametern besser zu überblicken. Die in Abbildung 4.3 gezeigten Plots beschreiben das Verhalten des Erwartungswerts doppelt binomialverteilter Zufallsvariablen in Abhängigkeit von $\pi \in [0; 1]$ und $0 < \phi < 3$. Wie sich zeigt, ist der Erwartungswert zwar nicht konstant in ϕ , verändert sich für kleine Änderungen in ϕ jedoch nur extrem geringfügig.

Das Verhalten der Standardabweichung der doppelten Binomialverteilung ist in den Abbildungen 4.4 und 4.5 dargestellt. Im Gegensatz zur multiplikativen Binomialverteilung zeigt sich hier, dass eine klare Trennung zwischen Unter- und Überdispersion rein anhand des Wertes von ϕ möglich ist. So besitzt die Verteilung für $\phi < 1$ eine größere Standardabweichung als die klassische Binomialverteilung, wohingegen sie für $\phi > 1$ eine kleinere Standardabweichung und somit Unterdispersion aufweist.

Dieser Sachverhalt ist auch in Abbildung 4.6 nochmals dargestellt, diesmal jedoch für die Varianz. Um die großen Wertebereiche darstellen zu können, wurde für die z-Achse abermals eine logarithmische Skalierung gewählt. Die doppelte Binomialverteilung weist also im Vergleich zur klassischen Binomialverteilung Unterdispersion auf, wenn $\phi > 1$ ist und Überdispersion, wenn der Wert von $\phi < 1$ ist.

Nun muss auch für die doppelte Exponentialfamilie erwähnt werden, dass die Summe zweier unabhängig und identisch doppelt binomialverteilter Zufallsvariablen nicht mehr doppelt binomialverteilt ist. Für die dafür benötigte momentenerzeugende Funktion gilt auch in diesem Fall

$$M_y(s) = \frac{c^*(\nu_1, \nu_2)}{c^*(\nu_1, \nu_2 + s)}$$

für die für ν_2 in y lineare, kanonische Exponentialfamilie. Daher ergibt sich aus

$$c^*(\nu_1, \nu_2) = \frac{1}{\sum_{j=0}^n \binom{n}{j} (j^j (n-j)^{n-j})^{1-\nu_1} e^{j\nu_2}}$$

für $M_y(s)$ im Fall dieser Verteilung

$$M(s) = \frac{\sum_{j=0}^n \binom{n}{j} (j^j (n-j)^{n-j})^{1-\nu_1} e^{j\nu_2} e^{js}}{\sum_{j=0}^n \binom{n}{j} (j^j (n-j)^{n-j})^{1-\nu_1} e^{j\nu_2}}$$

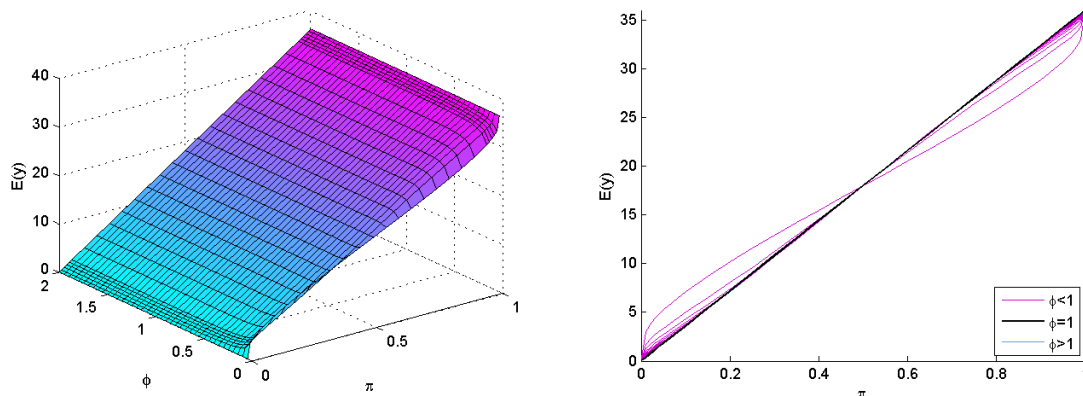


Abbildung 4.3.: Visualisierung des Erwartungswerts in Abhängigkeit von π und ϕ .

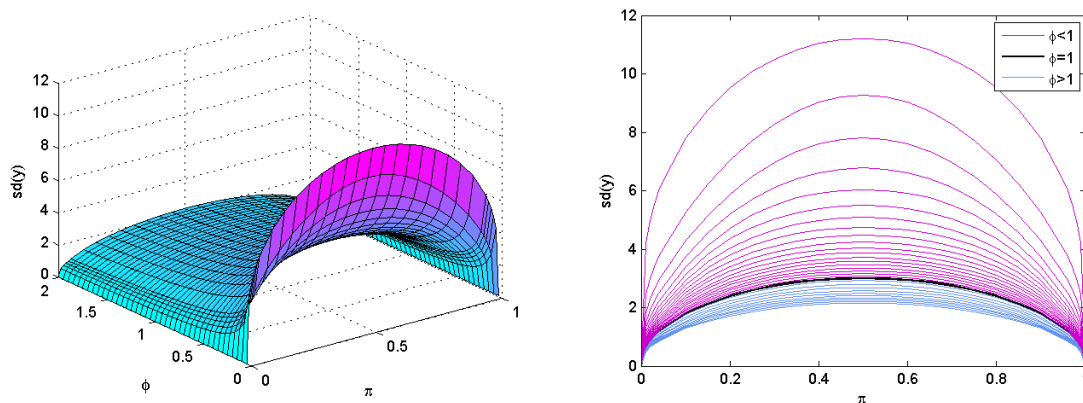


Abbildung 4.4.: Visualisierung der Standardabweichung in Abhängigkeit von π und ϕ .

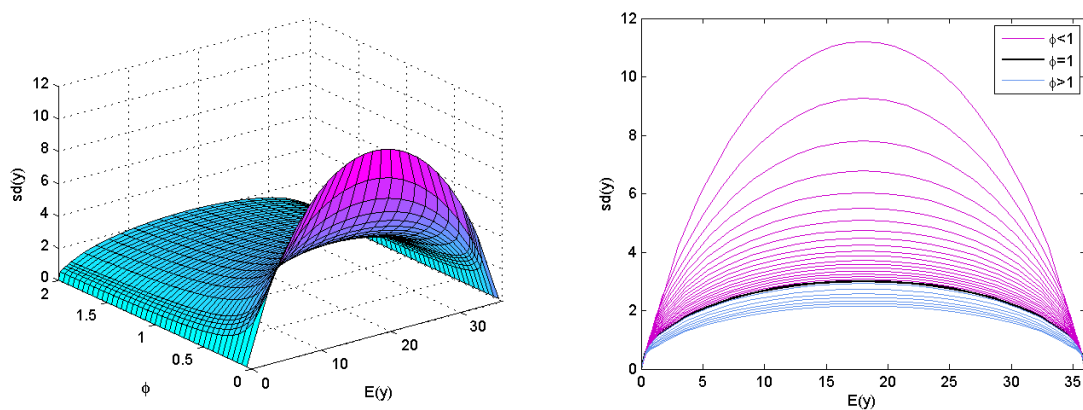


Abbildung 4.5.: Visualisierung der Standardabweichung in Abhängigkeit vom Erwartungswert und ϕ .

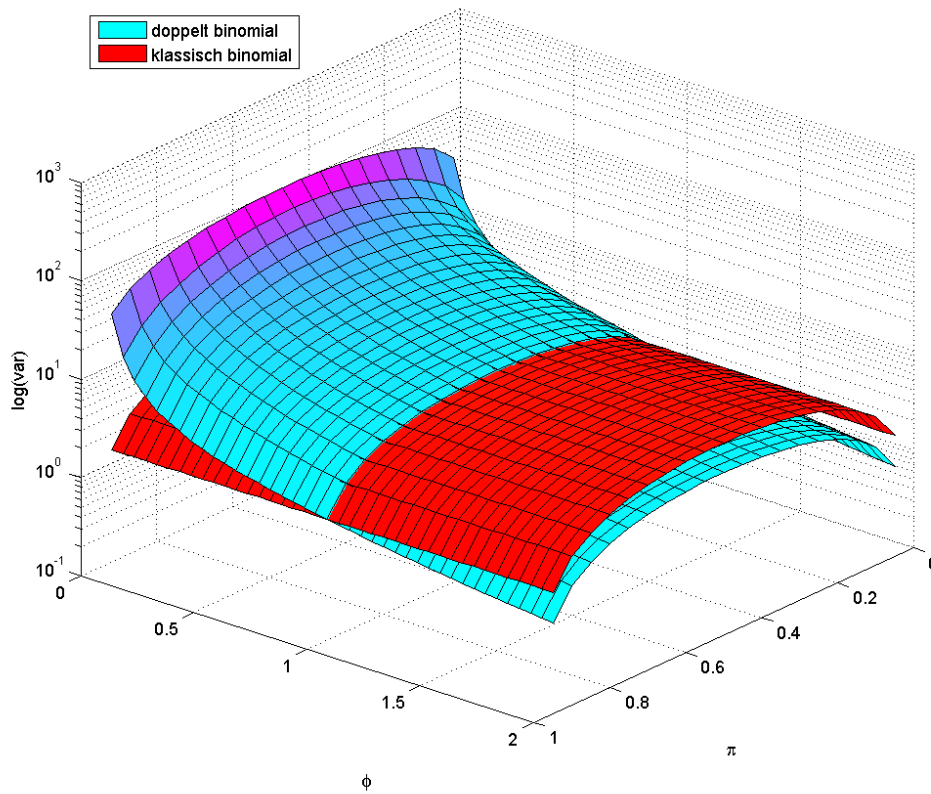


Abbildung 4.6.: Vergleich der Varianz der doppelten Binomialverteilung mit jener der klassischen Binomialverteilung.

Für die Summe zweier unabhängig und identisch doppel binomialverteilter Zufallsvariablen y_1 und y_2 folgt dann

$$M_{y_1+y_2}(s) = \left(\frac{\sum_{j=0}^n \binom{n}{j} (j^j (n-j)^{n-j})^{1-\nu_1} e^{j\nu_2} e^{js}}{\sum_{j=0}^n \binom{n}{j} (j^j (n-j)^{n-j})^{1-\nu_1} e^{j\nu_2}} \right)^2.$$

Diese resultierende Funktion ist keine momentenerzeugende Funktion der doppelten Binomialverteilung mehr. Daher ist auch diese Verallgemeinerung der Binomialverteilung nicht abgeschlossen bezüglich ihrer Faltung.

4.3 Die doppelte Binomialverteilung in der logistischen Regression

Für N doppelt binomialverteilte Beobachtungen y_i mit Parametern π_i und einem gemeinsamen ϕ ist die log-Likelihoodfunktion durch

$$l(\boldsymbol{\pi}, \phi | \mathbf{y}) = \sum_{i=1}^N \left[\log \binom{n_i}{y_i} + y_i \log y_i + (n_i - y_i) \log(n_i - y_i) \right. \\ \left. - \log \sum_{j=0}^{n_i} \binom{n_i}{j} \left(j^j (n_i - j)^{n_i - j} \right)^{1 - \phi} \left(\frac{\pi_i}{1 - \pi_i} \right)^{j\phi} \right. \\ \left. + y_i \phi \log \frac{\pi_i}{1 - \pi_i} - [y_i \log y_i + (n_i - y_i) \log(n_i - y_i)] \phi \right]$$

gegeben. Um die Erfolgswahrscheinlichkeiten $\boldsymbol{\pi}$ zu beschreiben, wird abermals der logistische Link

$$\log \frac{\pi_i}{1 - \pi_i} = \eta_i = \mathbf{x}_i^t \boldsymbol{\beta}$$

gewählt. Der zweite Parameter wird wie bei der multiplikativen Binomialverteilung anhand des log-linearen Links

$$\phi = \exp(\alpha)$$

modelliert, um $\phi > 0$ sicherzustellen. Daher gilt für die log-Likelihoodfunktion in diesem Fall

$$l(\boldsymbol{\beta}, \alpha | \mathbf{y}) = \sum_{i=1}^N \left[\log \binom{n_i}{y_i} + y_i \log y_i + (n_i - y_i) \log(n_i - y_i) \right. \\ \left. - \log \sum_{j=0}^{n_i} \binom{n_i}{j} \left(j^j (n_i - j)^{n_i - j} \right)^{1 - e^\alpha} e^{j e^\alpha \mathbf{x}_i^t \boldsymbol{\beta}} \right. \\ \left. + y e^\alpha \mathbf{x}_i^t \boldsymbol{\beta} - [y_i \log y_i + (n_i - y_i) \log(n_i - y_i)] \phi \right]. \quad (4.6)$$

Nun können die partiellen Ableitungen erster Ordnung nach β_k und α gebildet werden und lassen sich notieren als

$$\frac{\partial}{\partial \beta_k} l(\boldsymbol{\beta}, \alpha | \mathbf{y}) = \sum_{i=1}^N e^\alpha x_{ik} \left[y_i - \frac{\sum_{j=0}^{n_i} j d_j^{(\boldsymbol{\beta}, \alpha)}}{\sum_{j=0}^{n_i} d_j^{(\boldsymbol{\beta}, \alpha)}} \right]$$

für $k = 0, \dots, p - 1$ und

$$\frac{\partial}{\partial \alpha} l(\boldsymbol{\beta}, \alpha | \mathbf{y}) = \sum_{i=1}^N e^\alpha \left[y_i \mathbf{x}_i^t \boldsymbol{\beta} - y_i \log y_i - (n_i - y_i) \log(n_i - y_i) \right. \\ \left. - \frac{\sum_{j=0}^{n_i} [j \mathbf{x}_i^t \boldsymbol{\beta} - j \log j - (n_i - j) \log(n_i - j)] d_j^{(\boldsymbol{\beta}, \alpha)}}{\sum_{j=0}^{n_i} d_j^{(\boldsymbol{\beta}, \alpha)}} \right],$$

wobei der Normierungsterm in den Parametern $\boldsymbol{\beta}$ und α abermals zu $d_j^{(\boldsymbol{\beta}, \alpha)}$ zusammengefasst wurde mit

$$d_j^{(\boldsymbol{\beta}, \alpha)} := \binom{n_i}{j} \left(j^j (n_i - j)^{n_i - j} \right)^{1 - e^\alpha} e^{j e^\alpha \mathbf{x}_i^t \boldsymbol{\beta}}.$$

Für eine handlichere Darstellung der Ableitung der log-Likelihoodfunktion nach α soll der Ausdruck

$$a(y_i) := y_i \mathbf{x}_i^t \boldsymbol{\beta} - y_i \log y_i - (n_i - y_i) \log(n_i - y_i)$$

zusammenfassen und es folgt

$$\frac{\partial}{\partial \alpha} l(\boldsymbol{\beta}, \alpha | \mathbf{y}) = \sum_{i=1}^N e^\alpha \left[a(y) - \frac{\sum_{j=0}^{n_i} a(j) d_j^{(\boldsymbol{\beta}, \alpha)}}{\sum_{j=0}^{n_i} d_j^{(\boldsymbol{\beta}, \alpha)}} \right].$$

Außerdem lassen sich die partiellen Ableitungen zweiter Ordnung angeben, die in weiterer Folge zur Berechnung der Standardabweichungen der gefundenen Punktschätzer benötigt werden. Es sind dies

$$\frac{\partial^2}{\partial \beta_k \partial \beta_h} l(\boldsymbol{\beta}, \alpha | \mathbf{y}) = - \sum_{i=1}^N e^{2\alpha} x_{ik} x_{ih} \frac{\sum_{j=0}^{n_i} j^2 d_j^{(\boldsymbol{\beta}, \alpha)} \sum_{j=0}^{n_i} d_j^{(\boldsymbol{\beta}, \alpha)} - \left\{ \sum_{j=0}^{n_i} j d_j^{(\boldsymbol{\beta}, \alpha)} \right\}^2}{\left\{ \sum_{j=0}^{n_i} d_j^{(\boldsymbol{\beta}, \alpha)} \right\}^2},$$

$$\begin{aligned} \frac{\partial^2}{\partial \beta_k \partial \alpha} l(\boldsymbol{\beta}, \alpha | \mathbf{y}) = & \sum_{i=1}^N e^\alpha x_{ik} \left[y_i - \frac{\sum_{j=0}^{n_i} j d_j^{(\boldsymbol{\beta}, \alpha)}}{\sum_{j=0}^{n_i} d_j^{(\boldsymbol{\beta}, \alpha)}} \right. \\ & \left. - e^\alpha \frac{\sum_{j=0}^{n_i} j a(j) d_j^{(\boldsymbol{\beta}, \alpha)} \sum_{j=0}^{n_i} d_j^{(\boldsymbol{\beta}, \alpha)} - \sum_{j=0}^{n_i} j d_j^{(\boldsymbol{\beta}, \alpha)} \sum_{j=0}^{n_i} a(j) d_j^{(\boldsymbol{\beta}, \alpha)}}{\left\{ \sum_{j=0}^{n_i} d_j^{(\boldsymbol{\beta}, \alpha)} \right\}^2} \right] \end{aligned}$$

und

$$\begin{aligned} \frac{\partial^2}{\partial \phi^2} l(\boldsymbol{\beta}, \phi | \mathbf{y}) = & \sum_{i=1}^N e^\alpha \left[a(y) - \frac{\sum_{j=0}^{n_i} a(j) d_j^{(\boldsymbol{\beta}, \alpha)}}{\sum_{j=0}^{n_i} d_j^{(\boldsymbol{\beta}, \alpha)}} \right. \\ & \left. - e^\alpha \frac{\sum_{j=0}^{n_i} a(j)^2 d_j^{(\boldsymbol{\beta}, \alpha)} \sum_{j=0}^{n_i} d_j^{(\boldsymbol{\beta}, \alpha)} - \left\{ \sum_{j=0}^{n_i} a(j) d_j^{(\boldsymbol{\beta}, \alpha)} \right\}^2}{\left\{ \sum_{j=0}^{n_i} d_j^{(\boldsymbol{\beta}, \alpha)} \right\}^2} \right]. \end{aligned}$$

Somit wurden nun zwei Verallgemeinerungen der Binomialverteilung vorgestellt, die beide mit Unter- oder Überdispersion im Vergleich zur klassischen Binomialverteilung arbeiten können. Im folgenden Kapitel wird nun darauf eingegangen, wie mithilfe der jeweiligen log-Likelihoodfunktion und ihren partiellen Ableitungen Punktschätzer ermittelt werden können.

Simulationen und Code

Dieser Abschnitt beschreibt, wie mittels der Maximum Likelihood Methode die Modellparameter unter Annahme der doppelten sowie der multiplikativen Binomialverteilung geschätzt werden können.

Die für die Maximum Likelihood Schätzung benötigten log-Likelihoodfunktionen der beiden betrachteten Verteilungen sowie ihre ersten und zweiten Ableitungen wurden bereits in den beiden vorangegangenen Kapiteln aufgestellt. Um nun aus diesen die Parameterschätzer zu erhalten, wird ein Optimierungsverfahren benötigt, welches das Maxima der betrachteten log-Likelihoodfunktion berechnen kann. Daher beginnt dieses Kapitel auch mit einer Diskussion der verwendeten Optimierungsroutinen. Danach werden die in R erstellten Funktionen zum Erzeugen multiplikativ sowie doppelt binomialverteilter Zufallsvariablen vorgestellt.

Simulationsergebnisse sowohl unter multiplikativer als auch unter doppelter Binomialverteilung werden in den Abschnitten 5.3 und 5.4 ausführlich besprochen. Dabei wird in beiden Fällen mit einer Simulationsserie für Erwartungswert und Varianz der jeweiligen Verteilung begonnen. Danach werden die Ergebnisse einiger Monte-Carlo Simulationen aus einer größeren Simulationsreihe vorgestellt, um das Verhalten der Verteilungen innerhalb eines Regressionskontextes darzustellen.

Den Abschluss macht eine Vorstellung der für eine derartige Modellschätzung geschriebenen R-Funktionen.

Für diese Simulationsreihen sowie die späteren Modellschätzungen wurde die frei zugängliche und speziell auf Probleme der Statistik zugeschnittene Software R verwendet. Die gezeigten Ergebnisse wurden unter der 32-Bit Version von R 2.11.1 auf einer 64-Bit Maschine¹ erstellt. Die Laufzeiten für eine derartige Simulation liegen auf diesem Rechner je nach verwendeter Verteilungsannahme bei bis zu einigen Stunden. Die schiereren Datenmengen (vgl. Kapitel 6) stellen dabei für das Programm keine Stabilitätsprobleme dar. Besonders laufzeitintensiv ist vor allem die in Abschnitt 5.2 angesprochene Erzeugung von Zufallszahlen, da sie in R und nicht in besser dafür geeigneten Umgebungen durchgeführt wird.

¹Intel Core2 Duo 2,99 GHz E8400, 8 GB RAM, Windows 7 Professional (SP1)

5.1 Diskussion der Optimierungsroutine

Um mittels Maximum Likelihood Verfahren die Punktschätzer eines Modells zu erhalten, muss ein Maximierungsproblem gelöst werden.

Die von R zur Verfügung gestellte Funktion `optim()`² ist eine allgemein verwendbare Optimierungsroutine für nicht-lineare Funktionen in mehreren Variablen ohne Nebenbedingungen. Standardmäßig verwendet die Funktion ein von Nelder und Mead (1965) vorgestelltes Optimierungsverfahren, welches ohne die Verwendung von Gradienteninformationen ein Minimum der Zielfunktion finden kann. Laut der Beschreibung der Funktion `optim()` ist diese Methode obwohl stabil jedoch sehr langsam. Und da sich von den log-Likelihoodfunktionen der multiplikativen sowie der doppelten Binomialverteilung die ersten Ableitungen bilden lassen, wird hier anstelle der Standardmethode ein Optimierungsverfahren gewählt, welches neben der Zielfunktion auch den Gradienten derselben für die Suche nach dem Extremum verwendet.

Ein solches Verfahren, welches auch innerhalb der Optimierungsroutine `optim()` zur Verfügung steht, ist eine von Broyden, Fletcher, Goldfarb und Shanno vorgestellte Variante des Quasi-Newton-Verfahrens. Dieses Verfahren wird üblicherweise mit den Initialen der vier Autoren abgekürzt und als BFGS-Methode bezeichnet.

5.1.1 Abriss zum Newton- und Quasi-Newton-Verfahren

Das klassische Newton-Verfahren stellt eine grundlegende Methode zur Verfügung, eine Nullstelle einer (ein- oder mehrdimensionalen) Funktion $f(x)$ über die Iterationsvorschrift

$$x_{k+1} = x_k - \left(\frac{\partial}{\partial \theta} f(x_k) \right)^{-1} f(x_k)$$

anzunähern. Dabei steht x_k für den Näherungswert der Nullstelle im k -ten Iterationsschritt.

Die Suche nach einem lokalen Minimum der Funktion $f(x)$ entspricht nun der Suche nach einer Nullstelle der Ableitung $\frac{\partial}{\partial \theta} f(x)$ derselben Funktion. In diesem Fall spricht man vom Quasi-Newton-Verfahren, da dieselben Ideen wie im Newton-Verfahren verwendet werden, um nun jedoch anhand der Ableitungen erster und zweiter Ordnung ein Extremum der Funktion zu finden. Die Iterationsvorschrift ändert sich zu

$$x_{k+1} = x_k - \left(\frac{\partial^2}{\partial \theta^2} f(x_k) \right)^{-1} \frac{\partial}{\partial \theta} f(x_k). \quad (5.1)$$

In jedem Iterationsschritt werden der Gradient und die Inverse der Hessematrix der eigentlichen Funktion benötigt. Aufgrund der numerischen Instabilität der Berechnung der inversen Hessematrix wird die in Gleichung (5.1) verwendete Darstellung in der Praxis jedoch nicht direkt verwendet. Statt dessen wird aus dem Gleichungssystem (entstanden aus Gleichung (5.1) durch Umformen)

$$\frac{\partial^2}{\partial \theta^2} f(x_k) \Delta x_k = - \frac{\partial}{\partial \theta} f(x_k)$$

der sogenannte Newton-Korrekturterm Δx_k errechnet und in einem zweiten Schritt ergibt sich über den Zusammenhang

$$x_{k+1} = x_k + \Delta x_k$$

²Teil des integrierten `stats` Pakets aus R (Version 2.11.1)

der neue Wert x_{k+1} . Diese beiden Berechnungen werden solange wiederholt, bis Konvergenz eintritt. Die auf diese Art und Weise erzeugte Folge (x_k) konvergiert dann gegen einen Wert x_* , welcher der gesuchten Optimalstelle entspricht. Jedoch ist das Newton-Verfahren recht anfällig gegenüber der Wahl des Startwerts x_0 für den ersten Iterationsschritt. Für unangemessen gewählte Startwerte kann es unter Umständen passieren, dass das Verfahren trotz einer existierenden Lösung divergiert oder in einem Sattelpunkt hängen bleibt, welcher nicht der gesuchten Lösung entspricht. In der Praxis wird die Lösungssuche mittels Quasi-Newton-Verfahren außerdem durch die Berechnung von $\frac{\partial^2}{\partial \theta^2} f(x_k)$ erschwert, da deren berechnete Werte mitunter nicht in jeder Iteration eine symmetrische und positiv definite Matrix ergeben.

5.1.2 Approximation der Hessematrix mittels BFGS-Verfahren

Eine mögliche Alternative zur exakten Berechnung der Hessematrix ist eine Approximation derselben mittels geeigneten Verfahren. Allen diesen Verfahren gemein ist, dass dabei anstelle der Hessematrix in Gleichung (5.1) eine nach den jeweiligen Regeln aufgestellte symmetrische Matrix F_k tritt. Die Iterationsgleichung verändert sich somit zu

$$x_{k+1} = x_k - \alpha_k F_k \frac{\partial}{\partial \theta} f(x_k),$$

wobei α_k die Schrittweite der Iteration steuert.

Die Idee einer Schrittweite α_k stammt dabei aus dem globalen Newton-Verfahren. Sie erlaubt anstelle der festen Schrittweite von 1 eine variable Wahl der Schrittweite innerhalb jeder Iteration. So werden bessere Steuerungsmöglichkeiten für die Berechnung von x_{k+1} möglich.

Für die Berechnung von F_k als Approximation der inversen Hessematrix haben sich vor allem zwei Verfahren durchgesetzt, von denen eines die Inverse der Hessematrix, das andere die Hessematrix selber annähert. Diese direkte Approximation der Hessematrix geht auf ein Verfahren zurück, welches 1970 von Broyden, Fletcher, Goldfarb und Shanno in getrennten Arbeiten vorgestellt wurde, inzwischen jedoch unter dem gemeinsamen Namen aller vier Autoren bekannt ist. Die grundlegende Vorgehensweise dieser Methode ist, die Hessematrix startend aus einer Einheitsmatrix iterationsweise anzunähern. Wie Harzheim (2007) schreibt, basiert die Approximation bei dieser Methode auf der Idee, die Hessematrix ähnlich der Annäherung einer ersten Ableitung durch Differenzen der Funktionswerte iterativ zu errechnen. So kann die Hessematrix mittels Differenzen der ersten Ableitung approximiert werden.

Durch diese Vorgehensweise ist es somit nicht mehr notwendig, die Hessematrix in jedem Iterationsschritt explizit zu berechnen. Weiters garantiert dieses Verfahren, dass die Matrix, welche die Hessematrix annähern soll, in jeder Iteration symmetrisch und positiv definit ist. Und die sukzessive Konstruktionsmethode der Hessematrix verhindert in gewissem Maße weiters, dass die Schätzprozedur in einem lokalen Optimum hängen bleibt.

Aufgrund dieser Stärken wurde die BFGS-Methode für die Berechnung der Maxima der log-Likelihoodfunktionen gewählt. Auf dieser Optimierungsroutine basierende Simulationsergebnisse folgen in den Abschnitten 5.3 und 5.4. Zuvor sollen jedoch noch die weiteren für die Simulationen verwendeten R-Funktionen besprochen werden.

5.2 Beschreibung der verwendeten R-Funktionen

Um für die Berechnung der Parameterschätzer die Optimierungsroutine `optim()` verwenden zu können, wurden sowohl die log-Likelihoodfunktionen als auch die partiellen Ableitungen erster und zweiter Ordnung der beiden Verteilungen in R implementiert. Die log-Likelihoodfunktionen (3.18) und (4.6) sind dabei wie in Code 5.1 definiert. Die Optimierungsroutine nimmt den Parametervektor β und den zusätzlichen Parameter α als die ersten beiden Argumente des Funktionsaufrufs. Die Erfolgswahrscheinlichkeit π_i hängt in diesen Darstellungen somit bereits von einem linearen Prädiktor $\mathbf{x}_i^t \beta$ ab. Der zweite Parameter der Verteilung wird hier über den log-Link in die Funktion integriert. Bei der Größe \mathbf{y} handelt es sich um den Vektor der Erfolge, gefolgt von der Gruppengröße \mathbf{n} . Diese kann dabei sowohl skalar- als auch vektorwertig sein. Außerdem wird die Designmatrix \mathbf{X} der Funktion übergeben. Die Variable `ResponseDistr` erlaubt über die zulässigen Argumente `binomial`, `multbin` und `doublebin` die Wahl der gewünschten Verteilungsannahme, welche dann innerhalb der Funktion `logl()` ausgewählt wird.

```
logl <- function(beta, alpha, y, n, X, ResponseDistr)
```

Code 5.1: Definition der log-Likelihoodfunktion der verallgemeinerten Binomialverteilungen.

An und für sich könnte in den Optionen der R-Funktion `optim()` von Minimierung auf Maximierung der übergebenen Funktion umgestellt werden. Da `optim()` die übergebene Funktion jedoch nach ihrem ersten Argument minimiert, stehen außerdem die negativen log-Likelihoodfunktionen der drei Verteilungen als interne Funktionen für die Optimierung zur Verfügung. Dafür werden wie in Code 5.2 gezeigt die Parameter β und α zu einem gemeinsamen Parametervektor `par` zusammengefasst, um `optim()` zur Optimierung nach beiden Parametern zu zwingen. Ebenso existieren die negativen Gradientenfunktionen mit β und α im ersten Argument für denselben Zweck.

```
neg.logl <- function(par, y, n, X, ResponseDistr)
```

Code 5.2: Definition der negativen log-Likelihoodfunktion im Parametervektor (β, α) .

5.2.1 Abbruchkriterien der Optimierungsroutine

Die R-Funktion `optim()` verwendet eine Kombination mehrerer Abbruchkriterien, um die Suche nach dem Optimum zu beenden. So bricht die Optimierung standardmäßig nach spätestens 100 Iterationen ab, da in diesem Fall die vorgegebene, maximal durchzuführende Anzahl an Iteration erreicht wurde. Diese Anzahl kann jedoch mit dem Argument `maxit` von `optim()` geändert werden.

Ebenso wird die Iteration abgebrochen, wenn sich die absoluten Werte $f(x_k)$ und $f(x_{k+1})$ um weniger als einen vorgegebenen Wert ändern. Dieser Wert kann durch das Argument `abstol` von `optim()` gesteuert werden, der default-mäßig auf minus unendlich gesetzt ist.

Außerdem wird eine gefundene Lösung als optimal angesehen, wenn sich die relativen Werte der Zielfunktion um weniger als einen vorgegebenen Wert `reltol` ändern, der üblicherweise mit

10^{-8} angenommen wird. Trifft dieses Kriterium zu, so gilt also für x_k und x_{k+1}

$$\frac{|f(x_{k+1}) - f(x_k)|}{|f(x_{k+1})|} \leq \epsilon_{\text{rel}}$$

für die relative Toleranz ϵ_{rel} und x_{k+1} entspricht der gesuchten Optimallösung.

Diese Abbruchkriterien von Optimierungsproblemen sind zum Beispiel im bereits genannten Werk von Harzheim (2007) nachzulesen.

5.2.2 Zufallszahlengeneratoren

Um die log-Likelihoodfunktionen sowie die darauf aufbauende Parameterschätzung zu testen, wurden weiters Funktionen zur Erzeugung von multiplikativ beziehungsweise doppelt binomialverteilten Zufallsvariablen erstellt. Wie die von Lindsey in seinem R-Paket „rmutil“ vorgestellten C-Funktionen berechnen diese Funktionen aus N auf dem Einheitsintervall gleichverteilten Zufallsvariablen u_i über den Zusammenhang

$$u_i = F(y_i)$$

die gesuchten Zufallsvariablen y_i . Mit $F(y_i)$ sei hierbei die Verteilungsfunktion der entsprechenden Verteilung gemeint. Numerisch wird für diese Lösung die Nullstelle von $F(y_i) - u_i$ gesucht, eine Aufgabe, die rechenzeittechnisch recht aufwändig ist. Eine Effizienzsteigerung würde sich ergeben, wenn die Nullstellensuche in C durchgeführt werden würde. Diese ist im Moment jedoch noch nicht vorhanden, und so liegen die Funktionen **dgenbin**, **pgenbin**, **qgenbin** sowie **rngenbin** für die doppelte und die multiplikative Binomialverteilung bis jetzt nur in R vor.

```
dgenbin(x, n, prob, theta, ResponseDistr)
pgenbin(q, n, prob, theta, ResponseDistr)
qgenbin(p, n, prob, theta, ResponseDistr)
rngenbin(m, n, prob, theta, ResponseDistr)
```

Code 5.3: Wahrscheinlichkeitsfunktion, Verteilungsfunktion, Quantilsfunktion und Zufallszahlengenerator der verallgemeinerten Binomialverteilungen.

Code 5.3 zeigt die Aufrufe der Wahrscheinlichkeitsfunktion, der Verteilungsfunktion, der Quantilsfunktion sowie des Zufallszahlengenerators. Die Argumente dieser Funktionen wurden bewusst analog zu den in R vorhandenen Funktionen zur Erzeugung derselben Werte bei gängigeren Verteilungen gewählt. Außerdem kommt das zusätzliche Argument **theta** für den zweiten Parameter der Verteilung hinzu. Dabei sei hier und in weiterer Folge mit **theta** der zweite Parameter einer verallgemeinerten Binomialverteilung gemeint. Es handelt sich um einen Platzhalter, der je nach multiplikativer oder doppelter Binomialverteilung dem **omega** oder dem **phi** der jeweiligen Verteilungen entspricht. Weiters wird im Funktionsaufruf noch **ResponseDistr** für die Auswahl der klassischen, multiplikativen oder doppelten Binomialverteilung benötigt.

Die oben beschriebene Vorgehensweise, welche innerhalb der Funktion **rngenbin()** zur Erzeugung der Zufallsvariablen verwendet wird, ist als Inversionsmethode bekannt und wird unter anderem bei Fishman (2001) oder Banks, Carson, Barry und Nicol (2009) genauer erläutert.

5.3 Simulationen basierend auf multiplikativ binomialverteilten Zufallsvariablen

Mehrere Datensätze wurden erzeugt, um die erstellte log-Likelihood- und Gradientenfunktion sowie die Funktion zur Berechnung der Hessematrix in Zusammenhang mit der `optim()`-Prozedur zu testen. Den Beginn der Diskussion dieser Resultate soll hier die multiplikative Binomialverteilung machen, bevor in Abschnitt 5.4 auf die Ergebnisse der Simulation unter doppelter Binomialverteilungsannahme eingegangen wird.

5.3.1 Vergleiche von Erwartungswert und Varianz

Zu Beginn der Simulationsserien wurde überprüft, ob die Funktion zur Erzeugung von multiplikativ binomialverteilten Zufallsvariablen korrekt arbeitet.

Dafür wurden bei unterschiedlichen Parameterkombinationen von π und ω sowie verschiedene Werte von n Serien von multiplikativ binomialverteilten Zufallsvariablen erstellt. Aus diesen Datensätzen wurden Erwartungswert und Varianz geschätzt, um sie in weiterer Folge mit den theoretischen Werten der beiden Momente zu vergleichen. Die beiden theoretischen Momente für verschiedene Werte von n , π und ω wurden sowohl in Matlab als auch in R berechnet, um eventuelle Programmierfehler weiter ausschließen zu können. Die betrachteten Wertebereiche von n , π und ω sind

$$\begin{aligned} n &= \{3^2; 4^2; 6^2; 8^2; 11^2\} \\ \pi &= \{0,6; 0,8; 0,9; 0,95; 0,975; 0,99; 0,995\} \\ \text{und } \omega &= \{0,9; 0,95; 0,98; 1; 1,02; 1,05; 1,1\}. \end{aligned}$$

Für jede Kombination von n , π und ω wurden $N = 5000$ multiplikativ binomialverteilte Zufallsvariablen erzeugt und deren ersten beiden empirischen Momente berechnet. Diese Berechnungen wurden $M = 1000$ mal wiederholt und die resultierenden Mittel dieser tausend Erwartungswerte und Varianzen mit ihren theoretischen Resultaten verglichen.

n	ω	π	theoretisch		empirisch	
			E	Var	E	Var
64	1,02	0,6000	35,9733	9,7373	35,9645	9,7653
121	1,00	0,9750	117,9750	2,9494	117,9689	2,9489
121	1,00	0,8000	96,8000	19,3600	96,7942	19,3676
36	1,02	0,8000	26,6707	5,4553	26,6760	5,4648
121	0,98	0,6000	111,1751	14,4042	111,1702	14,4064

Tabelle 5.1.: Die fünf größten Unterschiede zwischen empirischen und theoretischen Erwartungswerten der multiplikativen Binomialverteilung.

Für die multiplikative Binomialverteilung listet Tabelle 5.1 die Werte einiger Parameterkombinationen mit großen Abweichungen zwischen Theorie und Empirie bezüglich den Erwartungswerten auf. Wie sich zeigt, unterscheiden sich die Erwartungswerte selbst in diesen Fällen

um weniger als 0,01. Auffallend dabei ist, dass unter diesen starken Unterschieden in den Erwartungswerten sowohl Fälle mit $\omega \neq 1$ als auch bei $\omega = 1$ auftreten. Dies läßt sich im Gegensatz dazu für die Varianzunterschiede nicht beobachten. Deren fünf größten Unterschiede sind in Tabelle 5.2 aufgelistet und zeigen rein Fälle mit $\omega \neq 1$. Weiters lassen sie erkennen, dass diese großen Unterschiede unabhängig von der Wahl von n und ω auftreten.

n	ω	π	theoretisch		empirisch	
			E	Var	E	Var
36	0,98	0,6000	23,3490	12,0525	23,3482	12,0939
36	0,95	0,6000	29,5158	12,1132	29,5139	12,1446
64	0,98	0,6000	46,3923	25,8978	46,3940	25,9279
64	1,02	0,6000	35,9733	9,7373	35,9645	9,7653
121	1,02	0,9500	98,1808	10,7156	98,1788	10,6896

Tabelle 5.2.: Die fünf größten Unterschiede zwischen empirischen und theoretischen Varianzen der multiplikativen Binomialverteilung.

Nicht gezeigt ist hier der weiter oben angesprochene Vergleich zwischen der Berechnung von Erwartungswert und Varianz aus den zur Verfügung stehenden analytischen Formeln sowohl in Matlab³ als auch in R. In beiden Programmen liefern die betrachteten Funktionen jedoch dieselben Ergebnisse und erlauben den naheliegenden Schluss, dass die Funktionen zur Berechnung der Momente korrekt arbeiten.

Aufgrund dieser Ergebnisse wird davon ausgegangen, dass die für die Berechnung von Erwartungswert und Varianz verwendeten Funktionen sowie jene zur Erzeugung von multiplikativ binomialverteilten Zufallsvariablen in der Lage sind, richtige Resultate zu liefern.

5.3.2 Simulationsergebnisse der Parameterschätzung mittels `optim()`

In allen folgenden Simulationsergebnissen in diesem Abschnitt gilt für den Parameter der Blockgröße $n = 36$. Nun werden für gegebene Parameter π und ω multiplikativ binomialverteilte Zufallsvariablen erzeugt, wobei die Erfolgswahrscheinlichkeiten π im Kontext eines iid Modells betrachtet werden. In diesem Modell besitzen somit alle y_i denselben Prädiktor $\eta_i = \beta_0$. Für diese Zufallsstichprobe wird eine Maximum Likelihood Schätzung durchgeführt und die resultierenden Parameterschätzer $\hat{\beta}_0$ und $\hat{\omega}$ mit den wahren Werten verglichen. Außerdem werden für ω hier die drei unterschiedlichen Werte $\omega = \{0,98; 1; 1,02\}$ betrachtet.

Für eine bestimmte Parameterkombination von β_0 und ω werden $M = 500$ Monte-Carlo Wiederholungen durchgeführt, wobei in jedem Durchlauf $N = 1000$ Zufallsvariablen erzeugt und zur Parameterschätzung verwendet werden. Von jeder Wiederholung werden die resultierenden Punktschätzer $\hat{\beta}_0$ und $\hat{\omega}$ sowie die Anzahl an benötigten Iterationen der Optimierungsroutine, sowie empirischer Mittelwert und empirische Varianz der Zufallsstichprobe aufgezeichnet.

Parameterwahl $\beta_0 = 1,5$ und $\omega = 1$ (Binomialverteilung)

Die Boxplotserien in Abbildung 5.1 zeigen, wie sich die Punktschätzer $\hat{\beta}_0$ und $\hat{\omega}$ aus 500 Wiederholungen verhalten, wenn für die beiden Parameter die Werte $\beta_0 = 1,5$ und $\omega = 1$ gewählt

³Matlab R2009b 64-bit Version

werden.

Das Mittel der 500 Punktschätzer für β_0 liegt bei 1,5040. Der wahre Wert β_0 wird also von den 500 Simulationsergebnissen $\hat{\beta}_0$ im Mittel geringfügig überschätzt. Die Standardfehler der $\hat{\beta}_0$ liegen in etwa bei 0,0950. Diese Werte wurden zum einen exakt aus der analytischen Darstellung der Hessematrix in $(\hat{\beta}, \hat{\alpha})$ berechnet. Zum anderen wurden sie auch numerisch aus der Approximation der Hessematrix der BFGS-Optimierung ermittelt. Ein Vergleich der beiden zeigt die hohe Übereinstimmung der Resultate, was auf eine korrekte Implementierung der direkten Berechnung der Hessematrix hindeutet.

Außerdem zeigt die rechte Boxplotserie derselben Abbildung die Punktschätzer $\hat{\omega}$. Hier ist zu erkennen, dass diese das wahre ω im Mittel gut erreichen, da $\hat{\omega}$ im Mittel bei 1,0000 liegt. Auch die aus den beiden unterschiedlichen Vorgehensweisen geschätzten Standardfehler von $\hat{\omega}$ stimmen überein und liegen im Mittel bei 0,0042.

Aufgrund dieser Resultate wird davon ausgegangen, dass die Parameterschätzung mittels der log-Likelihoodfunktion und ihrer Gradientenfunktion für $\beta_0 = 1,5$ und $\omega = 1$ funktioniert.

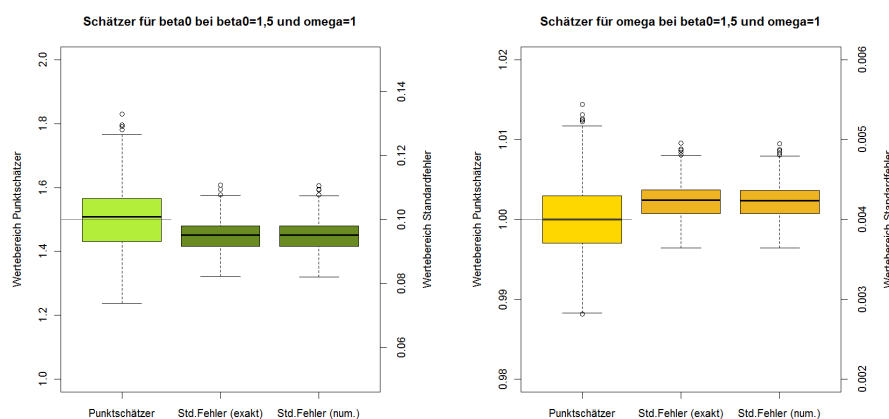


Abbildung 5.1.: Boxplots der 500 Punktschätzer $\hat{\beta}_0$ (links) und $\hat{\omega}$ (rechts) mit ihren Standardfehlern, exakt aus der Hessematrix und aus der numerischen Näherung der Hessematrix geschätzt. Die grauen Linien geben die wahren Parameterwerte aus der Erzeugung der Zufallsvariablen an.

Außerdem wurde mit den Vergleichen der Erwartungswerte und Standardabweichungen in Abbildung 5.2 die Funktionalität der R-Funktionen zur Berechnung der Zufallszahlen abermals überprüft. So liegt bei einem Wert von $\beta_0 = 1,5$ die Wahrscheinlichkeit für Erfolg bei

$$\pi = \frac{e^{1,5}}{1 + e^{1,5}} = 0,8176.$$

Die Elemente der betrachteten Zufallsstichproben besitzen in Folge dessen einen theoretischen Erwartungswert von 29,4 mit Varianz 5,37. Die Boxplots in Abbildung 5.2 zeigen, wie eng sich die empirischen Momente der 500 Wiederholungen um die theoretischen Momente gruppieren. Es sei noch darauf hingewiesen, dass der rechte Boxplot hier nicht die Varianzen, sondern die Standardabweichungen zeigt, deren theoretischer Wert bei 2,32 liegt.

Wie diese Diskussion gezeigt hat, ist im Fall einer Binomialverteilung ($\omega = 1$) eine sinnvolle Parameterschätzung möglich. In weiteren Schritten wird nun zuerst von $\omega = 1$ weggegangen

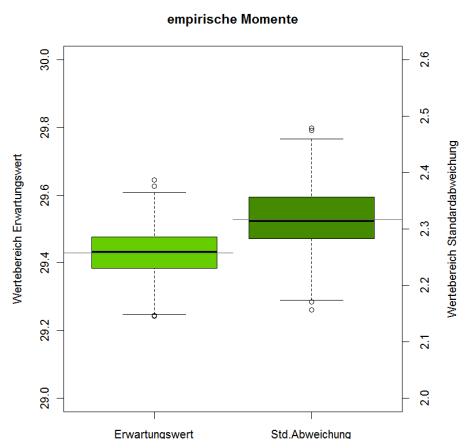


Abbildung 5.2.: Empirische Erwartungswerte und Standardabweichungen für Stichproben mit $\beta_0 = 1,5$ und $\omega = 1$. Die theoretischen Werte der Momente sind als graue Linien eingezeichnet.

und die Parameterschätzung für tatsächlich multiplikativ binomialverteilte Zufallsvariablen betrachtet. Außerdem wird das für die Konstruktion der Erfolgswahrscheinlichkeiten verwendete Modell auf ein Regressionsmodell mit Intercept und erklärender Größe erweitert, sodass die Parameterschätzung drei Parameter β_0 , β_1 und ω beinhaltet.

Aus Symmetriegründen werden die Simulationen nur für positive Parameter β_0 (und Erfolgswahrscheinlichkeiten zwischen $1/2$ und 1) betrachtet. Da die Ausführungen in Abschnitt 5.3.1 sich bereits ausführlich mit den empirischen Momenten der multiplikativen Binomialverteilung beschäftigt haben, wird in weiterer Folge auf Vergleiche derselben Art (siehe Abbildung 5.2) verzichtet.

In den folgenden Unterabschnitten werden nun Ergebnisse der Simulationen aus einigen der betrachteten Parameterkombinationen diskutiert.

Parameterwahl $\beta_0 = 3$ und $\omega < 1$ (Unterdispersion)

Bei dieser Wahl der Parameter mit $\pi = \exp(3)/(1+\exp(3)) = 0,9526$ und $\omega = 0,98$ ergibt sich bei der multiplikativen Binomialverteilung für den Erwartungswert $35,11$ mit einer Varianz von $0,9025$. Im Vergleich dazu liegt die Varianz der klassischen Binomialverteilung für $\pi = 0,9526$ bei $1,6264$. Im Gegensatz zur klassischen Binomialverteilung variieren diese Daten also schwächer als bei einer Binomialverteilung mit selbiger Erfolgswahrscheinlichkeit. Es liegt somit Unterdispersion in den multiplikativ binomialverteilten Daten vor.

Dabei ist auffallend, wie extrem sich die Varianz bereits für diese geringe Abweichung des Parameters ω von der 1 ändert. Aufgrund dieser Tatsache wird der Parameter ω in den weiteren Diskussionen der multiplikativen Binomialverteilung nicht weiter als in diesem Beispiel von der 1 abweichen. Für deutlich stärkere Abweichungen von der 1 , wie sie zum Beispiel in Abschnitt 5.4 für die doppelte Binomialverteilung betrachtet werden, ist die Betrachtung von multiplikativ verteilten Zufallsvariablen nicht mehr zielführend.

Für die Parameterwahl $\beta_0 = 3$ zeigen die Boxplotserien in Abbildung 5.3 die Punktschätzer $\hat{\beta}_0$ sowohl für den Fall $\omega = 0,98$ als auch für die klassische Binomialverteilung ($\omega = 1$, rechts).

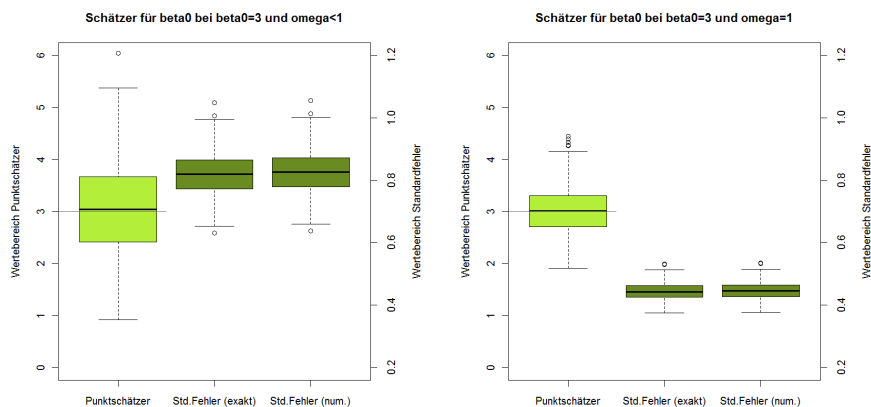


Abbildung 5.3.: Vergleich der Schätzer $\hat{\beta}_0$ für wahres $\beta_0 = 3$ bei $\omega = 0,98$ (links) beziehungsweise $\omega = 1$ (rechts).

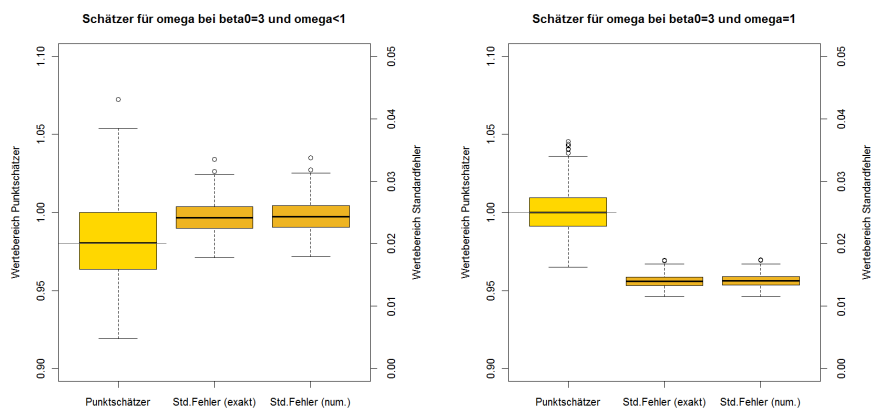


Abbildung 5.4.: Vergleich der Schätzer $\hat{\omega}$ bei $\beta_0 = 3$ für wahres $\omega = 0,98$ (links) beziehungsweise $\omega = 1$ (rechts).

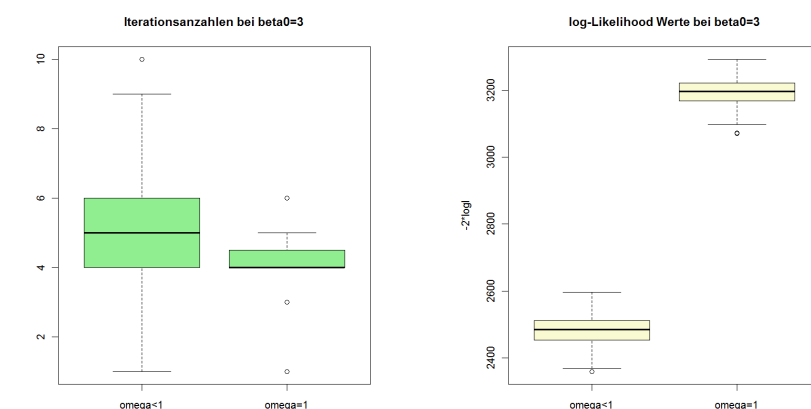


Abbildung 5.5.: Boxplots der benötigten Iterationen (links) und Vergleich der maximalen negativen doppelten log-Likelihoodwerte (rechts) für $\beta_0 = 3$ bei $\omega = 0,98$ und $\omega = 1$.

Auffallend ist, dass bei der Wahl von $\omega = 0,98$ die Punktschätzer für β_0 aus den 500 Wiederholungen im Mittel $\beta_0 = 3$ geringfügig überschätzen, da ihr Mittel bei 3,07 liegt. Außerdem weisen die Schätzungen nun eine deutlich breitere Standardabweichung auf. Der Standardfehler von $\hat{\beta}_0$ liegt für $\omega = 0,98$ im Mittel in etwa bei 0,8216. Im Vergleich dazu liegt der Standardfehler von $\hat{\beta}_0$ im Fall der klassischen Binomialverteilung bei nur 0,4436. Aufgrund der deutlich geringeren Variabilität multiplikativ binomialverteilter Zufallsvariablen bei $\omega = 0,98$ und $\beta_0 = 3$ beziehungsweise $\pi = 0,9526$ wird somit die Schätzung des Parameters β_0 unschärfer.

Abbildung 5.4 zeigt, wie sich die Punktschätzer $\hat{\omega}$ bei wahren $\omega = 0,98$ sowie $\omega = 1$ verhalten. Auch hier wird der wahre Wert ω leicht überschätzt, das Mittel der Schätzwerte liegt bei 0,9823. Abermals läßt sich auch ein Anwachsen des Standardfehlers ausmachen, wenn vom Fall der klassischen Binomialverteilung auf die multiplikative Verteilung mit $\omega < 1$ übergegangen wird. So liegt für $\omega = 1$ der Standardfehler bei 0,0140, wohingegen er sich bei $\omega = 0,98$ fast verdoppelt hat und nun 0,0243 beträgt. Im Mittel entspricht der resultierende Punktschätzer $\hat{\omega}$ jedoch auch bei $\omega < 1$ dem wahren Wert des Parameters.

Von besonderem Interesse ist in jedem Optimierungsproblem die Anzahl der Iterationen, die bis zum Erzielen der Konvergenz benötigt wurde. Die maximale Iterationsanzahl wurde für diese Simulationen auf 100 festgesetzt. Im Fall von $\omega = 1$ liegt die Anzahl an benötigten Iterationen im Mittel bei 4. Für $\omega < 1$ werden im Mittel 5 Iterationen bis zum Erzielen von Konvergenz benötigt. Dabei ist wichtig hervorzuheben, dass die Suche nach dem Optimum für $\omega \neq 1$ in den Werten der Punktschätzer einer klassischen logistischen Regression startet. Erst diese Wahl der Startwerte führt auf die genannte Anzahl an Iterationen. Die genauen Werte sind sowohl für $\omega = 1$ als auch $\omega < 1$ in den Boxplots aus Abbildung 5.5 (links) abzulesen. Es zeigt sich, dass die Iterationszahl bei $\omega < 1$ geringfügig größer ist, das Optimum aber selbst im schlechtesten Fall mit einer vernünftigen Anzahl an Iterationen erreicht werden kann.

Iterationsschritt	$-2 \cdot l(\hat{\beta}_0, \hat{\omega} \mathbf{y})$	$\hat{\beta}_0$	$\hat{\omega}$
1	2434,1270	3,7185	0,999984
2	2433,5160	2,8020	0,972893
3	2433,4590	3,0172	0,979156
4	2433,4580	3,0042	0,978781
5	2433,4580	3,0039	0,978771

Tabelle 5.3.: Negative doppelte log-Likelihoodwerte und Punktschätzer aus den einzelnen Iterationsschritten für tausend multiplikativ binomialverteilte Zufallsvariablen mit Parametern $\beta_0 = 3$ und $\omega = 0,98$.

Tabelle 5.3 zeigt für einen Satz von tausend multiplikativ binomialverteilten Zufallsvariablen mit $\pi = 0,9526$ und $\omega = 0,98$ den Verlauf der Iteration. Als Startwerte für die Optimierungsprozedur wurden dabei der Punktschätzer $\hat{\beta}_0$ aus einem logistischen Modell unter klassischer Binomialverteilungsannahme ($\omega = 1$) gewählt. Nach nur fünf Iterationen ist das Optimum bei $\hat{\beta}_0 = 3,0039$ und $\hat{\omega} = 0,9788$ mit $-2 \cdot l(\hat{\beta}_0, \hat{\omega} | \mathbf{y}) = 2433,46$ erreicht. Wird anstelle der Lösung des logistischen Modells die 0 als Startwert für den Parameter β_0 in der Optimierung gewählt, sind im allgemeinen etwas mehr Iterationen vonnöten. Für die hier betrachteten tausend Werte \mathbf{y} wären das konkret 16 anstelle von fünf Iterationen.

Eine Gegenüberstellung der maximalen negativen doppelten log-Likelihoodwerte ist in Abbildung 5.5 ebenfalls gezeigt. Für $\omega < 1$ werden kleinere $-2 \cdot l(\hat{\beta}_0, \hat{\omega} | \mathbf{y})$ Werte erzielt, aber in

beiden Fällen weisen die Werte von $-2 \cdot l(\hat{\beta}_0, \hat{\omega} | \mathbf{y})$ in etwa dieselben Standardabweichungen auf.

Parameterwahl $\beta_0 = 0$ und $\omega > 1$ (Unterdispersion)

Bei einer Wahl von $\beta_0 = 0$ liegt die Wahrscheinlichkeit für Erfolg bei genau $1/2$. Zusammen mit $\omega = 1,02$ liefert diese Parameterwahl multiplikativ verteilte Zufallsvariablen, welche eine deutlich kleinere Variabilität als in der klassischen Binomialverteilung aufweisen. Für $n = 36$ liegt die Varianz für $\omega = 1,02$ bei 6,6710. Im Vergleich dazu beträgt die Varianz im Fall der klassischen Binomialverteilung 9. Es liegt daher abermals ein Fall von Unterdispersion vor.

Die Ergebnisse für die Punktschätzer $\hat{\beta}_0$ sind in Abbildung 5.6 gezeigt und abermals in Relation zu den Ergebnissen der 500 Parameterschätzungen bei $\omega = 1$ gesetzt (rechte Boxplotserie). In beiden Fällen liegen die Punktschätzer $\hat{\beta}_0$ hier recht eng um den wahren Wert β_0 verteilt. Außerdem überschätzen die Punktschätzer $\hat{\beta}_0$ hier im Gegensatz zum Fall mit $\omega < 1$ den wahren Wert β_0 nicht. Ein Vergleich der Standardfehler bei unterschiedlichen ω zeigt, dass die Werte für $\omega > 1$ nur noch geringfügig größer sind als im Fall $\omega = 1$. So beträgt dieser bei $\omega > 1$ noch 0,0123 und fällt für $\omega = 1$ auf 0,0106. Im nicht gezeigten Fall von $\omega < 1$ liegt der Fehler nur noch bei 0,0085. Natürlich liefert die exakte Berechnung aus der analytischen Darstellung der Hessematrix in beiden Fällen dieselben Resultate wie eine Berechnung aus der numerischen Approximation der Hessematrix aus dem BFGS-Verfahren von `optim()`.

Wie die Boxplotserien für $\hat{\omega}$ aus Abbildung 5.7 erkennen lassen, können in beiden Fällen die wahren Werte von ω im Mittel korrekt geschätzt werden. Die mittleren Werte der Punktschätzer unterscheiden sich bei den beiden unterschiedlichen Werten von ω erst nach der vierten Nachkommastelle von ihren Sollwerten. Im Gegensatz zur zuvor betrachteten Situation mit $\beta_0 = 3$ und $\omega < 1$ ist der Standardfehler der resultierenden Punktschätzer $\hat{\omega}$ hier extrem klein. So beträgt dieser Fehler für $\omega > 1$ im Mittel 0,0035 und fällt für $\omega = 1$ sogar noch auf 0,0025. Eine Begründung dafür läßt sich in der Wahl des Parameters β_0 finden, der für dieses Beispiel eine Erfolgswahrscheinlichkeit aus dem Zentrum des Einheitsintervalls liefert. Auch für eine Parameterwahl von $\beta_0 = 0$ und $\omega = 0,98$ ist der Standardfehler der $\hat{\omega}$ nicht größer, er liegt im Mittel sogar bei nur 0,0016.

Abbildung 5.8 zeigt die Werte der negativen doppelten log-Likelihoodfunktion in den gefundenen Punktschätzern $\hat{\beta}_0$ und $\hat{\omega}$ sowohl für $\omega = 1,02$ als auch für $\omega = 1$. Außerdem sind die Iterationsanzahlen gezeigt, die in den einzelnen Wiederholungen benötigt wurden. Dabei traten bei $\omega = 1$ hier sechs Fälle auf, bei denen die Suche nach dem Optimum mehr als 20 Iterationen benötigte, wohingegen im Fall $\omega > 1$ keine derart langwierigen Modellschätzungen vorkamen. Der Median der benötigten Iterationen liegt für $\omega = 1$ bei 5 und für $\omega > 1$ bei 6.

Generell läßt sich also festhalten, dass die verwendete Schätzmethode auch für Werte von $\omega \neq 1$ in der Lage ist, ausgehend von einem passenden Startwert innerhalb weniger Iterationsschritte das gesuchte Optimum zu finden.

Parameterwahl $\beta_0 = 3$ und $\beta_1 = 1$ für verschiedene ω

Die nächste Simulation zeigt nun, wie sich die Ergebnisse verändern, wenn der lineare Prädiktor η zusätzlich zum Intercept auch von einer erklärenden Größe x abhängt. Die im betrachteten Modell $\eta = \beta_0 + \beta_1 x$ verwendete Größe x wurde als regelmäßige Sequenz von 601 Werten aus dem Intervall $[-3; 3]$ gewählt und jeder dieser Werte verzehnfacht. Auf diese Weise entsteht eine Sequenz von $N = 6010$ Werten, welche für alle nun folgenden Simulationsreihen als erklärende Größe x verwendet wird.

5.3. Simulationen basierend auf multiplikativ binomialverteilten Zufallsvariablen

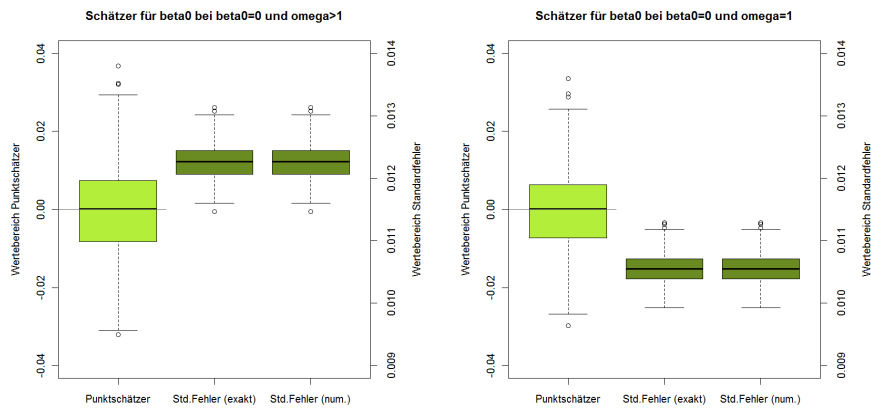


Abbildung 5.6.: Vergleich der Schätzer $\hat{\beta}_0$ für wahres $\beta_0 = 0$ bei $\omega = 1, 02$ (links) beziehungsweise $\omega = 1$ (rechts).

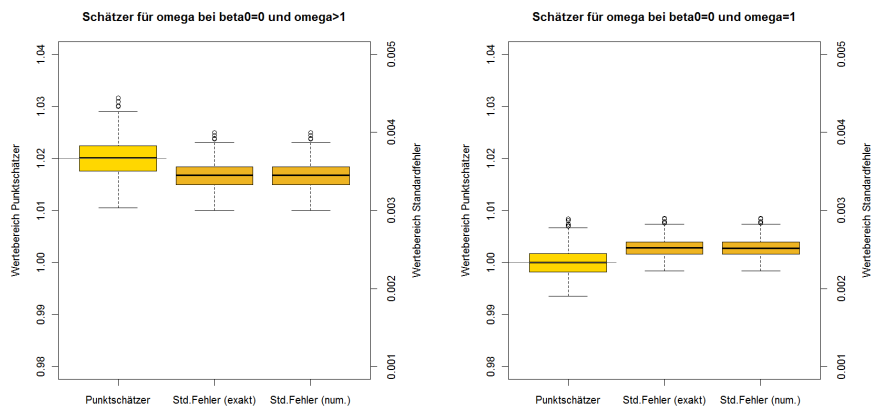


Abbildung 5.7.: Vergleich der Schätzer $\hat{\omega}$ bei $\beta_0 = 0$ für wahres $\omega = 1, 02$ (links) beziehungsweise $\omega = 1$ (rechts).

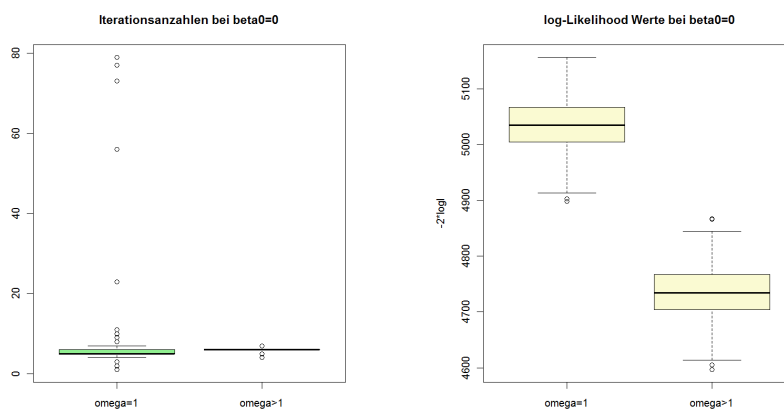


Abbildung 5.8.: Boxplots der benötigten Iterationen (links) und Vergleich der maximalen negativen doppelten log-Likelihoodwerte (rechts) für $\beta_0 = 0$ bei $\omega = 1, 02$ und $\omega = 1$.

Aus der Wahl der Parameter $n = 36$ und $\beta = (3, 1)$ lassen sich die Wertebereiche des Erwartungswerts sowie der Standardabweichung für die unterschiedlichen ω ermitteln. Diese Werte sind in den Plots in Abbildung 5.9 dargestellt. Für $\omega < 1$ liegt die zu erwartende Anzahl an Erfolgen etwas über jener der klassischen Binomialverteilung, wohingegen sie für $\omega > 1$ etwas darunter liegt.

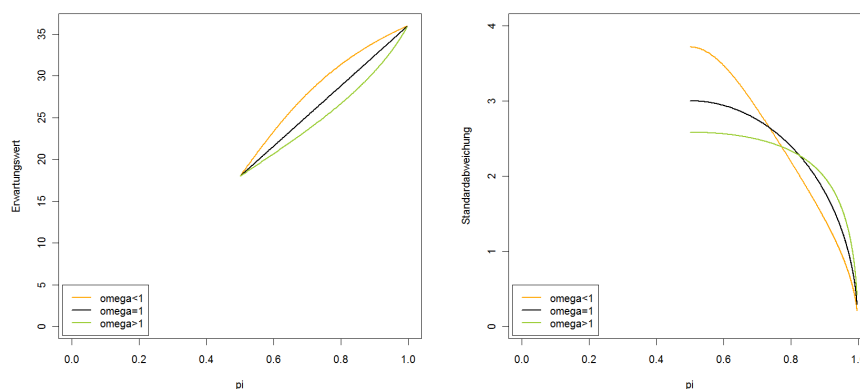


Abbildung 5.9.: Erwartungswerte (links) und Standardfehler (rechts) im Fall $\eta_i = 3 + 1 \cdot x_i$ für drei verschiedene Werte von ω und $x_i \in [-3; 3]$.

Die nun folgende Gegenüberstellung läßt erkennen, wie die Parameterschätzung für unterschiedliche Werte von ω funktioniert. So zeigen die drei Grafiken in den Abbildungen 5.10 bis 5.12, wie sich die Punktschätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ für verschiedene Werte von ω verhalten. Die Schätzungen der β_0 lassen erkennen, dass die Standardfehler der Punktschätzer eigentlich unabhängig von der Wahl von ω sind. Die Mittelwerte der Standardfehler dieser Punktschätzer liegen bei 0,0477 im Fall $\omega = 0,98$ und werden dann geringfügig größer, bis sie bei $\omega = 1,02$ bei 0,0498 liegen. Selbiges läßt sich für die Standardfehler der Punktschätzer $\hat{\beta}_1$ beobachten, welche für $\omega < 1$ im Mittel bei 0,0142, für $\omega > 1$ bei 0,0148 liegen. Dass die Standardfehler der Punktschätzer für Werte von ω in einem Bereich um die 1 nur geringfügig von der Wahl von ω abhängen, läßt sich auch für andere Parameterkombinationen von β_0 und β_1 beobachten, wird hier jedoch nicht extra dargestellt.

Abbildungen 5.13 bis 5.15 zeigen die Schätzergebnisse für die Punktschätzer $\hat{\omega}$ in den drei Fällen $\omega = 0,98$, $\omega = 1$ und $\omega = 1,02$. Die Mittelwerte der drei Serien an Punktschätzern weichen erst in der vierten Nachkommastelle von den jeweiligen Sollwerten ab. Die Punktschätzer stimmen also im Mittel sehr gut mit den wahren Werten überein. Die Standardfehler der $\hat{\omega}$ sind extrem klein, wobei auch diese für wachsendes ω etwas größer werden. Die genauen numerischen Werte sind 0,0012 bei $\omega < 1$, 0,0015 bei $\omega = 1$ und 0,0018 bei $\omega > 1$.

Zur Anzahl der Iterationen, die für eine erfolgreiche Schätzung benötigt werden, gibt Tabelle 5.4 einige Kennwerte an. Es zeigt sich, dass die Anzahl an benötigten Iterationen im Fall der klassischen Binomialverteilung ($\omega = 1$) im Mittel bei 5 liegt. Für $\omega \neq 1$ werden etwas mehr Iterationen benötigt. Die maximale Anzahl an Iterationen liegt jedoch auch in den Fällen von $\omega \neq 1$ bei nur 8. Eine Parameterschätzung via `optim()` ist also auch in den beiden Fällen von $\omega \neq 1$ möglich.

Ähnliche Iterationsanzahlen ergeben sich auch für andere Parameterkonstellationen, welche

5.3. Simulationen basierend auf multiplikativ binomialverteilten Zufallsvariablen

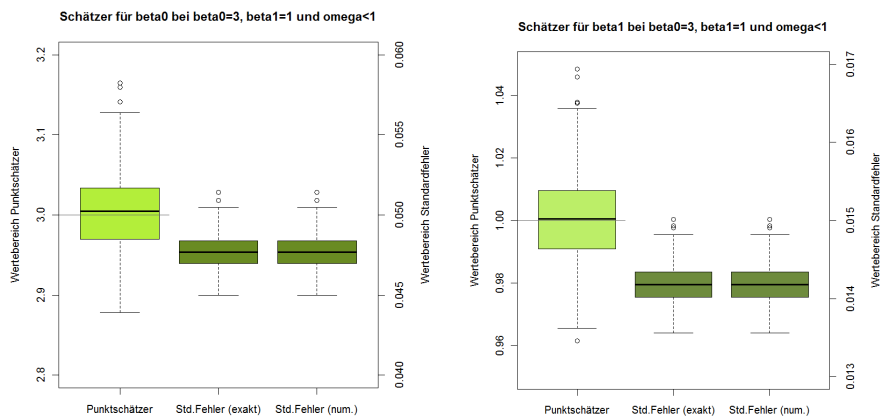


Abbildung 5.10.: Resultierende Punktschätzer $\hat{\beta}_0$ (links) und $\hat{\beta}_1$ (rechts) für wahres $\omega = 0,98$ sowie $\beta_0 = 3$ und $\beta_1 = 1$.

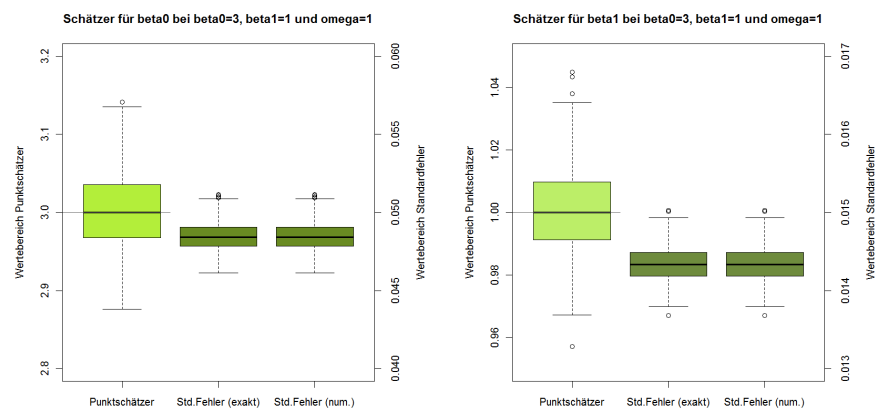


Abbildung 5.11.: Resultierende Punktschätzer $\hat{\beta}_0$ (links) und $\hat{\beta}_1$ (rechts) für wahres $\omega = 1$ sowie $\beta_0 = 3$ und $\beta_1 = 1$.

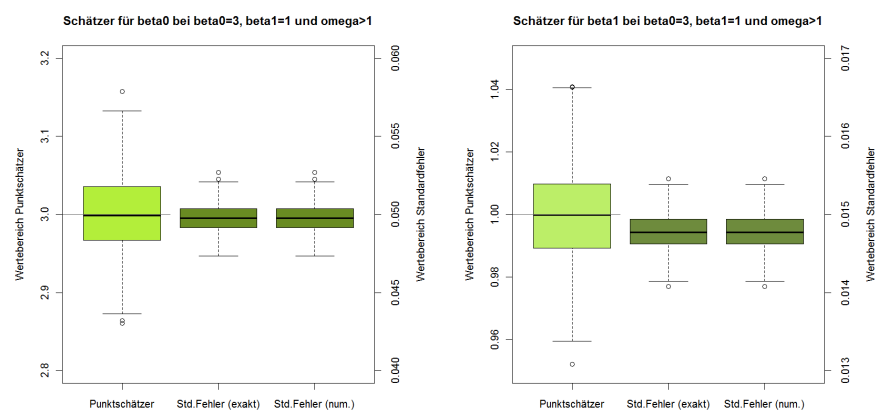


Abbildung 5.12.: Resultierende Punktschätzer $\hat{\beta}_0$ (links) und $\hat{\beta}_1$ (rechts) für wahres $\omega = 1,02$ sowie $\beta_0 = 3$ und $\beta_1 = 1$.

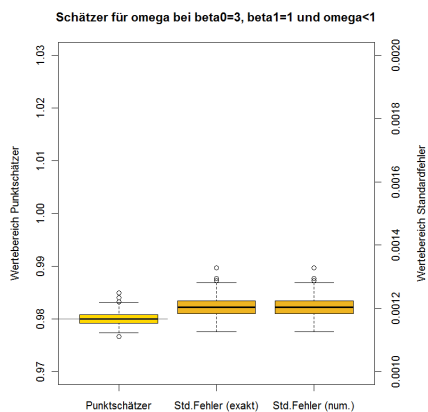


Abbildung 5.13.: Resultierende Punktschätzer $\hat{\omega}$ mit Standardfehlern im Fall von $\omega = 0,98$ bei $\beta_0 = 3$ und $\beta_1 = 1$.

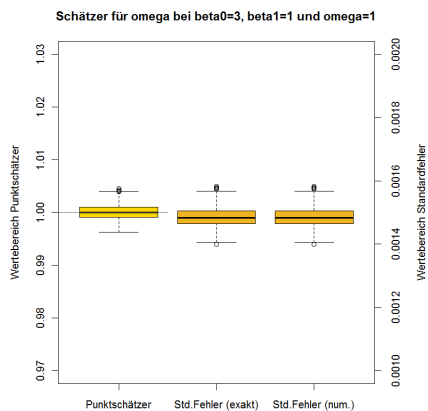


Abbildung 5.14.: Resultierende Punktschätzer $\hat{\omega}$ mit Standardfehlern im Fall von $\omega = 1$ bei $\beta_0 = 3$ und $\beta_1 = 1$.

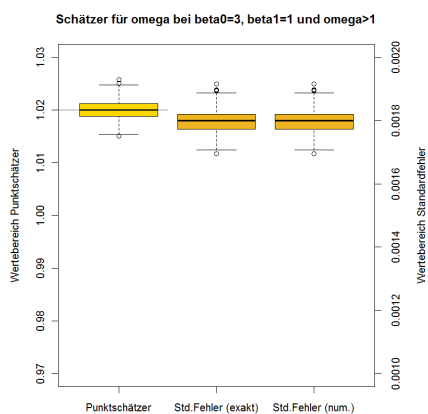


Abbildung 5.15.: Resultierende Punktschätzer $\hat{\omega}$ mit Standardfehlern im Fall von $\omega = 1,02$ bei $\beta_0 = 3$ und $\beta_1 = 1$.

ω	Minimum	1.Quantil	Median	Mean	3.Quantil	Maximum
0,98	8	8	8	8,0000	8	8
1,00	1	5	5	4,7560	5	6
1,02	7	7	7	7,0020	7	8

Tabelle 5.4.: Kennwerte zur Anzahl der Iterationen im Fall multiplikativ binomialverteilter Zufallsvariablen für $\beta_0 = 3$ und $\beta_1 = 1$ für verschiedene ω .

ebenfalls Erfolgswahrscheinlichkeiten π aus breiten Bereichen des Einheitsintervalls ermöglichen. Erst für Werte von β , welche die Erfolgswahrscheinlichkeiten extrem an die Ränder des Einheitsintervalls rücken, werden mehrere Iterationen notwendig. Ein derartiger Fall ist im nachfolgenden Abschnitt zu finden.

Parameterwahl $\beta_0 = 6$ und $\beta_1 = 1$ für verschiedene ω

Betrachtet man eine ähnliche Simulationskonfiguration, diesmal jedoch für $\beta_0 = 6$ und $\beta_1 = 1$, ergeben sich die folgenden Resultate: die Wahrscheinlichkeit für Erfolg liegt, ausgehend von $\beta = (6, 1)$ und einer erklärenden Variable x als regelmäßige Serie von 6010 Werten aus dem Intervall $[-3; 3]$, zwischen 0,95 und 1. Eine Darstellung der Wertebereiche der daraus resultierenden Erwartungswerte und ihren Standardabweichungen ist in Abbildung 5.16 zu finden. Außerdem sind die genauen Wertebereiche der beiden ersten Momente in Tabelle 5.5 wiedergegeben. So liegen bei der gewählten Parameterkombination die Erwartungswerte für die verschiedenen ω zwischen 33 und 36. Sowohl die Werte der Tabelle als auch die grafische Gegenüberstellung lassen erkennen, dass hier für $\omega > 1$ Überdispersion, für $\omega < 1$ hingegen Unterdispersion vorliegt.

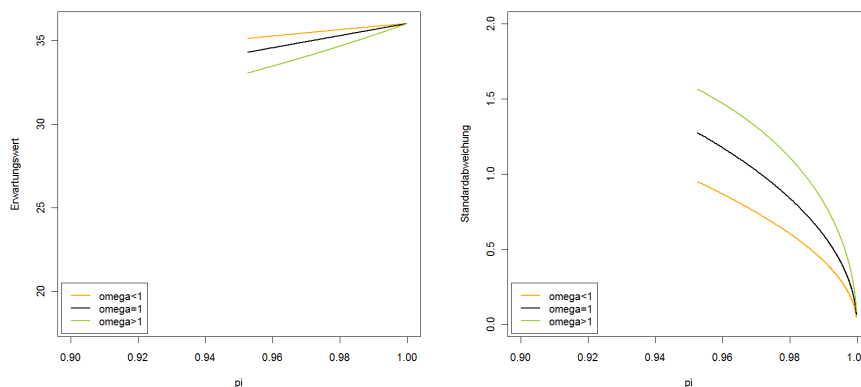


Abbildung 5.16.: Erwartungswerte (links) und Standardfehler (rechts) im Fall $\eta_i = 6 + 1 \cdot x_i$ für drei verschiedene Werte von ω und $x_i \in [-3; 3]$.

Eine Betrachtung des Optimierungsproblems für solch extreme Parameterkonstellationen ist von Interesse, da für die in Kapitel 6 betrachteten Datensätze mit ähnlichen Situationen zu rechnen ist.

Die Ergebnisse der Punktschätzungen $\hat{\beta}$ und $\hat{\omega}$ sind in den Abbildungen 5.17 bis 5.22 zu sehen. Zu erkennen ist, dass auch für diese extreme Wahl von Erfolgswahrscheinlichkeiten die Schätzung im Mittel angemessene Resultate liefert. Dennoch liegen vor allem die Punktschätzer

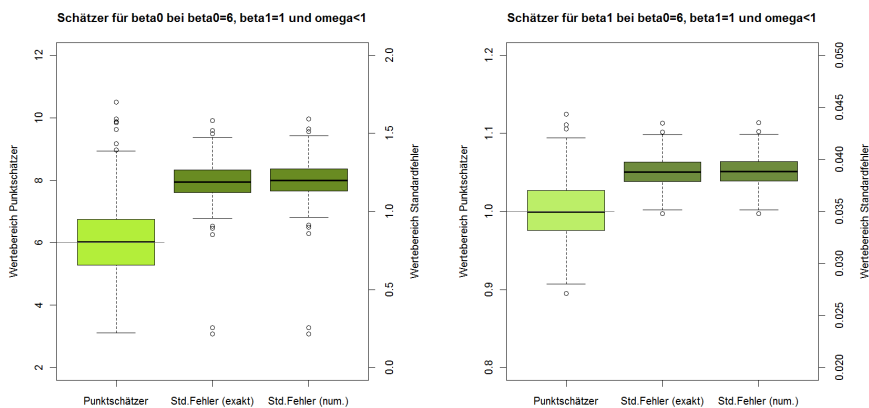


Abbildung 5.17.: Resultierende Punktschätzer $\hat{\beta}_0$ (links) und $\hat{\beta}_1$ (rechts) für wahres $\omega = 0,98$ sowie $\beta_0 = 6$ und $\beta_1 = 1$.

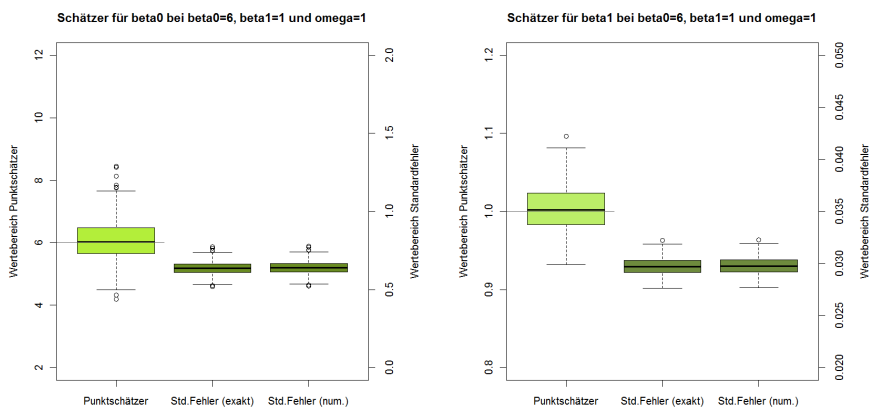


Abbildung 5.18.: Resultierende Punktschätzer $\hat{\beta}_0$ (links) und $\hat{\beta}_1$ (rechts) für wahres $\omega = 1$ sowie $\beta_0 = 6$ und $\beta_1 = 1$.

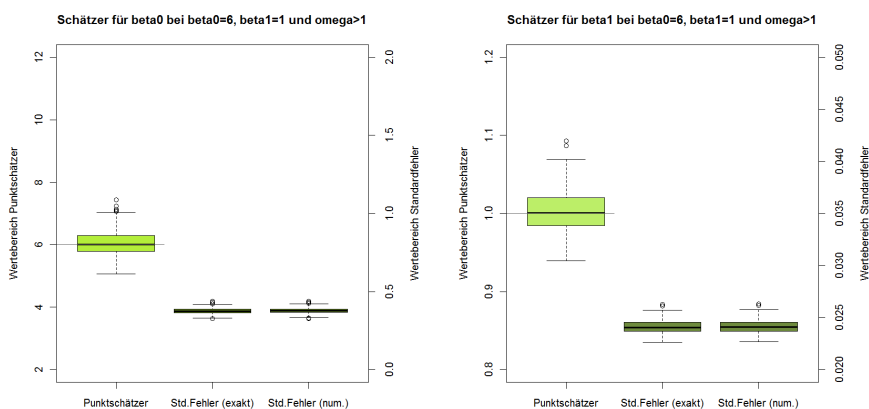


Abbildung 5.19.: Resultierende Punktschätzer $\hat{\beta}_0$ (links) und $\hat{\beta}_1$ (rechts) für wahres $\omega = 1,02$ sowie $\beta_0 = 6$ und $\beta_1 = 1$.

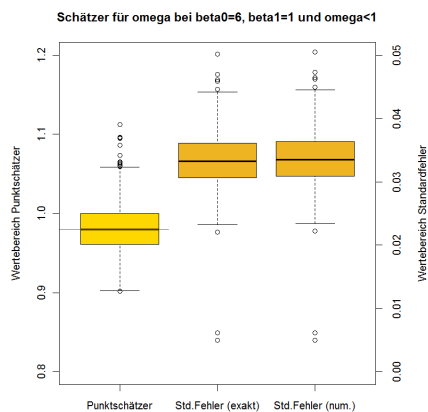


Abbildung 5.20.: Resultierende Punktschätzer $\hat{\omega}$ mit Standardfehlern im Fall von $\omega = 0,98$ bei $\beta_0 = 6$ und $\beta_1 = 1$.

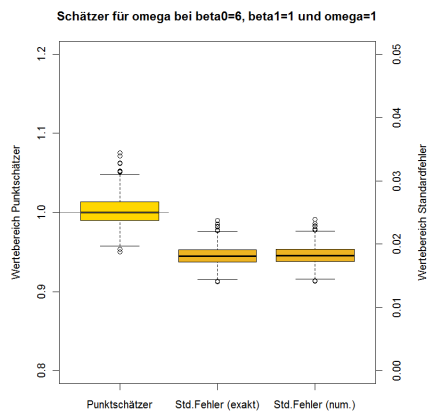


Abbildung 5.21.: Resultierende Punktschätzer $\hat{\omega}$ mit Standardfehlern im Fall von $\omega = 1$ bei $\beta_0 = 6$ und $\beta_1 = 1$.

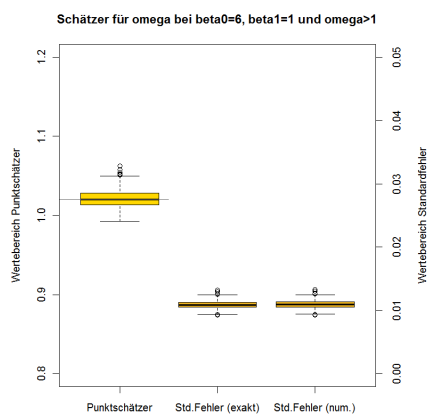


Abbildung 5.22.: Resultierende Punktschätzer $\hat{\omega}$ mit Standardfehlern im Fall von $\omega = 1,02$ bei $\beta_0 = 6$ und $\beta_1 = 1$.

ω	Erwartungswerte	Standardabweichungen
0,98	[35,11; 36,00]	[0,0468; 0,9500]
1,00	[34,29; 36,00]	[0,0667; 1,2753]
1,02	[33,06; 36,00]	[0,0942; 1,5652]

Tabelle 5.5.: Wertebereiche der ersten beiden Momente bei $\pi_i = \text{logit}(x_i^t \boldsymbol{\beta})$.

von β_0 im Mittel etwas über dem wahren β_0 . Die Mittelwerte der 500 Schätzungen liegen bei 6,13 für $\omega < 1$, bei 6,10 für $\omega = 1$ und bei 6,04 für $\omega > 1$. Im Gegensatz zum Fall $\beta_0 = 3$ zeigt sich hier, dass die Schätzung des Intercept-Parameters nun größerer Variabilität unterworfen ist. Das gilt besonders für den Fall $\omega < 1$, wo der Standardfehler im Mittel 1,1940 beträgt, lässt sich jedoch auch für $\omega > 1$ bei einem mittleren Standardfehler von 0,3764 in abgeschwächter Form beobachten.

Von deutlich größerem Interesse als die Schätzung der Konstanten β_0 ist jedoch jene des Steigungsparameters β_1 . Die Simulationsresultate der $\hat{\beta}_1$ stimmen im Mittel gut mit dem wahren Wert $\beta_1 = 1$ überein. Ihre Standardfehler sind deutlich geringer als jene der $\hat{\beta}_0$ und sind auch mit geringeren Variabilitäten behaftet als diese. Diese Vergleiche zeigen nun außerdem, dass die Standardfehler der $\hat{\beta}$ für wachsende Werte von ω kleiner werden. So beträgt der Standardfehler der $\hat{\beta}_1$ bei $\omega = 0,98$ im Mittel noch 0,0388 und fällt dann für $\omega = 1,02$ auf 0,0241. Bei dieser extremen Wahl der Erfolgswahrscheinlichkeiten hängen die Resultate der Parameterschätzung also etwas mehr vom zweiten Parameter, ω , ab. Für den hier nicht gesondert diskutierten Fall $\boldsymbol{\beta} = (6, 3)$ zum Beispiel, bei dem abermals Erfolgswahrscheinlichkeiten aus dem gesamten Einheitsintervall resultieren, ist diese Abhängigkeit der $\hat{\beta}$ von ω nicht mehr zu beobachten.

Die Ergebnisse für $\hat{\omega}$ in den Abbildungen 5.20 bis 5.22 zeigen, dass die wahren Werte von ω auch in diesem extremen Fall noch gut geschätzt werden können. Vor allem für $\omega < 1$ jedoch sind die Standardfehler mit einem Mittelwert von 0,0336 relativ groß. Sie fallen jedoch für wachsendes ω , bis sie bei $\omega = 1,02$ im Mittel bei 0,0110 liegen.

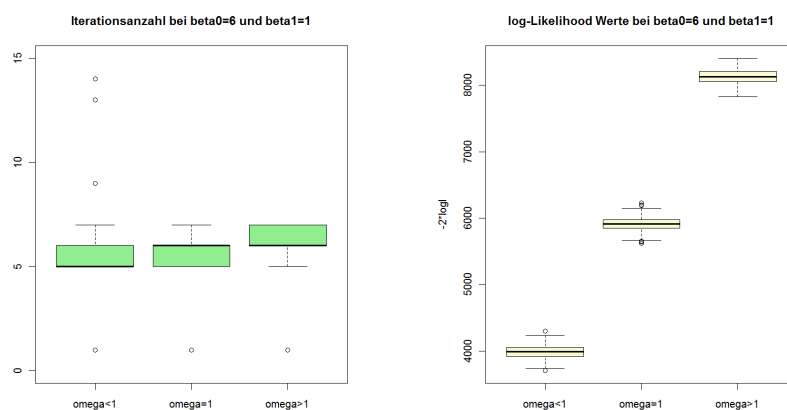


Abbildung 5.23.: Boxplots der benötigten Iterationen (links) und Vergleich der maximalen negativen doppelten log-Likelihoodwerte (rechts) bei $\omega = 0,98$, $\omega = 1$ und $\omega = 1,02$ sowie $\beta_0 = 6$ und $\beta_1 = 1$.

Im Gegensatz zur Simulationsserie mit $\beta_0 = 3$ liegt die Anzahl der Iterationen in diesem Fall für alle drei Wahlen von ω im Mittel bei 5 beziehungsweise 6, wie die in Abbildung 5.23 gezeigte Boxplotserie erkennen lässt. Es liegen nun zwar für $\omega < 1$ zwei Fälle vor, in denen mehr als 10 Iterationen bis zur Konvergenz benötigt wurden, der Median der Anzahl an Iterationen jedoch ist für $\omega \neq 1$ nun kleiner als im vorherigen Fall (vergleiche Tabelle 5.4).

Somit konnten auch im Fall eines Regressionsmodells mit $\eta_i = 6 + 1 \cdot x_i$ für $x_i \in [-3, 3]$ die Parameter des logistischen Modells mit logit-Link für die Wahrscheinlichkeiten π_i und log-linearem Link für den zweiten Parameter ω angemessen geschätzt werden.

Zusammengefasst soll festgehalten werden, dass die auf `optim()` aufbauende Maximum Likelihood Schätzung für die multiplikative Binomialverteilung gut funktioniert. Die Schätzungen konvergieren in allen Fällen, wenn als Startwerte für die Optimierung die Punktschätzer des logistischen Modells verwendet werden. Die gesuchten Punktschätzer können mit der verwendeten Methode und den geschriebenen Funktionen mit vernünftigem Aufwand gefunden werden.

5.4 Simulationen basierend auf doppelt binomialverteilten Zufallsvariablen

In diesem Abschnitt sollen nun die Simulationsergebnisse zur doppelten Binomialverteilung diskutiert werden. Den Anfang macht dabei abermals eine kurze Betrachtung der theoretischen Momente im Vergleich zu den simulierten, empirischen Momenten von Zufallsstichproben aus doppelt binomialverteilten Populationen mit unterschiedlichen Parameterwerten.

5.4.1 Vergleiche von Erwartungswert und Varianz

Für diese Vergleiche zwischen theoretischen und empirischen Momenten wurden für die Parameter π und ϕ sowie verschiedene Blockgrößen n ähnlich zur Simulation im Fall der multiplikativen Binomialverteilung die Werte

$$\begin{aligned} n &= \{3^2; 4^2; 6^2; 8^2; 11^2\} \\ \pi &= \{0,6; 0,8; 0,9; 0,95; 0,975; 0,99; 0,995\} \\ \text{und } \phi &= \{0,9; 0,95; 0,98; 1; 1,02; 1,05; 1,1\} \end{aligned}$$

gewählt. Für jede Kombination dieser Werte wurde in R eine Monte-Carlo Simulation durchgeführt, um die empirischen Momente der doppelten Binomialverteilung zu berechnen. Diese Berechnungen sollen zeigen, dass sowohl die Erzeugung von doppelt binomialverteilten Zufallszahlen als auch die Berechnung der theoretischen Erwartungswerte korrekt funktioniert. Dafür besteht wie bereits in Abschnitt 5.3.1 jede Simulationsserie aus 5000 doppelt binomialverteilten Zufallsvariablen, aus denen Erwartungswert und Varianz geschätzt werden können. Aus den 1000 Wiederholungen lassen sich dann der mittlere Erwartungswert sowie die mittlere Varianz errechnen.

Die größten Abweichungen zwischen den simulierten und den theoretisch berechneten Erwartungswerten sind in Tabelle 5.6 zu finden, jene der Varianzen in Tabelle 5.7.

n	ϕ	π	theoretisch		empirisch	
			E	Var	E	Var
121	0,98	0,8	96,8001	19,7552	96,7919	19,7311
121	0,90	0,8	96,8004	21,5116	96,7943	21,4564
121	1,10	0,6	72,6000	26,3999	72,5947	26,3828
121	1,02	0,8	96,7999	18,9803	96,7955	18,9625
16	1,02	0,6	9,5999	3,7643	9,5958	3,7584

Tabelle 5.6.: Die fünf größten Unterschiede zwischen empirischen und theoretischen Erwartungswerten der doppelten Binomialverteilung.

n	ϕ	π	theoretisch		empirisch	
			E	Var	E	Var
121	0,90	0,6	72,6001	32,2669	72,6012	32,1879
121	1,00	0,6	72,6000	29,0400	72,5959	28,9786
121	0,90	0,8	96,8004	21,5116	96,7943	21,4564
121	1,10	0,8	96,7998	17,5997	96,8049	17,5678
64	1,02	0,6	38,4000	15,0588	38,4045	15,0293

Tabelle 5.7.: Die fünf größten Unterschiede zwischen empirischen und theoretischen Varianzen der doppelten Binomialverteilung.

Wie diese Vergleiche der Erwartungswerte erkennen lassen, treten besonders deutliche Unterschiede zwischen exakter Berechnung und Simulation vor allem bei $n = 121$ auf, der größten betrachteten Gruppengröße. Die Wahl des Parameters ϕ läßt vor allem beim Vergleich der Erwartungswerte noch erkennen, dass bei $\phi \neq 1$ größere Unterschiede aufzutreten scheinen als bei $\phi = 1$. Außerdem zeigt sich, dass π vor allem für zentralere Werte zu größeren Unterschieden zwischen den Simulationsergebnissen und den berechneten Werten führt.

Allerdings handelt es sich selbst bei den hier gezeigten Differenzen von berechneten und simulierten Werten nur um geringfügige Abweichungen. Daher wird davon ausgegangen, die zur Berechnung der Momente sowie zur Generierung doppelt binomialverteilter Zufallsvariablen geschriebenen Funktionen arbeiten korrekt. Ausgehend davon werden nun Parameterschätzungen doppelt binomialverteilter Zufallsvariablen durchgeführt.

5.4.2 Simulationsergebnisse der Parameterschätzung mittels `optim()`

Wie bei der multiplikativen Binomialverteilung wird auch für die doppelte Binomialverteilung untersucht, wie sich die Parameterschätzung mit der Optimierungsroutine `optim()`, der log-Likelihoodfunktion und den ersten und zweiten partiellen Ableitungen derselben im Fall eines Regressionsmodells verhält. Begonnen wird dabei abermals mit dem iid Fall, im dem die Wahrscheinlichkeit für Erfolg nur von einem β_0 abhängt. Danach wird in einem weiteren Schritt diese Abhängigkeitsstruktur zu einem einfachen logistischen Modell mit Prädiktor $\beta_0 + \beta_1 x$ erweitert, wobei x wiederum als feste Sequenz von zehn mal 601 Werten aus dem Intervall $[-3, 3]$ definiert ist.

Parameterwahl $\beta_0 = 3$ und $\phi < 1$ (Überdispersion)

Den Anfang macht wie bereits erwähnt eine Monte-Carlo Simulation, bei der die Erfolgswahrscheinlichkeit π der Variablen nur von einem konstanten Term β_0 abhängt, der für dieses Beispiel auf $\beta_0 = 3$ gesetzt wird. Für den Parameter ϕ wird $\phi = 0,2$ gewählt. Außerdem werden zum Vergleich auch die Ergebnisse für $\phi = 1$, also den Fall der klassischen Binomialverteilung, wiedergegeben. Die Simulation besteht aus 500 Wiederholungen, wobei jede Wiederholung auf tausend doppelt binomialverteilten Zufallsvariablen basiert.

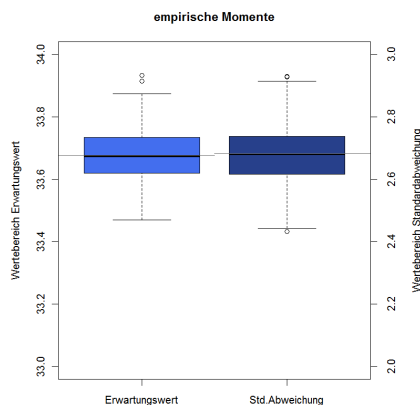


Abbildung 5.24.: Empirische Erwartungswerte und Standardabweichungen für Stichproben mit $\beta_0 = 3$ und $\phi = 0,2$. Die theoretischen Werte der Momente sind als graue Linien eingezeichnet.

Die in Abbildung 5.24 zusammengefassten Ergebnisse zeigen nochmals die empirischen Erwartungswerte und Varianzen aus den 500 Wiederholungen im Vergleich zu ihren theoretischen Werten. Sie dienen in erster Linie nochmals der Überprüfung der Berechnungsroutine der beiden Momente in R. Aufgrund der guten Übereinstimmung zwischen den 500 empirischen und dem wahren Erwartungswert 33,68 beziehungsweise der wahren Standardabweichung 2,6809 (graue Linien) wird davon ausgegangen, dass diese Berechnungsroutinen korrekt arbeiten.

Die Boxplotserien in Abbildung 5.25 zeigen die resultierenden $\hat{\beta}_0$ der Punktschätzung für $\phi = 0,2$ und zum Vergleich auch für die klassische Binomialverteilung bei $\phi = 1$ (rechte Boxplotserie). Wie bereits bei der multiplikativen Binomialverteilung besteht die erste Box jeder Serie aus den Punktschätzern $\hat{\beta}_0$, die zweite aus den aus der Hessematrix berechneten, exakten Standardfehlern und die dritte Box aus den im Zuge der Optimierung geschätzten Standardfehlern.

Wie bei der multiplikativen Binomialverteilung ist auch hier ein Anwachsen der Standardfehler der Punktschätzer im Fall $\phi \neq 1$ zu beobachten. Jedoch liegen die Punktschätzer aus den 500 Simulationen in beiden Fällen immer noch in einem engen Bereich um den wahren Wert β_0 verteilt. Die Standardfehler aus den beiden alternativen Berechnungsmethoden - ob über die exakte Hessematrix oder über die numerische Näherung - stimmen gut überein. Die Mittelwerte der Standardfehler für β_0 liegen für $\phi < 1$ bei 0,0867 und bei 0,0248 für $\phi = 1$.

Abbildung 5.26 zeigt, wie sich die Punktschätzer für den zweiten Parameter ϕ des Modells verhalten. Dabei sei darauf hingewiesen, dass die Skalen der beiden linken y-Achsen aufgrund der beiden unterschiedlichen Wertebereiche hier nicht denselben Bereich zeigen, beide jedoch

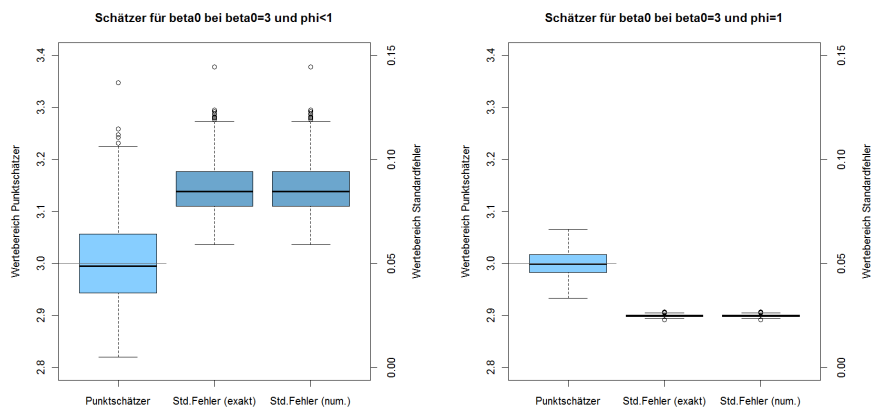


Abbildung 5.25.: Vergleich der Schätzer $\hat{\beta}_0$ für wahres $\beta_0 = 3$ bei $\phi = 0,2$ (links) beziehungsweise $\phi = 1$ (rechts).

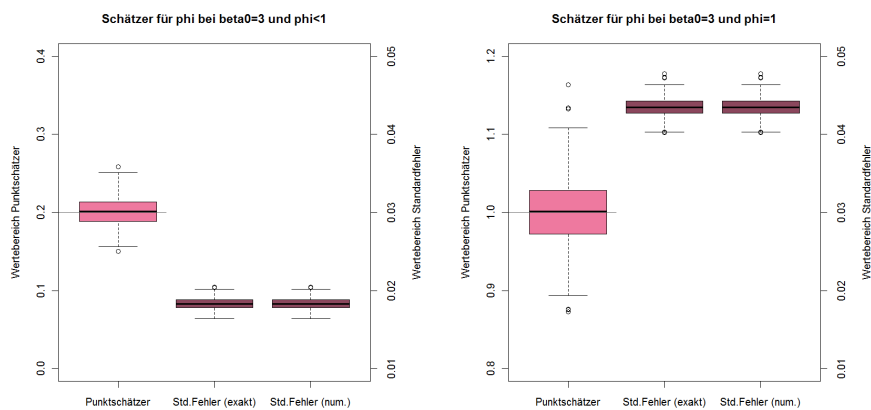


Abbildung 5.26.: Vergleich der Schätzer $\hat{\phi}$ bei $\beta_0 = 3$ für wahres $\phi = 0,2$ (links) beziehungsweise $\phi = 1$ (rechts).

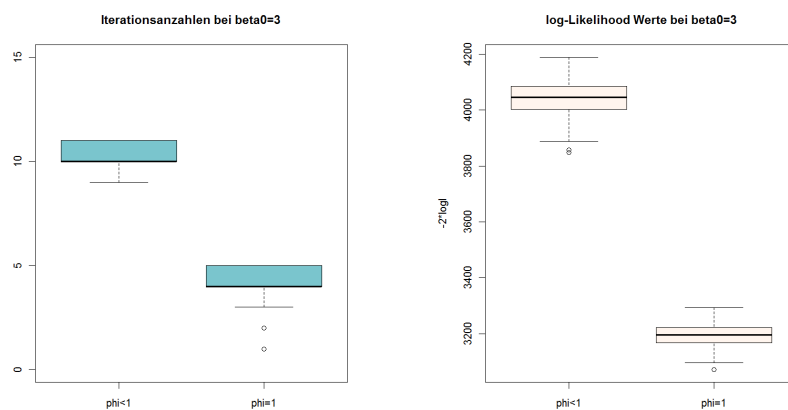


Abbildung 5.27.: Boxplots der benötigten Iterationen (links) und Vergleich der maximalen negativen doppelten log-Likelihoodwerte (rechts) für $\beta_0 = 3$ bei $\phi = 0,2$ und $\phi = 1$.

ein Intervall derselben Länge zeigen. Wie aus den Boxplotserien abzulesen ist, streuen die $\hat{\phi}$ der 500 Simulationen nur wenig um den wahren Wert ϕ , wobei diese Streuung für $\phi < 1$ sogar kleiner ist als im Fall der klassischen Binomialverteilung. Die Standardfehler von $\hat{\phi}$ betragen im Mittel 0,0183 bei $\phi = 0,2$ und 0,0435 bei $\phi = 1$.

Was also die Unterschiede zwischen $\phi < 1$ und $\phi = 1$ angeht, unterscheidet sich die Parameterschätzung für die doppelte Binomialverteilung von den Ergebnissen der multiplikativen Binomialverteilung. Dort war für Punktschätzer bei $\omega \neq 1$ ein deutlicher Anstieg des Standardfehlers des Parameterschätzers von β_0 zu beobachten, wohingegen dieser hier selbst bei $\phi < 1$ sehr gering ist. Über den Schätzer von ϕ ist zu sagen, dass dessen Standardfehler hier bei $\phi < 1$ sogar deutlich kleiner ist als im Fall der klassischen Binomialverteilung, wohingegen bei der multiplikativen Binomialverteilung das Gegenteil zu beobachten war.

Wie aus Abbildung 5.27 abzulesen liegt die Anzahl an benötigten Iterationen für dieses Beispiel im Mittel bei 4 für $\phi = 1$ und steigt für $\phi < 1$ auf 10. Im schlechtesten Fall benötigt die Optimierungsprozedur unter $\phi < 1$ aber auch nur 11 Iterationen, um das Optimum zu finden. Die Konvergenz der Parameterschätzung stellt somit mit den Standardeinstellungen der `optim()`-Routine kein Problem dar.

Zuletzt zeigt Abbildung 5.27 noch die log-Likelihoodwerte in den resultierenden Punktschätzern $\hat{\beta}_0$ und $\hat{\phi}$ für $\phi = 0,2$ und $\phi = 1$. Wie schon bei der multiplikativen Binomialverteilung läßt sich erkennen, dass die numerischen Werte, an denen in den Schätzungen das jeweilige Optimum aufgetreten ist, in Intervallen liegen, deren Breite nicht von der Wahl von ϕ abhängt.

Parameterwahl $\beta_0 = 4$ und $\phi > 1$ (Unterdispersion)

Nachdem nun bekannt ist, wie gut die Parameterschätzung für $\phi < 1$ funktioniert, soll nun das Augenmerk auf den Fall $\phi > 1$ gerichtet werden. Für diesen Vergleich wurde $\beta_0 = 4$ gewählt, was einer Erfolgswahrscheinlichkeit $\pi = 0,9820$ entspricht. Wie bereits erwähnt sind diese Betrachtungen am Rand des Einheitsintervalls von besonderem Interesse, da auch die in Kapitel 6 betrachteten Fälle eine Wahrscheinlichkeit für erfolgreiche Farbübertragung von fast 1 aufweisen werden. Für ϕ wurde der Wert $\phi = 1,8$ gewählt.

Der Vergleich der Punktschätzer $\hat{\beta}_0$ der Simulationsserie in Abbildung 5.28 zeigt für $\beta_0 = 4$ recht geringe Unterschiede zwischen dem Fall $\phi = 1,8$ und $\phi = 1$. Die aus der Optimierung der log-Likelihoodfunktion resultierenden Punktschätzer liegen in beiden Fällen in einem engen Bereich um den wahren Wert β_0 . Eine Eigenschaft, die bei der multiplikativen Binomialverteilung nicht so deutlich beobachtet werden konnte. Dort streuen die Punktschätzer für $\omega \neq 1$ bereits bei $\pi = 0,95$ stärker um den wahren Wert β_0 (vergleiche Boxplotserien in Abbildung 5.3) als hier mit $\pi = 0,98$. Bei den Standardfehlern der Punktschätzer $\hat{\beta}_0$ fällt auf, dass diese bei $\phi > 1$ sogar geringer sind als bei $\phi = 1$.

Auch für die Punktschätzer $\hat{\phi}$ zeigt sich, dass die Werte aus den 500 Simulationen im Mittel den wahren Wert ϕ gut erreichen können. Wie in Abbildung 5.29 zu sehen ist, gilt dies unabhängig von der Wahl von ϕ . Die Mittelwerte beider Punktschätzer stimmen in den ersten beiden Nachkommastellen mit ihren Sollwerten überein. Die Standardfehler liegen im Mittel bei 0,0697 für $\phi = 1$ und steigen dann geringfügig auf 0,0941 für $\phi = 1,8$.

Im Gegensatz zur klassischen Binomialverteilung bei $\phi = 1$ ist die Varianz einer doppelt binomialverteilten Stichprobe bei $\phi = 1,8$ kleiner. Der tatsächliche Wert beträgt 0,4272 bei $\phi = 1,8$ im Gegensatz zu 0,6359 bei der klassischen Binomialverteilung. Für $\phi = 1,8$ wurde somit ein Fall von Unterdispersion diskutiert. Für den weiter oben betrachteten Fall mit $\beta_0 = 3$

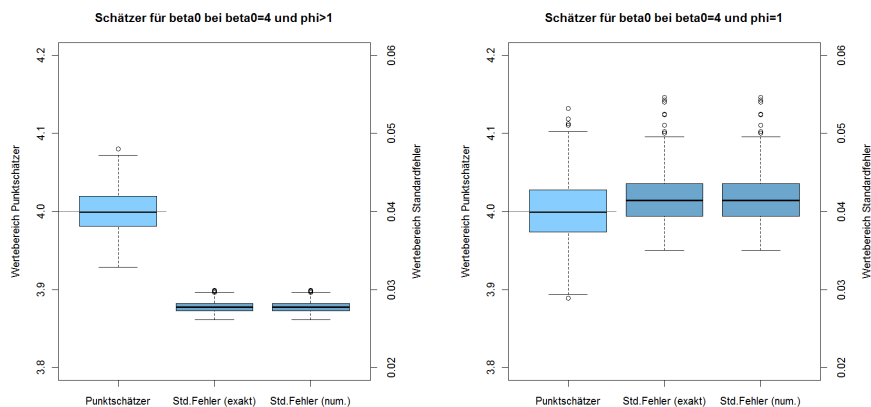


Abbildung 5.28.: Vergleich der Schätzer $\hat{\beta}_0$ für wahres $\beta_0 = 4$ bei $\phi = 1, 8$ (links) beziehungsweise $\phi = 1$ (rechts).

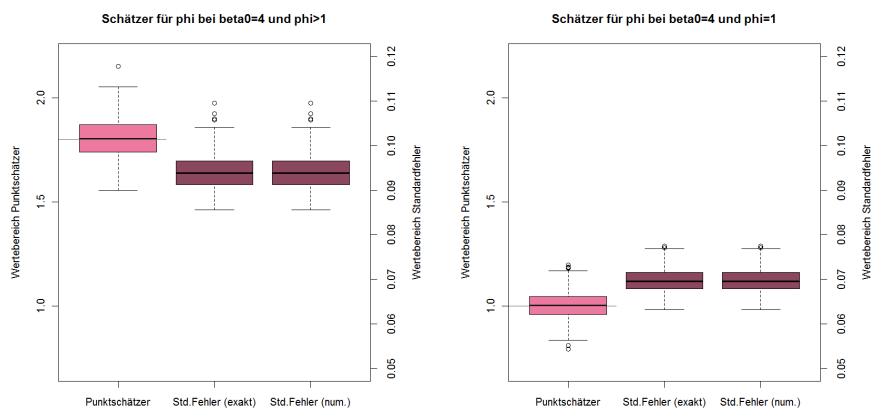


Abbildung 5.29.: Vergleich der Schätzer $\hat{\phi}$ bei $\beta_0 = 4$ für wahres $\phi = 1, 8$ (links) beziehungsweise $\phi = 1$ (rechts).

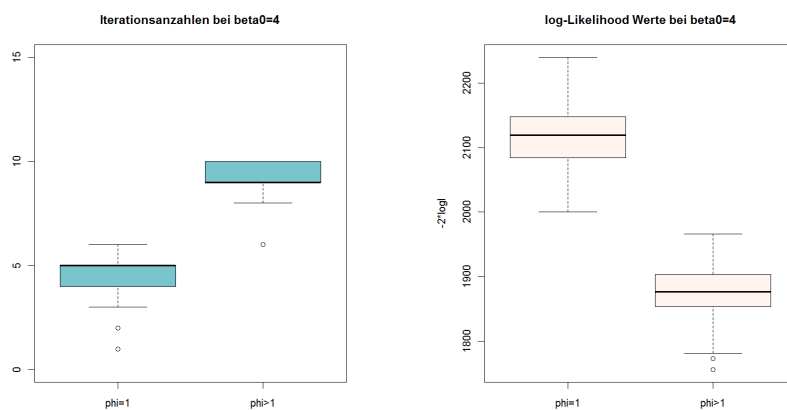


Abbildung 5.30.: Boxplots der benötigten Iterationen (links) und Vergleich der maximalen negativen doppelten log-Likelihoodwerte (rechts) für $\beta_0 = 4$ bei $\phi = 1, 8$ und $\phi = 1$.

und $\phi < 1$ beträgt die Varianz bei $\phi = 0,2$ und $\pi = 0,95$ immer noch 7,1874. Bei einer Stichprobe der klassischen Binomialverteilung mit demselben π hingegen beträgt die Varianz nur noch 1,6264. In diesem Fall ist es also durchaus gerechtfertigt, von deutlicher Überdispersion zu sprechen.

Abbildung 5.30 zeigt die Anzahl der benötigten Iterationen bis zur erzielten Konvergenz. Im Gegensatz zur multiplikativen Binomialverteilung zeigt sich hier, dass für $\phi > 1$ deutlich mehrere Iterationen benötigt werden, um das gesuchte Optimum zu finden. Außerdem zeigt sie noch die Wertebereiche der log-Likelihoodfunktion in den Lösungen $\hat{\beta}$ und $\hat{\phi}$ des Optimierungsproblems. Im Gegensatz zur multiplikativen Binomialverteilung, bei der diese bereits für ω geringfügig größer oder kleiner 1 deutlich unterschiedliche Werte geliefert haben, ist der Unterschied der Werte der log-Likelihoodfunktionen hier gering. Die Abweichungen der Verteilung für $\phi \neq 1$ gehen deutlich langsamer vonstatten als in der multiplikativen Binomialverteilung.

Nachdem nun gezeigt wurde, dass die Parameterschätzung für Modelle mit nur einem konstanten Term funktioniert, soll im nächsten Schritt auch eine erklärende Größe x in das betrachtete Modell aufgenommen werden.

Parameterwahl $\beta_0 = 3$ und $\beta_1 = 1$ für verschiedene ϕ

Die neu in das betrachtete Modell aufgenommene Größe x besteht aus einer festen Sequenz von 6010 Werten aus dem Intervall $[-3, 3]$, die für jede Simulation verwendet werden.

Dieses Simulationsbeispiel bewertet die Parameterschätzung unter dem Modell $\eta = 3 + 1 \cdot x$. Mit diesen Werten werden Erfolgswahrscheinlichkeiten π erzeugt, die zwischen 0,5 und 0,9975 liegen. Für ϕ werden die Werte $\phi = 0,2$, $\phi = 1$ und $\phi = 1,8$ betrachtet.

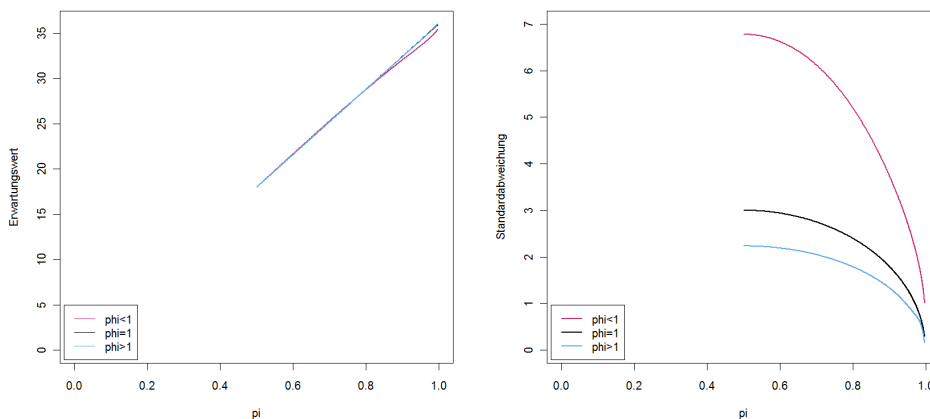


Abbildung 5.31.: Erwartungswerte (links) und Standardfehler (rechts) im Fall $\eta = 3 + 1 \cdot x$ für drei verschiedene Werte von ϕ und $x \in [-3; 3]$.

Erwartungswert und Standardabweichung von mit diesen Werten erzeugten Serien an Zufallsvariablen liegen in den in Tabelle 5.8 aufgelisteten Intervallen. Auch sind die Wertebereiche in Abbildung 5.31 gezeigt. Wie diese Vergleiche zeigen, greift die Wahl des Parameters ϕ kaum merkbar in den Erwartungswert ein. Nur für $\phi = 0,2$ lässt sich für große π eine geringe Abweichung von den beiden anderen betrachtete Situationen erkennen. Außerdem ändern sich die

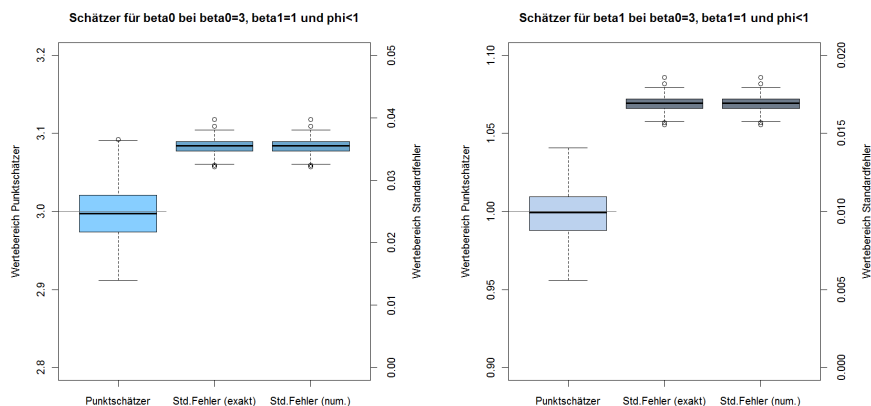


Abbildung 5.32.: Resultierende Punktschätzer $\hat{\beta}_0$ (links) und $\hat{\beta}_1$ (rechts) für wahres $\phi = 0,2$ sowie $\beta_0 = 3$ und $\beta_1 = 1$.

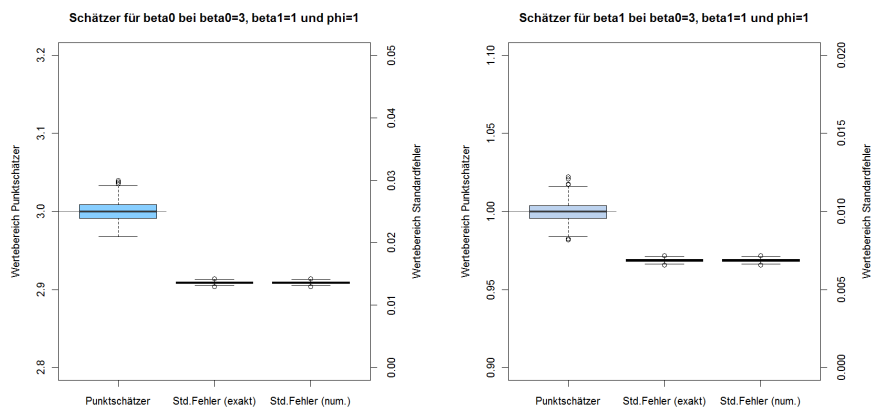


Abbildung 5.33.: Resultierende Punktschätzer $\hat{\beta}_0$ (links) und $\hat{\beta}_1$ (rechts) für wahres $\phi = 1$ sowie $\beta_0 = 3$ und $\beta_1 = 1$.

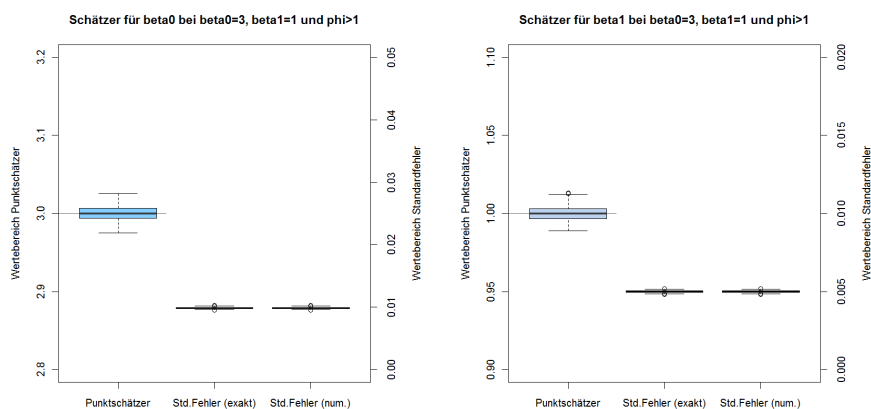


Abbildung 5.34.: Resultierende Punktschätzer $\hat{\beta}_0$ (links) und $\hat{\beta}_1$ (rechts) für wahres $\phi = 1,8$ sowie $\beta_0 =$ und $\beta_1 = 1$.

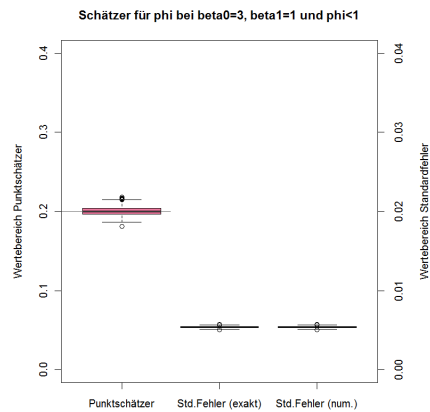


Abbildung 5.35.: Resultierende Punktschätzer $\hat{\phi}$ mit Standardfehlern im Fall von $\phi = 0,2$ bei $\beta_0 = 3$ und $\beta_1 = 1$.

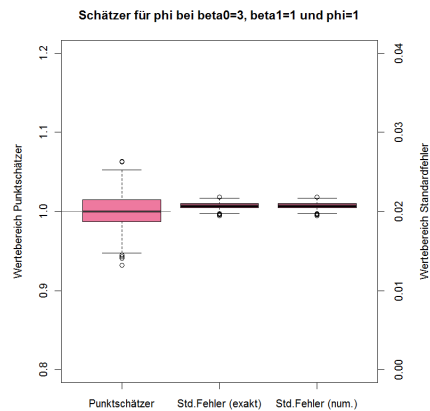


Abbildung 5.36.: Resultierende Punktschätzer $\hat{\phi}$ mit Standardfehlern im Fall von $\phi = 1$ bei $\beta_0 = 3$ und $\beta_1 = 1$.

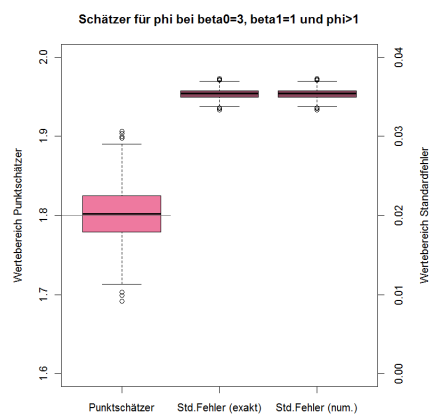


Abbildung 5.37.: Resultierende Punktschätzer $\hat{\phi}$ mit Standardfehlern im Fall von $\phi = 1,8$ bei $\beta_0 = 3$ und $\beta_1 = 1$.

ϕ	Erwartungswerte	Standardabweichungen
0,8	[18,00; 35,44]	[1,0163; 6,7785]
1,0	[18,00; 35,91]	[0,2980; 3,0000]
1,2	[18,00; 35,97]	[0,1654; 2,2359]

Tabelle 5.8.: Wertebereiche der ersten beiden Momente bei $\pi_i = \text{logit}(x_i^t \boldsymbol{\beta})$.

Standardabweichungen stärker als bei der für die multiplikativen Binomialverteilung betrachteten Situation (vgl. z.B. Grafik 3.5 in Kapitel 3 oder Tabelle 5.5 in Abschnitt 5.3.2). Hier jedoch ist nun eine scharfe Trennung von Über- und Unterdispersion allein anhand des Parameters ϕ möglich. Wie in Abbildung 5.31 zu sehen, variieren doppelt binomialverteilte Zufallsvariablen mit $\phi < 1$ deutlich stärker als im Fall der klassischen Binomialverteilung. Und für $\phi < 1$ variieren doppelt binomialverteilte Zufallsvariablen schwächer als im Fall $\phi = 1$.

Die Grafiken in den Abbildungen 5.32 bis 5.34 zeigen nun, wie sich die Punktschätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ für verschiedene Werte von ϕ verhalten. Die Punktschätzer $\hat{\beta}_0$ aus den 500 Durchgängen der Simulation liegen unabhängig der Wahl von ϕ in einem sehr engen Bereich um den wahren Wert $\beta_0 = 3$. Die Standardfehler der Punktschätzer betragen im Mittel 0,0355 bei $\phi = 0,2$ und sinken dann für wachsendes ϕ , bis bei $\phi = 1,8$ im Mittel der Wert 0,0099 erreicht wird.

Auch die Werte $\hat{\beta}_1$ stimmen im Mittel mit dem wahren Wert $\beta_1 = 1$ gut überein. Die Standardfehler dieses Punktschätzers werden wie im Fall von β_0 für $\phi < 1$ größer und für $\phi > 1$ etwas kleiner als im Fall der klassischen Binomialverteilung. Dasselbe konnte schon im Fall der multiplikativen Binomialverteilung beobachtet werden, im Gegensatz dazu sind die Unterschiede hier jedoch gering. So liegt der Standardfehler von $\hat{\beta}_1$ bei der Wahl $\phi = 0,2$ bei 0,0169 und fällt für $\phi = 1,8$ auf 0,0050.

Im Vergleich zur multiplikativen Binomialverteilung sind die Ergebnisse der Parameterschätzung bei der doppelten Binomialverteilung somit deutlich genauer. Was die Standardfehler der Punktschätzer unter verschiedenen ϕ angeht, sind diese im Fall der doppelten Binomialverteilung für $\phi \geq 1$ deutlich kleiner.

Die in den Abbildungen 5.35 bis 5.37 gezeigten Boxplotserien zeigen die Ergebnisse der Punktschätzung für $\hat{\phi}$. Dabei wurde für diese drei Plots die Skala der linken y-Achse abermals an den Wertebereich des jeweiligen ϕ angepasst. Zu erkennen ist, dass die Punktschätzung dieses Parameters für wachsendes ϕ unschärfer wird. Dennoch stimmen die Werte der $\hat{\phi}$ im Mittel gut mit dem wahren Wert ϕ überein und unterscheiden sich erst ab der dritten Nachkommastelle von ihren Sollwerten. Die dazugehörigen Standardfehler liegen bei 0,0054 für $\phi = 0,2$ und wachsen dann auf 0,0354 für $\phi = 1,8$ an.

Somit ist es für Erfolgswahrscheinlichkeiten zwischen 0,5 und 0,9975 möglich, Modelle auf Basis der doppelten Binomialverteilung zu schätzen. Für Fälle mit Erfolgswahrscheinlichkeiten weiter im Zentrum des Einheitsintervalls ist dies ebenso möglich, wird in dieser Arbeit aber nicht gesondert diskutiert. Ebenso kann das Intervall der Erfolgswahrscheinlichkeiten auch noch weiter an den Rand des Einheitsintervalls geschoben werden. Für $\boldsymbol{\beta} = (6, 1)$ wurde dies noch durchgeführt, ohne Probleme mit der Parameterschätzung zu bekommen. Jedoch wird an dieser Stelle auf eine ausführliche Darlegung von Simulationen mit weiteren Parameterkombinationen verzichtet.

Statt dessen werden im kommenden Abschnitt die R-Funktionen vorgestellt, welche zur Mo-

dellschätzung unter Annahme der multiplikativen oder der doppelten Binomialverteilung geschrieben wurden.

5.5 R-Code zur Modellierung unter multiplikativer sowie doppelter Binomialverteilungsannahme

Um Modellschätzungen unter Annahme von multiplikativer oder doppelter Binomialverteilung zugänglicher zu gestalten, wurden die R-Funktionen `gbmod()`, `gbsummary()` und `gbplot()` erstellt.

Die eigentliche Modellschätzung erfolgt dabei innerhalb der Funktion `gbmod()`, deren Aufruf ähnlich zu den in R üblichen Modellierungsfunktionen gestaltet ist. Wie aus der Funktionsdefinition in Code 5.4 abzulesen ist, benötigt sie zum einen eine Modellspezifikation der Art $\mathbf{y} \sim \mathbf{x1} * \mathbf{x2}$ und zum anderen muss angegeben werden, mit welcher Verteilungsannahme gearbeitet werden soll. Zulässige Argumente dafür sind `multbin`, `doublebin` und `binomial`. Die standardmäßig auf `FALSE` gesetzte Variable `trace` erlaubt es, einige Zeilen Ausgabe direkt von `gbmod()` zu erhalten.

Die Funktion `gbmod()` errechnet in einem ersten Schritt die Punktschätzer $\hat{\beta}$ eines logistischen Regressionsmodells (siehe Zeile 3). Der in den verallgemeinerten Binomialverteilungen hinzukommende zweite Parameter ω beziehungsweise ϕ wird mittels log-linearem Link modelliert. Es gilt also $\alpha = \log \omega$ im Fall der multiplikativen Binomialverteilung oder $\alpha = \log \phi$ bei doppelter Binomialverteilungsannahme. Die Punktschätzer eines logistischen Modells werden dann, zusammen mit einem Wert α für den zweiten Parameter, als Startwerte für die Optimierungsroutine verwendet. Dabei wird α so gewählt, dass die klassische Binomialverteilung vorliegt, also $\alpha = 0$ (Zeilen 5 - 9). Mittels `optim()` können dann die Punktschätzer des neuen Modells errechnet und aus der `result`-Liste ausgelesen werden. Die an `optim()` übergebenen Elemente sind dabei neben den Startwerten die negative log-Likelihoodfunktion sowie die negativen partiellen ersten Ableitungen der entsprechenden log-Likelihoodfunktion. Der Parameter `method` steuert die zu verwendende Optimierungsroutine, wobei die in diesem Fall angewandte Methode von Broyden, Fletcher, Goldfarb und Shanno bereits in Abschnitt 5.1.2 besprochen wurde. Weiters benötigt `optim()` den Vektor der erfolgreichen Beobachtungen `y`, die Gesamtanzahl aller Beobachtungen `n` einer Gruppe (Vektor oder Skalar) sowie die Designmatrix `X`. Die Variable `ResponseDistr` wird übergeben, damit in den beiden Funktionen `neg.logl()` und `neg.gradient()` die richtige Verteilung ausgewählt werden kann. Auch eine Anpassung der üblichen Argumente von `optim()` durch den Benutzer ist möglich.

Wird der Schalter `trace` auf `TRUE` gesetzt, so gibt die Funktion `gbmod()` einige Zeilen Modellzusammenfassung aus, die in Code 5.5 aufgelistet sind. Für dieses Beispiel wurde ebenfalls wieder eine regelmäßige Sequenz von 601 Werten aus dem Intervall $[-3; 3]$ verzehnfacht und mit diese 6010 Werte als erklärende Größe x angenommen. Der Parameter ω der multiplikativen Binomialverteilung wurde auf $\omega = 1,08$ gesetzt.

Bei dem von `gbmod()` zurückgegebenen Element handelt es sich um eine Liste von Informationen, welche in Anlehnung an jene von `glm()` die folgenden Einträge enthält:

`coefficients` Maximum Likelihood Schätzer von β
`theta` Maximum Likelihood Schätzer von $\exp(\alpha)$

```
gbmod <- function(formula, ResponseDistr, trace=FALSE, ...){
2   [...]
   glmmod <- glm(formula, family=binomial, x=TRUE, ...)
4   initCoefficients <- glmmod$coefficients
   if (ResponseDistr=="binomial") {
6     initAlpha <- NaN # the artificial case
   } else {
8     initAlpha <- 0 # start value for alpha
   }
10  startParams <- c(initCoefficients, initAlpha)
   [...]
12
14  result <- optim(startParams, fn=neg.logl, gr=neg.grad, method="BFGS",
   y=y, n=n, X=X, ResponseDistr=ResponseDistr, ...)

16  par <- result$par
   coefficients <- head(par,-1)
18  alpha <- tail(par,1)
   theta <- exp(alpha)

20
22  p <- length(par)
   N <- length(y)

24  hessian <- hess(coefficients, alpha, y, n, X, ResponseDistr)
   logl <- -2*logl(coefficients, alpha, y, n, X, ResponseDistr)
26  aic <- gb.logl + 2*p

28  [...]
   mod <- list([...])
30  class(mod) <- 'generalizedBinomial'
   return(mod)
32 }
```

Code 5.4: Kern der Funktionsdefinition von `gbmod()`.

```

> mod <- gbmod(cbind(y,n-y)~x1, "multbin", trace=TRUE)
2
Call:
4 gbmod(formula = cbind(y, n - y) ~ x1 , multbin )

6 Coefficients:
  (Intercept)          x1
8   3.025899    1.008110    1.081532

10 Degrees of Freedom: 6009 Total (i.e. Null); 6007 Residual
    -2*logl: 22758.2
12 AIC: 22764.2

14 Convergence reached after 12 Iterations.

```

Code 5.5: Überblickshafte Modellzusammenfassung im Fall von `trace=TRUE`.

pearson.residuals Vektor der Pearson Residuen
fitted.values Vektor der geschätzten Wahrscheinlichkeiten $\hat{\pi}_i$
family Verteilungsannahme (multiplikativ oder doppelt binomial)
linear.predictor Vektor der linearen Prädiktoren $\hat{\eta}_i$
aic AIC des Modells
iter Anzahl an Iterationen bis zur erzielten Konvergenz
df.residual Freiheitsgrade des betrachteten Modells
df.null Freiheitsgrade des Nullmodells
y Vektor der beobachteten Erfolge
n Vektor der Anzahl an Beobachtungen pro Zelle
X Designmatrix
converged Logisch; `TRUE` bzw. 1 wenn Modellschätzung konvergierte, `FALSE` bzw. 0 sonst
formula Die Modellformel in R-Notation, wobei die Responsevariable aus dem Vektor der Erfolge und der Misserfolge einer jeden Gruppe besteht.
glm.summary Ergebnisse der Schätzung unter dem binomialen Modell:
 coefficients Koeffizienten
 pearson.residuals Pearson Residuen
 fitted.values geschätzte Wahrscheinlichkeiten $\hat{\pi}_i$
 null.deviance Nulldevianz
 deviance Modelldevianz
 df.residual Freiheitsgrade
 aic Akaike Informationskriterium
hessian Hessematrix in den Punktschätzern $(\hat{\beta}, \hat{\alpha})$

Außerdem ist ein von `gbmod()` geschätztes Modell von der Klasse `generalizedBinomial` (siehe Zeile 30 in Code 5.4).

Die oben genannten Informationen werden von den beiden Funktionen `gbsummary()` und

`gbplot()` verwendet, um ähnlich den in R bereits vorhandenen Funktionen `summary()` und `plot()` Modellzusammenfassungen für ein mit `gbmod()` geschätztes Modell auszugeben. Dabei muss erwähnt werden, dass sowohl `gbsummary()` als auch `gbplot()` nur mit einem Objekt der Klasse `generalizedBinomial` funktionieren und es nicht möglich ist, die altbekannten R-Funktionen `summary()` und `plot()` auf ein Objekt der Klasse `generalizedBinomial` anzuwenden.

```

> mod <- gbmod(cbind(y,n-y)~x1,"multbin", trace=TRUE)
2 > gbsummary(mod)

4 Values from the binomial model:
  -2*logl: 23995.35
6   Deviance: 3953.071 on 6008 degrees of freedom
  Pearson X^2: 3609.018
8   AIC: 23999.35

10 Response distribution: multiplicative binomial distribution.

12 Pearson Residuals:
      Min.    1st Qu.    Median      Mean    3rd Qu.     Max.
14 -3.7020000 -0.6595000  0.0537800  0.0009023  0.6934000  3.6930000

16 Coefficients:
              Estimates Std.Errors t values P(>|t|)
18 (Intercept)    3.0259    0.0564    53.7 <2e-16 ***
   x1              1.0081    0.0176    57.4 <2e-16 ***
20 ---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

22 Coefficient of Parameter omega:
  Estimate Std.Error t value P(>|t|)
24  1.08153  0.00281   29.1 <2e-16 ***

26 Degrees of Freedom: 6009 Total (i.e. Null); 6007 Residual
28 -2*logl: 22758.2
  Pearson X^2: 6044.508
30 AIC: 22764.2

32 Convergence reached after 12 Iterations.

```

Code 5.6: Output der Funktion `gbsummary()`.

Ein beispielhafter Output von `gbsummary()` ist in Code 5.6 angegeben. Zum Vergleich mit einem logistischen Regressionsmodell unter klassischer Binomialverteilungsannahme sowie für Tests auf die Güte des Modells sind in den Zeilen 4 bis 8 der Wert der negativen doppelten log-Likelihoodfunktion, die Devianz, Pearson X^2 sowie der AIC-Wert dieses Modells angegeben. Im vorliegenden Beispiel ist die Devianz in etwa halb so groß wie die Anzahl der Freiheitsgrade, es liegt also deutliche Unterdispersion vor.

Danach zeigt der R-Code in den Zeilen 12 bis 14 die Zusammenfassung der Pearson Residuen

des Modells. Die Tabelle in den Zeilen 16 bis 21 gibt die aus der Optimierung resultierenden Parameterschätzer für β und ihre Standardfehler zurück. Der für den zusätzlichen Parameter resultierende Punktschätzer ist gesondert angeführt und in der Tabelle in den Zeilen 23 bis 25 zu finden. Dabei wird intern zwar mit dem Parameter α gerechnet, in Konformität mit der Definition der Verteilungen (siehe Formeln (3.8) und (4.2)) in der Zusammenfassung jedoch das Ergebnis für den entsprechenden Schätzer $\hat{\omega}$ beziehungsweise $\hat{\phi}$ angegeben. Der Standardfehler des Parameterschätzers errechnet sich dabei mit Hilfe der in Abschnitt 3.4.1 angesprochenen Delta-Methode. Der in Zeile 25 gezeigte Test testet mittels t-Test auf die Hypothese, der zusätzliche Parameter ω beziehungsweise ϕ sei 1.

Außerdem ist in Zeile 28 der Zusammenfassung 5.6 noch der negative doppelte log-Likelihoodwert des Modells unter Annahme einer generalisierten Binomialverteilung zu finden, welcher einen Vergleich des Modells mit jenem der klassischen logistischen Regression erlaubt. Ebenso ist mittels dem Pearson X^2 in Zeile 29 oder dem AIC-Wert in Zeile 30 ein Vergleich mit anderen Modellen möglich. Aufgrund der in Abschnitt 2.2.4 angesprochenen asymptotischen Nähe zur χ^2 -Verteilung erlaubt die Pearson X^2 Statistik darüber hinaus einen Test auf die Güte des Modells. Eine Devianz kann jedoch für solch ein Modell nicht mehr errechnet werden, da die dafür notwendige log-Likelihoodfunktion eines saturierten Modells nicht mehr aufgestellt werden kann.

Den Abschluss macht die in Zeile 32 angegebene Anzahl der Iterationen, die `optim()` für die Modellschätzung benötigte, wobei die maximale Anzahl standardmäßig bei 100 Iterationen liegt. Sollte die Modellschätzung nicht konvergieren, wird anstelle dieser Zeile der Text **Maximal number of iterations reached.** angezeigt.

Analog zur Funktion `gbsummary()`, welche der Funktion `summary()` aus R nachempfunden wurde, existiert auch eine Funktion `gbplot()`, welche ähnlich wie `plot()` arbeitet. Nach ihrem Aufruf mittels

```
> gbplot(mod, full=FALSE)
```

liefert sie eine Serie von Plots, welche in Abbildung 5.38 gezeigt sind. Sie zeigen die in diesen Modellen berechenbaren Pearson Residuen, geplottet gegen den linearen Prädiktor sowie als Normal-QQ-Plot. Außerdem sind die geschätzten Erwartungswerte sowie der Standardabweichungen des verallgemeinerten Modells gegen die Schätzungen derselben unter dem klassischen Binomialmodell gezeigt.

Für dieses generische Beispiel mit sortierten π_i lässt sich ein Vergleich der geschätzten Erwartungswerte und Standardabweichungen unter dem verallgemeinerten Modell sowie unter dem Binomialmodell auch über die in Abbildung 5.39 gezeigten Plots darstellen. Dabei sind in der linken Abbildung außerdem noch die beobachteten Werte eingezeichnet, um einen Vergleich mit der Realität zu erlauben. Für die vorliegenden Parameterkonstellationen ($\phi = 1,08$) liegt die geschätzte Standardabweichung unter multiplikativer Binomialverteilungsannahme deutlich unter jener des Binomialmodells. Diese beiden zusätzlichen Plots werden ausgegeben, wenn im Aufruf von `gbplot` der Schalter `full` auf `TRUE` gesetzt wird.

Dieses Kapitel hat gezeigt, wie die in den Kapiteln 3 und 4 aufgestellten ersten und zweiten Ableitungen der log-Likelihoodfunktionen in eine bestehende Optimierungsroutine eingebettet werden können. Eine ausführliche Diskussion von Simulationsergebnissen zeigte, dass die verwendeten Funktionen korrekt arbeiten und sinnvolle Parameterschätzer ermitteln können. R-Funktionen wurden vorgestellt, welche die Maximum Likelihood Schätzung unter doppel-

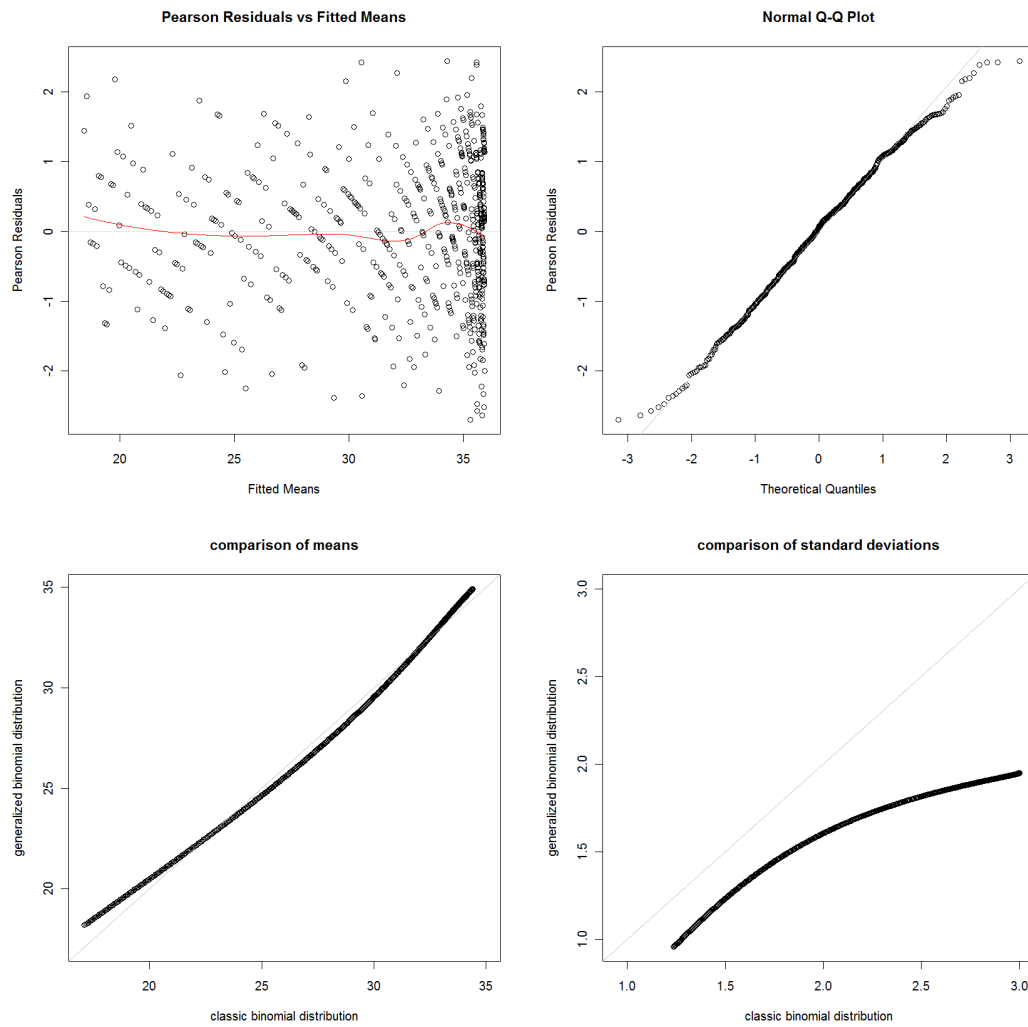


Abbildung 5.38.: Die vier Plots, welche beim Aufruf von `gplot()` für das vorliegende Modell erstellt werden.

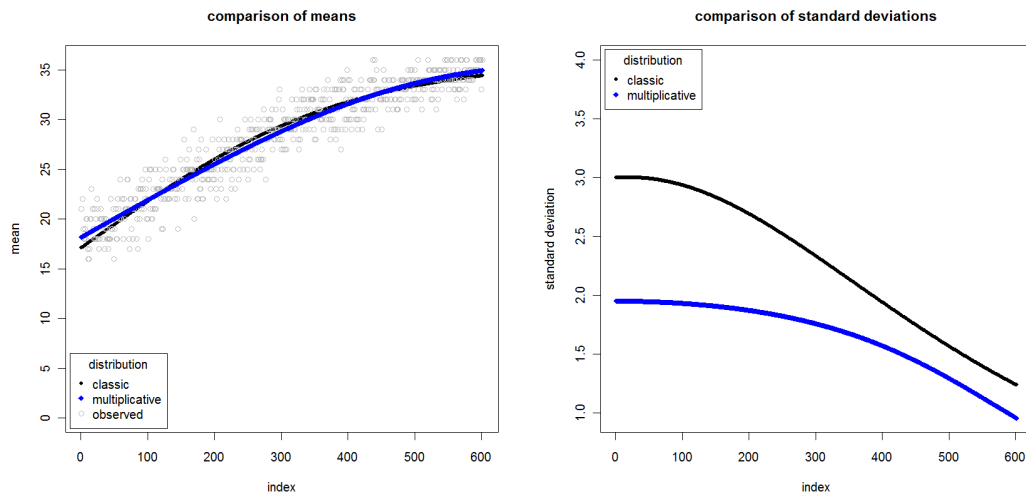


Abbildung 5.39.: Vergleich der geschätzten Erwartungswerte (links) und Standardabweichungen (rechts) unter klassischem und verallgemeinertem Binomialmodell.

ter oder multiplikativer Binomialverteilungsannahme mit wenigen Befehlen ermöglichen. Eine detailliertere Diskussion aller erstellten R-Funktionen ist in Anhang A zu finden.

Im letzten Teil der Arbeit werden besagte R-Funktionen nun verwendet, um die aus der Industrie stammenden Datensätze zu modellieren.

Modellierung der Farbübertragung

Nachdem in den vorangegangenen Kapiteln zwei Verteilungsfamilien besprochen wurden, welche eine Modellierung von verallgemeinert binomialverteilten Daten ermöglichen, wendet sich dieses Kapitel einigen realen Datensätzen zu, welche eine Variabilität größer oder kleiner jener der klassischen Binomialverteilung aufzuweisen scheinen.

Wie bereits zu Beginn angesprochen entstand diese Arbeit im Zuge des Projekts „Modellierung der Farbübertragung aus lokalen Papiereigenschaften“ am Institut für Papier-, Zellstoff- und Fasertechnik der Technischen Universität Graz. Dieses Kapitel beschäftigt sich daher mit der Modellierung eines Datensatzes, welcher im Zuge dieses Projekts erhoben wurde. Begonnen wird dabei in Abschnitt 6.1 mit einer ausführlichen Diskussion der vorliegenden Daten. Diese Einführung versucht, auch der Papier- und Druckindustrie fremden Personen eine grobe Vorstellung der vorliegenden Situation zu geben. Daher wird knapp auf die Papiereigenschaften Topografie und Formation eingegangen sowie ein Abriss einer bildanalytischen Signalverarbeitung skizziert. Ebenso wird das Flexodruckverfahren, welches für die Bedruckung der Proben verwendet wurde, vorgestellt.

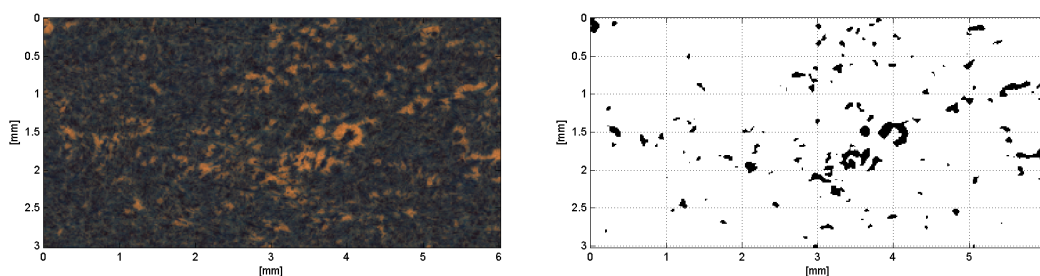


Abbildung 6.1.: Farbaufnahme eines Teils einer Probenfläche. Bedruckte Regionen sind blau, unbedruckte gelblich dargestellt. Im Binärbild desselben Ausschnittes sind die Fehlstellen in schwarz dargestellt.

Abschnitt 6.2 illustriert dann anhand des in Abbildung 6.1 gezeigten Ausschnittes einer Probe die Modellierung der Farbübertragung mit einem logistischen Regressionsmodell unter den

generalisierten Binomialverteilungen. Der Grundstein für diese Modellschätzungen wurde dabei bereits in der Diplomarbeit von Feirer (2008) gelegt, in der die Farbübertragung allein aufgrund der Topografie der Proben modelliert wurde. Dieser Teil beinhaltet viele Resultate, die unter anderem in Feirer, Hirn, Friedl und Bauer (2010b), Feirer, Hirn, Friedl und Bauer (2010a) oder Feirer, Hirn, Friedl und Bauer (2012) vorgestellt wurden. Anhand des kleinen Probenausschnittes wird gezeigt, inwiefern Formation und Topografie das Auftreten unbedruckter Stellen beeinflussen. Danach werden Modelle der vollen Proben des Datensatzes besprochen. Außerdem zeigt ein gemeinsames Modell aller vorhandenen Papiere deren Gemeinsamkeiten sowie Unterschiede auf.

Danach wird die in den mathematischen Kapiteln entwickelte Modellschätzung mit verallgemeinerten Binomialverteilungen in Abschnitt 6.3 auf die Papierdaten angewandt. Besonderes Augenmerk wird dabei auf Vergleiche von Modellen unter den drei unterschiedlichen Verteilungsannahmen gelegt und deren Unterschiede und Gemeinsamkeiten besprochen. Wie im Fall des logistischen Modells werden die Resultate anhand eines Ausschnitts einer Probe vorgestellt und erst danach die Ergebnisse der vollen Proben dargelegt.

Den Abschluss dieses Praxisteils macht ein Vergleich von Modellschätzungen unter unterschiedlichen Gruppengrößen n , um auf die in den Kapiteln 3 und 4 angesprochene Eigenschaft der beiden verallgemeinerten Binomialverteilungen, bezüglich Faltung nicht abgeschlossen zu sein, einzugehen.

6.1 Der Datensatz

6.1.1 Sackpapiere

Für die Modellierungen wurden sogenannte Sackpapiere (auch: Kraftpapiere, Kraftsackpapiere) untersucht. Diese Art von Papier ist besonders reißfest und wird daher zur Verpackung von so verschiedenen Gütern wie Zement oder Lebensmittel (Gebäck, Obst, Gemüse, . . .) verwendet. Aufgrund seiner speziellen Eigenschaften hat Sackpapier eine recht raue Oberflächenbeschaffenheit. Wie sich in Folge zeigen wird, wirkt sich diese raue Oberfläche des Papiers direkt auf das Druckbild aus.

Um das Druckbild eines Sackpapiers zu verbessern, kann eine Satinage des Papiers vorgenommen werden. Dabei wird das Papier zwischen zwei oder mehreren Walzen geglättet und so eine bessere Oberflächenbeschaffenheit erzielt. Aus diesem Grund besteht der betrachtete Datensatz aus Proben, welche mit unterschiedlichen Einstellungen satiniert wurden.

6.1.2 Flexodruck

Die untersuchten Sackpapiere wurden im Flexodruckverfahren bedruckt. Eine genaue Beschreibung des Verfahrens ist zum Beispiel in Kipphan (2000) zu finden, aus dem auch die in Abbildung 6.2 gezeigte Skizze stammt.

Das Besondere an diesem Hochdruckverfahren ist, dass mit einer weichen, elastischen Druckform gearbeitet wird. Die zu druckenden Stellen sind gegenüber den nicht druckenden Regionen auf der Druckplatte erhaben und werden mit einer Raster- oder Aniloxwalze gleichmäßig eingefärbt. Das Farbauftragsvolumen wird dabei über die Geometrie der Rasterwalze gesteuert. Im

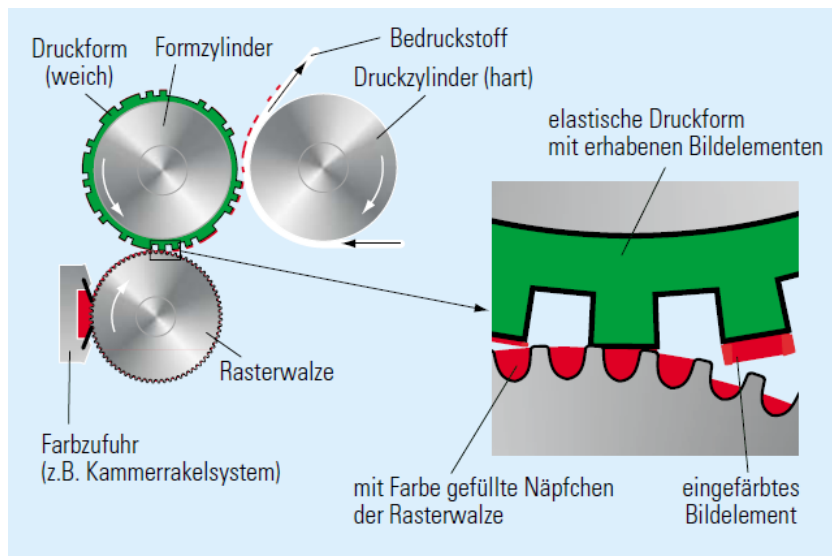


Abbildung 6.2.: Skizze der Funktionsweise des Flexodrucks, entnommen aus Kipphan (2000).

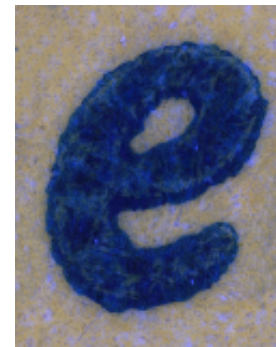


Abbildung 6.3.: Im Flexodruck typische Quetschränder.

Spalt zwischen Form- und Druckzylinder wird die Farbe auf den Bedruckstoff übertragen, wobei durch den Anpressdruck auf die Güte des resultierenden Druckbildes Einfluss genommen werden kann. Daher rührt auch das deutlichste Erkennungszeichen des Flexodrucks, Quetschränder entlang von Kanten, an denen sich mehr Druckfarbe als innerhalb der Farbfläche angesammelt hat (siehe Abbildung 6.3).

Moderne Maschinen bestehen oft aus 8 bis 10 Farbwerken. Jedoch gibt es vor allem im Flexodruck keine einheitlichen Lösungen, was die Bauweise der Maschinen ebenso betrifft wie die Beschaffenheit der flexiblen Druckplatten. Aufgrund der elastischen Druckform ist es mit dem Flexodruck möglich, die unterschiedlichsten Materialien wie Folien, Papiere, dicke Pappen, Wellpappen, Verpackungsmaterialien mit rauer Oberfläche oder Gewebe zu bedrucken.

Laut Moran (2011) ist Flexodruck der weltweit am schnellsten wachsende Analogdrucksektor mit einer geschätzten Wachstumsrate von 4 bis 5 % bis 2013. Demselben Artikel ist auch zu entnehmen, dass 92 % aller im Flexodruck bedruckten Erzeugnisse in der Verpackungssparte verwendet werden, wobei die restlichen 8 % unter anderem Sicherheitsdrucke oder gedruckte Elektronikprodukte ausmachen. Zu erwähnen sei außerdem noch, dass laut der schweizer Zeitschrift Pack Aktuell (Kreuter, 2009) über 60 % aller flexiblen Verpackungen mit Flexodruck bedruckt werden.

Einem Whitepaper von Heidelberg (*Verpackungsdruck*, 2008) ist zu entnehmen, dass die immer wichtiger werdende Rolle der Verpackung eines Gutes, welche den Endkunden zum Kauf verführen soll, einer der Hauptgründe für das Wachstum der Branche ist. Ebenso wird dort der vor allem in den Industrienationen vorherrschende Trend zu immer kleineren Haushalten genannt, welche als treibende Kraft zu vergleichsweise hohen Verpackungsanteilen führt. Für das Jahr 2007 kann derselbe Bericht den weltweiten Umsatz der Verpackungsindustrie mit 400 Milliarden Euro benennen. Für 2015 wird nach den im Bericht genannten Quellen¹ mit einem weltweiten Umsatz von 560 Milliarden Euro gerechnet, wobei 37 % auf den asiatischen,

¹Pira International und Schätzungen der IKB Deutsche Industriebank

35 % auf den europäischen und 25 % auf den nordamerikanischen Markt entfallen sollen.

Auch Sackpapier wird im Flexodruckverfahren bedruckt, da die weiche Druckform in der Lage ist, trotz der Unebenheiten der rauen Papieroberfläche eine gute Farbübertragung zu ermöglichen. Außerdem können im Flexodruck wasserbasierende Farben verwendet werden, die geruchlos und nicht gesundheitsschädlich und daher zum Bedrucken von Lebensmittelverpackungen besonders geeignet sind.

Aus diesen Gründen steigen auch für einen qualitativ minderwertigeren Druck wie den Flexodruck die Anforderungen an die Ergebnisse.

6.1.3 Formation

Für die vorliegenden Modellierungen wurden zwei Papiereigenschaften vermessen: die Topografie und die Formation.

Die Formation eines Papiers ist definiert als die *Art und Weise, wie Fasern verteilt, angeordnet und vermischt sind, um ein Papier zu bilden* (siehe DIN6730, Beuth, 2006). Üblicherweise werden Strukturgrößen zwischen 0,1 und 20 mm als Formation bezeichnet. Formation gibt also die lokale Masseverteilung von Fasern und Füllstoffen im Papier an. Diese Flockenstruktur im Blatt ist leicht sichtbar zu machen, indem ein Blatt Papier ins Licht gehalten wird.

In der Regel ist eine möglichst gute Formation über die gesamte Papierbahn erstrebenswert, da sie einen Einfluss auf wichtige Parameter wie Festigkeitseigenschaften, Bedruckbarkeit, Streichfähigkeit oder die optischen Eigenschaften eines Papiers hat. Dabei wird von schlechter Formation gesprochen, wenn die lokale Masseverteilung im Blatt ungleichmäßig ist, also stark ausgeprägte Flocken vorliegen. Bei der Herstellung von Sackpapier für Verpackungszwecke jedoch ist eine gewisse Unregelmäßigkeit im Papier erstrebenswert, da sie die Energieaufnahme-fähigkeit beim Reißen des Papiers erhöht.

Gemessen kann die Formation zum Beispiel durch Lichttransmission, β -Radiographie oder β -Strahlentransmission werden. Dem Institut für Papier-, Zellstoff- und Fasertechnik der TU Graz steht ein β -Radiograph von Fuji zur Verfügung, um die lokalen Masseschwankungen zu erheben. Bei dieser Messmethode wird wie bei Keller und Pawlak (2001) beschrieben eine fotosensitive Platte durch die Probe hindurch mit β -Strahlung für eine gewisse Zeit bestrahlt. Aufgrund der lokalen Schwankungen der Fasermasse des Papiers dringt unterschiedlich viel Strahlungsenergie durch die Probe hindurch und trifft auf den fotosensitiven Empfänger auf. Dieser wird eingescannt und anhand eines Kalibrierstreifens läßt sich dann eine Karte der lokalen Masseverteilungen errechnen. Die Auflösung dieser Messmethode liegt bei 50 μm pro Messpunkt.

Die Formation eines Ausschnitts aus Probe 201 (A) ist in Abbildung 6.4 gezeigt. Das Flächengewicht dieses Papiers liegt bei 90 g/m^2 , die lokalen Werte dieses Ausschnitts schwanken jedoch zwischen 65 und 100 g/m^2 . Diese Abbildung läßt weiters erkennen, warum die lokale Grammaturn beziehungsweise Formation auch als Wolkigkeit bezeichnet wird.

6.1.4 Topografie und gefilterte Topografie

Weiters wird die Topografie der Probe gemessen. Am Institut für Papier-, Zellstoff- und Fasertechnik steht für diese Aufgabe ein InfiniteFocus System von Alicona bereit. Die Messmethode

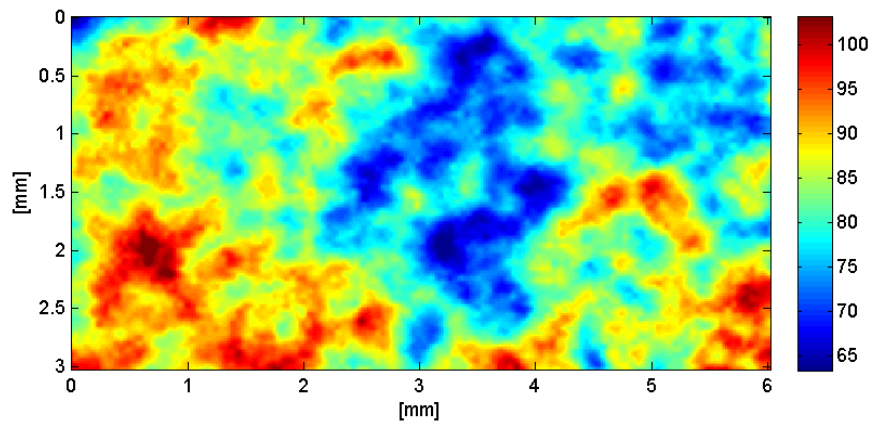


Abbildung 6.4.: Formation eines Papiers mit 90 g/m² Flächengewicht.

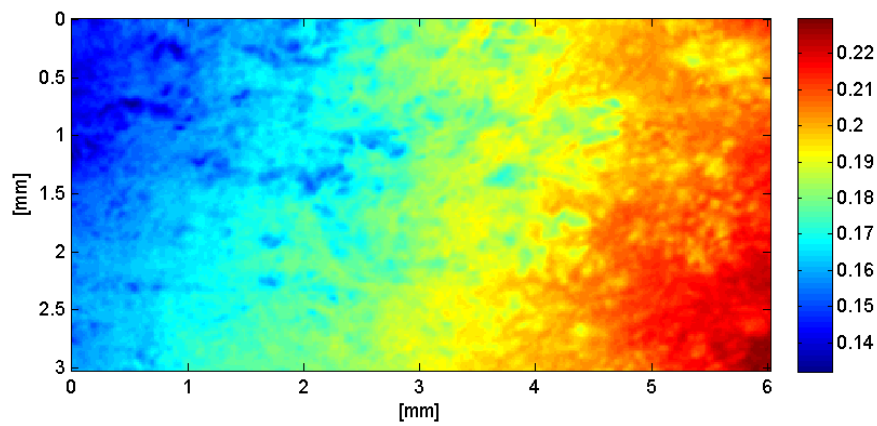


Abbildung 6.5.: Rohtopografie eines Papiers wie mit dem Mikroskop gemessen.

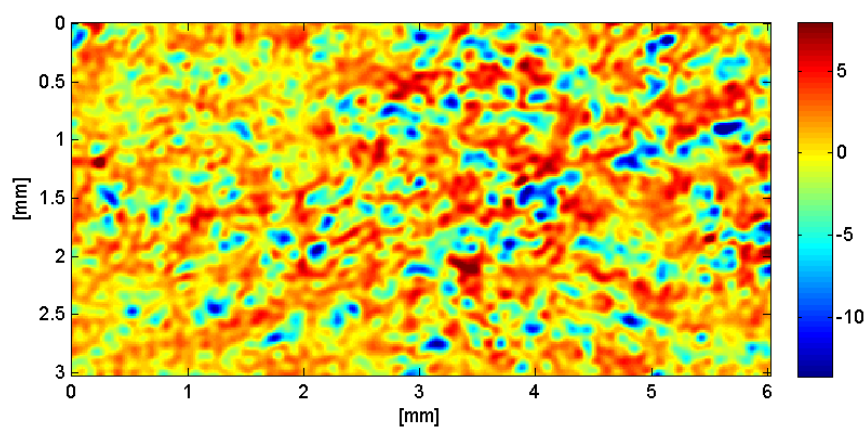


Abbildung 6.6.: Mit einem Bandpassfilter gefilterte Topografie von Abbildung 6.5. Der Bandpassfilter erhält Strukturen zwischen 100 und 3200 μm Größe.

basiert dabei auf dem Prinzip der Fokusvariation. Die Probe wird in unterschiedlichen Abständen zum Objektiv aufgenommen und aus dem resultierenden Bilderstapel kann dann durch unterschiedliche Schärfe- und Unschärfegrade das Höhenprofil errechnet werden. Eine genauere Beschreibung dieser Technik läßt sich in Wanske, Großmann und Scherer (2008) oder auf der Homepage des Herstellers² finden.

Die in dieser Arbeit verwendeten Proben wurden mit einer fünffachen Auflösung und unter diffuser Beleuchtung aufgenommen. Bei diesen Aufnahmeeinstellungen haben die resultierenden Aufnahmen der Probe eine Auflösung von 12,8 μm pro Pixel.

Diese gemessene Rohtopografie wird jedoch nicht direkt für die Modellierung verwendet. Grund dafür ist, dass die Messung der Rohtopografie wie in Abbildung 6.5 zu erkennen durch eine nicht-perfekte Planlage sowie eine Welligkeit der Probe gestört ist. Diese Störungen lassen sich jedoch mit geeigneten Filterungen aus der Messung entfernen.

Weitaus interessanter als die Frage nach dem Einfluss der Topografie auf das Druckbild ist für die Projektpartner jedoch eine Antwort darauf, wie groß die störenden Strukturen der Topografie tatsächlich sind. Dass die Topografie für das Auftreten unbedruckter Stellen verantwortlich ist, haben bereits vorangegangene Arbeiten bewiesen (siehe zum Beispiel Barros & Johansson, 2006). Viel mehr möchte man nun die genauen Größenordnungen jener Strukturen ermitteln, welche die Fehlstellen verursachen. Auch diese Frage läßt sich anhand von gefilterten Topografiemessungen beantworten.

Filterungen, ein Hilfsmittel aus der Signalverarbeitung, ermöglichen es, aus einer Messung interessante Strukturen herauszufiltern und andere gezielt zu löschen. So lassen sich mittels einer Bandpassfilterung Strukturen innerhalb eines definierten Passbandes erhalten, während gröbere sowie feinere Strukturen aus der Messung entfernt werden. Abbildung 6.6 zeigt an einem Ausschnitt von Probe 201 (A), wie sich die in Abbildung 6.5 gezeigte Rohtopografie durch eine Filterung, bei der nur Strukturen zwischen 100 und 3200 μm Größe erhalten bleiben, verändert.

Die in dieser Arbeit verwendeten Bandpassfilterungen werden im Frequenzraum der Fouriertransformation ausgeführt. Dafür wird die zu filternde Topografie mittels Fouriertransformation vom sogenannten Ortsraum in den Frequenzraum transformiert. Das Signal wird also in seine Sinus- und Kosinusfunktionen mit verschiedenen Amplituden aufgespalten. Da es sich bei den vorliegenden Daten um diskrete Messwerte handelt, wird genau genommen mit diskreten Fouriertransformationen gearbeitet. Die Funktion

$$F(u, v) = \frac{1}{\sqrt{NM}} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} f(x, y) \exp\left(-2\pi i \left(\frac{ux}{N} + \frac{vy}{M}\right)\right)$$

gibt die diskrete Fouriertransformation einer $N \times M$ Matrix mit Einträgen $f(x, y)$ an, aus der sich die Fouriertransformierte $F(u, v)$ ergibt. Klarerweise ist die Fouriertransformierte $F(u, v)$ wieder eine Matrix derselben Dimension wie das ursprüngliche Signal.

Da es sich hier um die diskrete Transformation handelt, beinhaltet die Transformierte $F(u, v)$ nicht sämtliche im Signal vorhandenen Frequenzen, sondern nur diejenigen, die im Ortsraum des Signals vorhanden sind. Für die Topografiemessungen bedeutet dies, dass nur Strukturen kleiner dem Probenausschnitt sowie Strukturen größer der Messauflösung vorliegen können. Noch feinere sowie gröbere Strukturen werden nicht erfasst und sind daher in dieser Messung nicht vertreten.

²www.alicon.at

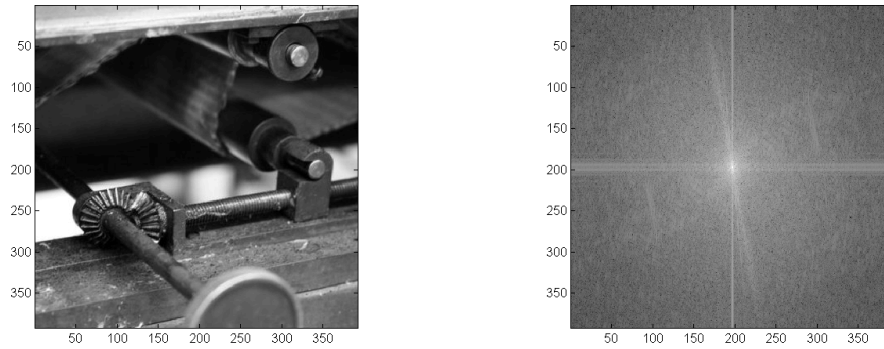


Abbildung 6.7.: Beispiel für ein Bild im Ortsraum (links) und im Frequenzraum (rechts).

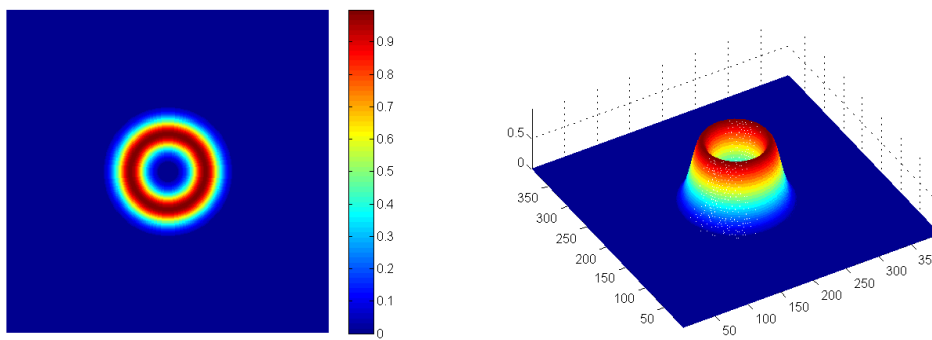


Abbildung 6.8.: Beispiel für einen Bandpassfilterkern im Frequenzraum.

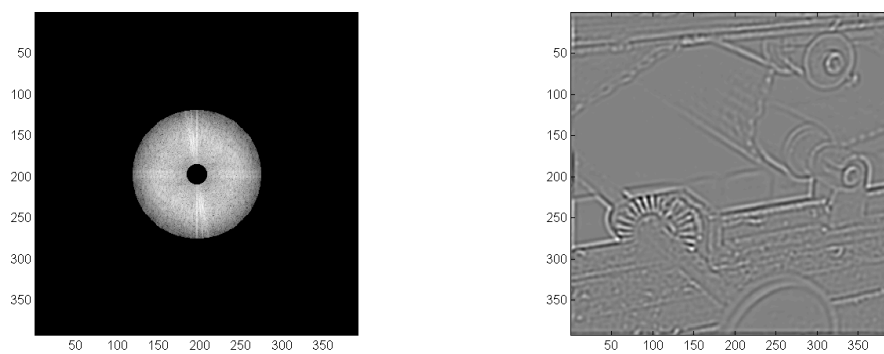


Abbildung 6.9.: Gefiltertes Bild im Frequenzraum (links) und im Ortsraum (rechts).

Für die Darstellung eines fouriertransformierten Signals wird meist der Betrag der komplexen Funktion $F(u, v)$ (beziehungsweise die logarithmische Transformation derselben) verwendet, das sogenannte Intensitätsbild. Gezeigt wird also eine Darstellung des Signals als Amplituden der einzelnen Frequenzen. Da das Spektrum einer Fouriertransformation eine periodische Funktion ist, also $F(u, v) = F(u + k \cdot N, v + l \cdot M)$ für $k, l \in \mathbb{Z}$, wird in Darstellungen der Transformierten der Ursprung $(0, 0)$ meist zentriert dargestellt. Im Zentrum sollten dabei die allertiefsten Strukturen des Signals vertreten sein. Nun limitiert die Größe des Bildes jedoch die vorliegenden Strukturgrößen. Strukturen, die so grob sind dass sie nicht auf dem Bild erfasst werden können, können auch im Frequenzraum nicht vorhanden sein. Daher bezieht sich der Wert $F(0, 0)$ nicht auf die tiefsten Frequenzen (Strukturen), sondern gibt den Mittelwert aller gemessenen Strukturen an. Je weiter vom Ursprung $(0, 0)$ entfernt eine Beobachtung liegt, desto feinere Strukturen repräsentiert sie. Dabei spricht man in der Signal- und Bildverarbeitung von hohen Frequenzen, wenn es sich um feine Strukturen, feine Muster oder hochfrequentes Rauschen handelt. Tiefe Frequenzen repräsentieren großflächig homogene Strukturen. Abbildung 6.7 zeigt ein Originalbild sowie das dazugehörige Frequenzspektrum. Der Ursprung des Koordinatensystems in der Darstellung des Bildes im Frequenzraum liegt dabei im Zentrum der gezeigten Region.

Da im Frequenzraum Strukturen nun in konzentrischen Kreisen um $F(0, 0)$ nach ihren Frequenzen geordnet vorliegen, lassen sich gezielt Signale gewisser Frequenzen herausfiltern oder löschen. Dies erfolgt mittels einem sogenannten Filterkern. Dieser wird mit dem fouriertransformierten Original punktweise multipliziert, erledigt also im Prinzip nichts anderes als eine Gewichtung der einzelnen Frequenzen. Je nach Design des Filterkerns ergeben sich unterschiedliche Filterresultate. So lassen sich zum Beispiel Tiefpassfilter konstruieren, indem die Gewichte im Zentrum 1 und am Rand 0 gesetzt werden. In so einem Fall dürfen nur sehr grobe Frequenzen (welche ja um das Zentrum liegen) den Filter passieren, wohingegen sehr hochfrequente, feine Informationen aus den Randregionen gelöscht werden. Ebenso lassen sich durch geeignetes Gewichten Hoch- oder Bandpassfilter konstruieren.

Da in dieser Arbeit Bandpässe von weiterem Interesse sind, soll hier nicht weiter auf die anderen Filter eingegangen werden. Abbildung 6.8 zeigt einen Bandpassfilterkern, der sowohl die größten und tiefsten Frequenzen im Zentrum als auch die feinen Strukturen, die in den Randbereichen zu finden sind, löscht. Nur Frequenzen innerhalb des Passbandes passieren den Filter.

Wird dieser Filterkern auf das Original aus Abbildung 6.7 angewandt, so entspricht das im Frequenzraum einer elementweisen Multiplikation der beiden Matrizen. Hier ist auch die Motivation für die Filterung im Frequenzraum überhaupt zu finden, da sie rein theoretisch auch direkt im Ortsraum durchgeführt werden könnte. Der große Vorteil der Filterung im Frequenzraum liegt darin, dass es sich - neben der Hin- und Rücktransformation - um eine einfache Multiplikation zweier Matrizen handelt, wohingegen eine Filterung im Ortsraum einer Faltung der zu den beiden Matrizen gehörenden linearen Abbildungen entspricht. Bereits für die vorliegenden Proben sind die Laufzeitvorteile einer Filterung im Frequenzraum deutlich. Abbildung 6.9 zeigt das Ergebnis der Matrixmultiplikation im Frequenzraum. Deutlich zu erkennen ist, dass nur noch ein Frequenzband an Informationen erhalten geblieben ist. Außerdem zeigt Abbildung 6.9 das bereits wieder zurücktransformierte Ergebnis der Filterung im Ortsraum. Sehr feine Strukturen (Gewinde der Schraube, Maserung der Stuhlung, ...) als auch sehr grobe (Helligkeitsunterschiede im Hintergrund) aus Abbildung 6.7 fehlen im rücktransformierten Bild in Abbildung 6.9.

Diese Diskussion beleuchtet nur einen kleinen Querschnitt durch die vielen Möglichkeiten an

Filterungstechniken. Eine gute Einführung in das Thema ist bei Gonzalez, Woods und Eddins (2004), detailliertere Informationen zum Beispiel bei Erhardt (2008), Burger und Burge (2006) oder Schüßler (2008) zu finden.

6.1.5 Papiereigenschaften der vermessenen Proben

Sechs verschiedene Papiere standen für die Modellierungen zur Verfügung. Sie wurden von Mondi Frantschach für diese Untersuchungen zur Verfügung gestellt und wurden in einer industriellen Flexodruckmaschine bedruckt. Daher ist es möglich, aus diesem Datensatz Schlüsse für weitere industrielle Flexodrucke zu ziehen, anstatt die Resultate auf Labordruckversuche einzuschränken.

Alle Papiere wurden unter konstanten Bedingungen im Vollton bedruckt, unterscheiden sich jedoch durch ihre Flächengewichte und die Glättungsverfahren. Tabelle 6.1 zeigt die bekannten Messwerte der einzelnen Papiere. Dabei läßt sich aus den beiden Messungen Bendtsen und PPS (Parker Print Surf) auf die Glätte, mit dem Glanz auf die Rauigkeit der einzelnen Papiere schließen.

Probe	101	102	103	201	202	203
Gewicht [g/m ²]	90	90	70	90	70	90
Bendtsen [ml/min]	225 (35)	350 (85)	460 (50)	145 (30)	355 (70)	440 (60)
PPS [μ m]	3,61 (0,1)	4,62 (0,08)	5,46 (0,06)	5,16 (0,11)	6,19 (0,08)	6,26 (0,11)
Glanz [%]	28,9 (0,8)	19,8 (0,4)	15,2 (0,4)	11,8 (1,4)	7 (0,9)	7,1 (0,7)
Dicke [μ m]	120 (3,45)	124 (3,14)	99 (4,41)	108 (1,86)	90 (1,17)	122 (2,05)

Tabelle 6.1.: Informationen über die 6 betrachteten Papiere. Werte in Klammern geben die Standardabweichungen der 20 Einzelmessungen an.

Sowohl Bendtsen als auch PPS sind Verfahren, welche auf Luftstrommessungen basieren. Dafür wird ein Messring auf die Probe gepresst und der Luftdurchfluss zwischen diesem Messring und dem Papier erhoben. Die Idee dahinter ist, dass der Messring auf einem rauen Papier schlechter aufsetzen kann als auf einem glatten und in Folge dessen mehr Luft durch den Messkanal strömen müsste. Daher ist aus diesen Werten abzulesen, dass vor allem Papier 203 extrem rau ist, gefolgt von den Papieren 103 und 202. Bei Papier 101 hingegen handelt es sich um ein besonders glattes Sackpapier. Auch der in der Tabelle aufgelistete Glanz erlaubt zu einem gewissen Maß Rückschlüsse auf die Glätte des Papiers.

Über die Papiere ist weiters bekannt, dass sie mit unterschiedlichen Glättungsverfahren behandelt wurden. Dabei unterscheiden sich die Proben der 100-er von jenen der 200-er Serie dahingehend, dass sie mit unterschiedlichen Technologien satiniert wurden. Die Unterschiede innerhalb der beiden Serien rühren von verschiedenen Einstellungen der Glättungsparameter, von denen jedoch keine weiteren Details bekannt sind.

Von jedem dieser Papierbögen wurden zwei Probeflächen ausgewählt, die alle von derselben Region der Druckplatte bedruckt wurden. Jede der Proben misst 5×5 cm². Für die spätere Registrierung der einzelnen Messungen sind die Proben durch Laserlöcher eindeutig markiert worden. Auf der in Abbildung 6.10 gezeigten Farbaufnahme von Probe 201 (A) sind besagte Markierungen in den Ecken zu erkennen. Anhand dieser Markierungen lassen sich die Messungen

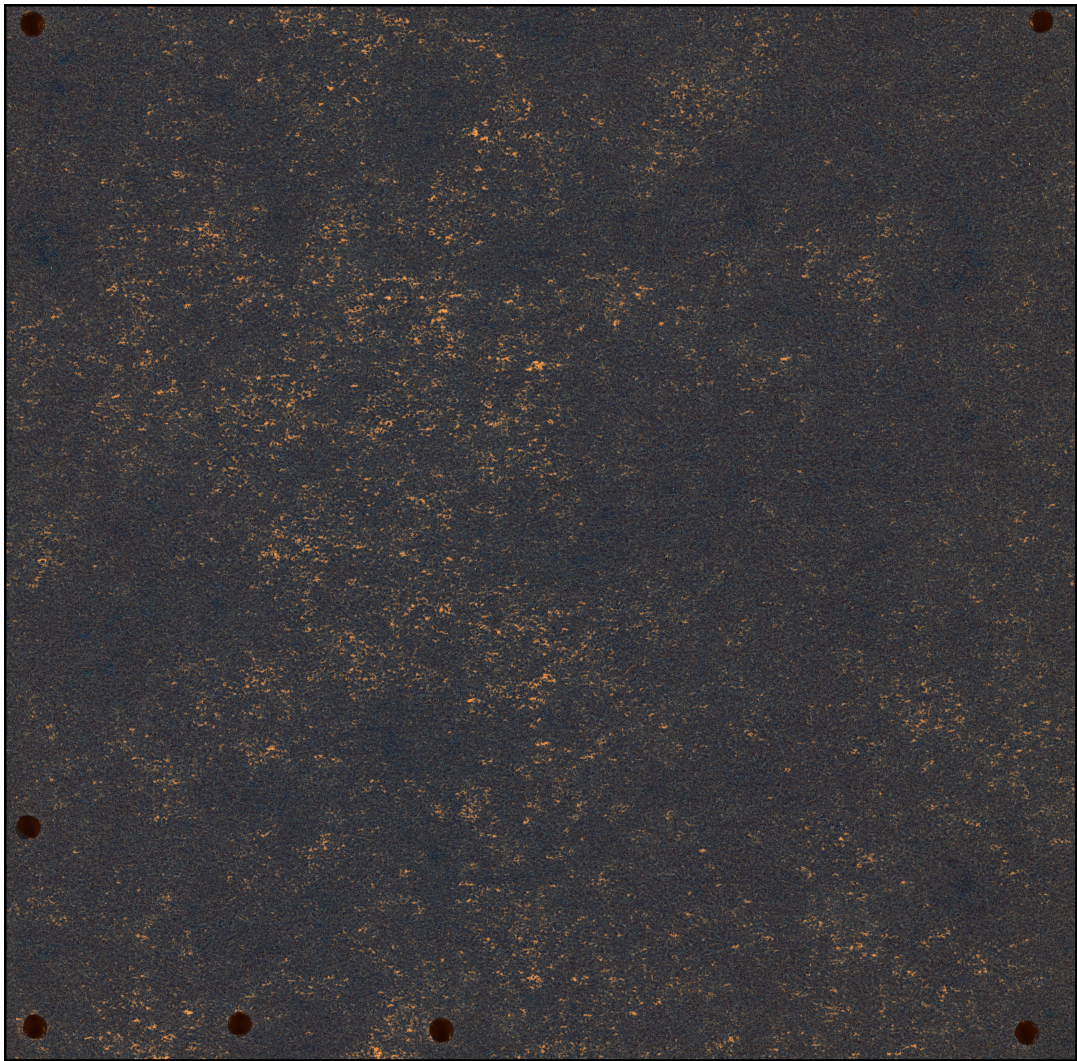


Abbildung 6.10.: Farbaufnahme einer der beiden Proben von Papier 201.

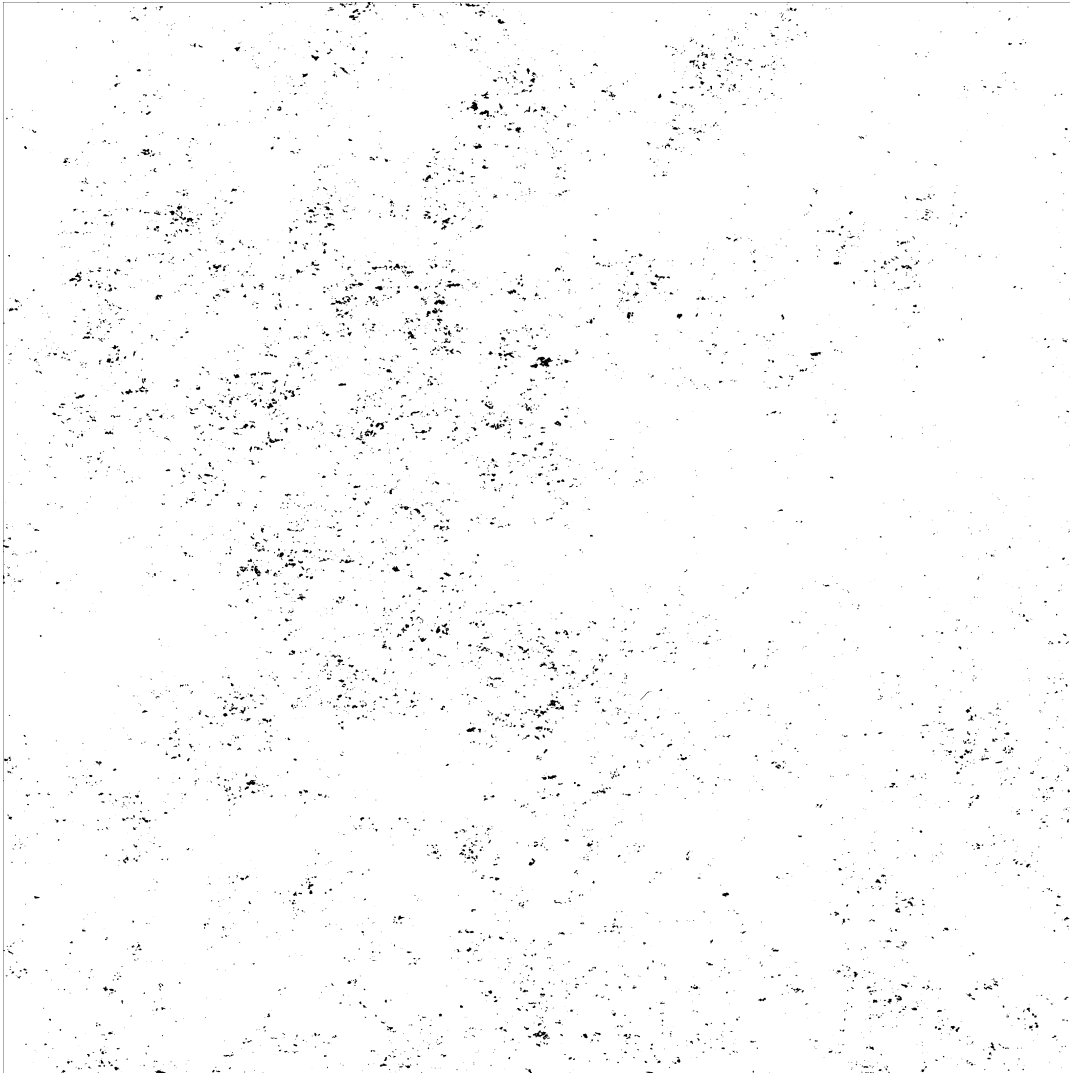


Abbildung 6.11.: Aus der automatischen Fehlstellenerkennung resultierende dichotomisierte Fläche bedruckter und unbedruckter Regionen der Aufnahme aus Abbildung 6.10. Unbedruckte Zonen sind schwarz eingezeichnet.

einer Probe überlagern und die unterschiedlichen Messwerte einer bestimmten Beobachtung eindeutig wiederfinden.

6.1.6 Druckqualität

Außerdem läßt sich eine Aussage darüber tätigen, wie gut beziehungsweise schlecht das Druckbild einer Probe ist. Dafür werden die in Tabelle 6.2 angegebenen Werte verwendet, welche die Qualität des Drucks der jeweiligen Probe beschreiben können.

Papier	101	102	103	201	202	203
Probe (A)	7,28	6,62	5,16	4,35	4,51	3,31
Probe (B)	6,93	5,63	4,74	3,31	5,11	3,20

Tabelle 6.2.: Druckqualitäten dq der untersuchten Papiere.

Für die Berechnung dieser Werte wurde ein aus der Psychologie bekanntes Gesetz herangezogen. Aus dem Gebiet der Psychophysik ist das Gesetz von Fechner (seltener auch als Weber-Fechner Gesetz) bekannt, welches den Zusammenhang zwischen einem Sinneseindruck S und dem dazugehörigen Reiz R beschreibt. Dabei basieren Fechners 1907 veröffentlichten Untersuchungen auf den von Weber bereits 1834 erkannten Relationen zwischen Reizen und den dadurch hervorgerufenen Sinneseindrücken. Dieser Zusammenhang ist laut Fechner logarithmischer Natur, wobei

$$S = k \log \frac{R}{R_0}$$

für eine Reizschwelle R_0 und eine Konstante k gilt. Jüngere Arbeiten haben diesen Zusammenhang zwischen Reiz und Sinneseindruck weiter verfeinert, werden hier jedoch nicht berücksichtigt. Dieser logarithmische Zusammenhang gilt vor allem für den Tast- und den Sehsinn, nicht hingegen für unsere Temperatur- und Schmerzempfindlichkeit. Eine detailliertere Beschreibung dieses Gesetzes ist zum Beispiel in Roeckelein (1998) oder im Kontext der Gehirnforschung bei Dehaene (2003) zu finden.

Aufbauend auf diesen Zusammenhang wurde nun die Druckqualität definiert als Logarithmus des Quotienten der Anzahl aller Beobachtungen sowie der unbedruckten Stellen. Formal notiert berechnet sich die Druckqualität als

$$dq = \log \frac{\# \text{ Beobachtungen}}{\# \text{ Fehlstellen}} = \log(\# \text{ Beobachtungen}) - \log(\# \text{ Fehlstellen}).$$

Auf diese Art und Weise steht nun ein Maß³ zur Verfügung, welches die globale Druckqualität einer gesamten Probe anzugeben vermag. Dabei wird dieses Maß sehr klein, wenn ein schlechter Druck mit vielen Fehlstellen vorliegt. Eine komplett unbedruckte Probe hätte Druckqualität 0. Je größer der Wert wird, umso besser ist das vorliegende Druckbild.

Den Zusammenhang zwischen der Druckqualität und der Rauigkeit der einzelnen Papiere (siehe die Werte aus Tabelle 6.1) zeigt Abbildung 6.12. Dieser läßt bereits erkennen, dass rauere Papiere einen schlechteren Druck liefern als glattere.

³Die Druckqualität ist jedoch kein Maß im Sinne der Maßtheorie.

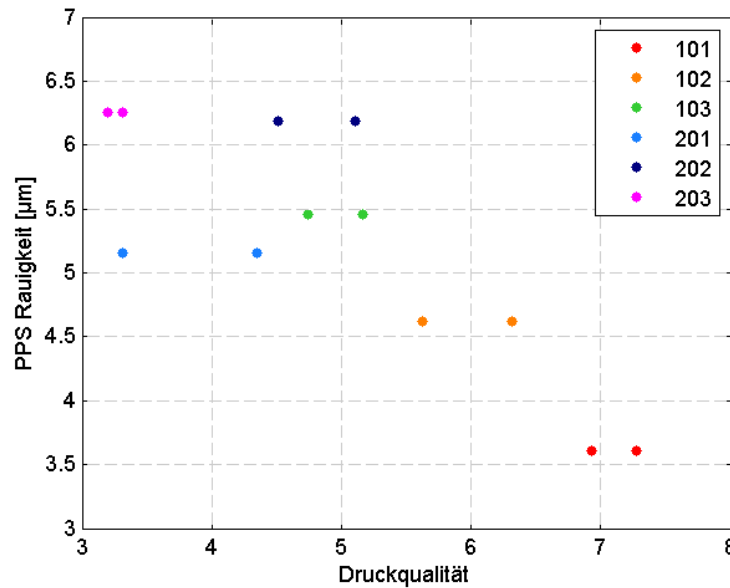


Abbildung 6.12.: Gegenüberstellung der Druckqualität der einzelnen Proben und der mithilfe des PPS gemessenen Rauigkeit der Proben.

Automatisierte Erkennung der Fehlstellen

Um die Zielvariable, bedruckte sowie unbedruckte Regionen, zu erhalten, muss diese Information aus der Farbaufnahme (Abbildung 6.10) extrahiert werden. Diese Trennung von bedruckten und unbedruckten Zonen erfolgt mithilfe eines halbautomatischen Segmentierungsalgorithmus, basierend auf einer von Donoser, Bischof und Wiltsche (2006) vorgestellten Prozedur. Bei der hier verwendeten Version werden dem Programm zwei Muster übergeben - ein Muster einer bedruckten sowie ein Muster einer unbedruckten Region. Diese Muster werden vor Beginn der Berechnungen aus der vorliegenden Probe ausgewählt. Anhand dieser kann der Algorithmus für jeden einzelnen Bildpunkt den Farbraumwert errechnen und mit jenen der beiden Muster vergleichen. Aufgrund dieser Vergleiche wird entschieden, ob ein Pixel bedruckt ist oder nicht. Abbildung 6.11 zeigt die resultierende, bereits registrierte binäre Segmentierungsmaske der Probe aus Abbildung 6.10.

In einem letzten Schritt werden die binären Einzelbeobachtungen noch in Blöcke gruppiert, um in späterer Folge eine Modellierung basierend auf der doppelten oder der multiplikativen Binomialverteilung zu ermöglichen. Für diese Gruppierungen wurde die Blockgröße von 6×6 Pixel gewählt, welche bei einer Auflösung von $15 \mu\text{m}$ pro Pixel einer Blockgröße von $90 \mu\text{m}$ Kantenlänge entspricht. Damit sind diese Blöcke gerade noch so fein, dass sie mit freiem Auge nicht erkannt werden können. Die Topografie innerhalb eines Blocks wird durch den Mittelwert der ursprünglichen 36 Topografiemessungen der Einzelbeobachtungen repräsentiert. Mit der Formation verhält es sich ebenso. Die Anzahl an erfolgreichen Beobachtungen innerhalb eines jeden Blocks kann in weiterer Folge zwischen 0 und der Blockgröße variieren. Abschnitt 6.3.3 werden noch die Resultate aus anderen Blockgrößen angesprochen, die ausführliche Diskussion der Modellierungsergebnisse jedoch beschränkt sich auf den bereits motivierten Fall $n = 36$.

Damit liegen nun sowohl Prädiktoren als auch Responsevariable vor und es kann ein logisti-

sches Modell für die absolute Häufigkeiten von Fehlstellen angewendet werden.

6.2 Binomiale Modelle

Nun soll ein Modell gefunden werden, welches das Auftreten von unbedruckten Regionen anhand der Topografie und der Formation des Papiers erklären kann. Da Visualisierungen der einzelnen Modellierungsergebnisse für die vollen, 25 cm^2 großen Probeflächen keine feinen Details mehr erkennen ließen, wird mit einem Modell begonnen, welches nur einen Ausschnitt von $3 \times 6 \text{ mm}^2$ einer Probe betrachtet. Mit der in Abbildung 6.13 gezeigten Auswahl aus Probe 201 (A) lassen sich diese Details besser erkennen.

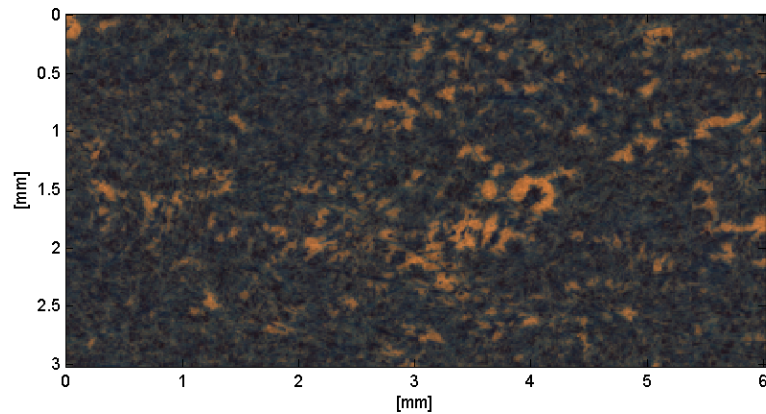


Abbildung 6.13.: Ein Ausschnitt aus Probe 201 (A).

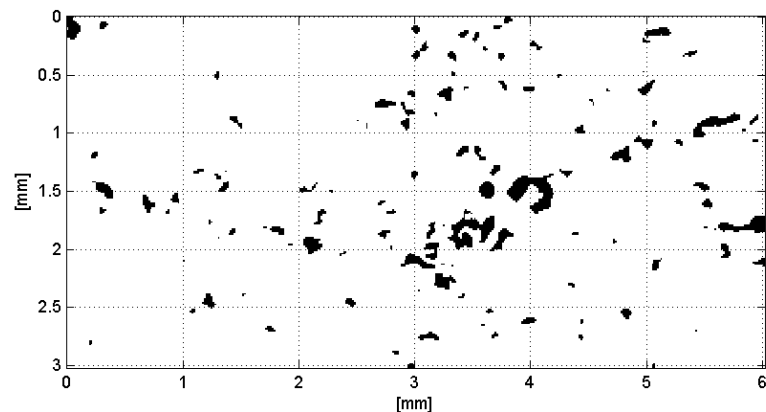


Abbildung 6.14.: Binäre Responsevariable erfolgreiche oder fehlgeschlagene Farbübertragung desselben Ausschnitts wie in Abbildung 6.13.

Modellierungsergebnisse für die zwölf $5 \times 5 \text{ cm}^2$ großen Probeflächen werden in Abschnitt 6.2.2 besprochen, auf grafische Darstellungen wird in diesen Fällen jedoch größtenteils verzichtet. Außerdem werden die Ergebnisse der logistischen Regression nur sehr knapp diskutiert. Eine weitaus

ausführlichere Auflistung und Darlegung dieser Resultate ist den jährlichen Forschungsberichten zum Projekt „Modellierung der Farbübertragung“ (2008 bis 2011) zu entnehmen.

6.2.1 Modellschätzung für einen Ausschnitt von Probe 201 (A)

Abbildungen 6.15 und 6.16 zeigen die beobachteten Werte der Prädiktoren Topografie und Formation des $3 \times 6 \text{ mm}^2$ großen Ausschnitts aus Probe 201 (A), welche das Verhalten der in Abbildung 6.14 gezeigten binären Responsevariablen — erfolgreiche oder fehlgeschlagene Farbübertragung — erklären sollen.

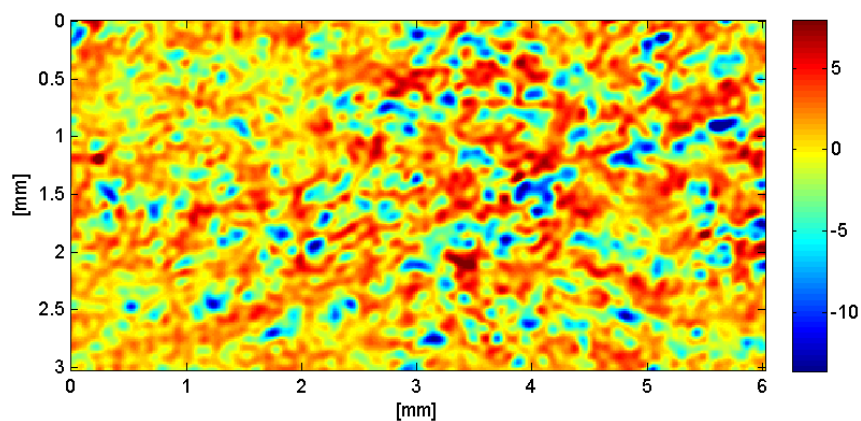


Abbildung 6.15.: Der Prädiktor gefilterte Topografie des Ausschnitts von Probe 201 (A) aus Abbildung 6.13.

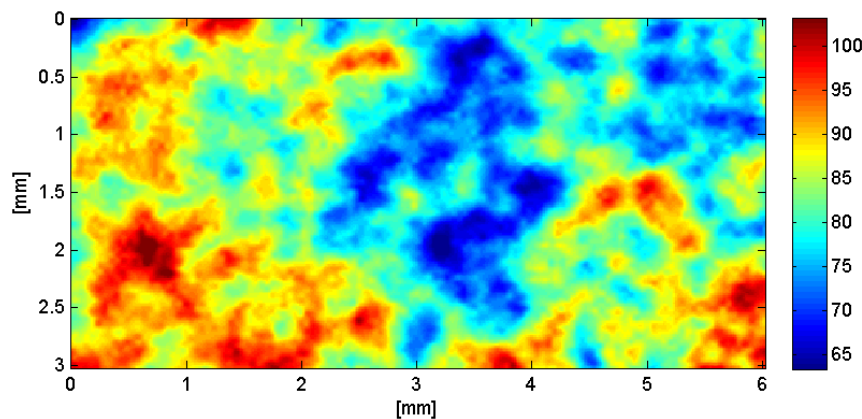


Abbildung 6.16.: Der Prädiktor Formation des Ausschnitts von Probe 201 (A) aus Abbildung 6.13.

Ein logistisches Regressionsmodell unter klassischer Binomialverteilungsannahme dieses Ausschnitts liefert die im Code 6.1 gezeigten Ergebnisse. Dabei wird hier mit 80601 Beobachtungen auf einer Fläche von 18 mm^2 gearbeitet. Der Anteil unbedruckter Pixel liegt in diesem Fall bei

3,6 %. Modelliert wird nun die Wahrscheinlichkeit für erfolgreiche Farbübertragung. Positive Parameter erhöhen diese Wahrscheinlichkeit, negative hingegen verringern sie.

```

1 > mod <- glm(cbind(successes,all-successes)~topography+formation, family=binomial)
  > summary(mod)
3
  Call:
5 glm(formula = cbind(y, 1 - y) ~ topography + formation, family = binomial)

7 Deviance Residuals:
      Min       1Q   Median       3Q      Max
9 -5.39609  0.02788  0.05168  0.11163  3.66531

11 Coefficients:
              Estimate Std. Error z value Pr(>|z|)
13 (Intercept) -4.196218  0.303396  -13.83  <2e-16 ***
   topography  0.656207  0.009403   69.79  <2e-16 ***
15 formation   0.122905  0.003961   31.03  <2e-16 ***
   ---
17 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

19 (Dispersion parameter for binomial family taken to be 1)

21      Null deviance: 24853  on 80600  degrees of freedom
   Residual deviance: 11103  on 80598  degrees of freedom
23 AIC: 11109

25 Number of Fisher Scoring iterations: 8

```

Code 6.1: Resultat eines Binomialmodells mit zwei Haupteffekten für den Ausschnitt aus Probe 201 (A).

Es zeigt sich, dass sowohl die Topografie als auch die Formation in diesem Haupteffektmodell eine wichtige Rolle für das Auftreten der Fehlstellen spielen. Dabei haben sowohl eine höhere Topografie als auch eine höhere Formation einen positiven Einfluss auf die erfolgreiche Bedruckung der Beobachtung. Dass das Modell deutlich besser ist als ein Nullmodell, zeigt die deutliche Reduktion der Devianz zwischen Nullmodell und dem aktuellen Modell. Trotzdem entspricht die Devianz mit 11103 in etwa nur einem Achtel der Anzahl der Freiheitsgrade des Modells. Laut Abschnitt 2.2.4 liegt somit ein potentieller Fall von Unterdispersion vor.

Modelle mit Wechselwirkungen zwischen den beiden Prädiktoren oder quadratischen Effekten werden etwas später noch angesprochen werden, hier wird nun mit der Vorstellung der Modellauswertungsmethode fortgefahren.

Nach einer erfolgreichen Modellschätzung lässt sich für jede Beobachtung die Wahrscheinlichkeit für erfolgreiche Farbübertragung schätzen und wie in Abbildung 6.17 gezeigt darstellen. Gut zu erkennen ist der Zusammenhang zwischen den Regionen, welche höchst wahrscheinlich nicht bedruckt werden und den - bildlich gesprochen - tiefen Canyons und Kratern in der Topografie aus Abbildung 6.15.

Nun unterliegt die erfolgreiche Bedruckung eines Papiers in Realität jedoch einem binären Entscheidungsprozess. Daher werden in einem weiteren Schritt trotz Informationsverlust die

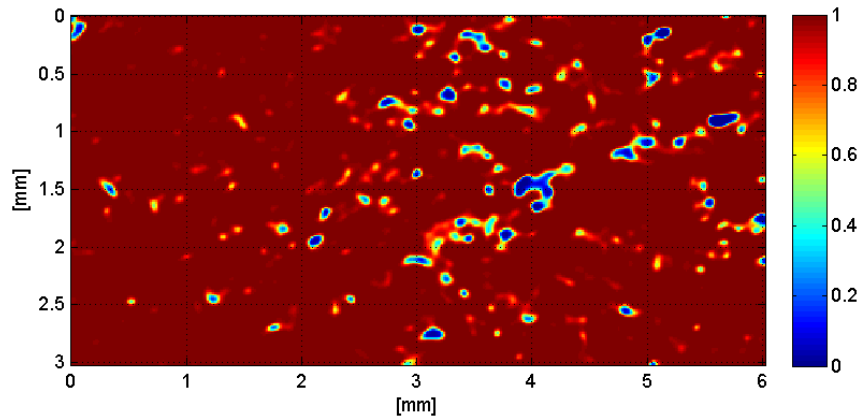


Abbildung 6.17.: Geschätzten Wahrscheinlichkeiten für erfolgreiche Farbübertragung auf dem Ausschnitt aus Probe 201 (A).

prognostizierten Wahrscheinlichkeiten diskretisiert, um zu einer binären Modellprognose zu gelangen. Diese abermals binären Informationen über die Bedruckbarkeit einer Probe sind für die Projektteilnehmer von größerem Interesse als die geschätzten Wahrscheinlichkeiten erfolgreicher Farbübertragung, da dadurch ein direkter Vergleich mit den beobachteten Responsevariablen möglich wird.

Für die Diskretisierung werden jene Beobachtungen, deren prognostizierte Wahrscheinlichkeiten für erfolgreiche Farbübertragung größer einem gewissen Schwellwert sind, als erfolgreich bedruckte Regionen angesehen. Beobachtungen mit prognostizierter Wahrscheinlichkeit kleiner dem Schwellwert hingegen werden als unbedruckt klassifiziert.

Zwei unterschiedliche Vorgehensweisen für die Wahl des Schwellwerts werden in Betracht gezogen. Für die eine Methode wird der Schwellwert zwischen erfolgreicher und fehlgeschlagener Farbübertragung derart gewählt, dass der Anteil an prognostizierten Fehlstellen gleich dem beobachteten Anteil ist. Bei der zweiten Methode wird der ideale Schwellwert basierend auf den von Barros und Johansson (2006) in ihrer Arbeit vorgestellten Kennwerten ermittelt. Dafür werden aus einem Vergleich zwischen Modellprognose und Realität die folgenden vier Fälle identifiziert:

negativer Treffer	korrekt erkannte Fehlstellen,
Fehlalarm	fälschlicherweise als unbedruckt charakterisierte Beobachtungen,
Verfehlung	fälschlicherweise nicht als unbedruckt klassifizierte Beobachtungen
positiver Treffer	korrekt erkannte bedruckte Regionen

Die vier Situationen sind auch in Tabelle 6.3 aufgeschlüsselt. Dabei sei zu bemerken, dass die Wahrscheinlichkeit für eine Verfehlung, also $P(\text{als bedruckt eingestuft} | \text{tatsächlich unbedruckt})$, dem Fehler 1. Art entspricht. Ebenso entspricht die Wahrscheinlichkeit eines Fehlalarms dem Fehler 2. Art.

Die Unterscheidung dieser vier Fälle erlaubt in weiterer Folge die Berechnung verschiedener Kennwerte. Unter anderem können mit Sensitivität und Relevanz zwei Werte berechnet werden, welche sich auf die unbedruckten Stellen der Probe konzentrieren. Bei beiden Werten handelt es

	unbedruckt eingestuft	bedruckt eingestuft
tatsächlich unbedruckt	negativer Treffer	Verfehlung
tatsächlich bedruckt	Fehlalarm	positiver Treffer

Tabelle 6.3.: Die vier bei einem Vergleich von Realität und Modellprognose möglichen Fälle.

sich um bedingte Wahrscheinlichkeiten, welche die korrekt erkannten, unbedruckten Regionen in ein Verhältnis zu den tatsächlich unbedruckten Beobachtungen und umgekehrt setzen. Die beiden Werte sind definiert als

$$\begin{aligned} \text{Sensitivität} &= P(\text{als unbedruckt eingestuft} \mid \text{tatsächlich unbedruckt}) \\ \text{sowie Relevanz} &= P(\text{tatsächlich unbedruckt} \mid \text{als unbedruckt eingestuft}). \end{aligned}$$

Ebenso können auch die Kennwerte Spezifität und Segreganz berechnet werden, welche sich auf die erfolgreich bedruckten Flächen beziehen. Diese Werte sind definiert als

$$\begin{aligned} \text{Spezifität} &= P(\text{als bedruckt eingestuft} \mid \text{tatsächlich bedruckt}) \\ \text{und Segreganz} &= P(\text{tatsächlich bedruckt} \mid \text{als bedruckt eingestuft}). \end{aligned}$$

Diese beiden Maße liefern in den hier vorliegenden Betrachtungen bedruckter und unbedruckter Regionen jedoch keine relevanten Informationen. Da in dieser Arbeit der Fokus auf die unbedruckten Regionen der Proben gerichtet wird, werden Spezifität und Segreganz nicht weiter betrachtet und statt dessen der Blick auf die beiden anderen Kennwerte gerichtet.

Dabei werden die beiden sich auf die unbedruckten Stellen konzentrierenden bedingten Wahrscheinlichkeiten Sensitivität und Relevanz mittels Wertetabellen wie in Tabelle 6.3 gezeigt über die Formeln

$$\begin{aligned} \text{Sensitivität} &= \frac{\text{negativer Treffer}}{\text{negativer Treffer} + \text{Verfehlt}} \\ \text{und Relevanz} &= \frac{\text{negativer Treffer}}{\text{negativer Treffer} + \text{Fehlalarm}} \end{aligned}$$

ermittelt. Der Verlauf der beiden Kennwerte für den Ausschnitt aus Probe 201 (A) ist für verschiedene Schwellwerte aus dem Einheitsintervall in Abbildung 6.18 dargestellt.

Aus diesen beiden Kennwerten läßt sich der ideale Schwellwert für die Diskretisierung ermitteln. So wird jener Wert, bei dem beide Wahrscheinlichkeiten gleichzeitig möglichst groß werden, als idealer Schwellwert angenommen werden.

Es wurde jedoch auch eine zweite und einfachere Methode der Diskretisierung in Betracht gezogen. Dafür wird der Anteil der in Realität unbedruckten Stellen, p (mit $p \in (0, 1)$), ermittelt. Danach wird der ideale Schwellwert derart gewählt, dass auch in der resultierenden Modellprognose derselbe Anteil an Stellen unbedruckt ist. Der gesuchte Schwellwert ist also nichts anderes als das p -Quantil der prognostizierten Wahrscheinlichkeiten. Dieser Wert ist leicht zu berechnen und so kann die auf diese Art und Weise erfolgende Diskretisierung der prognostizierten Wahrscheinlichkeiten schnell durchgeführt werden. Bei dieser Methode wird also der Anteil der beobachteten unbedruckten Regionen festgehalten und in der dichotomisierten Modellprognose reproduziert.

Ein Vergleich der beiden idealen Werte zeigt nun, dass die beiden Schwellwerte praktisch ident sind. Wird der Anteil an Fehlstellen konstant gehalten, liegt der Schwellwert im gezeigten

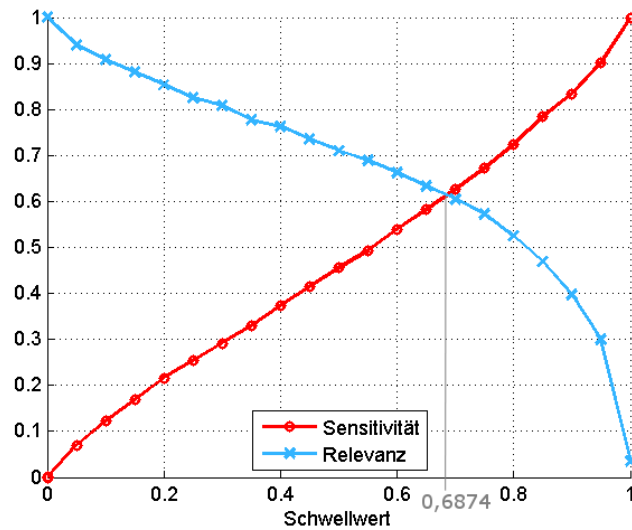


Abbildung 6.18.: Sensitivität und Relevanz des Ausschnitts von Probe 201 (A).

Beispiel bei 0,6874. Wird hingegen über Sensitivität und Relevanz der Schwellwert ermittelt, so ergibt sich ein Wert von 0,6853. Daher wird für den Rest der Diskussion dieser Anwendungsbeispiele über den Anteil unbedruckter Regionen diskretisiert, da diese Diskretisierung leichter umsetzbar ist. Sensitivität und Relevanz werden als Maß für die Güte der diskretisierten Prognose verwendet.

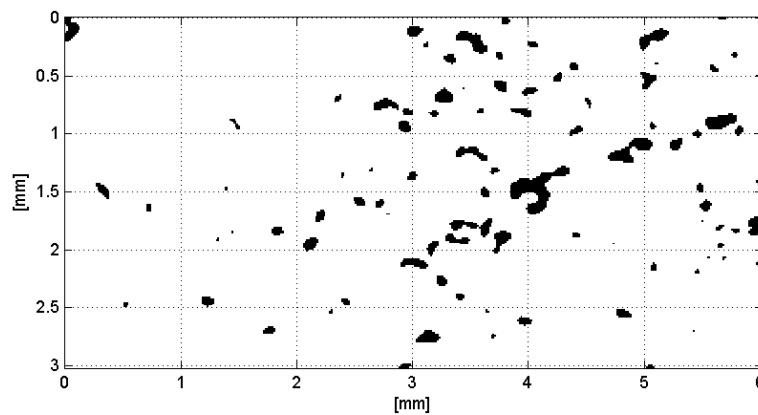


Abbildung 6.19.: Aus den geschätzten Wahrscheinlichkeiten in Abbildung 6.17 erhaltene diskretisierte Modellprognose.

Im vorliegenden Beispiel zeigt sich, dass bei einem idealen Schwellwert (basierend auf dem konstant gehaltenen Anteil unbedruckter Stellen) der Anteil korrekt erkannter Fehlstellen bei 61 % liegt (vergleiche Grafik 6.18). Das bedeutet, dass 61 % aller tatsächlichen Fehlstellen auch als solche erkannt wurden und dass umgekehrt auch 61 % aller als unbedruckt erkannten Stellen tatsächlich unbedruckt sind. Die diskretisierte Modellprognose ist in Abbildung 6.19 gezeigt.

Ohne die zusätzliche Information über die tatsächlich (un-)bedruckten Stellen der Probe jedoch ist diese Darstellung nicht allzu aussagekräftig. Daher wird statt dieser die in Abbildung 6.20 gezeigte Grafik verwendet, welche bereits eine Überlagerung der beobachteten Farbübertragung und der dichotomisierten Modellprognose zeigt.

Anhand solcher Darstellungen läßt sich nun gut ablesen, welche Fehlstellen erkannt werden konnten und welche nicht. Außerdem läßt sich dieser Plot in Relation mit jenem aus Abbildung 6.18 setzen und erkennen, dass für wachsenden Schwellwert der auch als Fehler 1. Art interpretierbare Anteil an Fehlalarmen steigt (rot). Zugleich wird der Fehler 2. Art, der Anteil an verfehlten unbedruckten Regionen (blau), für wachsenden Schwellwert immer kleiner.

Die Darstellung in Grafik 6.20 läßt nun erkennen, dass Fehlalarme vor allem um tatsächliche Fehlstellen herum auftreten. Eine Erklärung dafür läßt sich in der dünnflüssigen Flexo-Druckfarbe finden, welche vor dem Wegschlagen ein wenig in Richtung der zunächst nicht benetzten, noch unbedruckten Regionen spreitet und durch das Spreiten die tatsächlich unbedruckten Stellen verkleinert. Ebenso fällt auf, dass vor allem im linken Teil der Probenfläche etliche tatsächlich unbedruckte Stellen vorliegen, welche jedoch nicht als solche erkannt werden.

Interessante Ergebnisse lassen sich weiters erzielen, wenn Modelle betrachtet werden, welche nur eine der beiden erklärenden Größen als Prädiktor verwenden. So wurde für die in Abbildung 6.21 gezeigte Modellprognose nur die Formation als erklärende Größe verwendet. Das resultierende Bild zeigt, dass die Formation alleine die Farbübertragung nicht erklären kann.

Ein Modell, welche nur die Topografie als Prädiktor verwendet, liefert wie in Abbildung 6.22 gezeigt deutlich bessere Ergebnisse. Im Zentrum der Probenfläche jedoch werden sehr viele Fehlstellen nicht erkannt. Ein Vergleich mit der Modellprognose aus Abbildung 6.21 zeigt, dass dort eine Region sehr geringer Formation liegt, welche im Modell zu Abbildung 6.22 nicht berücksichtigt wird. Im untersten Drittel der Probenfläche prognostiziert ein reines Topografiemodell hingegen einige unbedruckte Stellen, welche in Realität jedoch erfolgreich bedruckt wurden. Wird die Formation zusätzlich in das Modell genommen, reduzieren sich diese Falschalarme deutlich.

In einem Modell, welches beide Papiereigenschaften Topografie und Formation als Prädiktoren verwendet (Abbildung 6.20), können die unbedruckten Stellen also deutlich besser erkannt werden. Dabei fällt auf, dass die Formation alleine keine große Aussagekraft über das Auftreten von unbedruckten Stellen besitzt. Die Topografie alleine kann durchaus sinnvolle Ergebnisse liefern, erkennt jedoch auch in großflächig erfolgreich bedruckten Bereichen Fehlstellen. Die Topografie steuert also die Form jeder einzelnen Fehlstelle, wohingegen die Formation die Probenfläche in kritischere und weniger kritische Bereiche einzuteilen vermag.

Die Sensitivitätswerte dieser drei sowie weiterer Modelle mit Wechselwirkungen und quadratischen Einflüssen sind in Tabelle 6.4 angegeben. Diese unterstreichen abermals das Ergebnis, die Formation alleine könne nur ein unzureichendes Abbild der Wirklichkeit wiedergeben. Die Übereinstimmung prognostizierter sowie tatsächlicher Fehlstellen liegt für diesen kleinen Probenausschnitt bei 24 %. Ein deutlich besseres Ergebnis erlaubt ein Modell mit dem Prädiktor Topografie, da die Sensitivität auf 58 % gestiegen ist. Der Anstieg der Sensitivität eines Modells mit beiden Prädiktoren gegenüber einem Modell ohne Formation liegt bei 3,6 %, liefert also noch einen geringen zusätzlichen Informationsgewinn.

Im Vergleich zu einem Haupteffektmodell verringert ein Interaktionsmodell die Devianz um 39 Einheiten bei einem zusätzlichen Freiheitsgrad. Die auf der diskretisierten Modellprognose basierende Sensitivität verändert sich im Vergleich zum Haupteffektmodell jedoch nicht mehr. Die Verwendung von quadratischen Effekten zusätzlich zu den beiden Haupteffekten hingegen

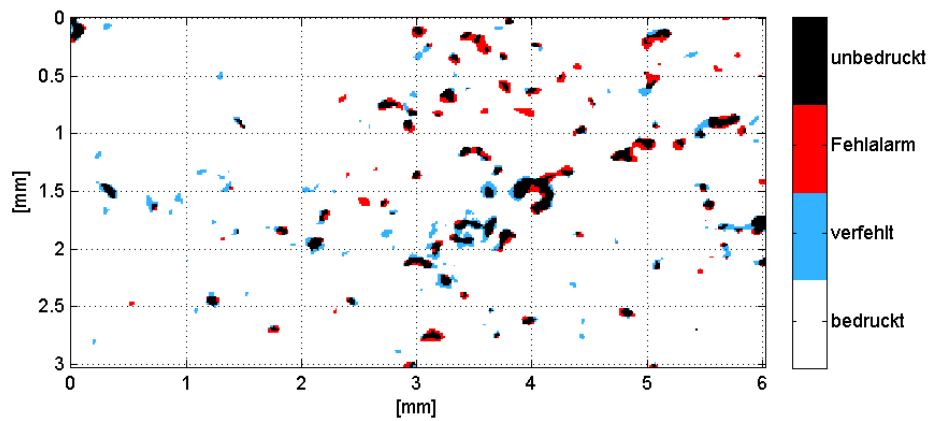


Abbildung 6.20.: Überlagerung der beobachteten Farbübertragung und der diskretisierten Modellprognose, basierend auf einem Modell mit den Prädiktoren Topografie und Formation.

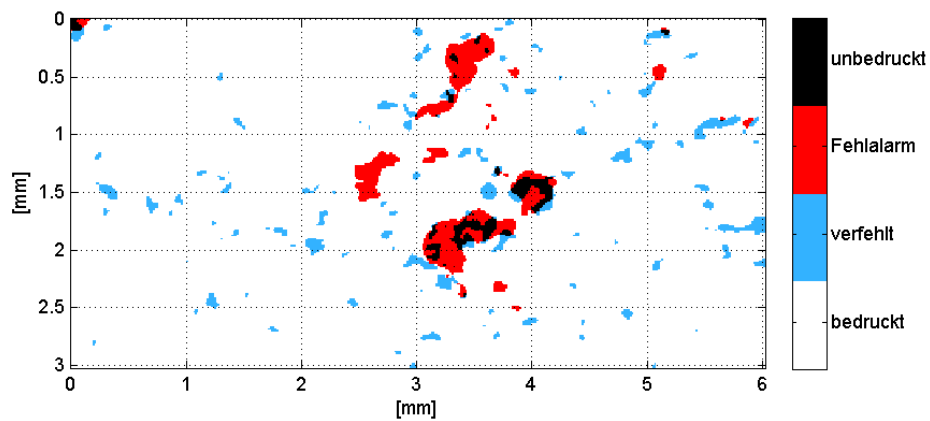


Abbildung 6.21.: Überlagerung der beobachteten Farbübertragung und der diskretisierten Modellprognose, wenn für die Modellschätzung nur der Prädiktor Formation verwendet wird.

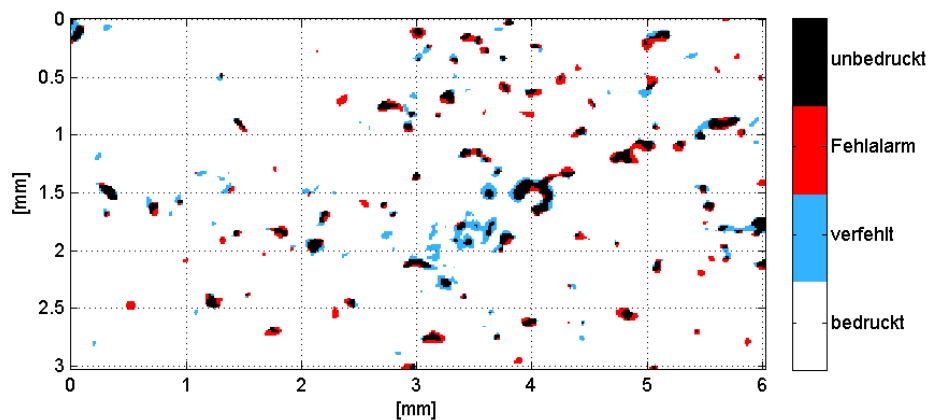


Abbildung 6.22.: Überlagerung der beobachteten Farbübertragung und der diskretisierten Modellprognose, wenn für die Modellschätzung nur der Prädiktor Topografie verwendet wird. 117

liefert eine Verringerung der Devianz von nur 16 Einheiten. Ebenso kann eine geringe Reduktion von 13 Einheiten beobachtet werden, wenn von einem Modell mit Haupteffekten und Wechselwirkung auf ein Modell mit zusätzlichen quadratischen Effekten übergegangen wird. Durch die Interaktion der beiden Haupteffekte wird also noch ein deutlicher Informationsgewinn erzielt, die quadratischen Effekte hingegen können nicht mehr viel zusätzliche Information beisteuern.

Der AIC erreicht den kleinsten Wert im Fall eines Modells mit Haupteffekten, ihrer Interaktion sowie den quadratischen Effekten. Jener des Modells ohne die quadratischen Effekte ist um 9 Einheiten größer. Dabei soll jedoch erwähnt werden, dass die zu den quadratischen Effekten gehörenden p-Werte klar gegen eine Signifikanz dieser beiden Prädiktoren sprechen.

Prädiktoren	Sensitivität	Devianz	Freiheitsgr.	AIC
form(ation)	0,2392	21080	80599	21084
topo(grafie)	0,5765	12184	80599	12188
form + topo	0,6130	11103	80598	11109
form + topo + form ² + topo ²	0,6158	11087	80596	11097
form + topo + form:topo	0,6151	11064	80597	11072
form + topo + form:topo + form ² + topo ²	0,6144	11051	80595	11063

Tabelle 6.4.: Kennwerte verschiedener Modelle mit unterschiedlichen Haupteffekten, Interaktionen und quadratischen Einflüssen für den Ausschnitt aus Probe 201 (A) mit 80601 Beobachtungen. Die Devianz des Nullmodells beträgt 24853.

Von allen betrachteten Modellen ist daher ein Modell mit Wechselwirkung und quadratischen Einflüssen zusätzlich zu den beiden Haupteffekten Formation und Topografie am besten dafür geeignet, die Realität zu beschreiben. Wie bereits angedeutet, ist jedoch ein Modell ohne quadratische Effekte nur unwesentlich schlechter als jenes mit quadratischen Einflüssen. Außerdem liegen die p-Werte dieser quadratischen Einflüsse mit $p = 0,4622$ für den quadratischen Einfluss der Formation und $p = 0,0003$ für jenen der Topografie in Anbetracht der Stichprobengröße doch in einem Bereich, in dem ihre Präsenz im Modell zumindest überdacht werden sollte. Da sich außerdem ein quadratischer Topografieeinfluss oder ein quadratischer Formationseffekt praktisch nicht interpretieren lassen, werden die quadratischen Effekte der beiden Prädiktoren für die weiteren Betrachtungen außer Acht gelassen.

Ein Vergleich von Devianz und Freiheitsgraden zeigt nun, dass die Modelle deutlich kleinere Devianzwerte liefern als von einem angemessenen Modell zu erwarten wäre. Daher sollen nun Modelle betrachtet werden, welche besser auf die vorliegenden Datensituation einzugehen vermögen als ein logistisches Modell mit Bernoulli-verteilter Responsegrößen dies vermag. Dafür werden die in den Kapiteln 3 und 4 eingeführten verallgemeinerten Binomialverteilungen herangezogen. Beide Verteilungen bringen einen zusätzlichen Parameter in die Modellierung ein, mit dem eine bessere Anpassung der Varianzstruktur an die vorliegende Datensituation möglich werden soll.

Bevor jedoch eine Modellschätzung unter multiplikativer oder doppelter Binomialverteilung in Angriff genommen wird, müssen die Originaldaten geblockt werden. Grund dafür ist die bereits angesprochene Eigenheit dieser beiden Verteilungen, sich für eine Anzahl von $n = 1$ Versuchen zur Bernoulli-Verteilung zu vereinfachen. Der zusätzliche Parameter, welcher eine bessere Anpassung des Modells an die Daten liefern sollte, ginge in diesem Fall verloren. Daher müssen die vorliegenden Daten in Blöcke zusammengefasst werden. Dafür wird zu Beginn eine

Blockgröße von 6×6 Pixel gewählt, da diese wie bereits in Abschnitt 6.1 erläutert knapp unter der Sichtbarkeitsgrenze liegt. In Abschnitt 6.3.3 wird noch auf die Wahl der Blockgröße eingegangen und die Resultate von Modellen mit größeren Blöcken angesprochen.

Nun werden die Modellierungsergebnisse dieser geblockten Daten, wiederum unter Annahme binomialverteilter Responsevariable, für den Ausschnitt 6.13 aus Probe 201 (A) dargelegt. Analog zu Tabelle 6.4 werden auch in Tabelle 6.5 die Devianz sowie der AIC verschiedener Modelle aufgelistet. Dabei sei darauf hingewiesen, dass für diese gruppierten Daten die Berechnung einer Sensitivität nicht mehr möglich ist, da diese auf einem Vergleich zwischen der diskretisierten Modellprognose und der dichotomisierten Farbaufnahme der Probe beruht. Da im Fall von gruppierten Daten für die Responsevariablen $y_i = \{0, \dots, 36\}$ gilt, kann ein derartiger Vergleich nun nicht mehr angestellt werden. Anstelle der ursprünglichen 80601 Beobachtungen umfasst der neue, aus diesen geblockten Beobachtungen basierende Datensatz 2484 Beobachtungen.

Prädiktoren	Devianz	AIC
form(ation)	11868	13049
topo(grafie)	6862	8043
form + topo	5473	6656
form + topo + form ² + topo ²	5437	6624
form + topo + form:topo	5335	6520
form + topo + form:topo + form ² + topo ²	5292	6481

Tabelle 6.5.: Auf geblockten Daten ($n = 36$) basierende Kennwerte verschiedener Modelle mit unterschiedlichen Haupteffekten, Interaktionen und quadratischen Einflüssen für den Ausschnitt aus Probe 201 (A) mit 2484 Beobachtungen. Die Devianz des Nullmodells beträgt 15894.

Abermals ist zu erkennen, dass ein Modell mit Interaktion einem reinen Haupteffektmodell überlegen ist. Wird von einem Wechselwirkungsmodell auf ein Modell übergegangen, welches außerdem noch die quadratischen Effekte der beiden Prädiktoren berücksichtigt, können sich sowohl die Devianz als auch der AIC nochmals deutlich verringern. Die praxisbezogene Sachlage jedoch spricht gegen die Verwendung quadratischer Effekte, da eine Interpretation dieser schwierig ist. Daher werden die quadratischen Einflüsse der beiden Haupteffekte in weiterer Folge nicht mehr in Betracht gezogen.

Die Modellzusammenfassung eines Wechselwirkungsmodells ist in Code 6.2 gezeigt. Neben den bereits genannten Kennwerten Devianz und AIC zeigt die Koeffiziententabelle, dass alle drei betrachteten Prädiktoren signifikant sind und die Modellschätzung innerhalb von 6 Iterationen konvergierte.

Aufgrund dieser Beobachtungen, welche auch für die übrigen Proben zutreffen, wird für den Rest dieser Arbeit das Augenmerk auf Wechselwirkungsmodelle gelegt. Quadratische Einflüsse werden nicht mehr berücksichtigt.

Der nächste Abschnitt wendet sich nun einem Vergleich der Modellierungsergebnisse für Modelle mit Haupteffekten und Interaktion der vollen, 25 cm^2 großen Probenflächen zu.

```

1 > mod <- glm(cbind(successes,all-successes)~topography*formation, family=binomial)
2 > summary(mod)
3
4 Call:
5 glm(formula = cbind(y, n - y) ~ topography * formation, family = binomial)
6
7 Deviance Residuals:
8     Min       1Q   Median       3Q      Max
9  -15.0685   0.1926   0.4263   0.8189   5.9339
10
11 Coefficients:
12             Estimate Std. Error z value Pr(>|z|)
13 (Intercept)    -8.990228   0.432980  -20.764 < 2e-16 ***
14 topography     -0.561530   0.092934   -6.042 1.52e-09 ***
15 formation       0.171369   0.005707   30.026 < 2e-16 ***
16 topography:formation 0.014280   0.001225   11.662 < 2e-16 ***
17 ---
18 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
19
20 (Dispersion parameter for binomial family taken to be 1)
21
22 Null deviance: 15893.6 on 2483 degrees of freedom
23 Residual deviance: 5335.3 on 2480 degrees of freedom
24 AIC: 6520.6
25
26 Number of Fisher Scoring iterations: 6

```

Code 6.2: Resultat eines Binomialmodells mit zwei Haupteffekten und deren Interaktion für den Ausschnitt aus Probe 201 (A).

6.2.2 Vergleich der Modellierungsergebnisse der zwölf Proben

Bei einer effektiven Probenfläche von $4,6 \times 4,6 \text{ cm}^2$, einer Auflösung von $15 \text{ }\mu\text{m}$ pro Pixel und einer Blockgröße von 6×6 Pixel besteht der Datensatz einer jeden Probe aus 261121 Beobachtungen, die auf etwa 10,5 Millionen Einzelmessungen basieren. Dass der effektive Datensatz etwas kleiner als die ursprünglich angesprochenen $5 \times 5 \text{ cm}^2$ ist liegt an der Filterung der Topografie, im Zuge derer aufgrund von Artefakten der Rand der Probe verworfen werden muss.

Wegen der immer noch großen Probenfläche wird hier auf bildliche Darstellungen der Modellprognosen verzichtet und statt dessen die wichtigsten Modellierungsergebnisse der zwölf Proben in Tabelle 6.6 aufgelistet. Alle betrachteten Modelle basieren auf den beiden Haupteffekten Topografie und Formation sowie deren Interaktion, besitzen also vier Parameter. Nur für Probe 103 (B) ist aufgrund messtechnischer Komplikationen die effektive Messfläche etwas kleiner und besteht in gruppierter Form aus nur 217156 Beobachtungen (9 Millionen Einzelbeobachtungen auf einer Fläche von $4,2 \times 4,2 \text{ cm}^2$).

Probe	101 (A)	101 (B)	102 (A)	102 (B)	103 (A)	103 (B)
Nulldevianz	76435	92807	168690	298020	377353	415396
Modelldevianz	56170	69758	65795	128017	250143	326494
AIC	60377	76088	75546	146872	282795	372061
Probe	201 (A)	201 (B)	202 (A)	202 (B)	203 (A)	203 (B)
Nulldevianz	870498	1707914	662638	415627	1802059	1831584
Modelldevianz	276588	709096	401163	233193	719375	843617
AIC	327714	841748	456332	266662	851360	997179

Tabelle 6.6.: Modellierungsergebnisse der zwölf Proben für Modelle basierend auf den Haupteffekten Topografie und Formation sowie deren Wechselwirkung. Mit Ausnahme von Probe 103 (B) bestehen alle Datensätze aus $N = 261121$ Beobachtungen, für Probe 103 (B) sind es $N = 217156$ Beobachtungen.

Tabelle 6.6 zeigt nun Null- sowie Modelldevianz jeder Probe sowie den AIC Wert. Die Modelldevianzen wurden auch für den in Abbildung 6.23 dargestellten Plot verwendet, sind dort jedoch in durch ihre Freiheitsgrade skaliertes Form gezeigt. Bereits in Abschnitt 2.2.4 wurde erwähnt, dass die Devianz zumindest asymptotisch aus einer χ^2_{N-p} Verteilung stammt. Die Devianz sollte also in etwa so groß wie die Anzahl der Freiheitsgrade des Modells, $N - p = \# \text{Beobachtungen} - \# \text{Modellparameter}$, sein. Daher erlaubt die Betrachtung des Quotienten der Modelldevianz und der Freiheitsgrade eine schnelle visuelle Entscheidung darüber, ob die Variabilität in den Daten deutlich größer jener der Binomialverteilung (Quotient größer 1) oder deutlich kleiner (Quotient kleiner 1) ist.

Für die vorliegenden Daten zeigt die Grafik in Abbildung 6.23 den Zusammenhang des Devianzquotienten mit der unter Abschnitt 6.1 vorgestellten Druckqualität. Wie aus der Gegenüberstellung abzulesen ist, liegt ein deutlicher Zusammenhang zwischen der Druckqualität und der Variabilität der Responsevariablen vor. Dabei variieren bei Proben mit gutem Druck die Daten geringer, als unter Binomialverteilungsannahme angenommen. Liegt jedoch ein schlechtes Druckbild vor, so ist die Devianz deutlich größer als die Anzahl der Freiheitsgrade des Modells,

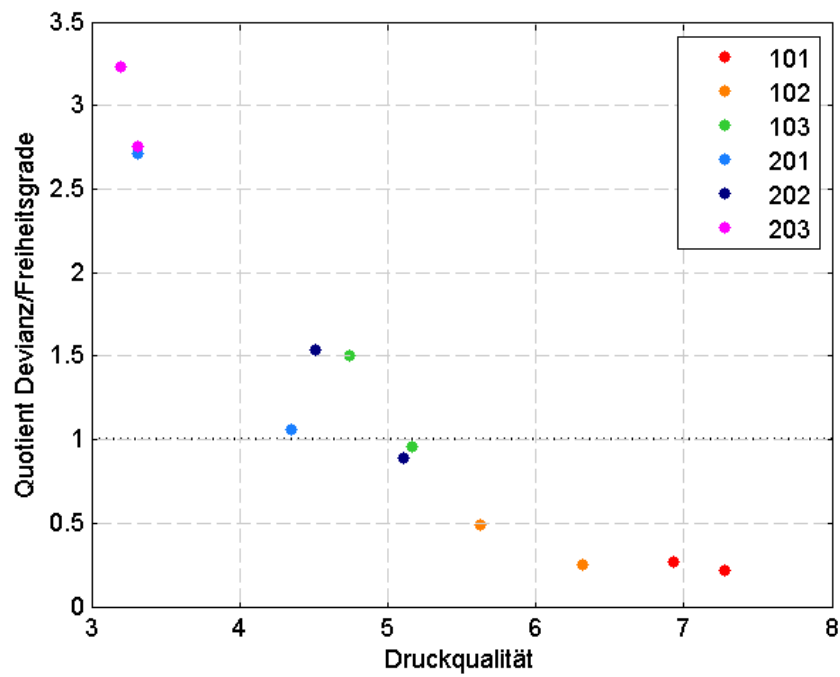


Abbildung 6.23.: Gegenüberstellung der Druckqualitäten und der resultierenden Modelldevianzen für die zwölf betrachteten Proben.

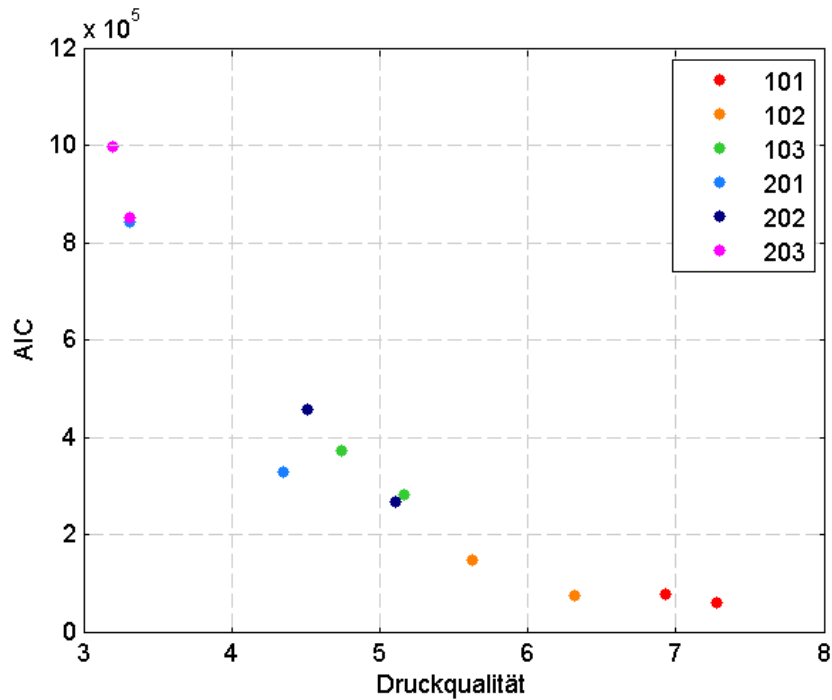


Abbildung 6.24.: Vergleich der Druckqualität mit den AIC Werten der zwölf Binomialmodelle.

die Daten variieren also stärker als von einem Modell unter Binomialverteilungsannahme erklärt werden kann.

Ein Vergleich mit den in Tabelle 6.1 aufgelisteten Werten zeigt, dass die Papiere 101 und 102 besonders glatt sind und ein sehr gutes Druckbild erlauben. Der Anteil unbedruckter Beobachtungen liegt bei diesen drei Papieren unter 5 %. Wie auch Tabelle 6.6 zu entnehmen ist, weisen die Devianzen dieser drei Modelle auf eine deutlich kleinere Variabilität in den Daten hin, als bei der klassischen Binomialverteilung zu erwarten wäre.

Das andere Extrem wird vor allem von den beiden Proben von Papier 203 sowie Probe 201 (B) gestellt. Deren unbedruckter Anteil liegt bei etwa 4 % und die Devianzen sind dreimal so groß wie erwartet. Dies weist auf eine deutlich größere Variabilität in den Daten hin, als vom Binomialmodell angenommen wird. Ein Blick auf die Papiercharakteristika in Tabelle 6.1 zeigt, dass Papier 203 extrem rau ist. Dass die beiden anderen Papiere mit ähnlichen Rauigkeitswerten (103 und 202) keine derartig große Variabilität aufweisen, könnte aufgrund der vorliegenden Informationen aus dem Flächengewicht der Proben stammen. Im Gegensatz zu den übrigen Proben wiegen diese beiden Doppelproben nur 70 g/m².

Ein Vergleich der AIC Werte gegen die Druckqualitäten der zwölf Proben ist in Abbildung 6.24 dargestellt. Korrigierter AIC oder BIC sind hier nicht gesondert geplottet, da sich deren numerische Werte nur minimal vom den AIC Werten unterscheiden.

Probe	Topografie		Formation		Wechselwirkung	
	Schätzer	Std.Fehler	Schätzer	Std.Fehler	Schätzer	Std.Fehler
101 (A)	3,3446	2,61·10 ⁻²	0,0386	1,58·10 ⁻³	-0,0361	2,92·10 ⁻⁴
101 (B)	2,6623	3,07·10 ⁻²	0,0007*	1,45·10 ⁻³	-0,0239	3,56·10 ⁻⁴
102 (A)	0,6476	2,94·10 ⁻²	0,0776	2,06·10 ⁻³	0,0018	3,56·10 ⁻⁴
102 (B)	0,8156	2,39·10 ⁻²	0,0630	1,40·10 ⁻³	-0,0005*	2,87·10 ⁻⁴
103 (A)	0,4821	9,55·10 ⁻³	0,0248	8,69·10 ⁻⁴	-0,0009	1,42·10 ⁻⁴
103 (B)	0,8645	9,37·10 ⁻³	-0,0068	6,45·10 ⁻⁴	-0,0077	1,36·10 ⁻⁴
201 (A)	-0,2001	1,33·10 ⁻²	0,1693	7,85·10 ⁻⁴	0,0099	1,67·10 ⁻⁴
201 (B)	-0,3979	8,82·10 ⁻³	0,1277	3,83·10 ⁻⁴	0,0116	1,08·10 ⁻⁴
202 (A)	0,0203	7,48·10 ⁻³	0,0817	6,94·10 ⁻⁴	0,0055	1,15·10 ⁻⁴
202 (B)	-0,0244	8,98·10 ⁻³	0,0913	9,42·10 ⁻⁴	0,0071	1,44·10 ⁻⁴
203 (A)	-0,0195*	6,43·10 ⁻³	0,1108	3,86·10 ⁻⁴	0,0059	7,75·10 ⁻⁵
203 (B)	-0,0857	7,00·10 ⁻³	0,0981	3,82·10 ⁻⁴	0,0064	8,29·10 ⁻⁵

Tabelle 6.7.: Punktschätzer mit Standardfehlern der Wechselwirkungsmodelle. p-Werte größer 10⁻⁵ sind mit * gekennzeichnet.

Bei den in Tabelle 6.7 gelisteten Punktschätzern der Topografie unterscheiden sich vor allem jene von Papier 101 deutlich von denen der restlichen Papiere, da sie erheblich größer sind als die übrigen. Dieses Papier wurde mit derselben Satinagetechnologie wie Papiere 102 und 103, jedoch unter anderen Einstellungen als diese geglättet. Laut den Messwerten aus Tabelle 6.1 handelt es sich dabei um das glatteste der vorliegenden Papiere, welches auch die beste Druckqualität erlaubt (siehe Tabelle 6.2). Die beiden Papiere 103 und 202 mit niedrigerem Flächengewicht heben sich nicht besonders von den anderen Papieren ab. Es ist jedoch eine deutliche Trennung bezüglich dem Punktschätzer Topografie zwischen den beiden Satinagetechnologien auszumachen. So sind die Punktschätzer $\hat{\beta}_{\text{Topografie}}$ der 100-er Serie positiv, jene

der 200-er Serie hingegen bis auf eine Ausnahme negativ.

Weiters zeigt sich, dass nur für Probe 102 (B) die Hypothese $\beta_{\text{Interaktion}} = 0$ nicht verworfen werden kann. In den elf restlichen Proben jedoch ist der zu einer derartigen Teststatistik gehörende p-Wert dermaßen klein, dass die Wechselwirkung zwischen den beiden Prädiktoren als relevant angesehen werden kann. Außerdem tragen die Punktschätzer im Fall der 100-er Serie mit Ausnahme von Probe 102 (A) ein negatives Vorzeichen, jene der 200-er Serie ein positives. Hier zeigt sich somit abermals eine Trennung in Abhängigkeit der angewandten Satinagetechnologie.

Ausgehend von diesen Beobachtungen sollen nun in weiterer Folge Modelle unter der Annahme der multiplikativen und der doppelten Binomialverteilung geschätzt werden.

6.3 Verallgemeinerte Binomialmodelle

6.3.1 Modellschätzung für einen Ausschnitt von Probe 201 (A)

Abermals soll mit einer Modellierung des in Abschnitt 6.2.1 vorgestellten Ausschnitts von Probe 201 (A) begonnen werden, der laut der Modellzusammenfassung in Code 6.2 eine deutliche Überdispersion aufweist. Diese soll nun mittels der multiplikativen beziehungsweise der doppelten Binomialverteilung in den Griff bekommen werden.

Um jedoch ein Modell unter doppelter oder multiplikativer Binomialverteilung schätzen zu können, müssen die vorliegenden Daten gruppiert werden. Wird dies nicht getan und statt dessen mit Gruppengröße $n = 1$ weiter gearbeitet, so vereinfachen sich beide Verallgemeinerungen der Binomialverteilung zur klassischen Binomialverteilung und die gewünschte Anpassung an größere oder kleinere Variabilität als jene der klassischen Binomialverteilung kann nicht erfolgen. Daher werden die vorliegenden Daten nun wie in Kapitel 6.1 beschrieben gruppiert, wobei für die Blockgröße zu Beginn 6×6 Pixel festgelegt wird. Am Ende dieses Abschnittes wird aufgrund der Beobachtungen der Kapitel 3 und 4, nach denen beide Verallgemeinerungen der Binomialverteilung nicht abgeschlossen bezüglich der Faltung sind, noch eine weitere Blockgröße betrachtet.

Modell unter klassischer Binomialverteilungsannahme

Zu Beginn zeigt Code 6.3 den etwas künstlichen Fall der Modellzusammenfassung eines Modells unter (klassischer) Binomialverteilungsannahme. Dabei werden dieses Mal jedoch nicht die R-Funktionen `glm()` und `summary()`, sondern die in Abschnitt 5.5 vorgestellten Funktionen `gbmod()` und `gbsummary()` verwendet. Selbst im Fall einer klassischen Binomialverteilungsannahme wird dabei innerhalb von `gbmod()` bewusst nicht auf die Modellierung mittels generalisiertem linearem Modell übergangen. Statt dessen wird die für die beiden verallgemeinerten Binomialverteilungen aufgestellte Vorgehensweise mit `optim()` nachvollzogen.

Die Zeilen 7 bis 11 sind in dieser Modellzusammenfassung abermals dem mit `glm()` geschätzten Binomialmodell gewidmet und geben die negative doppelte log-Likelihood, die Devianz, die Pearson X^2 Statistik sowie den AIC des Modells an. Wie in Zeile 28 des Codes abzulesen, wird die optimale Lösung der maximierten log-Likelihoodfunktion im Fall der Binomialverteilung klarerweise innerhalb des ersten Iterationsschritts von `optim()` gefunden. Da die Iteration mit

```

> binmod <- gbmod(cbind(succs,n-succs)~topography*formation, "binomial")
2 > gbsummary(binmod)

4 Call:
gbmod(formula = cbind(y, n - y) ~ topography * formation , binomial )
6
Values from the binomial model:
8     -2*logl: 6471.035
      Deviance: 5293.798 on 2480 degrees of freedom
10     Pearson X^2: 336529.6
      AIC: 6479.035
12
Response distribution: classic binomial distribution.
14
Pearson Residuals:
16      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
-572.3000   0.1331   0.2985  -0.1632   0.5872   9.7530
18
Coefficients:
20              Estimates Std.Errors t values P(>|t|)
(Intercept)      -9.09584    0.43356  -20.98 < 2e-16 ***
22 topography      -0.56535    0.09083   -6.22 5.7e-10 ***
formation         0.17275    0.00572   30.20 < 2e-16 ***
24 topography:formation 0.01424    0.00120   11.91 < 2e-16 ***
---
26 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

28 Convergence reached after 1 Iterations.

```

Code 6.3: Modellzusammenfassung für den Ausschnitt von Probe 201 (A) unter Annahme der klassischen Binomialverteilung.

den aus der Modellschätzung mit `glm()` resultierenden Punktschätzern startet, muss dies im Fall der klassischen Binomialverteilung bereits die optimale Lösung des mit `optim()` betrachteten Maximierungsproblems darstellen.

Modell unter multiplikativer Binomialverteilungsannahme

Aufgrund der bereits angesprochenen Überdispersion der Daten zeigt Code 6.4 nun die Modellzusammenfassung, wenn von einer multiplikativen Binomialverteilung ausgegangen wird.

Da für den Erwartungswert und die Varianz der multiplikativen Binomialverteilung komplexere Zusammenhänge als für die klassische Binomialverteilung gelten, haben sich die Punktschätzer des in Code 6.4 gezeigten Modells gegenüber dem Binomialmodell verändert. Sowohl Topografie als auch Formation sind jedoch wie nicht anders zu erwarten noch immer entscheidend, wenn es um die Beschreibung der Farbübertragung im Druck geht. Selbiges trifft auch für die Wechselwirkung der beiden Prädiktoren zu. Der zusätzliche Parameter wird auf $\hat{\omega} = 0,94$ mit einem Standardfehler von 0,0014 geschätzt. Der in den Zeilen 28 bis 30 gezeigte Test überprüft die Hypothese $H_0: \omega = 1$, verwirft diese im gezeigten Beispiel jedoch. Die aus den Daten geschätzte Verteilungsfunktion entspricht somit keiner klassischen Binomialverteilung mehr. In anderen Worten ist die klassische Binomialverteilung zur Modellierung des vorliegenden Datensatzes nicht angemessen.

Der Wert der negativen doppelten log-Likelihoodfunktion in den Punktschätzern $(\hat{\beta}, \hat{\omega})$ liegt in diesem Modell bei 5417, ist also im Vergleich zum Modell unter klassischer Binomialverteilungsannahme deutlich gesunken. Der Wert der Pearson X^2 Statistik beträgt nun 52532. Ein Vergleich mit dem Erwartungswert der entsprechenden χ^2 Verteilung deutet jedoch darauf hin, dass auch dieses Modell keine besonders gute Beschreibung der vorliegenden Stichprobe liefern kann. Im Gegensatz zur klassischen Binomialverteilungsannahme jedoch fand eine deutliche Modellverbesserung statt, wie der von 6479 auf 5428 gefallene AIC zeigt. Basierend auf diesen Entscheidungsgrößen ist das multiplikative Modell dem klassischen binomialen Modell somit vorzuziehen.

Modell unter doppelter Binomialverteilungsannahme

Natürlich soll auch die Zusammenfassung eines Modells unter Annahme der doppelten Binomialverteilung besprochen werden. Diese ist in Code 6.5 gezeigt.

Die hier resultierenden Punktschätzer unterscheiden sich aus demselben Grund wie bei der multiplikativen Binomialverteilung von jenen des Binomialmodells, wobei der Unterschied hier noch deutlicher ausfällt. Wie schon im multiplikativen Modell ändert das jedoch nichts an der bereits beobachteten Tatsache, dass sowohl die Topografie als auch die Formation wichtige Prädiktoren sind. Für den zusätzlichen Parameter ϕ liefert die Modellschätzung den Wert $\hat{\phi} = 0,06$ mit einem Standardfehler von 0,2005 Einheiten. Der im Output in den Zeilen 28 bis 30 durchgeführte Test testet im Fall der doppelten Binomialverteilung die Nullhypothese $H_0: \phi = 1$ gegen die zweiseitige Alternative, da die klassische Binomialverteilung innerhalb der doppelten Binomialverteilung bei $\phi = 1$ auftritt. Für das vorliegende Beispiel wird die Nullhypothese verworfen; die den Daten zugrunde liegende Verteilung entspricht also nicht jener der klassischen Binomialverteilung.

Der Wert der negativen doppelten log-Likelihood in den Parametern $(\hat{\beta}, \hat{\phi})$ liegt in diesem Modell bei 3887. Im Gegensatz zu den beiden anderen Modellen fand also abermals eine deutliche Verkleinerung statt. Auch der AIC, anhand dessen sich Modellvergleiche tätigen lassen,

```

> multmod <- gbmod(cbind(success,all-success)~topography*formation, "multbin")
2 > gbsummary(multmod)

4 Call:
gbmod(formula = cbind(y, n - y) ~ topography * formation , multbin )
6
Values from the binomial model:
8   -2*logl: 6471.035
   Deviance: 5293.798 on 2480 degrees of freedom
10  Pearson X^2: 336529.6
   AIC: 6479.035
12
Response distribution: multiplicative binomial distribution.
14
Pearson Residuals:
16   Min.    1st Qu.    Median      Mean    3rd Qu.     Max.
-212.00000   0.20970   0.39730   0.03084   0.65530   9.76800
18
Coefficients:
20             Estimates Std.Errors t values P(>|t|)
(Intercept)    -8.710842  0.299699  -29.1 <2e-16 ***
22 topography    -1.074825  0.059428  -18.1 <2e-16 ***
   formation      0.138809  0.004168   33.3 <2e-16 ***
24 topography:formation 0.017542  0.000789   22.2 <2e-16 ***
---
26 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

28 Coefficient of Parameter omega:
   Estimate Std.Error t value P(>|t|)
30  0.94153  0.00136    -43 <2e-16 ***

32 Degrees of Freedom: 2483 Total (i.e. Null); 2479 Residual
   -2*logl: 5417.715
34 Pearson X^2: 52532.24
   AIC: 5427.715
36
Convergence reached after 18 Iterations.

```

Code 6.4: Modellzusammenfassung für den Ausschnitt von Probe 201 (A) unter Annahme der multiplikativen Binomialverteilung.

```

1 > doublemod <- gbmod(cbind(success,all-success)~topography*formation,
+ "doublebin")
3 > gbsummary(doublemod)

5 Call:
  gbmod(formula = cbind(y, n - y) ~ topography * formation , doublebin )
7
  Values from the binomial model:
9     -2*logl: 6471.035
     Deviance: 5293.798 on 2480 degrees of freedom
11    Pearson X^2: 336529.6
     AIC: 6479.035

13 Response distribution: double binomial distribution.

15 Pearson Residuals:
17    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   -14.3000  0.2849  0.3701  0.1581  0.4660  5.7610

19 Coefficients:
21              Estimates Std.Errors t values P(>|t|)
   (Intercept)      -60.6710     9.1619   -6.62 4.3e-11 ***
23 topography         -7.8244     1.3057   -5.99 2.4e-09 ***
   formation          0.9549     0.1400    6.82 1.1e-11 ***
25 topography:formation  0.1250     0.0199    6.29 3.8e-10 ***
   ---
27 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

29 Coefficient of Parameter phi:
   Estimate Std.Error t value P(>|t|)
31    0.0615  0.2005  -4.68 3e-06 ***

33 Degrees of Freedom: 2483 Total (i.e. Null); 2479 Residual
   -2*logl: 3887.072
35 Pearson X^2: 4103.701
   AIC: 3897.072

37 Convergence reached after 34 Iterations.

```

Code 6.5: Modellzusammenfassung für den Ausschnitt von Probe 201 (A) unter Annahme der doppelten Binomialverteilung.

ist im Fall des doppelten Binomialverteilungsmodells deutlich kleiner als in den beiden anderen Modellen. Lag der AIC im multiplikativen Modell noch bei 5428, so beträgt der Wert im Modell unter doppelter Binomialverteilungsannahme nur noch 3897.

Die Anzahl an Iterationen ist mit 34 im Fall der doppelten Binomialverteilung jedoch höher als im Fall der multiplikativen Binomialverteilung. Diese Tatsache kann im direkten Vergleich der beiden Verteilungen an denselben Datensätzen oft beobachtet werden. Wichtig ist jedoch, dass die Parameterschätzung in beiden Fällen vor Erreichen der maximalen Iterationszahl konvergiert.

Somit deuten die Vergleiche der negativen log-Likelihoodwerte sowie der AICs darauf hin, dass das Modell unter doppelter Binomialverteilungsannahme für den vorliegenden Datensatz den anderen beiden Modellen deutlich überlegen ist.

Vergleiche der drei Modelle

Dieser Abschnitt sei nun der Gegenüberstellung der Resultate der drei Modellierungen gewidmet.

Die Abbildungen 6.25, 6.26 und 6.27 zeigen die gefitteten Wahrscheinlichkeiten unter den drei unterschiedlichen Binomialverteilungsannahmen. Auffallend ist dabei, dass im Vergleich zu den aus dem binomialen Modell geschätzten Wahrscheinlichkeiten ein Modell unter der multiplikativen Binomialverteilung sichtlich mehr Zonen besitzt, in denen die Wahrscheinlichkeit für erfolgreiches Bedrucken um $1/2$ liegt. Die kritischen Bereiche sind somit größer und unschärfer als im Binomialmodell.

Im Fall der doppelten Binomialverteilung erfolgt die Trennung zwischen sehr wahrscheinlicher und extrem unwahrscheinlicher erfolgreicher Farbübertragung hingegen viel schärfer als im Binomialmodell. Zonen mit prognostizierten Wahrscheinlichkeiten um 0,5 sind extrem selten, die meisten Regionen werden laut diesem Modell entweder mit großer Wahrscheinlichkeit erfolgreich bedruckt sein oder mit hoher Wahrscheinlichkeit nicht bedruckt werden.

Mit den Plots in den Abbildungen 6.28 und 6.29 wird versucht, für den Ausschnitt des Datensatzes einen Vergleich des Verhaltens der ersten beiden Momente der drei Verteilungen bildlich darzustellen. Zur Erstellung dieser Plots werden die Wertebereiche der beiden Prädiktoren Topografie und Formation in diskrete Intervalle unterteilt. Dabei wird die Unterteilung derart gewählt, dass auch in den Rändern noch mehrere Beobachtungen für die Berechnungen zur Verfügung stehen. Die Intervalle an den Rändern der Wertebereiche sind somit breiter als jene im Zentrum. Ausgehend von dieser Rasterung der beiden Prädiktoren wird dann für Beobachtungen innerhalb eines Topografie- sowie Formationsintervalls eine unabhängige und identische Verteilung angenommen.

Für den Plot in Abbildung 6.28 werden die Erwartungswerte der Beobachtungen innerhalb der auf diese Art und Weise entstandenen Blöcke geschätzt. Außerdem werden für die mittlere Topografie und die mittlere Formation eines jeden Blocks die Erwartungswerte unter den drei Modellen geschätzt.

Die in schwarz eingezeichnete Fläche zeigt die beobachteten Häufigkeiten erfolgreich bedruckter Beobachtungen in Abhängigkeit von Topografie und Formation. Wie sich erkennen lässt, sind diese über weite Wertebereiche der beiden Prädiktoren maximal. In diesen Bereichen werden also keine unbedruckten Stellen beobachtet. Für geringe Topografiewerte jedoch fallen diese beobachteten Häufigkeiten auf Werte kleiner der Blockgröße $n = 36$. Bei diesen niedrigen Topografien werden also unbedruckte Stellen beobachtet, wobei besonders viele bei niedrigen

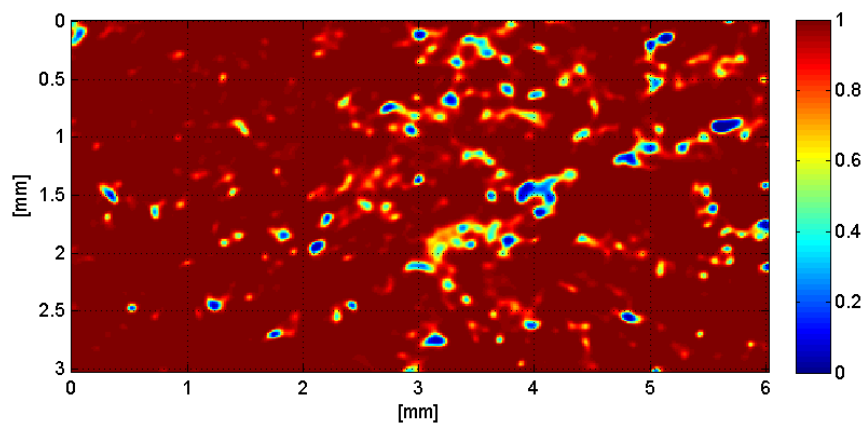


Abbildung 6.25.: Geschätzte Wahrscheinlichkeiten im Fall der klassischen Binomialverteilung.

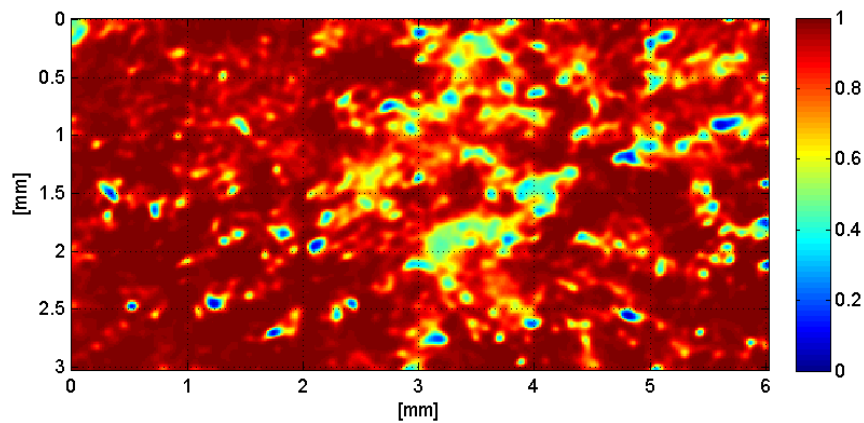


Abbildung 6.26.: Geschätzte Wahrscheinlichkeiten im Fall der multiplikativen Binomialverteilung.

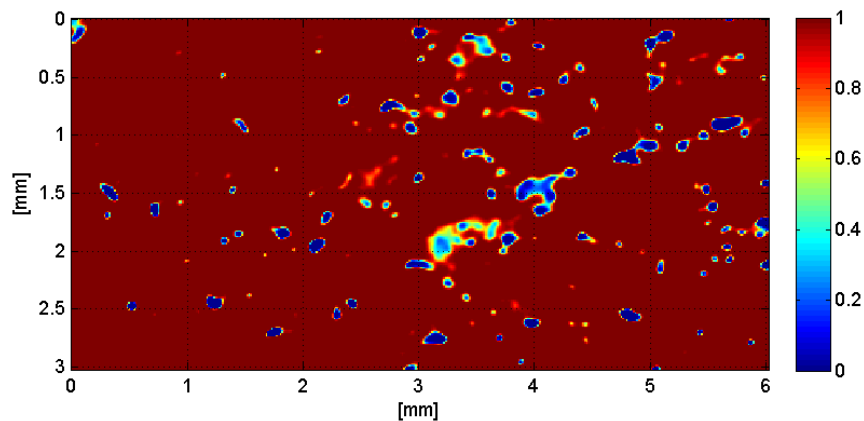


Abbildung 6.27.: Geschätzte Wahrscheinlichkeiten im Fall der doppelten Binomialverteilung.

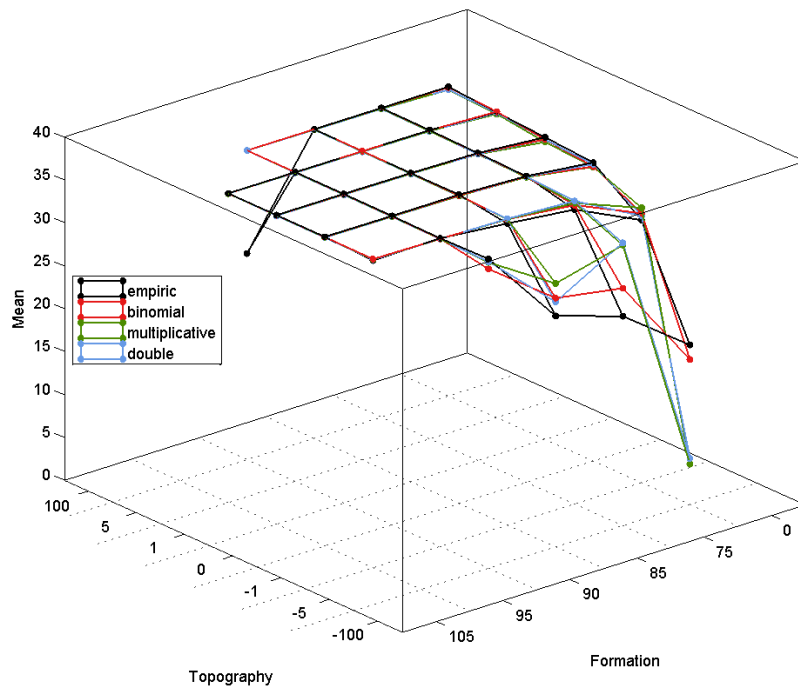


Abbildung 6.28.: Vergleich der aus den geschätzten Wahrscheinlichkeiten der drei Modelle errechneten Erwartungswerte.

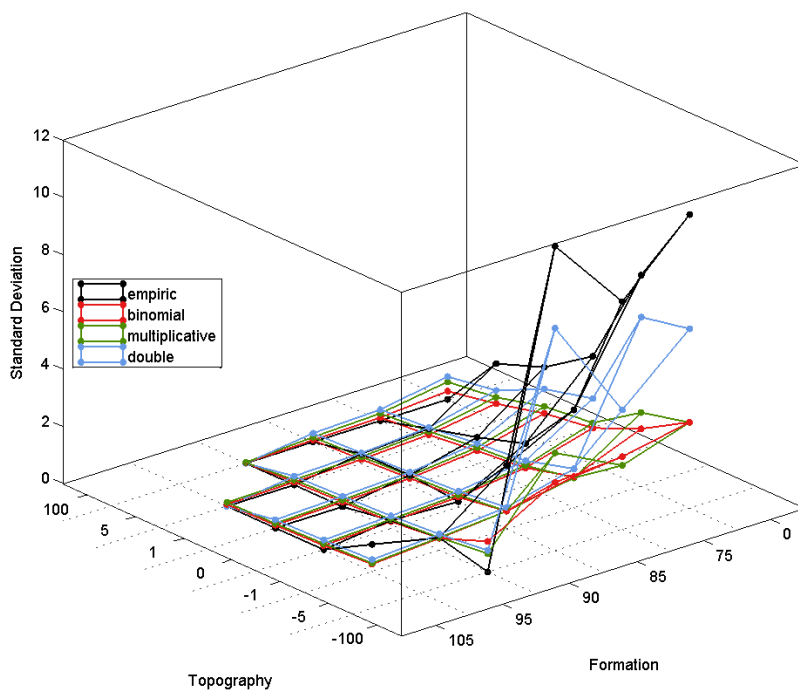


Abbildung 6.29.: Vergleich der aus den geschätzten Wahrscheinlichkeiten errechneten Standardabweichungen der drei unterschiedlichen Modelle.

Formationswerten vorkommen. Dieses Phänomen läßt sich durch physikalische Effekte aus dem Druckprozess erklären. Außerdem scheinen auch für extrem hohe Topografie- und Formationswerte unbedruckte Stellen beobachtet werden zu können. Diese Eigenschaft zeigt sich nicht nur an dem hier dargestellten Ausschnitt aus Probe 201 (A), sondern läßt sich für alle vorhandenen Probeflächen beobachten. Eine Erklärung für dieses Auftreten unbedruckter Stellen bei hohen Topografie- und Formationswerten liegt in der Messtechnik, mithilfe derer die Papiere aufgenommen wurden. Dabei tritt an Stellen sehr hoher Topografie mitunter Glanz auf, welcher in der Auswertungsprozedur als unbedruckt eingestuft wird, obwohl die Stelle womöglich dennoch erfolgreich bedruckt ist.

Wie die aus den drei Modellen geschätzten Häufigkeiten erfolgreicher Farbübertragung zeigen, liegen diese über große Wertebereiche der Prädiktoren bei 36, sind also ident dem beobachteten Wert. Für tiefe Topografie- und Formationswerte jedoch liegen jene aus dem Modell unter Annahme einer klassischen Binomialverteilung am nächsten an der beobachteten Situation. Die aus den beiden anderen Modellen geschätzten Erwartungswerte sind aufgrund ihres zweiten Parameters ω beziehungsweise ϕ weiter davon entfernt.

Interessant wird jedoch der Vergleich der resultierenden Standardabweichungen in Abbildung 6.29. Dieser zeigt, wie die geschätzten Standardabweichungen mit der tatsächlich beobachtbaren Situation übereinstimmen. Hier wird nun deutlich, warum die klassische Binomialverteilungsannahme aus einem Binomialmodell mitunter eine unangemessene Wahl ist. Im vorliegenden Beispiel erreichen die beobachteten Standardabweichungen Werte bis 11, wohingegen unter klassischer Binomialverteilungsannahme die Standardabweichung bei $n = 36$ selbst im besten Fall nicht größer als 3 werden kann. Aufgrund ihrer zusätzlichen Parameter jedoch können die multiplikative sowie die doppelte Binomialverteilung hier bessere Anpassungen erzielen.

Generell hat die Einführung eines zweiten Parameters in die für die Modellierung verwendete Verteilung zwar eine Verschlechterung gebracht, was den Erwartungswert betrifft. Die Variabilität der Daten kann dank dieses zweiten Parameters jedoch besser modelliert werden, als unter der klassischen Binomialverteilungsannahme möglich war.

Den Abschluss dieses Modellierungsteils machen nun Vergleiche der zwölf Modelle der vollen, $4,6 \times 4,6 \text{ cm}^2$ großen Probenflächen.

6.3.2 Modellierungsergebnisse der zwölf Proben

Den Beginn dieser Gegenüberstellungen macht ein Vergleich der in Tabelle 6.30 aufgelisteten AIC- und Sensitivitätswerte der zwölf Probenflächen. Die Anzahl an Beobachtungen beträgt mit Ausnahme von Probe 103 (B) 261121, für diese Probe liegen 217156 Beobachtungen vor. Außerdem muss erwähnt werden, dass für Probe 102 (B) die Modellschätzung unter doppelter Binomialverteilungsannahme nicht mehr funktionierte. Grund dafür sind numerische Probleme, da die für die Berechnung von log-Likelihood Funktion und deren Ableitungen notwendigen Ausdrücke dermaßen riesige Werte annehmen, dass in weiterer Folge die Maschinengenauigkeit überschritten wird.

Eine aussagekräftigere Gegenüberstellung der AIC-Werte als diese Tabelle zeigt der in Abbildung 6.31 dargestellte Scatterplot, welcher die drei aus den unterschiedlichen Verteilungsannahmen resultierenden AIC-Werte der zwölf Proben zeigt. Geplottet sind die Werte dabei gegen die Druckqualität der jeweiligen Probe. Somit läßt sich erkennen, wie geringfügig sich der AIC-Wert für gut bedruckte Proben ändert, wenn in der Modellschätzung von der klassischen

Verteilung Probe	binomial		multiplikativ	doppelt
	Devianz	AIC	AIC	AIC
101 (A)	56170	60377	55203	43372
101 (B)	69758	76088	71167	56544
102 (A)	65795	75546	73729	59485
102 (B)	128017	146872	140490	-
103 (A)	250143	282795	252896	179223
103 (B)	326494	372061	320278	224446
201 (A)	276588	327714	289538	210027
201 (B)	709096	841748	692499	468155
202 (A)	401163	456332	381643	258566
202 (B)	233193	266662	237137	168716
203 (A)	719375	851360	682999	462408
203 (B)	843617	997179	786099	526027

Abbildung 6.30.: AIC Werte der Wechselwirkungsmodelle unter den drei unterschiedlichen Verteilungsannahmen.

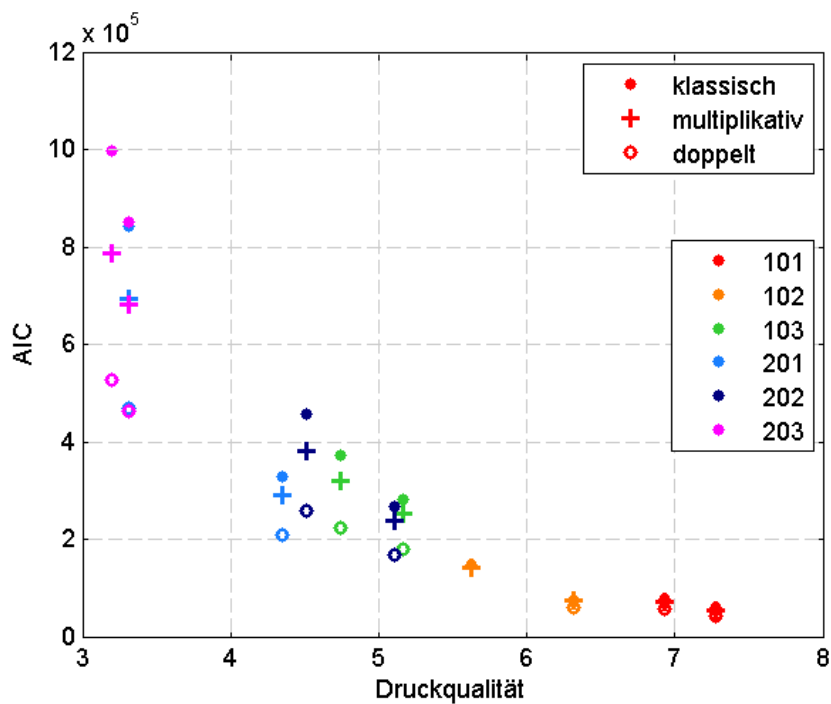


Abbildung 6.31.: AIC-Werte der drei Modelle unter verschiedenen Verteilungsannahmen aller zwölf Proben.

Binomialverteilungsannahme weggegangen wird. Die ist zum Beispiel bei beiden Proben 101 oder Probe 102 (A) der Fall. Für Proben schlechter Druckqualität hingegen wird der Unterschied in den drei unterschiedlichen AIC-Werte deutlich größer. Weiters läßt dieser Vergleich erkennen, dass die AIC-Werte unter doppelter Binomialverteilungsannahme immer die kleinsten der drei Verteilungsannahmen sind.

Ein Vergleich der AIC-Werte der zwölf Proben zeigt folglich, dass für gut bedruckte Proben (Druckqualität größer 6) ein Modell unter klassischer Binomialverteilungsannahme ebenso denkbar wäre wie unter verallgemeinerter Verteilungsannahme. Für Proben mit schlechterer Druckqualität jedoch deuten die AIC-Werte darauf hin, dass ein Modell mit doppelter Binomialverteilungsannahme die Daten besser beschreiben kann als die beiden anderen Modelle.

Da nun bereits oft vom zusätzlichen Parameter α beziehungsweise $\exp(\alpha) = \omega$ oder $\exp(\alpha) = \phi$ die Rede war, soll dieser Abschnitt mit einem Vergleich der Schätzungen dieses Parameters beschlossen werden. Die Schätzer $\hat{\omega}$ und ihre dazugehörigen Standardfehler der Modelle unter multiplikativer Binomialverteilungsannahme sind in Abbildung 6.32 gezeigt. Wie sich erkennen läßt, liegen alle Werte von ω im Intervall $[0,87; 0,97]$. Werte größer 1 wurden nicht beobachtet. Dabei sei hier nochmals auf die Grafik in Abbildung 3.7 verwiesen, welche eine Gegenüberstellung der Varianz bei der klassischen sowie der multiplikativen Binomialverteilung zeigt. Wie bereits bei der Vorstellung der multiplikativen Binomialverteilung in Kapitel 3 erwähnt, nehmen sowohl die Erfolgswahrscheinlichkeit π als auch der zweite Parameter, ω , Einfluss auf die Varianz der Verteilung.

Außerdem soll festgehalten werden, dass die Standardfehler der Punktschätzer $\hat{\omega}$ extrem gering sind. Sie sind dermaßen klein, dass die Fehlerbalken, obwohl in Abbildung 6.32 eingezeichnet, fast nicht zu erkennen sind.

Weiters sei darauf hingewiesen, dass rein aufgrund dieses Datensatzes kein Zusammenhang mit der Druckqualität vorzuliegen scheint. Die Punktschätzer von Probe 101 setzen sich zwar von den anderen Werten ab, abgesehen davon jedoch liegen die Punktschätzer aller Proben zwischen 0,9 und 0,97. Auch die beiden Proben aus leichterem Papier (103 und 202) zeichnen sich nicht von den schwereren ab.

Abbildung 6.33 zeigt nun die Gegenüberstellung der Punktschätzer des zweiten Parameters ϕ der doppelten Binomialverteilung, abermals geplottet gegen die Druckqualität. Auch hier sind wieder die Fehlerbalken der Punktschätzer eingezeichnet, jedoch sind sie nur bei Papieren 102 und 201 tatsächlich zu erkennen. Bei diesem Vergleich läßt sich nun ein gewisser Zusammenhang von Druckqualität und Realisation des Punktschätzers erkennen. So weisen vor allem Proben mit besonders schlechter Druckqualität einen sehr kleinen Parameterschätzer auf, wohingegen dieser für Proben besserer Druckqualität größer wird. Aufgrund der geringen Probendichte jedoch lassen sich hier noch keine klaren Aussagen tätigen.

6.3.3 Andere Blockgrößen

Nun kann wie in den Abschnitten 3.3 sowie 4.2 gezeigt weder bei der multiplikativen noch bei der doppelten Binomialverteilung davon ausgegangen werden, dass diese abgeschlossen bezüglich der Faltung von derartig verteilten Zufallsvariablen sind. Daher wird dieser Teil nun die Ergebnisse von Modellschätzungen unter anderen Wahlen der Größe n ansprechen. Bis jetzt wurde die Gruppengröße $n = 36$ betrachtet, nun sollen die Ergebnisse unter $n = 81$ (Block-

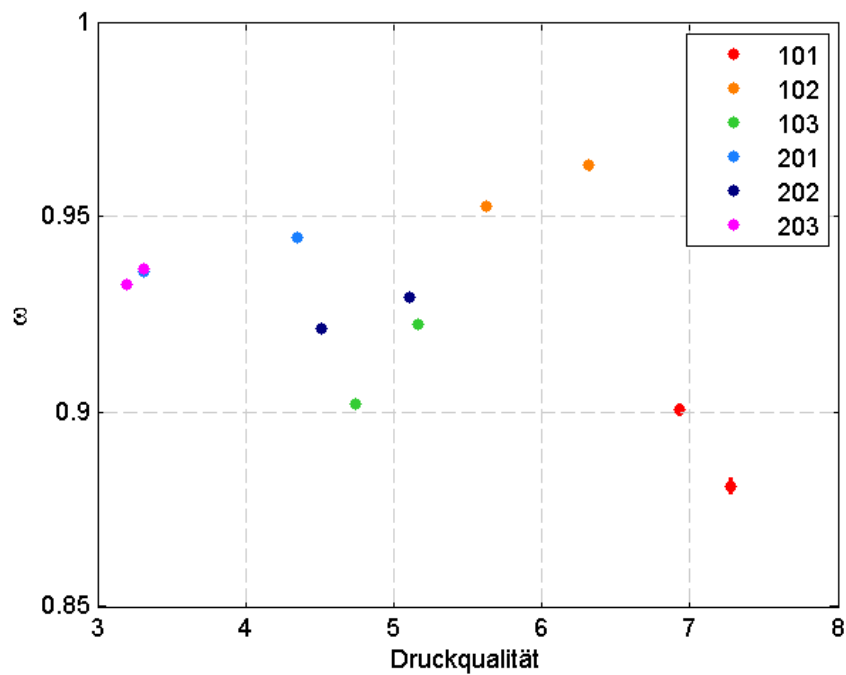


Abbildung 6.32.: Plot des zusätzlichen Parameters ω der Modelle unter multiplikativer Binomialverteilungsannahme inklusive Standardfehlern gegen die Druckqualität.

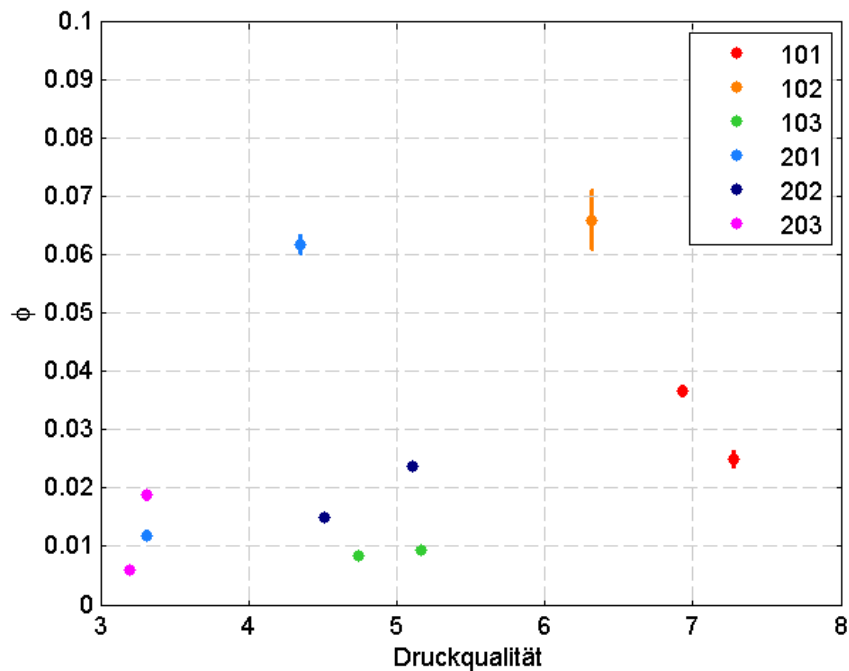


Abbildung 6.33.: Plot des zusätzlichen Parameters ϕ der Modelle unter doppelter Binomialverteilungsannahme inklusive Standardfehlern gegen die Druckqualität.

größe 9×9 Pixel) besprochen werden. Weiters wurden noch die Resultate für $n = 144$ und $n = 400$ betrachtet, diese werden hier jedoch nicht gesondert diskutiert.

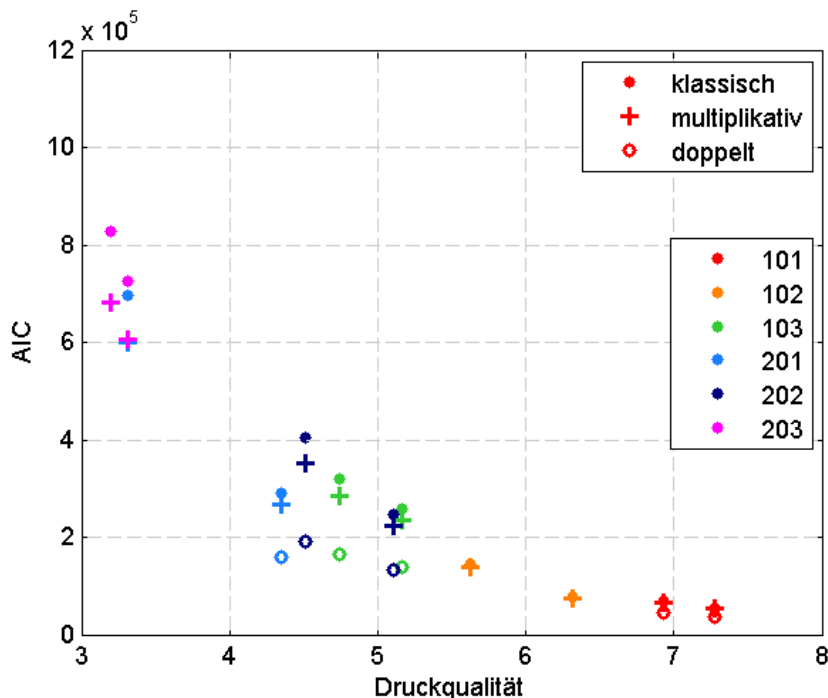


Abbildung 6.34.: Druckqualitäten gegen AIC-Werte unterschiedlicher Modelle im Fall $n = 81$.

Dafür sind in Abbildung 6.34 die AIC-Werte von Modellen unter den drei Verteilungsannahmen gegen ihre Druckqualitäten geplottet, und zwar für die Blockgröße $n = 81$. Zusammen mit dem äquivalenten Plot für $n = 36$ in Abbildung 6.31 läßt sich erkennen, wie die AIC-Werte aller Modelle bei größer werdendem n gleichmäßig kleiner werden. Es sei darauf hingewiesen, dass für die größer Blockgröße unter multiplikativer Verteilungsannahme der Responsevariable keine Komplikationen in der Modellschätzung auftreten, unter doppelter Binomialverteilungsannahme jedoch die Modellschätzung für die Papiere 102, 203 und 201 (A) nicht konvergiert. Bei diesen problematischeren Proben handelt es sich unter anderem um genau jene drei, welche auch die deutlich schlechtesten Druckqualitäten aufweisen.

Interessant sind auch Vergleiche der Punktschätzer $\hat{\omega}$ und $\hat{\phi}$ unter verschiedenen n . Dabei zeigt sich, wie für wachsendes n die Punktschätzer $\hat{\omega}$ immer weiter zur 1 wandern, der Binomialverteilung also ähnlicher werden. Dennoch muss erwähnt werden, dass die zu diesen Punktschätzern gehörenden Standardfehler dermaßen gering sind, dass ein Hypothesentest $H_0 : \omega = 1$ verworfen wird. Immer noch ist also eine generalisierte Binomialverteilung der klassischen vorzuziehen.

Bei der doppelten Binomialverteilung kann beobachtet werden, wie sich die $\hat{\phi}$ für wachsendes n der 0 annähern. Bereits bei $n = 36$ liegen die Punktschätzer $\hat{\phi}$ in der Nähe der 0 und nicht wie eher vermutet um die 1. Dieses Verhalten wird für wachsendes n nur noch deutlicher. Ein Blick auf die Wahrscheinlichkeitsfunktion der doppelten Binomialverteilung (Abbildung 4.2) jedoch erklärt dieses Verhalten. Für die vorliegenden Daten werden unter extrem vielen Erfolgen nur wenig Misserfolge beobachtet. Der Großteil der Masse der Wahrscheinlichkeiten liegt daher am

Rand bei $y = n$. Diese Situation kann in der doppelten Binomialverteilung nun für Parameterwerte für ϕ nur knapp größer der 0 wiedergegeben werden. Daher auch die auf den ersten Blick unerwarteten Ergebnisse für $\hat{\phi}$.

Dabei können bei größer werdender Blockgröße n numerische Probleme auftreten. So können die einzelnen Ausdrücke der Wahrscheinlichkeitsfunktionen sowie ihrer Ableitungen mitunter dermaßen große Werte annehmen, dass die Maschinengenauigkeit überschritten wird und aus diesem Grund keine Modellschätzung mehr möglich ist.

Damit schließt die Gegenüberstellung von Modellen unter unterschiedlichen Binomialverteilungsannahmen. Wie sich gezeigt hat, können die vorhandenen Daten mit den verallgemeinerten Binomialverteilungsannahmen geschätzt werden. Diese liefern zwar eine Verbesserung in der Schätzung der Varianzstruktur, erlauben jedoch keine Verbesserung der diskreten Modellprognosen.

Ein Ausblick über weitere mögliche Untersuchungen der hier vorgestellten Problemstellung ist in Kapitel 7 zu finden.

Zusammenfassung und Ausblick

In der vorliegenden Arbeit wurde das bereits in der Diplomarbeit (Feirer, 2008) betrachtete Thema der fehlerhaften Farbübertragung im Flexodruck abermals aufgegriffen und vertieft.

7.1 Zusammenfassung der praxisbezogenen Resultate

Neue Sackpapierproben wurden ausgewählt und vermessen, um das Phänomen der fehlerhaften Farbübertragung im Flexodruck mittels logistischem Modell zu untersuchen. Sowohl der Probenatz als auch die betrachteten Prädiktoren sind in Abschnitt 6.1 genau beschrieben. Dabei wurde neben dem in der Diplomarbeit verwendeten Prädiktor Topografie auch eine weitere Papiereigenschaft, die Formation, erhoben. Unterschiedliche Modelle wurden geschätzt und mit der beobachteten Situation verglichen, indem aus den prognostizierten Wahrscheinlichkeiten für erfolgreiche Farbübertragung diskrete Prognosen erstellt wurden. Eine genaue Beschreibung dieser Diskretisierung ist in Abschnitt 6.2.1 zu finden.

Modellschätzungen mit beiden Prädiktoren haben gezeigt, dass sowohl die Topografie als auch die Formation einen Einfluss auf das Auftreten von Fehlstellen nehmen. Dabei konnten die Modellschätzungen zeigen, dass die Topografie für die Form der einzelnen Fehlstellen verantwortlich ist. Die Formation hingegen unterteilt die gesamte Probenfläche in kritischere und weniger kritische Zonen, was die Farbübertragung im Druckwerk betrifft. So lassen sich Regionen höherer Formation besser bedrucken als niedrige. Das bestätigt auch ein Vergleich eines Modells mit Prädiktor Topografie mit einem Modell in den beiden Prädiktoren Topografie und Formation. Dieser zeigt, wie die zusätzliche Information der Formation vor allem die Prognosen in Zonen guter Bedruckbarkeit verbessert, indem sie die von der Topografie allein als unbedruckt klassifizierte Stellen dort doch noch als erfolgreich bedruckt einschätzt. Außerdem läßt ein rein auf der Formation basierendes Modell erkennen, dass diese als alleiniger Prädiktor keine Erklärungskraft besitzt. Erst als zusätzliche Information zur Topografie kann sie eine Modellschätzung verbessern.

Außerdem wurde die Wechselwirkung der beiden besagten Prädiktoren sowie ihre quadratischen Einflüsse untersucht. Aus statistischer Sichtweise sind sowohl die quadratischen Effekte als auch die Wechselwirkung der beiden Prädiktoren signifikant, um das Auftreten von Fehlstellen zu beschreiben. Aus papiertechnischer Sicht jedoch liegt das Hauptaugenmerk für die Auswahl signifikanter Prädiktoren auf dem Grad der Übereinstimmung der Fehlstellen zwischen der beobachteten Situation und der diskretisierten Modellprognose. Und dieser in Abschnitt 6.2.2 gezeigte Vergleich macht deutlich, dass eine diskretisierte Modellprognose weder durch eine Wechselwirkung der beiden Prädiktoren noch durch ihre quadratischen Effekte deutlich verbessert werden kann.

7.2 Zusammenfassung des theoretischen Teils

Es hat sich jedoch herausgestellt, dass Binomialmodelle für die Modellierung einer erfolgreichen Farbübertragung nicht ausreichend sind. Proben besonders guter Druckqualität neigen dazu, eine geringere Variabilität der Daten aufzuweisen, also vom Modell angenommen wird (siehe Abschnitt 6.2.2). Proben von extrem schlechter Druckqualität hingegen neigen dazu, eine größere Variabilität der Daten aufzuweisen, als von einem Binomialmodell beschrieben werden kann. Dies zeigt sich in einer deutlich größeren beziehungsweise kleineren Devianz als der erwartete Wert von in etwa der Anzahl der Freiheitsgrade des Modells.

Um auf diese Variabilität größer beziehungsweise kleiner jener der Binomialverteilung reagieren zu können, wurden zwei Verteilungsfamilien betrachtet. Beiden gemein ist, dass es sich um Verallgemeinerungen der Binomialverteilung handelt. In beiden Fällen wird ein zweiter Parameter eingeführt, welcher eine größere oder kleinere Varianz als jene der Binomialverteilung ermöglicht.

Die erste dieser beiden Verteilungen wurde 1978 von Altham unter der Bezeichnung multiplikative Verallgemeinerung der Binomialverteilung vorgestellt. Diese beruht auf der Idee, Abhängigkeiten zwischen den einzelnen Beobachtungen innerhalb einer Gruppe zu berücksichtigen. Eine der ursprünglichen Arbeit von Altham übernommene Herleitung dieser Verteilungsfamilie ist in Kapitel 3 zu finden. Da gezeigt werden kann, dass die multiplikative Binomialverteilung ein Mitglied der zweiparametrischen Exponentialfamilie ist, werden im selben Kapitel auch geschlossene Formeln für Erwartungswert und Varianz der Verteilung angegeben. Außerdem wird die Verteilung in den Kontext der generalisierten linearen Modelle eingebettet. Dabei wird für die Wahrscheinlichkeit erfolgreicher Farbübertragung wie im logistischen Modell üblich ein logistischer Link verwendet, um den Zusammenhang zwischen der Erfolgswahrscheinlichkeit π und dem linearen Prädiktor $\eta = \mathbf{x}^t \boldsymbol{\beta}$ herzustellen. Für den nun neu dazukommenden zweiten Parameter wird ein loglinearer Link verwendet. Außerdem werden die ersten und zweiten partiellen Ableitungen aufgestellt, anhand derer sich eine Maximum Likelihood Schätzung für die Parameter durchführen lässt.

Eine zweite Verteilung, welche ebenfalls eine Variabilität größer respektive kleiner jener der Binomialverteilung zu beschreiben vermag, wurde in allgemeinerer Form 1986 von Efron unter der Bezeichnung doppelte Exponentialfamilie vorgestellt. Für diese Arbeit wurde speziell die doppelte Binomialverteilung in Betracht gezogen, welche in Kapitel 4 ausführlich besprochen wird. Da auch diese Verteilung ein Mitglied der zweiparametrischen Exponentialfamilie ist, können Erwartungswert und Varianz der Verteilung ebenfalls wieder recht einfach in geschlossener

Form dargestellt werden. Auch kann die doppelte Binomialverteilung in den Kontext der generalisierten linearen Modelle eingebettet werden, indem ein logistischer Link für die Relation zwischen Erfolgswahrscheinlichkeit π und linearem Prädiktor $\eta = \mathbf{x}^t \boldsymbol{\beta}$ verwendet wird. Für den zweiten Parameter wird auch hier ein loglinearer Link benutzt. Partielle Ableitungen erster und zweiter Ordnung werden aufgestellt, um eine Parameterschätzung mittels Maximum Likelihood Methode zu vereinfachen.

Für beide Verteilungsfamilien wurden Monte-Carlo Simulationen durchgeführt, um die Parameterschätzung mittels Maximum Likelihood Methode, basierend auf den partiellen Ableitungen der Verteilungen, zu testen. Der Auswertung dieser Simulationsergebnisse ist der größte Teil des Kapitels 5 gewidmet. Dort werden für einige ausgewählte Parameterkombinationen die Resultate der Simulationen ausführlich besprochen. Ein kleinerer Teil des Kapitels ist der Beschreibung der für die Modellschätzung notwendigen R-Funktionen gewidmet, welche anhand der bereits angesprochenen Monte-Carlo Simulationen getestet wurden.

Diese Simulationen haben gezeigt, dass eine sinnvolle Modellschätzung unter multiplikativer oder doppelter Binomialverteilungsannahme möglich ist. In einem weiteren Schritt wurden daher Modelle für erfolgreiche Farbübertragung, basierend auf multiplikativer sowie doppelter Binomialverteilungsannahme, geschätzt (siehe Abschnitt 6.3). Für diesen Bedruckbarkeitsdatensatz zeigen Vergleiche von Modellen unter den drei unterschiedlichen Verteilungsannahmen, dass vor allem Proben schlechter Druckqualität unter verallgemeinerter Binomialverteilungsannahme deutlich bessere AIC-Werte liefern. Außerdem zeigt ein Vergleich der geschätzten Erwartungswerte, wie der sowohl im Erwartungswert der doppelten als auch der multiplikativen Binomialverteilung vorhandene zusätzliche Parameter die Schätzung negativ beeinflusst. Für die Varianz jedoch zeigt sich, dass die beiden Verallgemeinerungen der Binomialverteilung bessere Schätzungen abzugeben vermögen, als es bei der klassischen Binomialverteilung der Fall ist. Aufgrund des zusätzlichen Parameters können die beiden Verallgemeinerungen bessere Ergebnisse liefern, welche näher an der beobachteten Situation liegen als im Fall einer klassischen Binomialverteilungsannahme.

7.3 Ausblick

Weiterführende Arbeiten auf diesem Gebiet wären zum einen eine Erstellung eines R-Pakets für die Funktionen zur Schätzung eines generalisierten linearen Modells unter verallgemeinerter Binomialverteilungsannahme.

Außerdem soll die geschilderte Modellschätzung auf weitere Probensätze aus dem Flexodruck angewandt werden. Weiters soll die Modellklasse auch zum Einsatz kommen, um erfolgreiche beziehungsweise fehlgeschlagene Farbübertragung im Tiefdruck zu modellieren. Auch wäre eine Aufnahme weiterer Papiereigenschaften in den linearen Prädiktor interessant.

R-Funktionen

Es folgt eine Besprechung aller für die vorliegende Arbeit geschriebenen R-Funktionen.

A.1 Log-Likelihoodfunktionen

`logl`

Beschreibung

Die log-Likelihoodfunktion einer verallgemeinerten Binomialverteilung.

Funktionsaufruf

`logl(beta, alpha, y, n, X, ResponseDistr)`

Argumente

- beta** Werte für die Parameter $\beta_0 \dots, \beta_{p-1}$ eines Modells, an denen die log-Likelihoodfunktion ausgewertet werden soll.
- alpha** Wert für den zusätzlichen Parameter α der verallgemeinerten Binomialverteilungen, an dem die log-Likelihoodfunktion ausgewertet werden soll. Im Fall einer klassischen Binomialverteilung wird für diesen Wert zwar eine Eingabe erwartet, diese aber nicht für die Berechnung verwendet.
- y** Die Realisationen der y des Modells. Vektorwertig.

- n** Die Anzahl an Versuchen pro Zelle. Kann falls alle Zellen denselben Umfang aufweisen ein Skalar sein, sonst muss **n** ein Vektor derselben Länge wie **y** sein.
- X** Die Designmatrix des betrachteten Modells der Dimension $\text{length}(\mathbf{y}) \times \text{length}(\mathbf{beta})$.
- ResponseDistr** Textstring der Verteilungsannahme. Zulässig sind die drei Verteilungen **binomial**, **multbin** und **doublebin**.

Details

Die Berechnung der log-Likelihoodfunktion an **(beta, alpha)** für beobachtete **y**. Zur Berechnung des Normierungsterms wird auf die Funktionen **multbinNormConst** sowie **doublebinNormConst** zurückgegriffen. Für die jeweiligen Formeln wurden die Darstellungen (3.18) sowie (4.6) verwendet, welche bereits ein logistisches Modell für die Erfolgswahrscheinlichkeiten $\boldsymbol{\pi}$ und ein einfaches log-lineares für den zusätzlichen Parameter ω beziehungsweise ϕ verwendet.

Rückgabewerte

Der retournierte Wert entspricht der log-Likelihoodfunktion an der Stelle **(beta, alpha)** für beobachtete **y**.

grad

Beschreibung

Partielle erste Ableitungen der log-Likelihoodfunktion nach β_k für $k = 0, \dots, p - 1$ sowie nach α .

Funktionsaufruf

`grad(beta, alpha, y, n, X, ResponseDistr)`

Argumente

- beta** Werte für die Parameter $\beta_0, \dots, \beta_{p-1}$ eines Modells, an denen die log-Likelihoodfunktion ausgewertet werden soll.
- alpha** Wert für den zusätzlichen Parameter α der verallgemeinerten Binomialverteilungen, an dem die log-Likelihoodfunktion ausgewertet werden soll. Im Fall einer klassischen Binomialverteilung wird für diesen Wert zwar eine Eingabe erwartet, diese aber nicht für die Berechnung verwendet.

- y** Die Realisationen der Responsevariablen y des Modells. Vektorwertig.
- n** Die Anzahl an Versuchen pro Zelle. Kann falls alle Zellen denselben Umfang aufweisen ein Skalar sein, sonst muss **n** ein Vektor derselben Länge wie **y** sein.
- X** Die Designmatrix des betrachteten Modells der Dimension $\text{length}(\mathbf{y}) \times \text{length}(\mathbf{beta})$.
- ResponseDistr** Textstring der Verteilungsannahme. Zulässig sind die drei Verteilungen **binomial**, **multbin** und **doublebin**.

Rückgabewerte

Die Funktion gibt den Vektor der partiellen Ableitungen der log-Likelihoodfunktion zurück, besitzt also für verallgemeinerte Binomialverteilungen die Länge $p + 1$.

hess

Beschreibung

Hessematrix der verallgemeinerten Binomialverteilungen.

Funktionsaufruf

`hess(beta, alpha, y, n, X, ResponseDistr)`

Argumente

- beta** Werte für die Parameter $\beta_0 \dots, \beta_{p-1}$ eines Modells, an denen die log-Likelihoodfunktion ausgewertet werden soll.
- alpha** Wert für den zusätzlichen Parameter α der verallgemeinerten Binomialverteilungen, an dem die log-Likelihoodfunktion ausgewertet werden soll. Im Fall einer klassischen Binomialverteilung wird für diesen Wert zwar eine Eingabe erwartet, diese aber nicht für die Berechnung verwendet.
- y** Die Realisationen der Responsevariablen y des Modells. Vektorwertig.
- n** Die Anzahl an Versuchen pro Zelle. Kann falls alle Zellen denselben Umfang aufweisen ein Skalar sein, sonst muss **n** ein Vektor derselben Länge wie **y** sein.
- X** Die Designmatrix des betrachteten Modells der Dimension $\text{length}(\mathbf{y}) \times \text{length}(\mathbf{beta})$.
- ResponseDistr** Textstring der Verteilungsannahme. Zulässig sind die drei Verteilungen **binomial**, **multbin** und **doublebin**.

Details

Die zur Berechnung dieser Hessematrix verwendeten Formeln wurden in den Abschnitten 3.4 und 4.3 vorgestellt. Die jeweiligen Einträge der Hessematrix entsprechen folgenden Ableitungen:

$$H = \begin{pmatrix} \frac{\partial^2 l(\boldsymbol{\beta}, \alpha | \mathbf{y})}{\partial \beta_0 \partial \beta_0} & \cdots & \frac{\partial^2 l(\boldsymbol{\beta}, \alpha | \mathbf{y})}{\partial \beta_0 \partial \beta_{p-1}} & \frac{\partial^2 l(\boldsymbol{\beta}, \alpha | \mathbf{y})}{\partial \beta_0 \partial \alpha} \\ \vdots & \ddots & \vdots & \vdots \\ \frac{\partial^2 l(\boldsymbol{\beta}, \alpha | \mathbf{y})}{\partial \beta_{p-1} \partial \beta_0} & \cdots & \frac{\partial^2 l(\boldsymbol{\beta}, \alpha | \mathbf{y})}{\partial \beta_{p-1} \partial \beta_{p-1}} & \frac{\partial^2 l(\boldsymbol{\beta}, \alpha | \mathbf{y})}{\partial \beta_{p-1} \partial \alpha} \\ \frac{\partial^2 l(\boldsymbol{\beta}, \alpha | \mathbf{y})}{\partial \alpha \partial \beta_0} & \cdots & \frac{\partial^2 l(\boldsymbol{\beta}, \alpha | \mathbf{y})}{\partial \alpha \partial \beta_{p-1}} & \frac{\partial^2 l(\boldsymbol{\beta}, \alpha | \mathbf{y})}{\partial \alpha \partial \alpha} \end{pmatrix}$$

Rückgabewerte

Hessematrix der doppelten oder multiplikativen Binomialverteilung.

`neg.logl`, `neg.grad`

Beschreibung

Die negative log-Likelihoodfunktion sowie die negative Gradientenfunktion einer verallgemeinerten Binomialverteilung für die Maximum Likelihood Schätzung der Modellparameter.

Funktionsaufruf

`neg.logl(par, y, n, X, ResponseDistr)`
`neg.grad(par, y, n, X, ResponseDistr)`

Argumente

- par** Parametervektor $(\beta_0, \dots, \beta_{p-1}, \alpha)$ der log-Likelihoodfunktion. Im Fall der klassischen Binomialverteilung wird für α zwar eine Eingabe erwartet, diese aber für keine weiteren Berechnungen verwendet.
- y** Die Realisationen der Responsevariablen y des Modells. Vektorwertig.
- n** Die Anzahl an Versuchen pro Zelle. Kann falls alle Zellen denselben Umfang aufweisen ein Skalar sein, sonst muss **n** ein Vektor derselben Länge wie **y** sein.
- X** Die Designmatrix des betrachteten Modells der Dimension `length(y)` \times `length(beta)`.

ResponseDistr Textstring der Verteilungsannahme. Zulässig sind die drei Verteilungen **binomial**, **multbin** und **doublebin**.

Details

Die negative log-Likelihoodfunktion mit einer gemeinsamen Variable für die Modellparameter (**beta,alpha**) der verallgemeinerten Binomialverteilung. Existiert, damit ein Aufruf von **optim()** die log-Likelihoodfunktion sowohl nach **beta** als auch nach **alpha** maximiert beziehungsweise diese negative log-Likelihoodfunktion minimiert. Genauere Informationen dazu sind in Abschnitt 5.5 nachzulesen.

Rückgabewerte

Der retournierte Wert entspricht der negativen log-Likelihoodfunktion beziehungsweise den negativen partiellen Ableitungen der log-Likelihoodfunktion an (**beta, alpha**) für beobachtete **y**.

A.2 Modellierungsfunktionen

Die für eine Modellschätzung notwendigen Funktionen wurden bereits in Kapitel 5 besprochen. Hier sollen nur nochmals ihre Aufrufe aufgelistet werden. Für genauere Informationen sei auf Abschnitt 5.5 verwiesen.

gbmod

Beschreibung

Modellschätzung unter verallgemeinerter Binomialverteilungsannahme,

Funktionsaufruf

```
gbmod(formula, ResponseDistr, trace=FALSE, ...)
```

Argumente

formula Formel des Modells in der für R üblichen Notation.

ResponseDistr Textstring der Verteilungsannahme. Zulässig sind die drei Verteilungen **binomial**, **multbin** und **doublebin**.

trace Schalter für einigen Modellierungoutput.

Details

Schätzt mittels Maximum Likelihood Schätzung die Parameter eines Modells $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ und α , wenn für die Responsevariable eine verallgemeinerte Binomialverteilung angenommen wird. Für die Parameterschätzung werden die negative log-Likelihoodfunktion sowie die Gradienten der jeweiligen Verteilungsfamilie verwendet. Sie wird mittels der aus dem *stats* Paket stammenden R-Funktion `optim()` realisiert.

Rückgabewerte

`coefficients` Maximum Likelihood Schätzer von $\boldsymbol{\beta}$
`theta` Maximum Likelihood Schätzer von $\exp(\alpha)$
`pearson.residuals` Vektor der Pearson Residuen
`fitted.values` Vektor der geschätzten Wahrscheinlichkeiten $\hat{\pi}_i$
`family` Verteilungsannahme (multiplikativ oder doppelt binomial)
`linear.predictor` Vektor der linearen Prädiktoren $\hat{\eta}_i$
`aic` AIC des Modells
`iter` Anzahl an Iterationen bis zur erzielten Konvergenz
`df.residual` Freiheitsgrade des betrachteten Modells
`df.null` Freiheitsgrade des Nullmodells
`y` Vektor der beobachteten Erfolge
`n` Vektor der Anzahl an Beobachtungen pro Zelle
`X` Designmatrix
`converged` Logisch; 1 wenn Modellschätzung konvergierte, 0 sonst
`formula` Die Modellformel in R-Notation, wobei die Responsevariable aus dem Vektor der Erfolge und der Misserfolge einer jeden Gruppe besteht.
`glm.summary` Ergebnisse der Schätzung unter dem binomialen Modell:
 `coefficients` Koeffizienten
 `pearson.residuals` Pearson Residuen
 `fitted.values` geschätzte Wahrscheinlichkeiten $\hat{\pi}_i$
 `null.deviance` Nulldevianz
 `deviance` Modelldevianz
 `df.residual` Freiheitsgrade
 `aic` Akaike Informationskriterium
`hessian` Hessematrix in den Punktschätzern $(\hat{\boldsymbol{\beta}}, \hat{\alpha})$

`gbsummary`

Beschreibung

Erstellt die Modellzusammenfassung eines mit **gbmod** geschätzten Modells mit verallgemeinert binomialverteilten Responsevariablen.

Funktionsaufruf

```
gbsummary(mod)
```

Argumente

mod Ein mittels **gbmod** geschätztes Modell.

Details

Der erstellte Output entspricht Großteils jenem von **summary.lm** oder **summary.glm**. Dieser wird im Detail in Abschnitt 5.5 besprochen.

gbplot

Beschreibung

Vier Plots, die eine Übersicht über das geschätzte Modell geben sollen. Es sind dies ein Plot der Pearson Residuen gegen die geschätzten Erwartungswerte, ein QQ-Plot der Pearson Residuen sowie ein Vergleich der geschätzten Erwartungswerte und der geschätzten Standardabweichungen unter einem verallgemeinert binomialverteilten Modell gegen jene eines klassischen Binomialmodells.

Funktionsaufruf

```
gbplot(mod, full=FALSE)
```

Argumente

mod Ein mittels **gbmod** geschätztes Modell.
full Schalter, um zusätzliche Plots ausgeben zu lassen.

Details

Angelehnt an die R-Funktion **plot.lm** erstellt diese Funktion die vier genannten Plots der geschätzten Erwartungswerte sowie der Pearson Residuen. Sie sind in Abschnitt 5.5 gezeigt.

A.3 Normierungsterme

`multbinNormConst`, `doublebinNormConst`

Beschreibung

Berechnet die einzelnen Summanden des Normierungsausdrucks der doppelten sowie der multiplikativen Binomialverteilung.

Funktionsaufruf

```
multbinNormConst(n, eta, alpha)
multbinNormConst(k, n, p, omega)
doublebinNormConst(n, eta, alpha)
doublebinNormConst(k, n, p, phi)
```

Argumente

n	Anzahl an Versuchen. Vektor oder Skalar bei Aufruf mit 3 Argumenten, Skalar bei einem Aufruf mit 4 Argumenten.
eta	Linearer Prädiktor $\eta = \mathbf{x}^t \boldsymbol{\beta}$ unter dem betrachteten Modell.
alpha	$\alpha = \log(\omega)$, der zusätzliche Parameter der multiplikativen Binomialverteilung unter einem log-linearen Link. Skalar.
k	Anzahl an Erfolgen bei n Versuchen. Skalar.
p	Erfolgswahrscheinlichkeit. Skalar.
omega	Wert des zweiten Parameters der multiplikativen Binomialverteilung.
phi	Wert des zweiten Parameters der doppelten Binomialverteilung.

Details

Bei drei Argumenten werden mit dieser Funktion die einzelnen Summanden des Normierungsausdrucks errechnet. Dabei wird die in Abschnitt 3 vorgestellte Darstellung (3.18) beziehungsweise Darstellung (4.6) aus Kapitel 4 verwendet, welche bereits im linearen Prädiktor η sowie α vorliegt. Liegen mehrere Beobachtungen y_i vor, kann ein vektorwertiges **eta** übergeben werden. Sollte die Anzahl n der Versuche pro Zelle variieren, so kann auch **n** vektorwertig sein, muss dann jedoch dieselbe Länge wie **eta** aufweisen. Der zusätzliche Parameter $\alpha = \log(\omega)$ der multiplikativen Binomialverteilung unter dem log-linearen Link muss jedoch für alle Zellen gleich sein.

Werden vier Argumente übergeben, so wird die Summe der ersten $k+1$ Summanden des Normierungsausdrucks für skalarwertige **y**, **n**, **p** und **omega** ermittelt.

Rückgabewerte

normConst Im Fall eines Aufrufs mit den drei Argumenten **n**, **eta** und **alpha** eine Matrix der Dimension $\text{length}(\text{eta}) \times \text{max}(\mathbf{n})$. In Zeile i sind die Summanden des Normierungsausdrucks für η_i zu finden.

Im Fall eines Aufrufs mit vier Argumenten wird die Summe der ersten $k + 1$ Summanden des Normierungsausdrucks berechnet.

gbmean, **gbvariance**, **gbsd**

Beschreibung

Momente doppelter oder multiplikativ binomialverteilter Zufallsvariablen.

Funktionsaufruf

```
gbmean(n, p, theta, ResponseDistr)
gbvariance(n, p, theta, ResponseDistr)
gbsd(n, p, theta, ResponseDistr)
```

Argumente

- n** Anzahl an Versuchen. Skalar oder Vektor derselben Länge wie **p**.
- p** Erfolgswahrscheinlichkeit. Skalar oder Vektor.
- theta** Wert für den zweiten Parameter der verallgemeinerten Binomialverteilungen.
- ResponseDistr** Textstring der Verteilungsannahme. Zulässig sind die drei Verteilungen **binomial**, **multbin** und **doublebin**.

Details

Berechnet mit den in den Kapiteln 3 und 4 vorgestellten Formeln (3.11) und (3.12) sowie (4.4) und (4.5) für die ersten beiden Momente Erwartungswert und Varianz in π_i und ω beziehungsweise ϕ .

Rückgabewerte

Ein Vektor der berechneten Erwartungswerte, Varianzen oder Standardabweichungen der multiplikativen oder doppelten Binomialverteilung.

`dgenbin`, `pgenbin`, `qgenbin`, `rgenbin`

Beschreibung

Wahrscheinlichkeitsfunktion, Verteilungsfunktion, Quantilsfunktion und Zufallszahlengenerator der verallgemeinerten Binomialverteilungen.

Funktionsaufruf

```
dgenbin(y, n, p, theta, ResponseDistr)
pgenbin(q, n, p, theta, ResponseDistr)
qgenbin(r, n, p, theta, ResponseDistr)
rgenbin(cnt, n, p, theta, ResponseDistr)
```

Argumente

y, q Quantile. Vektor.
r Wahrscheinlichkeiten. Vektor.
cnt Anzahl an zu erzeugenden Zufallsvariablen.
n Anzahl an Versuchen. Skalar oder Vektor.
p Vektor der Erfolgswahrscheinlichkeiten.
theta Wert für den zweiten Parameter der verallgemeinerten Binomialverteilungen.
ResponseDistr Textstring der Verteilungsannahme. Zulässig sind die drei Verteilungen **binomial**, **multbin** und **doublebin**.

Details

Berechnet mit den in den Kapiteln 3 und 4 vorgestellten Formeln (3.11), (3.12), (4.4) und (4.5) für die ersten beiden Momente Erwartungswert und Varianz in π_i und ω beziehungsweise ϕ .

Rückgabewerte

pgenbin berechnet die Wahrscheinlichkeit $P(Y = y)$, **pgenbin** die Verteilungsfunktion in q , **qgenbin** das Quantil und **rgenbin** erzeugt multiplikativ beziehungsweise doppelt binomialverteilte Zufallsvariablen.

A.4 Funktionen für den Simulationsteil

`example1`, `example2`

Beschreibung

Monte-Carlo Simulation der Parameterschätzung verallgemeinert binomialverteilter Stichproben.

Funktionsaufruf

```
example1(beta, theta, ResponseDistr)
example2(beta, theta, ResponseDistr)
```

Argumente

beta Werte für die Parameter β_0 im Fall eines Intercept-Only Modells (**example1**) oder für (β_0, β_1) im Fall eines Modells mit einem Prädiktor x (**example2**).

theta Wert für den zweiten Parameter der verallgemeinerten Binomialverteilungen.

ResponseDistr Textstring der Verteilungsannahme. Zulässig sind die drei Verteilungen **binomial**, **multbin** und **doublebin**.

Details

Rahmen für die Monte-Carlo Schätzung der Parameter β und α . Hier wird die Anzahl an Wiederholungen ($M = 500$) und die Größe der Stichproben ($N = 1000$ bei **example1** und $N = 6010$ bei **example2**) festgelegt.

Bei **example2** erfolgt hier die Erzeugung der erklärenden Größe x als regelmäßige Sequenz von Werten aus dem Intervall $[-3; 3]$.

Die eigentliche Monte-Carlo Schätzung passiert in **estimate.M.times**.

Die von **estimate.M.times** zurückgegebene Tabelle wird danach noch im `.dat`-Format abgespeichert.

`estimate.M.times`

Beschreibung

Monte-Carlo Simulation der Parameterschätzung verallgemeinert binomialverteilter Stichproben.

Funktionsaufruf

```
estimate.M.times(M, N, n, p, theta, ResponseDistr, X)
```

Argumente

- M** Anzahl an Wiederholungen der Parameterschätzung.
- N** Anzahl an Zufallsvariablen pro Durchlauf.
- n** Anzahl an Versuchen. Skalar oder Vektor derselben Länge wie **p**.
- p** Erfolgswahrscheinlichkeiten der **N** Zufallsvariablen.
- theta** Wert für den zweiten Parameter der verallgemeinerten Binomialverteilungen.
- ResponseDistr** Textstring der Verteilungsannahme. Zulässig sind die drei Verteilungen **binomial**, **multbin** und **doublebin**.
- X** Designmatrix.

Details

Hier erfolgt die eigentliche Monte-Carlo Simulation. Für die Parameter **p** und **theta** werden **M** mal **N** multiplikativ oder doppelt binomialverteilte Zufallsvariablen mittels **rgenbin** erzeugt. Für diese werden dann unter Verwendung von **gmod** die Parameterschätzer $\hat{\beta}$ und $\hat{\alpha}$ geschätzt.

Rückgabewerte

Die Routine erstellt eine Tabelle mit **M** Zeilen, in denen die wichtigsten Resultate der Parameterschätzungen gesammelt werden.

Die Einträge entsprechen der Reihe nach:

- beta** Punktschätzer für β .
- alpha** Punktschätzer für α .
- theta** Punktschätzer für $\theta = \exp \alpha$.
- empir.mean** Mittelwerte der einzelnen Stichproben.
- empir.sd** Standardabweichungen der einzelnen Stichproben.
- logl** Maximierte log-Likelihoodfunktionswerte.
- no.of.iteration** Anzahl an Iterationsschritten von **optim**.

Literaturverzeichnis

- Agresti, A. (2002). *Categorical Data Analysis*. New York: Wiley-interscience.
- Aitkin, M. & Clayton, D. (1980). The fitting of exponential, Weibull and extreme value distributions to complex censored survival data using GLIM. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 29, 156-163.
- Aitkin, M., Francis, B. & Hinde, J. (2005). *Statistical Modelling in GLIM4* (2. Aufl.). Oxford: Oxford University Press.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. Petrov & F. Csàki (Hrsg.), *Proceedings of the second international symposium on information theory* (S. 267-281). Budapest.
- Altham, P. (1978). Two generalizations of the binomial distribution. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 27, 162-197.
- Annis, D. (2007). A note on quasi-likelihood for exponential families. *Statistics & Probability Letters*, 77, 431-437.
- Atwill, E., Mohammed, H., Scarlett, J. & McCulloch, C. (1995). Extending the interpretation and utility of mixed effects logistic regression models. *Preventive Veterinary Medicine*, 24, 187-201.
- Banks, J., Carson, J., Barry, L. & Nicol, D. (2009). *Discrete-Event System Simulation* (5. Aufl.). Upper Saddle River, New Jersey: Prentice Hall.
- Barros, G., Fahlerantz, C.-M. & Johansson, P.-Å. (2005). Topographic distribution of uncovered areas (UCA) in full tone flexographic prints. *TAGA Journal*, 2, 43 - 57.
- Barros, G. & Johansson, P.-Å. (2006). Prediction of uncovered area occurrence in flexography based on topography - a feasibility study. *Nordic Pulp and Paper Research Journal*, 21, 172 - 179.
- Beuth (Hrsg.). (2006). *DIN 6730:2006-05: Papier und Papp - Begriffe*. Berlin: Deutsches Institut für Normung e.V.
- Broyden, C. (1970). The convergence of a class of double-rank minimization algorithms. *Journal of the Institute of Mathematics and its Applications*, 6, 76-90.

- Burger, W. & Burge, M. (2006). *Digitale Bildverarbeitung*. Berlin: Springer.
- Casella, G. & Berger, R. (2002). *Statistical Inference*. Pacific Grove, California: Duxbury.
- Cox, D. (1972). The analysis of multivariate binary data. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 21, 113-120.
- Cox, F. (Hrsg.). (1993). *Modern Parasitology: A Textbook of Parasitology*. Oxford: Blackwell Science.
- Darroch, J. (1974). Multiplicative and additive interaction in contingency tables. *Biometrika*, 61, 207-214.
- Dauer, M., Feirer, V., Hirn, U., Trimmel, M. & Wind, E. (2011). *Modellierung der Farbübertragung aus lokalen Papiereigenschaften - Berichte 4. Forschungsjahr* (Bericht). Technische Universität Graz: Institut für Papier-, Zellstoff- und Fasertechnik.
- Dauer, M., Feirer, V., Hirn, U. & Wind, E. (2009). *Modellierung der Farbübertragung aus lokalen Papiereigenschaften - Berichte 2. Forschungsjahr* (Bericht). Technische Universität Graz: Institut für Papier-, Zellstoff- und Fasertechnik.
- Dauer, M., Feirer, V., Hirn, U. & Wind, E. (2010). *Modellierung der Farbübertragung aus lokalen Papiereigenschaften - Berichte 3. Forschungsjahr* (Bericht). Technische Universität Graz: Institut für Papier-, Zellstoff- und Fasertechnik.
- Davison, A. (2003). *Statistical Models*. Cambridge: Cambridge University Press.
- Dehaene, S. (2003). The neural basis of the Weber-Fechner law: a logarithmic mental number line. *TRENDS in Cognitive Science*, 7, 145-147.
- Dey, D., Gelfand, A. & Peng, F. (1997). Overdispersed generalized linear models. *Journal of Statistical Planning and Inference*, 64, 93-107.
- Dobson, A. & Barnett, A. (2008). *An Introduction to Generalized Linear Models* (3. Aufl.). Boca Raton, Florida: Chapman & Hall/CRC.
- Donoser, M., Bischof, H. & Wiltsche, M. (2006). Color blob segmentation by MSER analysis. In *2006 IEEE international conference on image processing*.
- Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, 81, 709-721.
- Engel, B. & te Brake, J. (1993). Analysis of embryonic development with a model for under- or overdispersion relative to binomial variation. *Biometrics*, 49, 269-279.
- Erhardt, A. (2008). *Einführung in die Digitale Bildverarbeitung*. Wiesbaden: Vieweg+Teubner / GWV Fachverlage GmbH.
- Ernst & Young. (2007). *Competitiveness of the european graphic industry - prospects for the EU printing sector to respond to its structural and technological challenges* (Branchenreport). European Communities.

- Feirer, V. (2008). *Statistische Modellierung des Auftretens von Fehlstellen im Flexodruck*. Diplomarbeit, Technische Universität Graz.
- Feirer, V., Hirn, U., Friedl, H. & Bauer, W. (2010a). Ink transmission in flexo printed sack paper - a statistical modelling approach. In *Proceedings PTS Symposium CP1010: „Papier und Bedruckbarkeit“*. München.
- Feirer, V., Hirn, U., Friedl, H. & Bauer, W. (2010b). Predicting local ink coverage in flexo printed packaging paper. In *Progress in Paper Physics Seminar*. Montreal.
- Feirer, V., Hirn, U., Friedl, H. & Bauer, W. (2012). A statistical approach to the modelling of ink transmission on flexo-printed sack paper. In *Advances in printing and media technology* (Bd. XXXVIII).
- Feirer, V., Hirn, U., Mettänen, M., Wind, E. & Wirleitner, U. (2008). *Modellierung der Farbübertragung aus lokalen Papiereigenschaften - Berichte 1. Forschungsjahr* (Bericht). Technische Universität Graz: Institut für Papier-, Zellstoff- und Fasertechnik.
- Fishman, G. (2001). *Discrete-Event Simulation: Modeling, Programming, and Analysis*. New York: Springer.
- Fletcher, R. (1970). A new approach to variable metric algorithms. *The Computer Journal*, 13, 317-322.
- Gelfand, A. & Dalal, S. (1990). A note on overdispersed exponential families. *Biometrika*, 77, 55-64.
- Goldfarb, D. (1970). A family of variable-metric methods derived by variational means. *Mathematics of Computation*, 24, 23-26.
- Gonzalez, R., Woods, R. & Eddins, S. (2004). *Digital Image Processing using Matlab*. Upper Saddle River, New Jersey: Pearson Prentice Hall.
- Gregersen, Ø., Jonsen, P. & Helle, T. (1995). Small-scale topographical variations of newsprint surfaces and their effects on printing ink transfer distribution. *Journal of Pulp and Paper Science*, 21, 331-336.
- Hardy, I. (Hrsg.). (2002). *Sex Ratios. Concepts and Research Methods*. Cambridge: Cambridge University Press.
- Harzheim, L. (2007). *Strukturoptimierung - Grundlagen und Anwendungen*. Frankfurt am Main: Wissenschaftlicher Verlag Harri Deutsch.
- Jankowski, P. (2010). *Druckgewerbe* (Branchenreport Nr. WZ 18.1). Deutscher Sparkassen- und Giroverband e.V.
- Jørgensen, B. (1987). Exponential dispersion models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 49, 127-162.
- Jørgensen, B. (1993). *The Theory of Linear Models*. London: Chapman & Hall.

- Keller, D. & Pawlak, J. (2001). β -radiographic imaging of paper formation using storage phosphor screens. *Journal of Pulp and Paper Science*, *27*, 117-123.
- King, G. (1989). Variance specification in event count models: From restrictive assumptions to a generalized estimator. *American Journal of Political Science*, *33*, 762-784.
- Kipphan, H. (2000). *Handbuch der Printmedien*. Berlin: Springer.
- Klaas, I., Enevoldsen, C., Ersbøll, A. & Tölle, U. (2005). Cow-related risk factors for milk leakage. *Journal of Dairy Science*, *88*, 128-136.
- Krackow, S. & Tkadlec, E. (2001). Analysis of brood sex ratios: implications of offspring clustering. *Behavioral Ecology and Sociobiology*, *50*, 293-301.
- Kreuter, J. (2009). Flexodruck wird noch schneller und besser. *Pack Aktuell*, *19/20*, 18. Verfügbar unter <http://www.packaktuell.ch>
- Kupper, L. & Haseman, J. (1978). The use of a correlated binomial model for the analysis of certain toxicological experiments. *Biometrics*, *34*, 69-76.
- Lee, Y. & Nelder, J. (2000). The relationship between double-exponential families and extended quasi-likelihood families, with application to modelling Geissler's human sex ratio data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *49*, 413-419.
- Lindsey, J. (o.J.). *rmutil: Utilities for nonlinear regression and repeated measurements*. R-Package. (erhältlich unter <http://www.commanster.eu/rcode.html>), zuletzt aufgerufen am 27. 7. 2011)
- Lindsey, J. (1996). *Parametric Statistical Inference*. Oxford: Clarendon Press.
- Lindsey, J. (1999). A review of some extensions to generalized linear models. *Statistics in Medicine*, *18*, 2223-2236.
- Lindsey, J. & Altham, P. (1998). Analysis of the human sex ratio by using overdispersion models. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *47*, 149-157.
- Lovison, G. (1998). An alternative representation of Altham's multiplicative-binomial distribution. *Statistics and Probability Letters*, *36*, 416-420.
- Molenberghs, G. & Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer.
- Moran, R. (2011). Modern-day demographics: What's printed flexo & what drives the business? *Flexo - The Flexography Technology Source*, *36*, 8-19.
- Nelder, J. & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, *7*, 308-313.
- Nelder, J. & Pregibon, D. (1987). An extended quasi-likelihood function. *Biometrika*, *74*, 221-232.

- Nelder, J. & Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 135, 370-384.
- Rawlings, J. O., Pantula, S. G. & Dickey, D. A. (1998). *Applied Regression Analysis - A Research Tool*. New York: Springer.
- Roeckelein, J. (1998). *Dictionary of Theories, Laws, and Concepts in Psychology*. Westport, Connecticut: Greenwood Press.
- Schüßler, H. (2008). *Digitale Signalverarbeitung 1 - Analyse diskreter Signale und Systeme* (5. Aufl.). Berlin: Springer.
- Seeber, G. (1997). Overdispersed exponential regression models. *Computational Statistics*, 12, 209-218.
- Shanno, D. (1970). Conditioning of quasi-Newton methods for function minimization. *Mathematics of Computation*, 24, 647-656.
- Verpackungsdruck* (Whitepaper). (2008). Heidelberg. Verfügbar unter <http://www.heidelberg.com>
- Wanske, M., Großmann, H. & Scherer, S. (2008). Messtechnische Bewertung von Tissue Produkten mit dem optischen Messsystem InfiniteFocus. *Wochenblatt für Papierfabrikation*, -, 473-477.
- Wedderburn, R. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, 61, 439-447.
- Witting, H. (1985). *Mathematische Statistik: Parametrische Verfahren bei festem Stichprobenumfang*. Stuttgart: Teubner.