

Technische Universität München

ZENTRUM MATHEMATIK

**Verallgemeinerte Poissonregression  
und daraus abgeleitete  
Zero-Inflated und Zero-Hurdle  
Regressionsmodelle**

Diplomarbeit

von

Daniel Stekeler

Themenstellerin: Prof. Dr. Claudia Czado  
Betreuer: Prof. Dr. Claudia Czado,  
Dipl. Math. Susanne Gschlößl  
Abgabetermin: 29. Oktober 2004

Hiermit erkläre ich, dass ich die Diplomarbeit selbstständig angefertigt und nur die angegebenen Quellen verwendet habe.

Garching, den 29. Oktober 2004

# Danksagung

Zuallererst gilt mein Dank Frau Prof. Claudia Czado für die zahlreichen Anregungen im Laufe der Erstellung meiner Diplomarbeit. Desweiteren möchte ich mich bei Herrn Aleksey Min für die gewissenhafte Überprüfung der Aussagen im theoretischen Teil dieser Arbeit bedanken. Ebenso danke ich Frau Susanne Gschlößl für ihre Unterstützung bei der Anfertigung der Simulationsstudie und der Datenauswertung in Kapitel 9.



# Inhaltsverzeichnis

<b>Einleitung</b>	<b>xi</b>
<b>1 Die verallgemeinerte Poissonverteilung</b>	<b>1</b>
1.1 Definition und grundlegende Eigenschaften . . . . .	1
1.2 Verhalten bei Parametervariation . . . . .	2
1.3 Erwartungswert und Varianz . . . . .	5
1.3.1 Erwartungswert . . . . .	5
1.3.2 Erzeugende Funktion . . . . .	6
1.3.3 Momenterzeugende Funktion . . . . .	7
1.3.4 Momente . . . . .	8
1.3.5 Varianz . . . . .	12
1.3.6 Dispersionsparameter . . . . .	13
1.4 Faltungseigenschaft . . . . .	13
1.5 Modelle die zur verallgemeinerten Poissonverteilung führen . . . . .	15
1.5.1 Der verallgemeinerte Poissonprozess . . . . .	15
1.5.2 Galton-Watson-Bienymé-Verzweigungsprozess . . . . .	18
1.6 Grenzverteilungen . . . . .	22
1.6.1 Normalverteilung . . . . .	22
1.6.2 Inverse Gaußverteilung . . . . .	25
1.7 Erzeugung von Zufallsvariablen . . . . .	26
1.7.1 Invertierungsmethode . . . . .	27
1.7.2 Verzweigungsmethode . . . . .	28
1.7.3 Approximationsmethode . . . . .	29
<b>2 Vergleich von verallgemeinerter Poisson- und negativer Binomialverteilung</b>	<b>31</b>
2.1 Die negative Binomialverteilung . . . . .	31
2.1.1 Erwartungswert und Varianz . . . . .	31
2.1.2 Faltungseigenschaft . . . . .	33
2.1.3 Negative Binomialverteilung mit stetigen Parametern . . . . .	34
2.1.4 Grenzverteilung . . . . .	35
2.1.5 Gamma-Poisson-Mischmodell . . . . .	36
2.2 Vergleich der verallgemeinerten Poisson- und negativen Binomialverteilung bei Übereinstimmung von Erwartungswert und Varianz . . . . .	37
<b>3 Einführung in die Theorie der Regressionsmodelle</b>	<b>43</b>
3.1 Lineare Modelle . . . . .	44
3.2 Verallgemeinerte lineare Modelle . . . . .	45
3.3 Maximum-Likelihood-Schätzung . . . . .	46

3.3.1	Maximum-Likelihood-Schätzer . . . . .	47
3.3.2	Eigenschaften von Maximum-Likelihood-Schätzern . . . . .	49
3.3.3	Die Likelihoodfunktion in verallgemeinerten linearen Modellen . . . . .	49
3.4	Messung der Anpassungsqualität . . . . .	50
3.4.1	Devianzresiduen . . . . .	52
3.4.2	Akaike-Informationskriterium . . . . .	53
<b>4</b>	<b>Modellierung von Zähldaten</b>	<b>55</b>
4.1	Das Poisson-Regressionsmodell . . . . .	55
4.2	Abweichende Länge der Beobachtungszeiträume . . . . .	56
4.2.1	Offset in der Poissonregression . . . . .	56
4.2.2	Modellbeschreibung einer Poissonregression mit Offset . . . . .	57
4.2.3	Natürliche Linkfunktion . . . . .	58
4.3	Ursachen von Misspezifikation . . . . .	59
4.3.1	Verknüpfungskomponente . . . . .	59
4.3.2	Unbeobachtete Heterogenität . . . . .	60
4.3.3	Nullenüberschuss . . . . .	61
4.3.4	Messfehler in den Kovariablen . . . . .	62
4.3.5	Abhängigkeiten im Erzeugungsprozess der Daten . . . . .	63
4.3.6	Selektivität . . . . .	63
4.3.7	Underreporting . . . . .	63
4.4	Behandlung von Überdispersion in Zähldaten . . . . .	64
4.4.1	Beibehaltung der Poisson-Verteilungsannahme . . . . .	64
4.4.2	Abänderung der Verteilungsannahme . . . . .	64
<b>5</b>	<b>Das verallgemeinerte Poisson-Regressionsmodell</b>	<b>65</b>
5.1	Modellbeschreibung . . . . .	65
5.2	Responseverteilung . . . . .	66
5.2.1	Die eingeschränkte verallgemeinerte Poissonverteilung . . . . .	67
5.2.2	Umparametrisierung der verallgemeinerten Poissonverteilung . . . . .	68
5.3	Maximum-Likelihood-Schätzung . . . . .	71
5.3.1	Likelihoodfunktion . . . . .	71
5.3.2	Scorefunktion . . . . .	76
5.3.3	Fisher-Information . . . . .	80
5.4	Offset in der verallgemeinerten Poissonregression . . . . .	82
5.5	Schätzprozeduren . . . . .	84
5.5.1	Dispersionsparameter als Störparameter . . . . .	84
5.5.2	Momentenschätzer für $\varphi$ . . . . .	84
5.5.3	Zweistufige Schätzprozedur . . . . .	85
5.5.4	Simultane Schätzung von $\varphi$ und $\beta$ . . . . .	86
5.6	Messung der Anpassungsqualität . . . . .	87
5.6.1	Unskalierte Devianzresiduen . . . . .	87
5.6.2	Akaike-Informationskriterium . . . . .	88
<b>6</b>	<b>Zero-Inflated-Regressionsmodelle</b>	<b>89</b>
6.1	Zero-Inflated-Verteilungen . . . . .	89
6.2	Zero-Inflated-Poissonregression (ZIP-Regression) . . . . .	94
6.2.1	Zero-Inflated-Poissonverteilung (ZIP-Verteilung) . . . . .	94

6.2.2	Spezifikation eines ZIP-Regressionsmodells mit individuellen Auswahlparametern . . . . .	95
6.2.3	Die Likelihoodfunktion im ZIP-Regressionsmodell mit individuellen Auswahlparametern . . . . .	96
6.3	Verallgemeinerte Zero-Inflated-Poissonregression mit gemeinsamem Auswahlparameter (ZIGP $\mu$ -Regression) . . . . .	98
6.3.1	Verallgemeinerte Zero-Inflated-Poissonverteilung (ZIGP-Verteilung) . . . . .	98
6.3.2	Vergleich von verallgemeinerter Poisson- und ZIGP-Verteilung . . . . .	100
6.3.3	Spezifikation des ZIGP-Regressionsmodells mit gemeinsamem Auswahlparameter . . . . .	101
6.3.4	Die Likelihoodfunktion im ZIGP-Regressionsmodell mit gemeinsamem Auswahlparameter . . . . .	101
6.3.5	Messung der Anpassungsqualität des ZIGP-Regressionsmodells mit gemeinsamem Auswahlparameter . . . . .	103
6.4	Verallgemeinerte Zero-Inflated-Poissonregression mit individuellen Auswahlparametern (ZIGP $\mu\omega$ -Regression) . . . . .	103
6.5	Verallgemeinerte Zero-Inflated-Poissonregression mit individuellen Auswahlparametern aber konstantem Parameter $\mu$ (ZIGP $\omega$ -Regression) . . . . .	105
<b>7</b>	<b>Hürden-Regressionsmodelle</b>	<b>109</b>
7.1	Kombinierte Verteilungen . . . . .	109
7.2	Kombinierte Regressionsmodelle mit Hürde Null . . . . .	109
7.2.1	Verteilung in Null-Hürdenmodellen . . . . .	109
7.2.2	Dispersion und Null-Hürdenmodelle . . . . .	111
7.2.3	Maximum-Likelihood-Schätzung in Null-Hürdenmodellen . . . . .	112
7.3	Zero-Hurdle-Poissonregression (ZHP-Regression) . . . . .	113
7.3.1	Zero-Hurdle-Poissonverteilung (ZHP-Verteilung) . . . . .	113
7.3.2	Spezifikation des ZHP-Regressionsmodells . . . . .	114
7.3.3	Maximum-Likelihood-Schätzung im ZHP-Regressionsmodell . . . . .	115
7.4	Verallgemeinerte Zero-Hurdle-Poissonregression (ZHGP-Regression) . . . . .	116
7.4.1	Verallgemeinerte Zero-Hurdle-Poissonverteilung (ZHGP-Verteilung) . . . . .	116
7.4.2	Vergleich von verallgemeinerter Poisson- und ZHGP-Verteilung . . . . .	118
7.4.3	Spezifikation des ZHGP-Regressionsmodells . . . . .	118
7.4.4	Maximum-Likelihood-Schätzung im ZHGP-Regressionsmodell . . . . .	120
7.4.5	Messung der Anpassungsqualität des ZHGP-Regressionsmodells . . . . .	121
<b>8</b>	<b>Simulationsstudie zu den Maximum-Likelihood-Schätzern</b>	<b>123</b>
8.1	Eigenschaften von Schätzern . . . . .	123
8.2	Verallgemeinerte Poissonregression . . . . .	127
8.2.1	Datensimulation . . . . .	127
8.2.2	Güte der Maximum-Likelihood-Schätzer . . . . .	129
8.2.3	Asymptotische Verteilung der Maximum-Likelihood-Schätzer . . . . .	130
8.3	ZIGP $\mu$ -Regression . . . . .	137
8.3.1	Datensimulation . . . . .	137
8.3.2	Güte der Maximum-Likelihood-Schätzer . . . . .	139
8.4	ZIGP $\mu\omega$ -Regression . . . . .	140
8.4.1	Datensimulation . . . . .	140

8.4.2	Güte der Maximum-Likelihood-Schätzer . . . . .	143
8.5	ZIGP $\omega$ -Regression . . . . .	147
8.5.1	Datensimulation . . . . .	147
8.5.2	Güte der Maximum-Likelihood-Schätzer . . . . .	147
8.6	ZHGP-Regression . . . . .	150
8.6.1	Datensimulation . . . . .	150
8.6.2	Güte der Maximum-Likelihood-Schätzer . . . . .	151
8.7	Missspezifikation . . . . .	152
<b>9</b>	<b>Anwendungsbeispiele</b>	<b>155</b>
9.1	Patent-Datensatz . . . . .	155
9.1.1	Datenbeschreibung . . . . .	155
9.1.2	Explorative Datenanalyse . . . . .	155
9.1.3	Poissonregression . . . . .	156
9.1.4	Verallgemeinerte Poissonregression . . . . .	159
9.1.5	ZIGP $\mu$ -Regression . . . . .	160
9.1.6	ZIP-Regression . . . . .	162
9.1.7	ZIGP $\mu\omega$ -Regression . . . . .	163
9.1.8	ZIGP $\omega$ -Regression . . . . .	164
9.1.9	ZHGP-Regression . . . . .	165
9.1.10	ZHP-Regression . . . . .	166
9.1.11	Zusammenfassung . . . . .	167
9.2	KFZ-Datensatz . . . . .	169
9.2.1	Datenbeschreibung . . . . .	169
9.2.2	Poissonregression . . . . .	170
9.2.3	Verallgemeinerte Poissonregression . . . . .	171
9.2.4	ZIGP $\mu$ -Regression . . . . .	171
9.2.5	ZIP-Regression . . . . .	173
9.2.6	ZIGP $\mu\omega$ -Regression . . . . .	174
9.2.7	ZIGP $\omega$ -Regression . . . . .	175
9.2.8	ZHGP-Regression . . . . .	176
9.2.9	ZHP-Regression . . . . .	177
9.2.10	Zusammenfassung . . . . .	178
<b>10</b>	<b>Zusammenfassung</b>	<b>179</b>
<b>A</b>	<b>Grundlegende Verteilungen und Verteilungsfamilien</b>	<b>181</b>
A.1	Familie der exponentiellen Verteilungen . . . . .	181
A.2	Poissonverteilung . . . . .	181
A.3	Gammaverteilung . . . . .	183
<b>B</b>	<b>Stochastische Prozesse</b>	<b>185</b>
B.1	Poissonprozess . . . . .	186
<b>C</b>	<b>S-Plus Funktionen</b>	<b>189</b>
C.1	Funktionen für die verallgemeinerte Poissonverteilung . . . . .	189
C.1.1	Wahrscheinlichkeitsfunktion <code>dgpois</code> . . . . .	189
C.1.2	Verteilungsfunktion <code>pgpois</code> . . . . .	190
C.1.3	Quantilsfunktion <code>qgpois</code> . . . . .	191



---

C.1.4	Erzeugung von Zufallsvariablen <code>rgpois</code> . . . . .	192
C.2	Funktionen für die verallgemeinerte Poissonregression . . . . .	193
C.2.1	Erzeugung verallgemeinert poissonverteilter Regressionsdaten . . . . .	193
C.2.2	Maximum-Likelihood-Schätzung im verallgemeinerten Poisson-Regressionsmodell . . . . .	194
C.3	Funktionen für die ZIGP-Regression . . . . .	204
C.3.1	Erzeugung von ZIGP $\mu$ -Regressionsdaten . . . . .	204
C.3.2	Maximum-Likelihood-Schätzung im ZIGP $\mu$ -Regressionsmodell . . . . .	205
C.3.3	Erzeugung von ZIGP $\mu\omega$ -Regressionsdaten . . . . .	211
C.3.4	Maximum-Likelihood-Schätzung im ZIGP $\mu\omega$ -Regressionsmodell . . . . .	212
C.3.5	Erzeugung von ZIGP $\omega$ -Regressionsdaten . . . . .	219
C.3.6	Maximum-Likelihood-Schätzung im ZIGP $\omega$ -Regressionsmodell . . . . .	220
C.4	Funktionen für die ZHGP-Regression . . . . .	226
C.4.1	Erzeugung von ZHGP-Regressionsdaten . . . . .	226
C.4.2	Maximum-Likelihood-Schätzung im ZHGP-Regressionsmodell . . . . .	227
<b>D</b>	<b>Überblick über die verwendeten Regressionsmodelle</b>	<b>235</b>
<b>E</b>	<b>Abkürzungen</b>	<b>237</b>
E.1	Abkürzungen der verwendeten Verteilungsmodelle . . . . .	237
E.2	Abkürzungen der verwendeten Regressionsmodelle . . . . .	237
	<b>Abbildungsverzeichnis</b>	<b>239</b>
	<b>Tabellenverzeichnis</b>	<b>241</b>
	<b>Literaturverzeichnis</b>	<b>243</b>
	<b>Index</b>	<b>247</b>



# Einleitung

Liegen Daten vor, deren Zielgröße nur ganzzahlige, nicht negative Werte annimmt, wie etwa die Schadenanzahl eines Versicherungsportfolios, so spricht man von Zähldaten. Bei der Anpassung von Regressionsmodellen an Zähldaten wird üblicherweise die Poissonverteilung verwendet. Eine charakteristische Eigenschaft dieser Verteilung ist die Übereinstimmung von Erwartungswert und Varianz. Diese ist für Zähldaten aus der Praxis jedoch längst nicht immer gegeben. Häufig übersteigt die Stichprobenvarianz den empirischen Erwartungswert - ein Phänomen welches als Überdispersion bezeichnet wird. Die Poissonverteilungsannahme ist mit wachsender Überdispersion ungeeignet die Daten zu modellieren. Lawless (1987) schlägt deshalb die Verwendung der negativen Binomialverteilung als angenommene Responseverteilung vor. Diese ist hinsichtlich des Zusammenhangs zwischen Erwartungswert und Varianz flexibler als die Poissonverteilung, da die Varianz den Erwartungswert übersteigen kann. Somit ist es möglich Überdispersion durch das Regressionsmodell abzubilden.

In der Praxis kommen nicht nur überdispersionierte Zähldaten vor. Es existiert auch der seltenere Fall der Unterdispersion. Ein Verteilungsmodell, welches sowohl Über- als auch Unterdispersion modellieren kann und sich bei übereinstimmenden ersten Momenten zur Poissonverteilung reduziert, stellen Consul and Jain (1973) in Form der verallgemeinerten Poissonverteilung vor.

Bevor wir uns mit dem Regressionsmodell beschäftigen, welches auf der Annahme verallgemeinert poissonverteilter Responsedaten beruht, wollen wir in Kapitel 1 die Definition dieser Verteilung wiedergeben und grundlegende Eigenschaften herleiten. Anschließend werden zwei stochastische Prozesse betrachtet, mit denen sich die verallgemeinerte Poissonverteilung konstruieren lässt. Einer dieser Prozesse wird von Famoye (1997) zur Erzeugung verallgemeinert poissonverteilter Zufallsvariablen benutzt. Diese und zwei weitere Methoden zur Generierung solcher Zufallsvariablen werden am Ende von Kapitel 1 vorgestellt. Eine umfassende Betrachtung der verallgemeinerten Poissonverteilung bietet Consul (1989).

Da bei Überdispersion meist die negative Binomialverteilung zur Modellierung von Regressionsdaten verwendet wird, werden wir diese in Kapitel 2, für den Fall von übereinstimmenden ersten Momenten, mit der verallgemeinerten Poissonverteilung vergleichen, bevor wir uns Regressionsmodellen zuwenden. Zunächst werden dabei in Kapitel 3 die theoretischen Grundlagen verallgemeinerter linearer Modelle, der Maximum-Likelihood-Schätzung, sowie der Diskrepanzmessung wiederholt.

In Kapitel 4 wird anschließend auf die Besonderheiten bei der Modellierung von Zähldaten eingegangen. In aller Kürze wird das Poisson-Regressionsmodell vorgestellt, wobei die Verwendung unterschiedlicher Beobachtungszeiträume Berücksichtigung findet. Überdispersion in Zähldaten hat ihre Ursache meist in Missspezifikation. In Abschnitt 4.3 werden daher einige Quellen für Missspezifikation erläutert.

Schließlich werden wir in Kapitel 5, aufbauend auf dem Vorschlag von Consul and Famoye (1992), ein verallgemeinertes Poisson-Regressionsmodell aufstellen und für den Fall abweichender Beobachtungszeiträume erweitern. Zur Parameterschätzung werden verschiedene, auf dem

Maximum-Likelihood-Konzept basierenden Schätzprozeduren vorgestellt. Zunächst wird die Likelihoodfunktion jedoch hinsichtlich ihrer Verwendbarkeit zur Parameterschätzung untersucht. Zur Messung der Anpassungsqualität wird die Devianzfunktion und die Akaike-Information für das verallgemeinerte Poisson-Regressionsmodell hergeleitet.

Einen verbreiteten Spezialfall von Überdispersion stellt Nullenüberschuss dar. Zur Modellierung dieser Missspezifikationsquelle existieren hauptsächlich zwei Konzepte: “zero inflated models” und “zero hurdle models”. Äquivalent zu dem Zero-Inflated Poissonmodell, wie es Mullahy (1997) verwendet, wird in Kapitel 6 ein verallgemeinertes Zero-Inflated-Poissonmodell, basierend auf einer verallgemeinerten Poisson-Verteilungsannahme, konstruiert. Eine Erweiterung des Zero-Inflated-Poissonmodells ist bei Lambert (1992) zu finden. Diese Erweiterung übernehmen wir für das verallgemeinerte Zero-Inflated-Poissonmodell.

Einen weiteren Ansatz zur Modellierung von Nullenüberschuss stellt Winkelmann (2003) in Form der Zero-Hurdle-Modelle vor. Durch Verwendung der verallgemeinerten Poissonverteilung als angenommene Responseverteilung, wandeln wir in Kapitel 7 das bestehende Zero-Hurdle-Poissonmodell ab. Sowohl für die entwickelten verallgemeinerten Zero-Inflated-Poissonmodelle wie auch für das verallgemeinerte Zero-Hurdle-Poissonmodell wird auf die Maximum-Likelihood-Schätzung und die Messung der Anpassungsgüte eingegangen.

Zu den Regressionsmodellen aus Kapitel 5 - 7 wurden für die Statistiksoftware S-Plus Funktionen zur Maximum-Likelihood-Schätzung programmiert. Diese sind in Anhang C zu finden. Welche Güte die damit bestimmten Schätzer aufweisen, soll in Kapitel 8 untersucht werden. Dazu werden Daten simuliert, deren Zielgrößen gemäß den betrachteten Modellen verteilt sind. Abschließend werden die Schätzroutinen in Kapitel 9 auf zwei Beispieldatensätze angewandt.

# Kapitel 1

## Die verallgemeinerte Poissonverteilung

### 1.1 Definition und grundlegende Eigenschaften

Die verallgemeinerte oder generalisierte Poissonverteilung<sup>1</sup>, wie sie von Consul and Jain (1973) vorgeschlagen wird, besitzt die beiden Parameter  $\theta$  und  $\lambda$ . Für  $\lambda > 0$  übersteigt die Varianz den Erwartungswert, für  $\lambda < 0$  ist der Erwartungswert größer als die Varianz. Ist  $\lambda = 0$  so entspricht die verallgemeinerte Poissonverteilung der gewöhnlichen Poissonverteilung - Erwartungswert und Varianz stimmen in diesem Fall überein.

**Definition 1.1 (Verallgemeinerte Poissonverteilung)** Sei  $X$  eine Zufallsvariable mit diskreter Verteilung, welche auf  $\mathbb{N}$  definiert ist.  $P_x(\theta, \lambda)$  bezeichne die Wahrscheinlichkeit  $P(X = x)$ , dass  $X$  den nicht-negativen ganzzahligen Wert  $x$  annimmt.  $X$  ist verallgemeinert poissonverteilt mit Parametern  $\theta$  und  $\lambda$ , falls

$$P_x(\theta, \lambda) = \begin{cases} \frac{\theta(\theta+x\lambda)^{x-1}}{x!} e^{-(\theta+x\lambda)} & x = 0, 1, 2, \dots \\ 0 & \text{für } x > m \text{ wenn } \lambda < 0 \end{cases} \quad (1.1)$$

und 0 in allen anderen Fällen (d.h. für  $x \notin \mathbb{N}$ ). Für die Parameter gelten dabei folgende Einschränkungen:

- $\theta > 0$
- $\max\{-1, -\frac{\theta}{m}\} \leq \lambda \leq 1$

wobei  $m$  die größte natürliche Zahl ist, sodass  $\theta + m\lambda > 0$  ist, falls  $\lambda < 0$

Wir schreiben  $X \sim GP(\theta, \lambda)$ .

Die Beschränkung von  $\lambda$  nach unten soll die Existenz einer gewissen Anzahl von Klassen<sup>2</sup> garantieren, da diese bei gleichbleibendem  $\theta$  mit kleiner werdendem  $\lambda$  abnimmt. So wird von Consul (1989)  $m \geq 4$  gewählt was dazu führt, dass für  $\lambda < 0$  mindestens 5 Klassen mit positiver Eintrittswahrscheinlichkeit vorhanden sind.

<sup>1</sup>engl.: generalized poisson distribution - GPD

<sup>2</sup>Als Klasse wird in diesem Zusammenhang ein ganzzahliges Ereignis  $X=x$  bezeichnet

Unter Verwendung der Jensen'schen Formel (Consul (1989) / Jensen (1902))

$$\sum_{i=0}^{\infty} \frac{\theta(\theta + i\lambda)^{i-1}}{i!} e^{-i\lambda} = e^{\theta} \quad \text{für } \lambda > 0 \text{ und } \theta > 0 \quad (1.2)$$

lässt sich für positive Werte von  $\lambda$  zeigen, dass sich die Eintrittswahrscheinlichkeiten der möglichen Werte für  $X$  zu eins addieren:

$$\begin{aligned} \sum_{x=0}^{\infty} P_x(\theta, \lambda) &= \sum_{x=0}^{\infty} \frac{\theta(\theta + x\lambda)^{x-1}}{x!} e^{-(\theta+x\lambda)} \\ &= e^{-\theta} \sum_{x=0}^{\infty} \frac{\theta(\theta + x\lambda)^{x-1}}{x!} e^{-x\lambda} \\ &\stackrel{(1.2)}{=} e^{-\theta} e^{\theta} = 1 \end{aligned}$$

Da sich die verallgemeinerte Poissonverteilung für  $\lambda = 0$  zur gewöhnlichen Poissonverteilung vereinfacht, gilt  $\sum_{x=0}^{\infty} P_x(\theta, 0) = 1$ . Damit ist für nicht-negatives  $\lambda$  nachgewiesen, dass es sich bei  $P_x(\theta, \lambda)$  um eine Wahrscheinlichkeitsfunktion handelt.

Auf einen ähnlichen Beweisansatz für  $\sum_{x=0}^{\infty} P_x(\theta, \lambda) = 1$  greifen Lerner et al. (1997), unter Verwendung der Lagrangeentwicklung zurück. Ein alternativer Beweis für diese Gleichung mit Hilfe des Euler'schen Differenzenlemmas ist in Tuenter (2000) zu finden.

Für negatives  $\lambda$  addieren sich die Wahrscheinlichkeiten nicht zu eins. Da alle Ereignisse  $X > m$  die Eintrittswahrscheinlichkeit  $P_x(\theta, \lambda) = 0$  haben, ist (1.1) keine vollständige Verteilung.

Abhängig von den Werten von  $\theta$  und  $\lambda$  ist die Summe

$$F_m(\theta, \lambda) = \sum_{x=0}^m P_x(\theta, \lambda)$$

der Eintrittswahrscheinlichkeiten bis zum Punkt ab dem alle Eintrittswahrscheinlichkeiten Null sind entweder kleiner oder größer 1. Aus diesem Grund kann der Normalisierungsfaktor  $\frac{1}{F_m(\theta, \lambda)}$  eingeführt werden, mit dem  $P_x(\theta, \lambda)$  multipliziert wird.

$$\Rightarrow \tilde{P}_X(\theta, \lambda) := \frac{1}{F_m(\theta, \lambda)} P_X(\theta, \lambda) \quad (1.3)$$

Den Effekt dieses Faktors auf Erwartungswert und Varianz der verallgemeinerten Poissonverteilung sowie auf deren Eintrittswahrscheinlichkeiten wird in Consul and Shoukri (1985) untersucht. Für die meisten praktischen Anwendungen sind demnach die Werte von  $\theta$ ,  $\lambda$  und  $m$  dergestalt, dass der Normalisierungsfaktor vernachlässigt werden kann.

## 1.2 Verhalten bei Parametervariation

Um uns ein Bild vom Verhalten der verallgemeinerten Poissonverteilung für verschiedene Parameterkonstellationen machen zu können, sind im folgenden die Eintrittswahrscheinlichkeiten für  $x = 0, 1, 2, \dots$  bei wechselnden Werten von  $\theta$  und  $\lambda$  grafisch veranschaulicht.

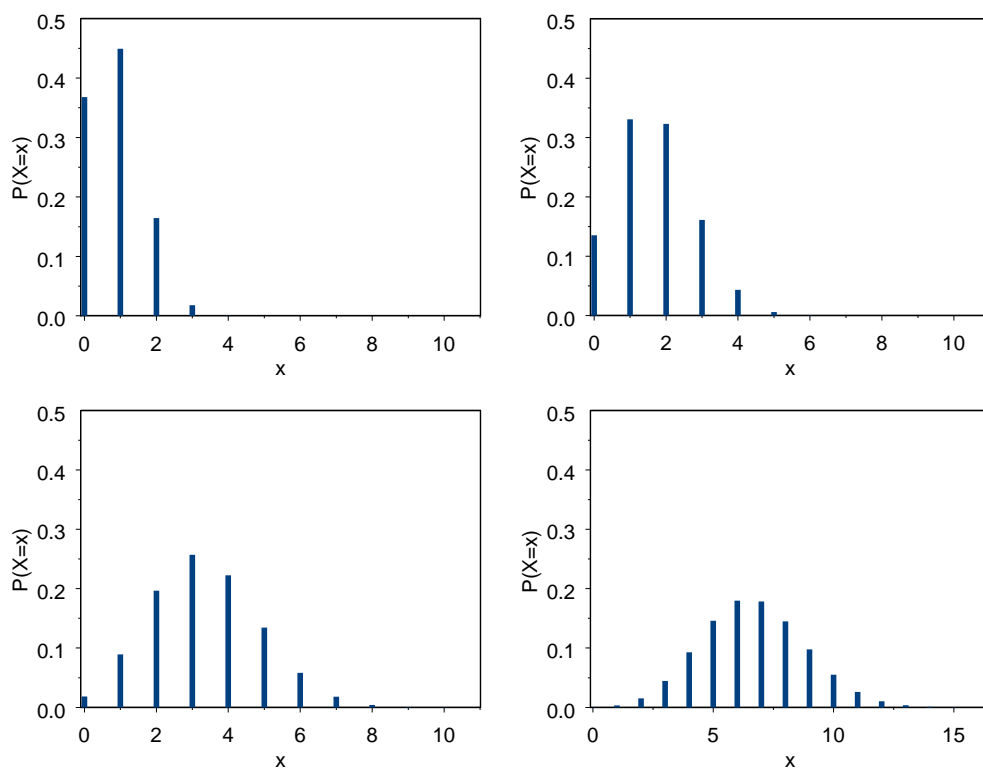


Abbildung 1.1: Eintrittswahrscheinlichkeiten für die Zufallsvariable  $X \sim GP(\theta, \lambda)$  mit  $\lambda = -0.2$  und  $\theta = 1, 2, 4, 8$

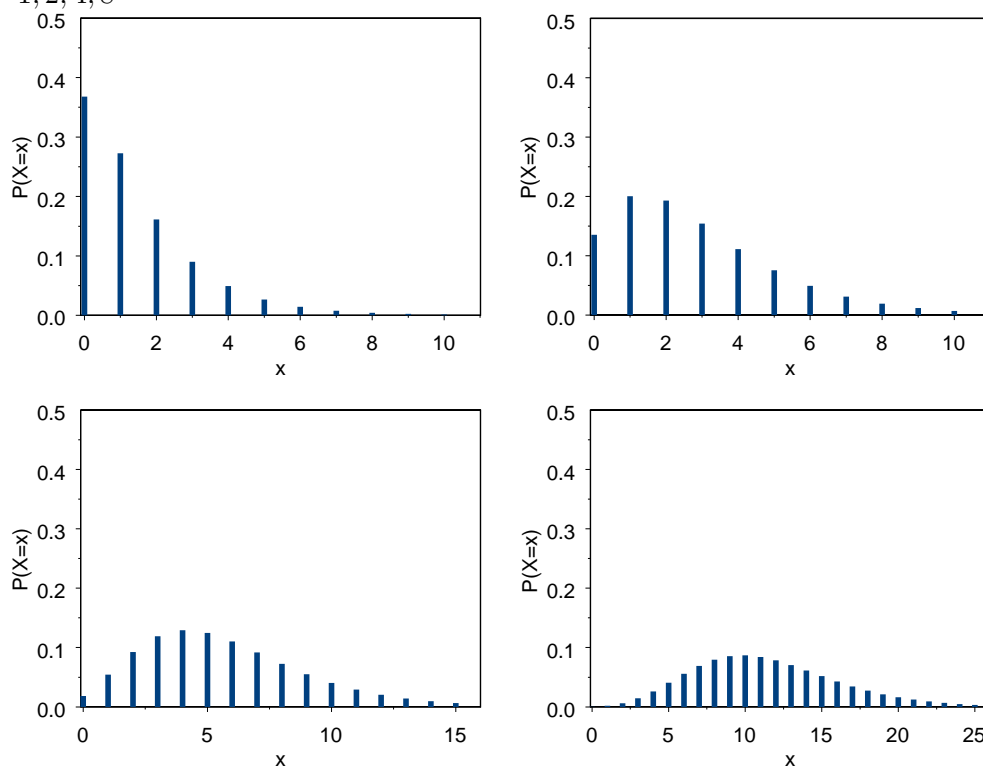


Abbildung 1.2: Eintrittswahrscheinlichkeiten für die Zufallsvariable  $X \sim GP(\theta, \lambda)$  mit  $\lambda = 0.5$  und  $\theta = 1, 2, 4, 8$

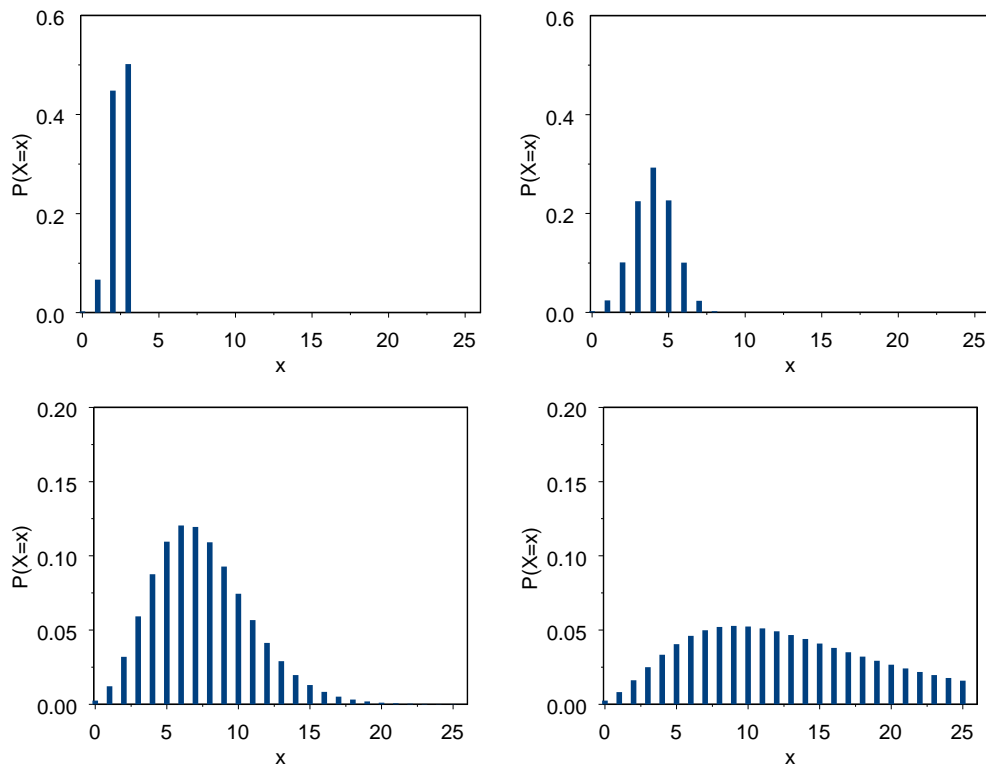


Abbildung 1.3: Eintrittswahrscheinlichkeiten für die Zufallsvariable  $X \sim GP(\theta, \lambda)$  mit  $\theta = 6$  und  $\lambda = -1.5, -0.5, 0.2, 0.6$

Zunächst wird der Parameter  $\lambda$  als Konstante behandelt um die Veränderung der Eintrittswahrscheinlichkeiten für wachsendes  $\theta$  betrachten zu können. In Abbildung 1.1 ist  $\lambda = -0.2$ , in Abbildung 1.2 ist  $\lambda = 0.5$ . Für  $\theta$  wurden in der oberen Zeile jeweils die Werte 1 und 2, in der unteren Zeile 4 und 8 gewählt. In beiden Abbildungen lässt sich erkennen, dass für kleines  $\theta$  links-Steilheit vorliegt. Mit zunehmendem  $\theta$  nimmt die Asymmetrie jedoch ab. Die sich verringernde Kuppelhöhe korrespondiert mit einer Zunahme der Streuung. Für großes  $\theta$  scheint sich das Schaubild der Glockenkurve einer Normalverteilung anzunähern. In Abschnitt 1.6 wird gezeigt werden, dass bezüglich  $\theta$  tatsächlich asymptotische Normalität vorliegt.

Sei nun  $\theta$  konstant und  $\lambda$  ändere sich. In Abbildung 1.3 wurde  $\theta = 6$  gewählt. Es ist zu erkennen, dass mit steigendem  $\lambda$  die Kuppelhöhe ab und somit die Varianz zunimmt. Der rechte Schwanz der Verteilung wird länger, der Erwartungswert erhöht sich. Dieser Verhaltensweise wird bei der Betrachtung verallgemeinerter Poissonregression eine entscheidende Rolle zukommen. Sie ermöglicht uns die Modellierung von überdispersionierten Zählraten.

Nicht zuletzt ist aus allen drei Schaubildern zu erkennen, dass die verallgemeinerte Poissonverteilung für alle  $\theta$  und alle  $\lambda$  unimodal ist. Ein formaler Beweis hierfür findet sich in Consul and Famoye (1986).

**Bemerkung:**

Bei der Betrachtung der Abbildungen ist zu beachten, dass sich bei den Abbildungen 1.1 und 1.2 der Maßstab der horizontalen Achse, bei Abbildung 1.3 der Maßstab der vertikalen Achse zwischen den einzelnen Boxen ändert.



### 1.3 Erwartungswert und Varianz der verallgemeinerten Poissonverteilung

#### 1.3.1 Erwartungswert der verallgemeinerten Poissonverteilung

**Satz 1.2** Für die Zufallsvariable  $X \sim GP(\theta, \lambda)$  ist der Erwartungswert durch

$$E(X) = \frac{\theta}{1 - \lambda} \quad (1.4)$$

gegeben.

*Beweis:*

Bezeichne  $\mu(\theta, \lambda)$  den Erwartungswert der verallgemeinert poissonverteilten Zufallsvariable  $X$  mit Parametern  $\theta$  und  $\lambda$

$$\mu(\theta, \lambda) = \sum_{x=0}^{\infty} x P_x(\theta, \lambda)$$

Da das Ereignis  $X=0$  keinen Beitrag zum Erwartungswert liefert, kann die Summation bei  $x = 1$  anstatt bei  $x = 0$  begonnen werden.

$$\begin{aligned} \mu(\theta, \lambda) &= \sum_{x=1}^{\infty} x P_x(\theta, \lambda) \\ &= \sum_{x=1}^{\infty} x \frac{\theta(\theta + x\lambda)^{x-1}}{x!} e^{-(\theta+x\lambda)} \\ &= \theta \sum_{x=1}^{\infty} \frac{x(\theta + x\lambda)^{x-1}}{x!} e^{-(\theta+x\lambda)} \end{aligned} \quad (1.5)$$

Die Substitution  $s := x - 1$  ergibt:

$$\begin{aligned} \mu(\theta, \lambda) &= \theta \sum_{s=0}^{\infty} \frac{(s+1)(\theta + s\lambda + \lambda)^s}{(s+1)!} e^{-(\theta+s\lambda+\lambda)} \\ &= \theta \sum_{s=0}^{\infty} \frac{(\theta + \lambda + s\lambda)^s}{s!} e^{-(\theta+\lambda+s\lambda)} \end{aligned} \quad (1.6)$$

Wir fassen nun den Ausdruck  $\theta + \lambda$  zu  $\tilde{\theta}$  zusammen.

$$\begin{aligned} \mu(\theta, \lambda) &= \theta \sum_{s=0}^{\infty} \frac{(\tilde{\theta} + s\lambda)^s}{s!} e^{-(\tilde{\theta}+s\lambda)} \\ &= \theta \sum_{s=0}^{\infty} \frac{(\tilde{\theta} + s\lambda)^{s-1}}{s!} (\tilde{\theta} + s\lambda) e^{-(\tilde{\theta}+s\lambda)} \\ &= \theta \left( \sum_{s=0}^{\infty} \frac{\tilde{\theta}(\tilde{\theta} + s\lambda)^{s-1}}{s!} e^{-(\tilde{\theta}+s\lambda)} + \frac{\lambda}{\tilde{\theta}} \sum_{s=0}^{\infty} s \frac{\tilde{\theta}(\tilde{\theta} + s\lambda)^{s-1}}{s!} e^{-(\tilde{\theta}+s\lambda)} \right) \\ &\stackrel{(1.1)}{=} \theta \left( \sum_{s=0}^{\infty} P_s(\tilde{\theta}, \lambda) + \frac{\lambda}{\tilde{\theta}} \sum_{s=0}^{\infty} s P_s(\tilde{\theta}, \lambda) \right) \end{aligned} \quad (1.7)$$

Da  $P_s$  die Wahrscheinlichkeitsfunktion einer Wahrscheinlichkeitsverteilung ist, ist der erste Summand in der Klammer gleich 1. Im zweiten Summanden finden wir den Ausdruck für den Erwartungswert einer  $GP(\tilde{\theta}, \lambda)$  verteilten Zufallsvariable. Daher gilt:

$$\mu(\theta, \lambda) = \theta \left( 1 + \frac{\lambda}{\tilde{\theta}} \mu(\tilde{\theta}, \lambda) \right) \quad (1.8)$$

Im folgenden ersetzen wir in (1.8)  $\tilde{\theta}$  wieder durch  $\theta + \lambda$  und führen die oben beschriebenen Umformungen für  $\mu(\theta + \lambda, \lambda)$  durch.

$$\begin{aligned} \mu(\theta, \lambda) &= \theta + \frac{\theta\lambda}{\theta + \lambda} \mu(\theta + \lambda, \lambda) \\ &= \theta + \frac{\theta\lambda}{\theta + \lambda} \left( (\theta + \lambda) + \frac{(\theta + \lambda)\lambda}{\theta + 2\lambda} \mu(\theta + 2\lambda, \lambda) \right) \\ &= \theta + \theta\lambda + \frac{\theta\lambda^2}{\theta + 2\lambda} \mu(\theta + 2\lambda, \lambda) \end{aligned}$$

Wird diese Vorgehensweise iterativ fortgesetzt, so erhalten wir eine geometrische Reihe. Für  $\lambda$  sind nur Werte kleiner 1 zugelassen. Daher konvergiert diese Reihe und der Erwartungswert lässt sich in einer einfachen Form schreiben (siehe z.B. Königsberger, 1999, S.60).

$$\begin{aligned} \mu(\theta, \lambda) &= \theta + \theta\lambda + \theta\lambda^2 + \dots \\ &= \theta \sum_{k=0}^{\infty} \lambda^k = \frac{\theta}{1 - \lambda} \end{aligned}$$

□

### 1.3.2 Erzeugende Funktion

**Definition 1.3 (Erzeugende Funktion)** Die diskrete Zufallsvariable  $X$  nehme die Werte  $x=0,1,2,\dots$  mit Wahrscheinlichkeiten  $P_x := P(X = x)$  an. Die Funktion

$$G_X(u) := E(u^X) = \sum_{x=0}^{\infty} P_x u^x, \quad u > 0 \quad (1.9)$$

heißt dann erzeugende Funktion<sup>3</sup> von  $X$ .

**Lemma 1.4 (Lagrange'sches Invertierungstheorem)** Sei  $g(t)$  mit  $g(0) \neq 0$  eine analytische Funktion in  $t$ , die in eine (konvergente) Potenzreihe entwickelt werden kann. Falls  $|ug(t)| \leq |t|$  gilt, lässt sich jede analytische Funktion  $f(t)$  in eine Potenzreihe in  $u$  entwickeln:

$$f(t) = f(0) + \sum_{k=1}^{\infty} \frac{u^k}{k!} \frac{\partial^{k-1}}{\partial t^{k-1}} \left[ g(t)^k \frac{\partial f(t)}{\partial t} \right] \Big|_{t=0} \quad (1.10)$$

*Beweis:* siehe z.B. Hofbauer

<sup>3</sup>In Anlehnung an die englische Bezeichnung 'probability generating function' (pgf) wird die erzeugende Funktion in diesem Zusammenhang oft auch als wahrscheinlichkeitserzeugende Funktion bezeichnet.

**Satz 1.5 (Erzeugende Funktion der verallgemeinerten Poissonverteilung)** Sei  $X \sim GP(\theta, \lambda)$ . Dann wird die erzeugende Funktion der Zufallsvariablen  $X$  durch

$$G_X(u) = E(u^X) = e^{\theta(t-1)} \quad (1.11)$$

mit  $t = ue^{\lambda(t-1)}$  bestimmt.

*Beweis:*

Setze  $f(t) := e^{\theta(t-1)}$ ,  $g(t) := e^{\lambda(t-1)}$  sowie  $u := \frac{t}{g(t)}$ . Damit gilt:

$$|ug(t)| = \left| \frac{t}{g(t)} g(t) \right| \leq |t|$$

Die Funktionen  $f$  und  $g$  sind analytisch, d.h. sie sind in jedem  $t \in \mathbb{N}$  differenzierbar. Damit lässt sich das Lagrangesche Invertierungstheorem anwenden und die Funktion  $f(t)$  kann unter Verwendung der Funktion  $g(t)$  um Null entwickelt werden.

$$\begin{aligned} e^{\theta(t-1)} &= e^{-\theta} + \sum_{k=1}^{\infty} \frac{u^k}{k!} \frac{\partial^{k-1}}{\partial t^{k-1}} \left[ e^{k\lambda(t-1)} \theta e^{\theta(t-1)} \right] \Big|_{t=0} \\ &= e^{-\theta} + \sum_{k=1}^{\infty} \frac{u^k}{k!} \theta e^{-(\theta+k\lambda)} \frac{\partial^{k-1}}{\partial t^{k-1}} \left[ e^{(\theta+k\lambda)t} \right] \Big|_{t=0} \\ &= e^{-\theta} + \sum_{k=1}^{\infty} \frac{u^k}{k!} \theta e^{-(\theta+k\lambda)} \left[ (\theta + k\lambda)^{k-1} e^{(\theta+k\lambda)t} \right] \Big|_{t=0} \\ &= e^{-\theta} + \sum_{k=1}^{\infty} u^k \frac{\theta(\theta + k\lambda)^{k-1}}{k!} e^{-(\theta+k\lambda)} \\ &= \sum_{k=0}^{\infty} u^k \frac{\theta(\theta + k\lambda)^{k-1}}{k!} e^{-(\theta+k\lambda)} \end{aligned} \quad (1.12)$$

Der Ausdruck  $\frac{\theta(\theta+k\lambda)^{k-1}}{k!} e^{-(\theta+k\lambda)}$  in (1.12) entspricht der Eintrittswahrscheinlichkeit von  $X=k$  für verallgemeinert poissonverteiltes  $X$ . Mit (1.1) lässt sich (1.12) somit umschreiben zu

$$e^{\theta(t-1)} = \sum_{k=0}^{\infty} u^k P_k(\theta, \lambda) = E(u^X)$$

womit die Behauptung folgt. □

### 1.3.3 Momenterzeugende Funktion

**Definition 1.6 (Momenterzeugende Funktion)** Die diskrete Zufallsvariable  $X$  nehme die Werte  $x = 0, 1, 2, \dots$  mit Wahrscheinlichkeiten  $P_x := P(X = x)$  an. Die Funktion

$$M_X(v) := E(e^{vX}) = \sum_{x=0}^{\infty} P_x e^{vx}, \quad s > 0 \quad (1.13)$$

heißt dann momenterzeugende Funktion<sup>4</sup> von  $X$ .

<sup>4</sup>moment generating function (mgf)

**Satz 1.7 (Momenterzeugende Funktion der verallgemeinerten Poissonverteilung)** .  
 Sei  $X \sim GP(\theta, \lambda)$ . Dann wird die momenterzeugende Funktion der Zufallsvariablen  $X$  durch

$$M_X(v) = E(e^{vX}) = e^{\theta(e^s - 1)} \quad (1.14)$$

mit  $s = v + \lambda(e^s - 1)$  bestimmt.

*Beweis:*

Ersetzen wir in dem Ausdruck für die erzeugende Funktion (1.11)  $u$  durch  $e^v$  so erhalten wir

$$\begin{aligned} E(u^X) &= E(e^{vX}) \\ &= e^{\theta(t-1)} \end{aligned} \quad (1.15)$$

wobei  $t = e^{v+\lambda(t-1)}$  ist. Nun führen wir die Größe  $s := \log t$  ein. Durch Einsetzen in (1.15) ergibt sich

$$E(e^{vX}) = e^{\theta(e^s - 1)} \quad \text{mit } s = v + \lambda(e^s - 1)$$

und somit die momenterzeugende Funktion von  $X$ .

□

### 1.3.4 Momente der verallgemeinerten Poissonverteilung

**Definition 1.8 (Momente)** Vorausgesetzt die entsprechenden Erwartungswerte existieren, so definiert man für  $k=1,2,\dots$

i)  $k$ -tes nicht zentriertes Moment:

$$\mu_k = \sum_{x=0}^{\infty} x^k P_x(\theta, \lambda) \quad (1.16)$$

ii)  $k$ -tes zentriertes Moment:

$$\nu_k = \sum_{x=0}^{\infty} (x - \mu)^k P_x(\theta, \lambda) \quad (1.17)$$

mit  $\mu_1 = E(X)$

Für  $\lambda < 1$  existieren alle Momente der verallgemeinerten Poissonverteilung (Consul, 1989, S.49).

**Satz 1.9 (Rekursionsgleichung für nicht zentrierte Momente)** Für die nicht zentrierten Momente  $\mu_k$  einer verallgemeinert poissonverteilten Zufallsvariable  $X \sim GP(\theta, \lambda)$  gilt die Rekursionsgleichung

$$(1 - \lambda)\mu_{k+1} = \theta\mu_k + \theta \frac{\partial \mu_k}{\partial \theta} + \lambda \frac{\partial \mu_k}{\partial \lambda} \quad (1.18)$$

*Beweis:*

Die momenterzeugende Funktion von  $X$  ist gemäß Satz 1.7 gegeben durch

$$M_X(v) = e^{\theta(e^s - 1)}$$

mit

$$s = v + \lambda(e^s - 1) \quad (1.19)$$

*Teil1: Ableitungen von s*

Die Funktion s ist von v und  $\lambda$  abhängig, nicht jedoch von  $\theta$ . Daher gilt  $\frac{\partial s}{\partial \theta} = 0$ . Wir bilden nun die partiellen Ableitungen von s nach v und  $\lambda$  und vereinfachen diese.

$$\begin{aligned} \frac{\partial s}{\partial v} : \quad \frac{\partial s}{\partial v} &= \frac{\partial}{\partial v}(v + \lambda(e^s - 1)) \\ &= 1 + \frac{\partial \lambda(e^s - 1)}{\partial v} \\ &= 1 + \lambda \frac{\partial e^s}{\partial v} \\ &= 1 + \lambda e^s \frac{\partial s}{\partial v} \end{aligned} \tag{1.20}$$

Durch Umformung von (1.20) lässt sich  $\frac{\partial s}{\partial v}$  explizit darstellen:

$$\begin{aligned} \frac{\partial s}{\partial v} - \lambda e^s \frac{\partial s}{\partial v} &= 1 \\ \Leftrightarrow (1 - \lambda e^s) \frac{\partial s}{\partial v} &= 1 \\ \Rightarrow \frac{\partial s}{\partial v} &= \frac{1}{1 - \lambda e^s} \end{aligned} \tag{1.21}$$

$$\begin{aligned} \frac{\partial s}{\partial \lambda} : \quad \frac{\partial s}{\partial \lambda} &= \frac{\partial}{\partial \lambda}(v + \lambda(e^s - 1)) \\ &= \frac{\partial}{\partial \lambda}(\lambda(e^s - 1)) \\ &= (e^s - 1) + \lambda \left( \frac{\partial(e^s - 1)}{\partial \lambda} \right) \end{aligned} \tag{1.22}$$

Auch hier liefern Umstellungen der Gleichung (1.22) die explizite Form von  $\frac{\partial s}{\partial \lambda}$ :

$$\begin{aligned} \frac{\partial s}{\partial \lambda} - \lambda e^s \frac{\partial s}{\partial \lambda} &= e^s - 1 \\ \Leftrightarrow (1 - \lambda e^s) \frac{\partial s}{\partial \lambda} &= e^s - 1 \\ \Rightarrow \frac{\partial s}{\partial \lambda} &= \frac{e^s - 1}{1 - \lambda e^s} \end{aligned} \tag{1.23}$$

Wir multiplizieren nun (1.23) mit  $\lambda$  und formen die rechte Seite so um, dass wir einen Bruch mit dem selben Nenner erhalten wie in Gleichung (1.21).

$$\begin{aligned} \lambda \frac{\partial s}{\partial \lambda} &= \frac{\lambda e^s - \lambda}{1 - \lambda e^s} \\ &= \frac{\lambda e^s - \lambda + 1 - 1}{1 - \lambda e^s} \\ &= \frac{-(1 - \lambda e^s) + (1 - \lambda)}{1 - \lambda e^s} \\ &= -1 + \frac{1 - \lambda}{1 - \lambda e^s} \end{aligned} \tag{1.24}$$

Multiplizieren wir nun (1.21) mit  $1 - \lambda$  und subtrahieren von dem Ergebnis die Gleichung (1.24), so erhalten wir die partielle Differentialgleichung

$$(1 - \lambda) \frac{\partial s}{\partial v} - \lambda \frac{\partial s}{\partial \lambda} = 1 \quad (1.25)$$

*Teil 2: Ableitungen von  $M_X$*

Ebenso wie für  $s$  bilden wir die partiellen Ableitungen von  $M_X$  nach  $\theta$ ,  $\lambda$  sowie  $v$  und formen diese um. Dabei verwenden wir die eben hergeleiteten Ableitungen von  $s$ , insbesondere das Resultat  $\frac{\partial s}{\partial \theta} = 0$ .

$$\begin{aligned} \frac{\partial M_X(v)}{\partial \theta} : \quad \frac{\partial M_X(v)}{\partial \theta} &= \frac{\partial e^{\theta(e^s-1)}}{\partial \theta} \\ &= e^{s\theta(e^s-1)} - e^{\theta(e^s-1)} \end{aligned} \quad (1.26)$$

Multiplikation von (1.26) mit  $\theta$  führt auf:

$$\begin{aligned} \theta \frac{\partial M_X(v)}{\partial \theta} &= e^s \theta e^{\theta(e^s-1)} - \theta e^{\theta(e^s-1)} \\ &= e^s \theta M_X(v) - \theta M_X(v) \end{aligned}$$

Umstellung dieser Gleichung ergibt:

$$\theta e^s M_X(v) = \theta \frac{\partial M_X(v)}{\partial \theta} + \theta M_X(v) \quad (1.27)$$

$$\begin{aligned} \frac{\partial M_X(v)}{\partial \lambda} : \quad \frac{\partial M_X(v)}{\partial \lambda} &= \frac{\partial e^{\theta(e^s-1)}}{\partial \lambda} \\ &= e^{\theta(e^s-1)} \frac{\partial \theta(e^s-1)}{\partial \lambda} \\ &= e^{\theta(e^s-1)} \theta e^s \frac{\partial s}{\partial \lambda} \\ &= M_X(v) \theta e^s \frac{\partial s}{\partial \lambda} \end{aligned} \quad (1.28)$$

$$\begin{aligned} \frac{\partial M_X(v)}{\partial v} : \quad \frac{\partial M_X(v)}{\partial v} &= M_X(v) \frac{\partial \theta(e^s-1)}{\partial v} \\ &= M_X(v) \theta \frac{\partial e^s}{\partial v} \\ &= M_X(v) \theta e^s \frac{\partial s}{\partial v} \end{aligned} \quad (1.29)$$

*Teil 3: Zusammenhang zwischen Momenten  $\mu_k$  und momenterzeugenden Funktion  $M_X$*

Unter Verwendung der Definition (1.16) kann der Ausdruck (1.13) für die momenterzeugende Funktion auch wie folgt dargestellt werden:

$$\begin{aligned} M_X(v) &= E(e^{Xv}) = \sum_{x=0}^{\infty} e^{xv} P_x(\theta, \lambda) \\ &= \sum_{x=0}^{\infty} \left( \sum_{k=0}^{\infty} \frac{(xv)^k}{k!} \right) P_x(\theta, \lambda) \end{aligned}$$

$$\begin{aligned}
&= \sum_{k=0}^{\infty} \left( \sum_{x=0}^{\infty} \frac{(xv)^k}{k!} P_x(\theta, \lambda) \right) \\
&= \sum_{k=0}^{\infty} \frac{v^k}{k!} \left( \sum_{x=0}^{\infty} x^k P_x(\theta, \lambda) \right) = \sum_{k=0}^{\infty} \frac{v^k}{k!} \mu_k
\end{aligned} \tag{1.30}$$

Mit der Gleichung (1.30) lässt sich eine Alternativdarstellung für die Ableitung (1.29) bestimmen.

$$\begin{aligned}
\frac{\partial M_x(v)}{\partial v} &= \frac{\partial}{\partial v} \sum_{k=0}^{\infty} \frac{v^k}{k!} \mu_k \\
&= \sum_{k=0}^{\infty} k \frac{v^{k-1}}{k!} \mu_k \\
&= \sum_{k=1}^{\infty} k \frac{v^{k-1}}{k!} \mu_k = \sum_{k=0}^{\infty} \frac{v^k}{k!} \mu_{k+1}
\end{aligned} \tag{1.31}$$

#### Teil 4: Rekursionsgleichung

Um auf die Rekursionsgleichung (1.18) zu kommen multiplizieren wir die Gleichung (1.25) mit  $\theta e^s M_X(v)$ .

$$\theta e^s M_X(v) = \theta e^s M_X(v) (1 - \lambda) \frac{\partial s}{\partial v} - \theta e^s M_X(v) \lambda \frac{\partial s}{\partial \lambda} \tag{1.32}$$

Auf der rechten Seite von (1.32) sind gemäß (1.28) und (1.29) Ausdrücke für die partiellen Ableitungen  $\frac{\partial M_X}{\partial v}$  und  $\frac{\partial M_X}{\partial \lambda}$  enthalten. Ersetzen wir diese, dann gilt:

$$\theta e^s M_X(v) = (1 - \lambda) \frac{\partial M_X(v)}{\partial v} - \frac{\partial M_X(v)}{\partial \lambda} \lambda$$

Die linke Seite kann unter Verwendung der Gleichung (1.27) ebenfalls ersetzt werden.

$$\begin{aligned}
\Rightarrow \quad & \theta \frac{\partial M_X(v)}{\partial \theta} + \theta M_X(v) = (1 - \lambda) \frac{\partial M_X(v)}{\partial v} - \lambda \frac{\partial M_X(v)}{\partial \lambda} \\
\Leftrightarrow \quad & (1 - \lambda) \frac{\partial M_X(v)}{\partial v} = \theta M_X(v) + \theta \frac{\partial M_X(v)}{\partial \theta} + \lambda \frac{\partial M_X(v)}{\partial \lambda}
\end{aligned} \tag{1.33}$$

Auf der linken Seite von (1.33) ersetzen wir nun  $\frac{\partial M_X}{\partial v}$  durch den Ausdruck (1.31) und auf der rechten Seite  $M_X$  durch den Ausdruck (1.30). Wir erhalten:

$$\begin{aligned}
(1 - \lambda) \sum_{k=0}^{\infty} \frac{v^k}{k!} \mu_{k+1} &= \theta \sum_{k=0}^{\infty} \frac{v^k}{k!} \mu_k + \theta \frac{\partial}{\partial \theta} \sum_{k=0}^{\infty} \frac{v^k}{k!} \mu_k + \lambda \frac{\partial}{\partial \lambda} \sum_{k=0}^{\infty} \frac{v^k}{k!} \mu_k \\
&= \sum_{k=0}^{\infty} \frac{v^k}{k!} \left( \theta \mu_k + \theta \frac{\partial \mu_k}{\partial \theta} + \lambda \frac{\partial \mu_k}{\partial \lambda} \right)
\end{aligned} \tag{1.34}$$

$$\Leftrightarrow \sum_{k=0}^{\infty} \frac{v^k}{k!} \left( (1 - \lambda) \mu_{k+1} - \theta \mu_k - \theta \frac{\partial \mu_k}{\partial \theta} - \lambda \frac{\partial \mu_k}{\partial \lambda} \right) = 0 \tag{1.35}$$

Für  $v > 0$  gilt somit:

$$(1 - \lambda)\mu_{k+1} - \theta\mu_k - \theta \frac{\partial \mu_k}{\partial \theta} - \lambda \frac{\partial \mu_k}{\partial \lambda} = 0$$

Woraus durch Umformung die Behauptung folgt. □

**Bemerkung:** In (Consul, 1989, S.51) ist die Rekursionsgleichung

$$\nu_{k+1} = \frac{k\theta}{(1 - \lambda)^3} \nu_{k-1} + \frac{1}{1 - \lambda} \left[ \frac{d}{dt} \nu_k(t) \right] \Big|_{t=1}$$

mit  $k = 1, 2, 3, \dots$  für die zentrierten Momente  $\nu_k$  zu finden.

### 1.3.5 Varianz der verallgemeinerten Poissonverteilung

**Satz 1.10** Für die Zufallsvariable  $X \sim GP(\theta, \lambda)$  ist die Varianz durch

$$\text{Var}(X) = \frac{\theta}{(1 - \lambda)^3} \tag{1.36}$$

gegeben.

*Beweis:*

Aus der Definition der zentrierten Momente folgen unmittelbar die Identitäten  $\mu_1 = E(X)$  und  $\mu_2 = E(X^2)$ . Unter Verwendung des Satzes von Steiner (Råde and Westergren, 1997, S.410) wonach  $\text{Var}(X) = E(X^2) + E(X)^2$  ist, gilt somit der Zusammenhang

$$\text{Var}(X) = \mu_2 - \mu_1^2 \tag{1.37}$$

zwischen den zentrierten Momenten und der Varianz. Mit dem Erwartungswert  $E(X) = \frac{\theta}{1 - \lambda} = \mu_1$  ist das erste Moment der verallgemeinerten Poissonverteilung bekannt. Damit kann die Rekursionsgleichung (1.18) zur Bestimmung des zweiten Moments verwendet werden.

$$\begin{aligned} (1 - \lambda)\mu_2 &= \theta\mu_1 + \theta \frac{\partial \mu_1}{\partial \theta} + \lambda \frac{\partial \mu_1}{\partial \lambda} \\ &\stackrel{(1.4)}{=} \frac{\theta^2}{1 - \lambda} + \frac{\lambda}{1 - \lambda} + \frac{\lambda\theta}{(1 - \lambda)^2} \\ &= \theta \frac{\theta(1 - \lambda) + (1 - \lambda) + \lambda}{(1 - \lambda)^2} \\ &= \theta \frac{\theta - \theta\lambda + 1 - \lambda + \lambda}{(1 - \lambda)^2} \\ \Rightarrow \mu_2 &= \theta \frac{1 - \lambda\theta + \theta}{(1 - \lambda)^3} \end{aligned} \tag{1.38}$$

Wir setzen nun das Resultat (1.38) in die Gleichung (1.37) ein:

$$\begin{aligned} \text{Var}(X) &= \mu_2 - \mu_1^2 \\ &= \theta \frac{1 - \lambda\theta + \theta}{(1 - \lambda)^3} - \frac{\theta^2}{(1 - \lambda)^2} \\ &= \theta \frac{1 - \lambda\theta + \theta - \theta + \lambda\theta}{(1 - \lambda)^3} = \frac{\theta}{(1 - \lambda)^3} \end{aligned}$$

□



### 1.3.6 Dispersionsparameter

Vergleicht man den Erwartungswert und die Varianz der verallgemeinerten Poissonverteilung, so erkennt man, dass

$$\text{Var}(X) = \frac{1}{(1-\lambda)^2} E(X) \quad (1.39)$$

gilt. Auf Grund dieses Zusammenhangs wird der Dispersionsparameter  $\varphi$  definiert, der die Abweichung des Erwartungswertes von der Varianz angibt.

$$\varphi := \frac{1}{1-\lambda} \quad (1.40)$$

Da sich der Wert des Parameters  $\lambda$  im halboffenen Intervall  $[\max\{-1, -\frac{\theta}{m}\}, 1)$  befindet, ist  $\varphi$  auf  $[a, \infty)$  mit  $a := \max\{\frac{1}{2}, \frac{1}{1+\frac{\theta}{m}}\}$  definiert. Unter Verwendung von (1.39) lässt sich somit schreiben:

$$\text{Var}(X) = \varphi^2 E(X)$$

Für  $\varphi < 1$  (gleichbedeutend mit  $\lambda < 0$ ) ist die Varianz kleiner, für  $\varphi > 1$  (d.h.  $\lambda > 0$ ) größer als der Erwartungswert. Für  $\varphi = 1$  stimmen Varianz und Erwartungswert überein - es liegt Äquidispersion vor.

Der Dispersionsparameter  $\varphi$  ist genau dann eins, wenn  $\lambda$  Null ist. In diesem Fall stimmt die  $GP(\theta, \lambda)$ -Verteilung mit der  $Poi(\theta)$ -Verteilung überein. Der Wert  $\lambda = -2$ , welcher ebenfalls zur Gleichheit von Erwartungswert und Varianz führen würde, ist nicht zugelassen, da in (1.1)  $\lambda \geq \max\{-1, -\frac{\theta}{m}\}$  vorausgesetzt wird.

## 1.4 Faltungungseigenschaft

**Satz 1.11 (Summe zweier verallgemeinert poissonverteilter Zufallsvariablen)** *Seien  $X \sim GP(\theta_1, \lambda)$  und  $Y \sim GP(\theta_2, \lambda)$  zwei unabhängige Zufallsvariablen. Damit ist die Zufallsvariable  $Z := X + Y$  verallgemeinert poissonverteilt mit Parametern  $\theta_1 + \theta_2$  und  $\lambda$  ( $Z \sim GP(\theta_1 + \theta_2, \lambda)$ ).*

*Beweis:*

Da es sich bei  $P_x(\theta, \lambda)$  um eine Wahrscheinlichkeitsfunktion<sup>5</sup> handelt, gilt:

$$\begin{aligned} \sum_{x=0}^{\infty} P_x(\theta, \lambda) &= \sum_{x=0}^{\infty} \frac{\theta(\theta + x\lambda)^{x-1}}{x!} e^{-(\theta+x\lambda)} = 1 \\ \Rightarrow \sum_{x=0}^{\infty} e^\theta P_x(\theta, \lambda) &= \sum_{x=0}^{\infty} \frac{\theta(\theta + x\lambda)^{x-1}}{x!} e^{-x\lambda} = e^\theta \end{aligned}$$

<sup>5</sup>für negatives  $\lambda$  können anstatt  $P_x$  die normalisierten Eintrittswahrscheinlichkeiten  $\tilde{P}_x$  (1.3) verwendet werden.

Mit der elementaren Multiplikationsregel für Potenzen mit gleicher Basis folgt:

$$\begin{aligned} & \sum_{z=0}^{\infty} \frac{(\theta_1 + \theta_2)(\theta_1 + \theta_2 + x\lambda)^{z-1}}{z!} e^{-z\lambda} = e^{\theta_1 + \theta_2} = e^{\theta_1} e^{\theta_2} \\ & = \left( \sum_{i=0}^{\infty} \frac{\theta_1(\theta_1 + i\lambda)^{i-1}}{i!} e^{-i\lambda} \right) \left( \sum_{j=0}^{\infty} \frac{\theta_2(\theta_2 + j\lambda)^{j-1}}{j!} e^{-j\lambda} \right) \end{aligned} \quad (1.41)$$

Es werden nun die beiden Terme im rechten Teil der Gleichung (1.41) ausmultipliziert. Anschließend wird die entstehende Doppelsumme umgeordnet.

$$\begin{aligned} & \left( \sum_{i=0}^{\infty} \frac{\theta_1(\theta_1 + i\lambda)^{i-1}}{i!} e^{-i\lambda} \right) \left( \sum_{j=0}^{\infty} \frac{\theta_2(\theta_2 + j\lambda)^{j-1}}{j!} e^{-j\lambda} \right) \\ & = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{\theta_1(\theta_1 + i\lambda)^{i-1}}{i!} e^{-i\lambda} \frac{\theta_2(\theta_2 + j\lambda)^{j-1}}{j!} e^{-j\lambda} \\ & = \sum_{z=0}^{\infty} \sum_{i=0}^z \frac{\theta_1(\theta_1 + i\lambda)^{i-1}}{i!} \frac{\theta_2(\theta_2 + (z-i)\lambda)^{z-i-1}}{(z-i)!} e^{-z\lambda} \\ & = \sum_{z=0}^{\infty} \sum_{i=0}^z \frac{\theta_1\theta_2(\theta_1 + i\lambda)^{i-1}(\theta_2 + (z-i)\lambda)^{z-i-1}}{i!(z-i)!} e^{-z\lambda} \end{aligned} \quad (1.42)$$

Für die Summanden aus dem linken Teil der Gleichung (1.41) und dem rechten Teil der Gleichung (1.42) muss Gleichheit gelten.

$$\frac{(\theta_1 + \theta_2)(\theta_1 + \theta_2 + z\lambda)^{z-1}}{z!} = \sum_{i=0}^z \frac{\theta_1\theta_2(\theta_1 + i\lambda)^{i-1}(\theta_2 + (z-i)\lambda)^{z-i-1}}{i!(z-i)!} \quad (1.43)$$

Damit können wir einen Ausdruck für die Eintrittswahrscheinlichkeiten

$$P(X + Y = z) = \sum_{i=0}^z P_i(\theta_1, \lambda) P_{z-i}(\theta_2, \lambda) \quad (1.44)$$

der Zufallsvariable  $Z = X + Y$  bestimmen. In die allgemeine Faltungsformel (1.44) werden dazu die konkreten Eintrittswahrscheinlichkeiten für die verallgemeinert poissonverteilten Zufallsvariablen  $X$  und  $Y$  eingesetzt und unter Verwendung des Resultates (1.43) umgeformt.

$$\begin{aligned} P(Z = z) & = \sum_{i=0}^z \frac{\theta_1(\theta_1 + i\lambda)^{i-1}}{i!} e^{-(\theta_1 + i\lambda)} \frac{\theta_2(\theta_2 + (z-i)\lambda)^{z-i-1}}{(z-i)!} e^{-(\theta_2 + (z-i)\lambda)} \\ & = \sum_{i=0}^z \theta_1\theta_2 \frac{(\theta_1 + i\lambda)^{i-1}(\theta_2 + (z-i)\lambda)^{z-i-1}}{i!(z-i)!} e^{-(\theta_1 + \theta_2 + z\lambda)} \\ & \stackrel{(1.43)}{=} e^{-(\theta_1 + \theta_2 + z\lambda)} \frac{(\theta_1 + \theta_2)(\theta_1 + \theta_2 + z\lambda)^{z-1}}{z!} = P_z(\theta_1 + \theta_2, \lambda) \end{aligned}$$

Wir erhalten wiederum die Formel (1.1) für die Parameter  $\theta_1 + \theta_2$  und  $\lambda$ , womit  $Z \sim GP(\theta_1 + \theta_2, \lambda)$  gezeigt ist.

□

Durch Induktion folgt weiterhin:

**Korollar 1.12**  $X_i \sim GP(\theta_i, \lambda)$ ,  $i = 1, \dots, n$  seien  $n$  unabhängige Zufallsvariablen. Dann gilt:

$$\sum_{i=1}^n X_i \sim GP\left(\sum_{i=1}^n \theta_i, \lambda\right) \quad (1.45)$$

## 1.5 Modelle die zur verallgemeinerten Poissonverteilung führen

### 1.5.1 Der verallgemeinerte Poissonprozess

$N(t)$  bezeichne die Anzahl gewisser Ereignisse, die im Intervall  $[0, t]$  auftreten. Damit nimmt die zufällige Größe  $N(t)$  die Werte  $0, 1, 2, \dots$  an. Wir definieren  $\pi(k, t)$  als die Wahrscheinlichkeit, dass bis zum Zeitpunkt  $t$  das betrachtete Ereignis  $k$ -mal eingetreten ist.

$$\pi(k, t) := P(N(t) = k) \quad (1.46)$$

$$\Rightarrow \sum_{k=0}^{\infty} \pi(k, t) = 1 \quad \forall t \geq 0$$

Somit ist  $\{\pi(k, t)\}$  für jeden Wert von  $t$  eine Wahrscheinlichkeitsfunktion,  $\{N(t), t \in T\}$  folglich ein stochastischer Zählprozess (s. Anhang B).

**Definition 1.13 (Verallgemeinerter Poissonprozess)** *Der stochastische Zählprozess  $\{N(t), t \in T\}$  heißt verallgemeinerter Poissonprozess, wenn folgende Axiome erfüllt sind (Consul, 1989, S.35 ff.):*

- *Axiom 1:*  
Der Prozess startet zum Zeitpunkt 0 im Zustand 0  
 $\Rightarrow P(N(0) = 0) = \pi(0, 0) = 1$   
 $\Rightarrow \pi(x, 0) = 0 \quad \forall x > 0$
- *Axiom 2:*  
 $0 < P(N(t) > 0) < 1$
- *Axiom 3:*  
Ein direkter Übergang ist nur von einem Zustand  $k$  in den nächst höheren Zustand  $k+1$  möglich. D.h. in einem hinreichend kleinen Intervall der Länge  $h$  tritt höchstens ein Ereignis auf.  
 $\Rightarrow P(N(h) \geq 2) = \sum_{x=2}^{\infty} \pi(x, h) = o(h)$
- *Axiom 4:*  
Die Wahrscheinlichkeit genau eines Ereignisses in einem kleinen Intervall der Länge  $h$  hängt nicht nur von der Intervalllänge  $h$  und dem Parameter  $\theta$  (vgl. Poissonprozess), sondern auch vom Zustand  $x$  des Prozesses zum Zeitpunkt  $t$ , sowie vom nachfolgenden Zustand ab. Dementsprechend gelten folgende Beziehungen

$$P(N[t+h] - N[t] = 0 | N[t] = x) = 1 - (\theta + x\lambda)h + o(h) \quad (1.47)$$

$$P(N[t+h] - N[t] = 1 | N[t] = x - 1) = \frac{(\theta + (x-1)\lambda)he^{-\lambda t}}{(1 - \lambda(\theta + x\lambda)^{-1})^{x-1}} + o(h) \quad (1.48)$$

**Satz 1.14 (Verteilung des verallgemeinerten Poissonprozesses)**  $\{N(t), t \geq 0\}$  sei ein verallgemeinerter Poissonprozess. Die Wahrscheinlichkeit  $P(N(t) = x) = \pi(x, t)$  ist dann durch Wahrscheinlichkeitsfunktion der verallgemeinerten Poissonverteilung bestimmt:

$$\pi(x, t) = P_x(\theta t, \lambda t) = \frac{\theta t (\theta t + x \lambda t)^{x-1}}{x!} e^{-(\theta t + x \lambda t)} \quad (1.49)$$

*Beweis:*

Um  $\pi(x, t)$  zu bestimmen, betrachten wir zunächst den Fall  $x = 0$ . Die Wahrscheinlichkeit  $\pi(0, t+h)$ , dass im Intervall der Länge  $t+h$  kein Ereignis auftritt, ist nach den Axiomen 3 und 4 gleich dem Produkt der Wahrscheinlichkeiten  $\pi(0, t)$  (kein Ereignis im Intervall der Länge  $t$ ) und  $\pi(0, h)$  (kein Ereignis im Intervall der Länge  $h$ ).

$$\begin{aligned} \pi(0, t+h) &= \pi(0, t) P(N[t+h] - N[t] = 0 | N[t] = 0) \\ &\stackrel{(1.47)}{=} \pi(0, t) [1 - \theta h - o(h)] \end{aligned} \quad (1.50)$$

Wir subtrahieren von (1.50)  $\pi(0, t)$ , formen das Ergebnis um und dividieren durch  $h$ :

$$\begin{aligned} \pi(0, t+h) - \pi(0, t) &= \pi(0, t) [1 - \theta h - o(h)] - \pi(0, t) \\ &= -\theta h \pi(0, t) - o(h) \pi(0, t) \\ \Leftrightarrow \frac{\pi(0, t+h) - \pi(0, t)}{h} &= -\theta \pi(0, t) - \frac{o(h)}{h} \pi(0, t) \end{aligned} \quad (1.51)$$

Anschließend lassen wir in (1.51)  $h$  gegen 0 streben und erhalten somit eine lineare Differentialgleichung 1.Ordnung.

$$\frac{\partial \pi(0, t)}{\partial t} = -\theta \pi(0, t) \quad (1.52)$$

Differentialgleichung (1.52) besitzt die Lösung  $\pi(0, t) = ce^{-\theta t}$  (Råde and Westergren, 1997). Aus Axiom 1 folgt die Anfangswertbedingung  $\pi(0, 0) = 1$ , welche auf den Wert  $c = 1$  führt. Somit gilt:

$$\pi(0, t) = e^{-\theta t} \quad (1.53)$$

Kommen wir nun zum Fall  $x \neq 0$ .  $\pi(x, t+h)$  ist die Wahrscheinlichkeit, dass bis zum Zeitpunkt  $t+h$  genau  $x$  Ereignisse auftreten. Diese ist gleich der Summe der Wahrscheinlichkeiten, dass bis zum Zeitpunkt  $t$  genau  $x$  Ereignisse eintraten und im anschließenden Intervall  $(t, t+h]$  keines, sowie der Wahrscheinlichkeit, dass bis zum Zeitpunkt  $t$  genau  $x-1$  Ereignisse eintraten und im Intervall  $(t, t+h]$  genau eines.

$$\begin{aligned} \pi(x, t+h) &= \pi(x, t) P(N[t+h] - N[t] = 0 | N[t] = x) \\ &\quad + \pi(x-1, t) P(N[t+h] - N[t] = 1 | N[t] = x-1) \\ &\stackrel{(1.47), (1.48)}{=} \pi(x, t) [1 - (\theta + x \lambda) h - o(h)] \\ &\quad + \pi(x-1, t) \left[ \frac{(\theta + (x-1)\lambda) h e^{-\lambda t}}{(1 - \lambda(\theta + x \lambda)^{-1})^{x-1}} + o(h) \right] \end{aligned} \quad (1.54)$$

Umformen von (1.54) führt auf

$$\begin{aligned}
& \pi(x, t+h) - \pi(x, t) + (\theta + x\lambda)h\pi(x, t) + o(h)\pi(x, t) \\
&= \pi(x-1, t) \frac{(\theta + (x-1)\lambda)he^{-\lambda t}}{(1 - \lambda(\theta + x\lambda)^{-1})^{x-1}} + o(h)\pi(x-1, t) \\
\Rightarrow & \pi(x, t+h) - \pi(x, t) + (\theta + x\lambda)h\pi(x, t) \\
&= \pi(x-1, t) \frac{(\theta + (x-1)\lambda)he^{-\lambda t}}{(1 - \lambda(\theta + x\lambda)^{-1})^{x-1}} + o(h)[\pi(x-1, t) - \pi(x, t)] \\
\Rightarrow & \frac{\pi(x, t+h) - \pi(x, t)}{h} + (\theta + x\lambda)\pi(x, t) \\
&= \pi(x-1, t) \frac{(\theta + (x-1)\lambda)e^{-\lambda t}}{(1 - \lambda(\theta + x\lambda)^{-1})^{x-1}} + \frac{o(h)}{h}[\pi(x-1, t) - \pi(x, t)] \tag{1.55}
\end{aligned}$$

Für  $h \rightarrow 0$  folgt aus (1.55):

$$\frac{\partial \pi(x, t)}{\partial t} + (\theta + x\lambda)\pi(x, t) = \pi(x-1, t) \frac{(\theta + (x-1)\lambda)e^{-\lambda t}}{(1 - \lambda(\theta + x\lambda)^{-1})^{x-1}} \tag{1.56}$$

mit  $x = 1, 2, 3, \dots$

Für  $x = 1$  ergibt sich aus (1.56) die gewöhnliche lineare Differentialgleichung für  $t$

$$\begin{aligned}
\frac{\partial \pi(1, t)}{\partial t} + (\theta + \lambda)\pi(1, t) &= \pi(0, t)\theta^{-\lambda t} \\
&= \theta e^{-(\theta+\lambda)t} \tag{1.57}
\end{aligned}$$

Diese ist von der Form  $y' + a(t)y = b(t)$ . Die allgemeine Lösung für solch ein Problem lautet:

$$y = e^{-A(t)} \left[ \int e^{A(t)} b(t) dt + c \right]$$

mit  $A(t) = \int a(t) dt$ ,  $a(t) = \theta + \lambda$  und  $b(t) = \theta e^{-(\theta+\lambda)t}$  (Råde and Westergren, 1997, S.199).

$$\begin{aligned}
\Rightarrow \pi(1, t) &= \exp\left\{-\int (\theta + \lambda) dt\right\} \left[ \int \exp\left\{\int (\theta + \lambda) dt\right\} \theta \exp\{-(\theta + \lambda)t\} dt + c \right] \\
&= \exp\{-(\theta + \lambda)t\} \left[ \int \exp\{(\theta + \lambda)t\} \exp\{-(\theta + \lambda)t\} \theta dt + c \right] \\
&= \exp\{-(\theta + \lambda)t\} (\theta t + c) \tag{1.58}
\end{aligned}$$

Wegen der Anfangswertbedingung  $\pi(1, 0) = 0$  ist  $c = 0$ . (1.58) vereinfacht sich zu

$$\pi(1, t) = \theta t e^{-(\theta+\lambda)t} \tag{1.59}$$

Für  $x=2,3,\dots$  kann unter Verwendung der Randwertbedingungen  $\pi(x, 0) = 0 \quad \forall x > 0$  durch Induktion

$$\pi(x, t) = \frac{\theta(\theta + \lambda x)^{x-1} t^x e^{-\theta t - x\lambda t}}{x!} \tag{1.60}$$

gezeigt werden (Consul, 1989, S.38). Da  $\pi(x, t)$  als  $P(N(t) = k)$  definiert ist, folgt aus (1.60)

$$\begin{aligned} P(N(t) = x) &\stackrel{(1.60)}{=} \frac{\theta(\theta + \lambda x)^{x-1} t^x e^{-\theta t - x\lambda t}}{x!} \\ &= \frac{\theta(\theta + \lambda x)^{x-1} t t^{x-1}}{x!} e^{-\theta t - x\lambda t} \\ &= \frac{\theta t (\theta t + x\lambda t)^{x-1}}{x!} e^{-(\theta t + x\lambda t)} \end{aligned}$$

Damit folgt die Anzahl der Ereignisse  $N(t)$  in einem Intervall der Länge  $t$  der verallgemeinerten Poissonverteilung mit Parametern  $(\theta t, \lambda t)$ .

□

### 1.5.2 Verteilung der Anzahl an Nachkommen im Galton-Watson-Bienymé-Verzweigungsprozess

Charakteristisch für einen Verzweigungsprozess<sup>6</sup> ist die Übertragung eines bestimmten Merkmals auf gewisse Mitglieder einer Gruppe ausgehend von einer bestimmten Anzahl von Merkmalsträgern zum Start(zeit)punkt. Individuen, welche das Merkmal besitzen, verbreiten es auf Mitglieder anderer Gruppen. Dieser Prozess endet entweder wenn das Merkmal ausgestorben ist oder von der gesamte Population aufgewiesen wird.

$X_k$  bezeichne im Folgenden jeweils die Anzahl der Merkmalsträger in der  $k$ -ten Generation<sup>7</sup>. Der Verzweigungsprozess startet in der Stammgeneration (nullte Generation) mit  $X_0$  Merkmalsträgern.  $X_0$  wird durch eine Startverteilung bestimmt.

Der Übergang von Generation  $k$  zu Generation  $k+1$  wird für alle  $k=0,1,2,\dots$  durch dieselbe Wahrscheinlichkeitsverteilung gesteuert. Dabei gibt die Verteilung die Wahrscheinlichkeit wieder, an wie viele Individuen aus Generation  $k+1$  ein einzelner Merkmalsträger aus  $X_k$  das Merkmal weiter gibt. Die Anzahl  $X_{k+1}$  der Merkmalsträger in der  $k+1$ -ten Generation ist daher von  $X_k$ , d.h. der Anzahl der Merkmalsträger in der vorangegangenen Generation abhängig. Startverteilung und Übergangsverteilung müssen nicht übereinstimmen.

Von besonderem Interesse bei der Betrachtung solcher Verzweigungsprozesse ist die Wahrscheinlichkeitsverteilung welcher die Gesamtzahl  $Y_k = X_0 + X_1 + \dots + X_k$  aller Merkmalsträger bis zur  $k$ -ten Generation folgt. Werden über den Verzweigungsprozess gewisse Annahmen gemacht, so kann gezeigt werden, dass die Gesamtzahl der Merkmalsträger  $Y_n$  zum Zeitpunkt  $n$ , an dem der Prozess abbricht, verallgemeinert poissonverteilt ist (siehe Consul and Shoukri, 1988).

Der zu betrachtende Verzweigungsprozess erfülle nun die folgenden Eigenschaften:

- (i) Die Anzahl der Individuen innerhalb einer Gruppe ist groß
- (ii) Die Wahrscheinlichkeit, dass ein Individuum innerhalb der Gruppe das Merkmal annimmt ist klein
- (iii) Jeder Merkmalsträger wird für einen kurzen Zeitraum zum Verbreiter des Merkmals
- (iv) Die Anzahl der Individuen einer Gruppe, in welcher jedes Mitglied das Merkmal aufweist, wird sich vergrößern

<sup>6</sup>Verzweigungsprozesse werden z.B. von Resnick (1992) beschrieben. Der einfache oder Galton-Watson-Bienymé-Verzweigungsprozess wird dort auf Seite 18 ff. charakterisiert.

<sup>7</sup>Jede Generation entspricht einer Gruppe

Consul trifft die Aussage, dass  $X_0$  auf Grund der Eigenschaften (i) und (ii) poissonverteilt mit Intensität  $\theta$  sein muss (Consul, 1989, S.40). Die erzeugende Funktion der Startverteilung, durch die  $X_0$  bestimmt ist, lautet daher:

$$f(s) := E(s^{X_0}) = e^{\theta(s-1)} \quad (1.61)$$

Jeder Merkmalsträger in der Stammgeneration  $X_0$  wird zum Verbreiter des Merkmals. Jedes Individuum in der Stammgeneration welches das Merkmal besitzt, gibt dieses an  $\nu$  Individuen der ersten Generation weiter. Wegen der Eigenschaften (iii) und (iv) muss auch  $\nu$  poissonverteilt mit Intensität  $\lambda$  sein (Consul, 1989, S.41). Somit ist die erzeugende Funktion der Übergangswahrscheinlichkeit gegeben durch:

$$g(s) := E(s^\nu) = e^{\lambda(s-1)} \quad (1.62)$$

Alle Individuen, welche das Merkmal von einem Merkmalsträger aus  $X_0$  erhalten haben, bilden die erste Generation. Diese Individuen können das Merkmal wiederum an die folgende Generation übertragen, wobei jeder dieser Übertragungsvorgänge durch die zu (1.62) gehörende Wahrscheinlichkeitsverteilung gesteuert wird. Dieser Prozesse setzt sich bis zum Aussterben des Merkmals fort.

Annahme:  $P(X_0 = 1) = 1$ .

Im Folgenden weichen wir von der Verteilungsannahme (1.61) ab und setzen  $P(X_0 = 1) = 1$  voraus. Der Verzweigungsprozess startet also mit genau einem Merkmalsträger. Daher erfolgt die Übertragung von der Startgeneration zur ersten Generation von nur einem Individuum aus.

$$\Rightarrow g_0(s) := E(s^{X_0}) = s \quad (1.63)$$

$$g_1(s) := E(s^{X_1}) = g(s) \quad (1.64)$$

Für  $n=2,3,\dots$  lautet die erzeugende Funktion von  $X_n$ :

$$g_n(s) = \sum_{k=0}^{\infty} s^k P(X_n = k) \quad (1.65)$$

Consul and Shoukri (1988) zeigen, dass sich der Ausdruck (1.65) der erzeugenden Funktion so umformen lässt, dass wir

$$\begin{aligned} g_n(s) &= \sum_{l=0}^{\infty} g(s)^l P(X_{n-1} = l) \\ &= g_{n-1}(g(s)) \end{aligned} \quad (1.66)$$

erhalten. Setzt man die Identität  $g_2(s) = g_1(g(s)) = g(g_1(s))$  fort, so folgt aus Gleichung (1.66):

$$\begin{aligned} g_n(s) &= g_{n-1}(g(s)) \\ &= g(g_{n-1}(s)) \end{aligned} \quad (1.67)$$

Sobald  $P(X_n = 0) = 1$  für ein  $n \in \mathbb{N}$  gilt, bricht der Verzweigungsprozess ab. Damit dies geschieht, muss die erwartete Anzahl an Merkmalsträgern in der folgenden Generation jeweils kleiner sein als die Anzahl in der gegenwärtigen Generation. Dies bedeutet, dass die erwartete

Anzahl an Individuen in der folgenden Generation, die von einem Individuum aus der gegenwärtigen Generation das Merkmal erhalten, kleiner eins ist. Da der Generationenübergang durch die Poissonverteilung gesteuert wird, muss für den Parameter  $\lambda$  die Ungleichung:

$$\lambda = E(X_n) < 1$$

gelten. Der Erwartungswert ist stets gleich der ersten Ableitung der erzeugenden Funktion ausgewertet an der Stelle eins. Für die erzeugende Funktion folgt daraus in diesem Fall:

$$g'(1) = \lambda < 1$$

Es wird nun angenommen, dass das Merkmal nach der  $n$ -ten Generation ausgestorben ist:  $P(X_{n+i} = 0) = 1 \forall i \geq 1$ . Der Verzweigungsprozess stoppt folglich nach der  $n$ -ten Generation. Unser Ziel ist es daher, die Verteilung von  $Y_n = X_0 + X_1 + \dots + X_n$  zu bestimmen, wobei zunächst die Annahme  $X_0 = 1$  aufrecht erhalten wird.

Sei  $G_n(s)$  die erzeugende Funktion von  $Y_n$ . Die erzeugende Funktion  $G_1(s)$  von  $Y_1$  lässt sich somit wie folgt bestimmen:

$$\begin{aligned} G_1(s) &:= E(s^{Y_1}) = E(s^{X_0+X_1}) \\ &= E(s^{X_0})E(s^{X_1}) \\ &= sg(s) \end{aligned} \tag{1.68}$$

Für die Zufallsgrößen  $X_1, X_2, \dots, X_n$  welche durch den selben Mechanismus erzeugt werden, wird die Summe  $Z_n := X_1 + \dots + X_n$  eingeführt.  $Z_n$  ist wiederum eine zufällige Größe, deren Verteilung durch die erzeugende Funktion  $H_n(s)$  bestimmt sei. Wegen (1.64) gilt:

$$H_1(s) = E(s^{X_1}) = g(s) \tag{1.69}$$

Jeder Merkmalsträger aus  $X_1$  kann Merkmalsträger in der zweiten Generation erzeugen. Somit ergibt sich für die erzeugende Funktion von  $Z_2$  (Consul and Shoukri, 1988):

$$\begin{aligned} H_2(s) &\stackrel{(1.69)}{=} g(sH_1(s)) \stackrel{(1.68)}{=} g(G_1(s)) \\ \Rightarrow H_n(s) &= g(G_{n-1}(s)) \quad \forall n \in \mathbb{N} \geq 2 \end{aligned} \tag{1.70}$$

Die erzeugende Funktion von  $Y_n = 1 + Z_n$  ist demnach

$$\begin{aligned} G_n(s) &= E(s^{Y_n}) = sE(s^{Z_n}) \\ &= sH_n(s) \\ &= sg(G_{n-1}(s)) \end{aligned} \tag{1.71}$$

Für die erzeugenden Funktionen  $G_n(s)$  bilden wir nun den Grenzwert für  $n \rightarrow \infty$ .

$$\begin{aligned} G(s) &:= \lim_{n \rightarrow \infty} G_n(s) \\ &= sg\left(\lim_{n \rightarrow \infty} G_{n-1}(s)\right) \\ &= sg(G(s)) \end{aligned} \tag{1.72}$$

Unter Verwendung von (1.62) und der Substitution  $G(s) = t$  ergibt sich aus (1.72) die Gleichung

$$\begin{aligned} t &= sg(t) \\ &= se^{\lambda(t-1)} \end{aligned} \tag{1.73}$$



Consul and Shoukri (1988) wenden nun Lagrangeentwicklung an. Sie entwickeln  $t$  als Funktion von  $s$  und erhalten dadurch die Wahrscheinlichkeitsverteilung

$$P(Y = j | X_0 = 1) = \frac{(j\lambda)^{j-1}}{j!} e^{-j\lambda} \quad (1.74)$$

für die Gesamtanzahl der Merkmalsträger des Verzweigungsprozesses, wenn dieser mit einem einzigen Individuum startet. Die Verteilung mit Wahrscheinlichkeitsfunktion (1.74) wird als Borelverteilung bezeichnet.

Wir verändern die getroffenen Annahmen nun dahingehend, dass die Stammgeneration aus einer bekannten Anzahl von  $N \in \mathbb{N}$  Individuen bestehe. In diesem Fall gilt  $P(X_0 = N) = 1$ , so dass sich unter Berücksichtigung von (1.61) die erzeugende Funktion (1.63) zu  $g_0(s) = s^N$  abändert. Aus (1.74) wird:

$$P(Y = j | X_0 = N) = \frac{N}{(j - N)!} j^{j-N-1} \lambda^{j-N} e^{-\lambda j}, \quad j = N, N + 1, \dots \quad (1.75)$$

(Consul, 1989, S.43). Die Verteilung mit Verteilungsfunktion (1.75) wird als Borel-Tanner-Verteilung bezeichnet.

Heben wir nun die Annahmen über den Wert von  $X_0$  auf und betrachten die Anzahl der Merkmalsträger in der Stammgeneration als unbekannt und zufällig. Die Zufallsvariable  $X_0$  besitze nun Poissonverteilung mit der erzeugenden Funktion (1.61). Es gilt folglich:

$$P(X_0 = N) = \frac{\theta^N}{N!} e^{-\theta}, \quad N \in \mathbb{N} \quad (1.76)$$

Die Verteilung der Gesamtanzahl an Merkmalsträger  $Y$  erhalten wir somit, indem wir (1.75) mit der Poisson-Eintrittswahrscheinlichkeit (1.76) multiplizieren und über alle  $N \in \mathbb{N}$  summieren.

$$\begin{aligned} P(Y = j) &= \sum_{N=0}^{\infty} P(Y = j | X_0 = N) P(X_0 = N) \\ &= \sum_{N=0}^{\infty} \frac{N}{(j - N)!} j^{j-N-1} \lambda^{j-N} e^{-\lambda j} \frac{\theta^N}{N!} e^{-\theta} \\ &= \frac{\theta(\theta + \lambda j)^{j-1}}{j!} e^{-(\theta + \lambda j)}, \quad j \in \mathbb{N} \end{aligned} \quad (1.77)$$

Da (1.77) mit der Wahrscheinlichkeitsfunktion (1.1) übereinstimmt, besitzt die Anzahl  $Y$  von Merkmalsträgern im gesamten Verzweigungsprozess verallgemeinerte Poissonverteilung mit Parametern  $\theta$  und  $\lambda$ .

Fassen wir die Erkenntnisse von Consul and Shoukri (1988) zusammen, so lässt sich folgende Aussage treffen:

Ist die Zufallsvariable  $X_0 \sim Poi(\theta)$

unabhängig von den Zufallsvariablen  $X_k \sim Poi(\lambda) \quad \forall k \in \mathbb{N}$

und gilt für die  $X_k$ ,  $k \in \mathbb{N}$  oben beschriebene Abhängigkeitsstruktur, so folgt:

$$Y = \sum_{k=0}^{\infty} X_k \sim GP(\theta, \lambda), \quad \text{falls } \lambda < 1$$

## 1.6 Grenzverteilungen der verallgemeinerten Poissonverteilung

### 1.6.1 Normalverteilung als Grenzverteilung

**Definition 1.15 (Kumulantenfunktion & Kumulante)** Falls die momenterzeugende Funktion der Zufallsvariable  $X$  existiert, so nennt man

$$\psi_X(s) := \ln [E(e^{sX})] = \ln [M_X(s)] \quad (1.78)$$

Kumulantenfunktion<sup>8</sup> von  $X$ , falls eine Entwicklung der Form

$$\psi_X(s) = \sum_{k=0}^{\infty} \kappa_k \frac{s^k}{k!} \quad (1.79)$$

in einer Umgebung um 0 existiert.

Die Größe  $\kappa_k$  wird hierbei als  $k$ -te Kumulante<sup>9</sup> der Zufallsvariable  $X$  bezeichnet.

**Lemma 1.16 (Eigenschaft der Kumulanten)**  $\psi_X$  sei die Kumulantenfunktion der Zufallsvariable  $X$ . Für die Kumulanten von  $X$  gilt:

$$\kappa_k = \psi_X^{(k)}(0), \quad k = 0, 1, 2, \dots \quad (1.80)$$

*Beweis:*

(Stirzaker, 1994, S.189)

**Bemerkung:**

Insbesondere folgt aus Lemma 1.16:

$$\begin{aligned} \kappa_1 &\stackrel{(1.80)}{=} \psi_X'(0) \stackrel{(1.13),(1.78)}{=} \left. \frac{\partial}{\partial s} \ln[M_X(s)] \right|_{s=0} = \left. \frac{1}{M_X(s)} M_X'(s) \right|_{s=0} \\ &= M_X'(0) \stackrel{(1.16)}{=} \mu_1 = E(X) \end{aligned} \quad (1.81)$$

**Satz 1.17 (Kumulantenfunktion der verallgemeinerten Poissonverteilung)** Die Zufallsvariable  $X$  folge der Verteilung  $GP(\theta, \lambda)$ . Dann lautet die Kumulantenfunktion von  $X$

$$\psi_X(v) = \theta(e^s - 1) \quad (1.82)$$

mit  $s = v + \lambda(e^s - 1)$ .

*Beweis:*

Wir setzen die momenterzeugende Funktion von  $X$  (1.14) in Definition 1.15 ein:

$$\begin{aligned} \psi_X(v) &= \ln [E(e^{vX})] \\ &\stackrel{(1.14)}{=} \ln [M_X(v)] = \theta(e^s - 1) \end{aligned}$$

wobei  $s$  durch (1.19) als  $v + \lambda(e^s - 1)$  bestimmt ist.

□

<sup>8</sup>auch: kumulantenenerzeugende Funktion, in Anlehnung an die engl. Bezeichnung 'cumulant generating function'

<sup>9</sup>auch: Semiinvariante

**Satz 1.18 (Kumulanten der verallgemeinerten Poissonverteilung)** *Es gelte  $X \sim GP(\theta, \lambda)$ . Für die Kumulanten der Zufallsvariable  $X$  gilt dann die Rekursionsgleichung*

$$(1 - \lambda)\kappa_{k+1} = \lambda \frac{\partial \kappa_k}{\partial \lambda} + \theta \frac{\partial \kappa_k}{\partial \theta}, \quad k = 1, 2, 3, \dots \quad (1.83)$$

wobei die erste Kumulante  $\kappa_1 = \frac{\theta}{1-\lambda}$  ist.

*Beweis:*

Gemäß der allgemeinen Übereinstimmung des ersten Moments  $\mu_1$  mit der ersten Kumulante  $\kappa_1$  (1.81) gilt unter Verwendung des Erwartungswertes (1.4):

$$\kappa_1 = \frac{\theta}{1 - \lambda} \quad (1.84)$$

Die Kumulantenfunktion  $\psi_X(v)$  hängt von  $v$ ,  $\theta$  und  $\lambda$  ab, da  $s$  eine Funktion von  $v$  und  $\lambda$  ist. Für die genannten Variablen bilden wir die partiellen Ableitungen von  $\psi_X(v)$ .

$$\frac{\partial \psi_X(v)}{\partial \theta} = \frac{\partial}{\partial \theta} [\theta(e^s - 1)] = e^s - 1 \quad (1.85)$$

$$\frac{\partial \psi_X(v)}{\partial v} = \frac{\partial}{\partial v} [\theta(e^s - 1)] = \theta e^s \frac{\partial s}{\partial v} \quad (1.86)$$

$$\frac{\partial \psi_X(v)}{\partial \lambda} = \frac{\partial}{\partial \lambda} [\theta(e^s - 1)] = \theta e^s \frac{\partial s}{\partial \lambda} \quad (1.87)$$

In Abschnitt 1.3.4 wurde die Gleichung (1.25) hergeleitet. Diese multiplizieren wir mit  $\theta(1 + \frac{\partial \psi_X(v)}{\partial \theta})$ :

$$\begin{aligned} 1 &= (1 - \lambda) \frac{\partial s}{\partial v} - \lambda \frac{\partial s}{\partial \lambda} \\ \Rightarrow \theta \left(1 + \frac{\partial \psi_X(v)}{\partial \theta}\right) &= \theta \left(1 + \frac{\partial \psi_X(v)}{\partial \theta}\right) (1 - \lambda) \frac{\partial s}{\partial v} - \theta \left(1 + \frac{\partial \psi_X(v)}{\partial \theta}\right) \lambda \frac{\partial s}{\partial \lambda} \end{aligned} \quad (1.88)$$

Setzen wir das Resultat (1.85) in (1.88) ein so erhalten wir:

$$\begin{aligned} \theta \left(1 + \frac{\partial \psi_X(v)}{\partial \theta}\right) &= \theta (1 + (e^s - 1)) (1 - \lambda) \frac{\partial s}{\partial v} - \theta (1 + (e^s - 1)) \lambda \frac{\partial s}{\partial \lambda} \\ \Rightarrow \theta \left(1 + \frac{\partial \psi_X(v)}{\partial \theta}\right) &= \theta e^s \frac{\partial s}{\partial v} (1 - \lambda) - \theta \frac{\partial s}{\partial \lambda} \lambda e^s \end{aligned} \quad (1.89)$$

Nun können wir auch die Resultate (1.86) und (1.87) verwenden. Wir setzen diese in (1.89) ein.

$$\Rightarrow \theta + \theta \frac{\partial \psi_X(v)}{\partial \theta} = (1 - \lambda) \frac{\partial \psi_X(v)}{\partial v} - \lambda \frac{\partial \psi_X(v)}{\partial \lambda}$$

Umstellen dieser Gleichung führt auf:

$$(1 - \lambda) \frac{\partial \psi_X(v)}{\partial v} - \theta = \lambda \frac{\partial \psi_X(v)}{\partial \lambda} + \theta \frac{\partial \psi_X(v)}{\partial \theta} \quad (1.90)$$

Die Rekursionsformel (1.83) erhalten wir, indem wir in der Gleichung (1.90) die Kumulantenfunktion  $\psi_X(v)$  durch ihre Potenzreihenentwicklung (vgl. Definition 1.15) ersetzen und nach den Koeffizienten von  $\frac{v^k}{k!}$  auflösen.

Wir bestimmen nun mit der eben hergeleiteten Formel (1.83) und dem Wert der ersten Kumulante (1.84) rekursiv die zweite, dritte und vierte Kumulanten der verallgemeinerten Poissonverteilung:

$$\begin{aligned}
 (1-\lambda)\kappa_2 &= \lambda \frac{\partial \kappa_1}{\partial \lambda} + \theta \frac{\partial \kappa_1}{\partial \theta} = \lambda \frac{\theta}{(1-\lambda)^2} + \theta \frac{1}{1-\lambda} \\
 &= \frac{\lambda\theta + \theta - \lambda\theta}{(1-\lambda)^2} \\
 \Rightarrow \kappa_2 &= \frac{\theta}{(1-\lambda)^3} \tag{1.91}
 \end{aligned}$$

$$\begin{aligned}
 (1-\lambda)\kappa_3 &= \lambda \frac{\partial \kappa_2}{\partial \lambda} + \theta \frac{\partial \kappa_2}{\partial \theta} = \lambda \frac{3\theta}{(1-\lambda)^4} + \theta \frac{1}{(1-\lambda)^3} \\
 &= \frac{3\theta\lambda + \theta(1-\lambda)}{(1-\lambda)^4} \\
 \Rightarrow \kappa_3 &= \frac{\theta(1+2\lambda)}{(1-\lambda)^5} \tag{1.92}
 \end{aligned}$$

$$\begin{aligned}
 (1-\lambda)\kappa_4 &= \lambda \frac{\partial \kappa_3}{\partial \lambda} + \theta \frac{\partial \kappa_3}{\partial \theta} = \lambda \frac{\theta(8\lambda+7)}{(1-\lambda)^6} + \theta \frac{1+2\lambda}{(1-\lambda)^5} \\
 &= \frac{\lambda\theta(8\lambda+7) + \theta(1+2\lambda)(1-\lambda)}{(1-\lambda)^6} \\
 &= \theta \frac{8\lambda^2 + 7\lambda + 1 + \lambda + 2\lambda - 2\lambda^2}{(1-\lambda)^6} \\
 \Rightarrow \kappa_4 &= \theta \frac{1+8\lambda+6\lambda^2}{(1-\lambda)^7} \tag{1.93}
 \end{aligned}$$

□

**Satz 1.19 (Normalverteilung als Grenzverteilung)** *Es gelte  $X \sim GP(\theta, \lambda)$ .  $\mu$  und  $\sigma^2$  seien der Erwartungswert bzw. die Varianz der Zufallsvariable  $X$ .  $Y$  sei eine standard normalverteilte Zufallsvariable.*

$$\Rightarrow Z := \frac{X - \mu}{\sigma} \xrightarrow{\mathcal{L}} Y \quad \forall \lambda \tag{1.94}$$

für  $\theta \rightarrow \infty$ .

Dabei bezeichnet  $\xrightarrow{\mathcal{L}}$  Konvergenz in Verteilung (siehe z.B. Råde and Westergren, 1997, S.411)

*Beweis:*

Die Kumulantenfunktion der Zufallsvariable  $Z$  lautet:

$$\begin{aligned}
 \psi_Z(t) &= \ln[E(e^{tZ})] = \ln[E(e^{t\frac{X-\mu}{\sigma}})] \\
 &= \ln[E(e^{\frac{tX}{\sigma}}) e^{-\frac{t\mu}{\sigma}}] \\
 &= -\frac{t\mu}{\sigma} + \ln[E(e^{\frac{tX}{\sigma}})] \\
 &= -\frac{t\mu}{\sigma} + \psi_X\left(\frac{t}{\sigma}\right) \tag{1.95}
 \end{aligned}$$

In (1.95) ersetzen wir die Kumulantenfunktion  $\psi_X$  nun durch ihre Reihendarstellung (1.79).

$$\begin{aligned}\psi_Z(t) &= -\frac{t\mu}{\sigma} + \sum_{k=1}^{\infty} \kappa_k \frac{\left(\frac{t}{\sigma}\right)^k}{k!} \\ &= -\frac{t\mu}{\sigma} + \frac{\kappa_1 t}{1! \sigma} + \frac{\kappa_2 t^2}{2! \sigma^2} + \frac{\kappa_3 t^3}{3! \sigma^3} + \frac{\kappa_4 t^4}{4! \sigma^4} + \dots\end{aligned}\quad (1.96)$$

$\kappa_1$  entspricht dem Erwartungswert, also  $\mu$ . Somit heben sich der erste und der zweite Summand in (1.96) gegenseitig auf. Der dritte Summand vereinfacht sich zu  $\frac{t^2}{2}$ , da gemäß (1.91)  $\kappa_2$  gleich der Varianz  $\sigma^2$  ist. Die weiter oben bestimmten Kumulanten (1.92) und (1.93) werden nun ebenfalls in (1.96) eingesetzt:

$$\psi_Z(t) = \frac{t^2}{2} + \frac{t^3 \theta(1+2\lambda)}{3! (1-\lambda)^5} \frac{1}{\sigma^3} + \frac{t^4 \theta(1+8\lambda+6\lambda^2)}{4! (1-\lambda)^7} \frac{1}{\sigma^4} + \dots\quad (1.97)$$

Wie zu Beginn festgelegt, ist  $\sigma^2$  gleich der Varianz der verallgemeinert poissonverteilten Zufallsvariable X.

$$\Rightarrow \sigma = \sqrt{\frac{\theta}{(1-\lambda)^3}}\quad (1.98)$$

Setzen wir den Wert (1.98) von  $\sigma$  in (1.97) ein, so gilt:

$$\begin{aligned}\psi_Z(t) &= \frac{t^2}{2} + \frac{t^3 \theta(1+2\lambda)}{3! (1-\lambda)^5} \frac{1}{\left(\sqrt{\frac{\theta}{(1-\lambda)^3}}\right)^3} + \frac{t^4 \theta(1+8\lambda+6\lambda^2)}{4! (1-\lambda)^7} \frac{1}{\left(\frac{\theta}{(1-\lambda)^3}\right)^2} + \dots \\ &= \frac{t^2}{2} + \frac{t^3 \theta(1+2\lambda)}{3! (1-\lambda)^5} \frac{(1-\lambda)^3}{\theta} \sqrt{\frac{(1-\lambda)^3}{\theta}} + \frac{t^4 \theta(1+8\lambda+6\lambda^2)}{4! (1-\lambda)^7} \frac{(1-\lambda)^6}{\theta^2} + \dots \\ &= \frac{t^2}{2} + \frac{t^3}{3!} \frac{1+2\lambda}{\sqrt{\theta(1-\lambda)}} + \frac{t^4}{4!} \frac{1+8\lambda+6\lambda^2}{\theta(1-\lambda)} + \dots\end{aligned}\quad (1.99)$$

Führt man die Berechnung für höhere Kumulanten durch, dann wird noch deutlicher ersichtlich, dass jeweils in dem Term, welcher mit der k-ten Kumulante korrespondiert, die Größe  $\theta^{\frac{k-2}{2}}$  im Nenner steht.

Für  $\theta \rightarrow \infty$  gilt daher:

$$\psi_Z(t) = \frac{t^2}{2}$$

Die Kumulantenfunktion der Normalverteilung lautet für  $Y \sim N(\mu, \sigma^2)$  (Kredler, 1998, S.93):

$$\psi_Y(t) = \mu t + \frac{1}{2} \sigma^2 t^2$$

Für  $\mu = 0$  und  $\sigma^2 = 1$  stimmen  $\psi_Y(t)$  und der Grenzwert von  $\psi_Z(t)$  für  $\theta \rightarrow \infty$  überein. Die Zufallsvariable Z konvergiert somit in Verteilung gegen die Standardnormalverteilung.

□

## 1.6.2 Inverse Gaußverteilung als Grenzverteilung

Eine weitere Grenzverteilung der verallgemeinerten Poissonverteilung ist die Inverse Gaußverteilung. Consul (1989, S.56) beweist folgenden Satz:

**Satz 1.20** *Es gelte  $X \sim GP(\theta, \lambda)$ .  $\sigma^2$  sei die Varianz der Zufallsvariable  $X$ . Die Zufallsvariable*

$$Z := \frac{X}{\sigma}$$

*folgt asymptotisch der Inversen Gaußverteilung mit Erwartungswert  $\nu$  und Varianz 1, falls  $\theta \rightarrow \infty$  und  $\lambda \rightarrow 1$  unter Einhaltung der Gleichung*

$$\frac{\theta}{1 - \lambda} = \nu^2$$

## 1.7 Erzeugung von Zufallsvariablen

Für Testzwecke werden bei der Untersuchung verallgemeinert poissonverteilter Daten Zufallsvariablen benötigt, welche der verallgemeinerten Poissonverteilung folgen. Von Programmpaketen werden üblicherweise nur Zufallsgeneratoren für die gängigen Verteilungen zur Verfügung gestellt. Solche Zufallsgeneratoren liefern Pseudozufallsvariablen. Diese sind zwar deterministisch, werden aber von gewöhnlichen statistischen Tests, welche die Zufälligkeit untersuchen als zufällige Größen akzeptiert.

Bei der Generierung von (Pseudo-)Zufallsvariablen unterscheidet man verteilungsspezifische Methoden, welche spezielle Eigenschaften der zu untersuchenden Verteilung ausnützen und nicht-verteilungsspezifische Methoden. Die einfachste nicht-verteilungsspezifische Methode besteht in der Verwendung der sog. Smirnovtransformation<sup>10</sup>. Dazu wird ein Zufallsgenerator benötigt, welcher gleichverteilte Zufallsvariablen liefert.

**Satz 1.21 (Smirnovtransformation)** *Sei  $X$  gleichverteilte Zufallsvariable auf  $[0, 1]$  und  $F(t)$  eine stetige Verteilungsfunktion. Dann folgt die Zufallsvariable  $\eta = F^{-1}(X)$  der Wahrscheinlichkeitsverteilung mit Verteilungsfunktion  $F(t)$ .*

*Beweis:*

Falls  $F^{-1}$  existiert, sei  $\tilde{F}(t)$  die Verteilungsfunktion von  $\eta = F^{-1}(X)$

$$\begin{aligned} \Rightarrow \tilde{F}(t) &= P(\eta < t) \\ &= P(F^{-1}(X) < t) \end{aligned}$$

Als Verteilungsfunktion ist  $F$  monoton steigend. Daher gilt:

$$\begin{aligned} P(F^{-1}(X) < t) &= P(F(F^{-1}(X)) < F(t)) \\ &= P(X < F(t)) \end{aligned}$$

Da  $X$  gleichverteilt auf dem Intervall  $[0, 1]$  ist, folgt weiterhin:

$$P(X < F(t)) = F(t)$$

Damit besitzt die Verteilung, welcher  $\eta$  folgt, die Verteilungsfunktion  $\tilde{F}(t) = F(t)$ .

□

<sup>10</sup>auch: probability integral transformation

Die Verteilungsfunktion der verallgemeinerten Poissonverteilung lautet:

$$F(X) = P(X \leq x) = \sum_{i=0}^x P_i(\theta, \lambda)$$

Da es sich bei der verallgemeinerten Poissonverteilung um eine diskrete Verteilung handelt, lässt sich die Umkehrung der Verteilungsfunktion nicht direkt bestimmen. Für  $U \sim \text{unif}[0, 1]$  gilt jedoch

$$\begin{aligned} P(F(x-1) < U \leq F(x)) &= \int_{F(x-1)}^{F(x)} 1 du \\ &= F(x) - F(x-1) \\ &= P_x(\theta, \lambda) \end{aligned}$$

Damit ist  $x$  Realisierung einer verallgemeinert poissonverteilten Zufallsvariable  $X$ , sofern die Ungleichung

$$F(x-1) < U \leq F(x) \tag{1.100}$$

für die auf  $[0, 1]$  gleichverteilte Zufallsvariable  $U$  erfüllt ist.

### 1.7.1 Invertierungsmethode

Bei der Invertierungsmethode wird zunächst eine Zufallsvariable  $U \sim \text{unif}[0, 1]$  erzeugt und anschließend die eben beschriebene Transformation von  $U$  in eine verallgemeinert poissonverteilte Zufallsvariable  $X$  vorgenommen.

Dazu muss über einen Suchalgorithmus ein  $x$  gesucht werden, für welches die Ungleichung (1.100) erfüllt ist. Der einfachster Suchalgorithmus ist die sequentielle Suche. Hier wird mit  $x = 0$  begonnen. In jedem nun folgenden Iterationsschritt, wird  $F(x)$  und  $F(x-1)$  mit  $U$  verglichen. Die Größe  $x$  wird solange um 1 erhöht, bis (1.100) erfüllt ist.

Um die rechenintensive Bestimmung von  $P_x(\theta, \lambda)$  in jedem Iterationsschritt des Verfahrens zu vermeiden, schlägt Famoye (1997) vor, die rekurrente Beziehung zwischen den Eintrittswahrscheinlichkeiten der verallgemeinerten Poissonverteilung auszunutzen.

**Satz 1.22** Mit  $P_0(\theta, \lambda) = \frac{\theta \theta^{-1}}{1!} e^{-\theta} = e^{-1}$  gilt für die Eintrittswahrscheinlichkeiten der verallgemeinerten Poissonverteilung die Rekursionsgleichung

$$P_x(\theta, \lambda) = \frac{\theta + (x-1)\lambda}{x} \left( 1 + \frac{\lambda}{\theta + (x-1)\lambda} \right)^{x-1} e^{-\lambda} P_{x-1}(\theta, \lambda) \tag{1.101}$$

für  $x \leq 1$ .

*Beweis:*

$$\begin{aligned} & \frac{\theta + (x-1)\lambda}{x} \left( 1 + \frac{\lambda}{\theta + (x-1)\lambda} \right)^{x-1} e^{-\lambda} P_{x-1}(\theta, \lambda) \\ = & \frac{\theta + (x-1)\lambda}{x} \left( 1 + \frac{\lambda}{\theta + (x-1)\lambda} \right)^{x-1} e^{-\lambda} \frac{\theta(\theta + (x-1)\lambda)^{x-2}}{(x-1)!} e^{-(\theta+(x-1)\lambda)} \end{aligned}$$

$$\begin{aligned}
&= \frac{\theta(\theta + (x-1)\lambda)^{x-1}}{x!} \left( \frac{\theta + \lambda x}{\theta + (x-1)\lambda} \right)^{x-1} e^{-(\theta+\lambda x)} \\
&= \frac{\theta(\theta + x\lambda)^{x-1}}{x!} e^{-(\theta+\lambda x)} \\
&= P_x(\theta, \lambda)
\end{aligned}$$

□

Die Rekursionsformel (1.101) ist numerisch nicht stabil. Es droht Auslöschung (Exponentenunterlauf) durch Rundungsfehler. Um diese Fehlerquelle zu begrenzen, sollte mit doppelter Genauigkeit gerechnet werden.

### Algorithmus 1:

- Initialisierung:
  1.  $x \leftarrow 0$
  2.  $p \leftarrow e^{-\theta}$
  3.  $s \leftarrow p$
  4. Erzeuge eine Realisierung  $u$  von  $U \sim \text{unif}[0, 1]$
- Iteration:
  1. Wiederhole 2. - 5. solange  $u < p$
  2.  $x \leftarrow x + 1$
  3.  $c \leftarrow \theta + (x-1)\lambda$
  4.  $p \leftarrow \frac{c}{x} \left(1 + \frac{\lambda}{c}\right)^{x-1} e^{-\lambda} p$
  5.  $s \leftarrow s + p$
- Ausgabe:  $x$

### 1.7.2 Verzweigungsmethode

In Abschnitt 1.5.1 wurde beschrieben, auf welche Weise eine verallgemeinert poissonverteilte Zufallsvariable über einen Galton-Watson-Bienymé-Verzweigungsprozess gewonnen werden kann. Dies machen wir uns nun zu nutze.

$X_0 \sim \text{Poi}(\theta)$  sei die Anzahl der Merkmalsträger zu Beginn des Prozesses,  $X_k$  die Anzahl der Merkmalsträger in der  $k$ -ten Generation. Jedes Individuum, welches das Merkmal besitzt, überträgt es auf die nachfolgende Generation gemäß der Poissonverteilung mit Parameter  $\lambda < 1$ . Der Ausbreitungsprozess endet nach der  $n$ -ten Generation mit dem Aussterben des Merkmals. Somit besitzt  $Y_n = X_0 + X_1 + \dots + X_n$  die Verteilung  $GP(\theta, \lambda)$ .

### Algorithmus 2:

- Initialisierung:
  1. Erzeuge eine Realisierung  $x$  von  $X \sim \text{Poi}(\theta)$
  2.  $y \leftarrow x$
  3. Gehe zur Ausgabe, falls  $x \leq 0$



- Iteration:
  1. Erzeuge eine Realisierung  $z$  von  $Z \sim Poi(x\lambda)$  <sup>11</sup>
  2.  $y \leftarrow y + z$
  3.  $x \leftarrow z$
  4. Wiederhole 1. - 3. solange  $x > 0$
- Ausgabe:  $y$

### Bemerkungen:

- (i) Mit Algorithmus 2 können nur Zufallsvariablen erzeugt werden, wenn  $\lambda$  positiv ist.
- (ii) Die beiden iterativen Algorithmen 1 und 2 wurden von Famoye (1997) hinsichtlich ihrer Geschwindigkeit untersucht. Dazu wurde jeweils die benötigte Rechenzeit zur Erzeugung von 1000 verallgemeinert poissonverteilter Zufallszahlen gemessen. Diese Messungen wurden für verschiedene Parameterkombinationen von  $\theta$  und  $\lambda$  vorgenommen. Bei der Implementierung eines Generators für verallgemeinert poissonverteilte Pseudozufallszahlen, bietet es sich auf Grund dieser Untersuchungen an, die verwendete Methode zur Generierung der Zufallszahl an die Parameterwerte anzupassen. So wird in der S-Plus Implementierung in Anhang C.1.1 für positives  $\lambda$  die Verzweigungsmethode und für negatives  $\lambda$  die Invertierungsmethode zur Erzeugung der Zufallsgröße verwendet.

### 1.7.3 Approximationsmethode

Famoye (1997) stellt eine weitere Möglichkeit vor, verallgemeinert poissonverteilte Zufallszahlen zu erzeugen. Dieser beruht auf einem Ansatz über Grenzverteilungen. In Abschnitt 1.6 wurde gezeigt, dass die verallgemeinerte Poissonverteilung als Grenzverteilung für große Werte von  $\theta$  die Normalverteilung besitzt. Mit der standardnormalverteilten Zufallsvariable

$$Z \sim N(0, 1)$$

sowie

$$\begin{aligned} \mu &= E(Y) = \frac{\theta}{1 - \lambda} \\ \sigma &= \sqrt{Var(Y)} = \sqrt{\frac{\theta}{(1 - \lambda)^3}} \end{aligned}$$

kann die Größe

$$X := \max\{0, \lfloor \mu + \sigma Z + 0.5 \rfloor\} \quad (1.102)$$

für großes  $\theta$  und  $\lambda$  im angegebenen Bereich als Approximation an die  $GP(\theta, \lambda)$ -verteilte Zufallsvariable  $Y$  herangezogen werden. Der Ausdruck, welcher sich innerhalb der Gaußklammer  $\lfloor \cdot \rfloor$  befindet, wird auf die nächst kleinere, ganze Zahl abgerundet. Auf Grund der Definition 1.1 sind für  $Y$  nur positive Werte zulässig. Daher wird durch Maximumbildung die Größe  $X$  in (1.102) auf positive Werte eingeschränkt. Weiterhin handelt es sich bei der verallgemeinerten Poissonverteilung um eine diskrete Verteilung. Durch die Gaußklammer wird sichergestellt, dass die Realisierungen von  $X$  ganzzahlig sind. Damit trotz der Beschränkung auf positive Werte

<sup>11</sup>Jeder der  $x$  Merkmalsträger überträgt das Merkmal gemäß Poissonverteilung mit Erwartungswert  $\lambda$

der Abschnitt der Gaußkurve von  $-0.5$  bis  $0$  berücksichtigt wird, verschiebt man diese durch Addition von  $0.5$  nach rechts.

**Algorithmus 3:**

- Initialisierung:

1.  $\mu \leftarrow \theta(1 - \lambda)^{-1}$
2.  $\sigma \leftarrow \sqrt{\theta(1 - \lambda)^{-3}}$

- Berechnung:

1. Erzeuge Realisierung  $y$  einer standard normalverteilten Zufallsvariable  $Y$
2.  $x \leftarrow \max(0, \lfloor \mu + \sigma y + 0.5 \rfloor)$

- Ausgabe:  $x$

Um den Approximationsfehler zu messen vergleicht Famoye (1997) die Verteilungsfunktion der verallgemeinerten Poissonverteilung ( $P(X \leq x)$ ,  $X \sim GP(\theta, \lambda)$ ) mit der Verteilungsfunktion der Normalverteilung ( $P(Z \leq z)$  mit  $z = \frac{x - \mu + 0.5}{\sigma}$ ) für den korrespondierenden Wert zu  $x$ . Für negatives  $\lambda$  ist die Abweichung zwischen  $P(X \leq x)$  und  $P(Z \leq z)$  bereits für  $\theta \geq 10$  hinreichend klein. Steigt  $\lambda$  an, so vergrößert sich auch die Differenz. Die Normalverteilungsapproximation kann jedoch auch für  $\lambda \in [0, 0.2]$  verwendet werden, falls  $\theta \geq 30$ .

Der obige Algorithmus ist auf Grund seiner sehr einfachen Struktur wesentlich schneller als die ersten beiden Algorithmen zur Generierung von Zufallszahlen. Er eignet sich daher vor allem in Situationen, in denen die Rechenzeit für die Erzeugung der Zufallszahlen knapp bemessen ist. Die eben genannten Einschränkungen haben jedoch zur Folge, dass er nur sehr eingeschränkt einsetzbar ist.

## Kapitel 2

# Vergleich von verallgemeinerter Poisson- und negativer Binomialverteilung

### 2.1 Die negative Binomialverteilung

Bei der Modellierung von Zählraten ist die Verwendung der Poissonverteilung bei unter- oder überdispersionierten Daten ab einem gewissen Dispersionsgrad nicht mehr sinnvoll. Für diesen Fall wurde mit der verallgemeinerten Poissonverteilung in Kapitel 1 eine Verteilungsalternative zur Poissonverteilung behandelt. Weitaus verbreiteter für die Modellierung von Zählraten mit Überdispersion ist jedoch die negative Binomialverteilung.

Um einen Vergleich mit der verallgemeinerten Poissonverteilung vornehmen zu können, soll im Folgenden zunächst die negative Binomialverteilung definiert und ihre grundlegenden Eigenschaften genannt werden.

**Definition 2.1 (Negative Binomialverteilung mit ganzzahligem Parameter)** Eine Zufallsvariable  $X$  mit der Wahrscheinlichkeitsfunktion

$$P_x(r, p) = P(X = x) = \begin{cases} \binom{x+r-1}{x} p^r (1-p)^x & x = 0, 1, 2, \dots \\ 0 & \text{sonst} \end{cases} \quad (2.1)$$

heißt negativ binomialverteilt mit Parametern  $r \in \mathbb{N}$  und  $p \in (0, 1)$ . Kurz  $X \sim NB^*(r, p)$ .

Das Ereignis  $X=x$  lässt sich in diesem Zusammenhang als Anzahl der Fehlversuche bis zum  $r$ -ten Erfolg bei einer Reihe unabhängiger Bernoulli-Versuchen mit Erfolgswahrscheinlichkeiten  $p$  interpretieren.

#### 2.1.1 Erwartungswert und Varianz

**Satz 2.2 (Momenterzeugende Funktion der negativen Binomialverteilung mit  $r \in \mathbb{N}$ )** Sei  $X \sim NB^*(r, p)$ . Dann wird die momenterzeugende Funktion der Zufallsvariablen  $X$  durch

$$M_X(v) = E(e^{vX}) = \left( \frac{p}{1 - (1-p)e^v} \right)^r \quad (2.2)$$

mit  $v < |\ln(1-p)|$  bestimmt.

*Beweis:* (siehe z.B. Johnson et al., 1993)

Erwartungswert und Varianz lassen sich über die nicht zentrierten Momente (vgl. Definition 1.8) bestimmen. Dabei gilt die Gleichung

$$\mu_k = M^{(k)}(0) = \frac{\partial^k M(0)}{\partial v^k}, \quad k \in \mathbb{N} \quad (2.3)$$

Mit der ersten Ableitung der momenterzeugenden Funktion können wir nun unter Verwendung von Gleichung (2.3) den Erwartungswert und die Varianz der negativen Binomialverteilung bestimmen.

**Satz 2.3 (Erwartungswert und Varianz der negativen Binomialverteilung mit  $r \in \mathbb{N}$ )**

Sei  $X \sim NB^*(r, p)$ . Erwartungswert und Varianz der Zufallsvariable  $X$  lauten dann:

$$E(X) = r \frac{1-p}{p} \quad (2.4)$$

$$\text{Var}(X) = r \frac{1-p}{p^2} \quad (2.5)$$

*Beweis:*

$$\begin{aligned} E(X) &= \mu_1 = M'(0) \\ &= r \left( \frac{p}{1 - (1-p)e^v} \right)^{r-1} \left[ \frac{0 - p(0 - (1-p)e^v)}{(1 - (1-p)e^v)^2} \right] \Big|_{v=0} \\ &= r \left( \frac{p}{1 - (1-p)} \right)^{r-1} \frac{p(1-p)}{(1 - (1-p))^2} \\ &= r \left( \frac{p}{1 - (1-p)} \right)^r \frac{1-p}{1 - (1-p)} \\ &= r \left( \frac{p}{1-p} \right)^r \frac{p}{1-p} = r \frac{1-p}{p} \end{aligned} \quad (2.6)$$

Um die Varianz zu bestimmen berechnen wir nun das zweite nicht zentrierte Moment:

$$\begin{aligned} \mu_2 &= M''(0) \\ &= \frac{\partial}{\partial v} \left[ r \left( \frac{p}{1 - (1-p)e^v} \right)^{r-1} \frac{p(1-p)e^v}{(1 - (1-p)e^v)^2} \right] \Big|_{v=0} \\ &= r \left[ (r-1) \left( \frac{p}{1 - (1-p)e^v} \right)^{r-2} \left( \frac{p(1-p)e^v}{(1 - (1-p)e^v)^2} \right)^2 + \left( \frac{p}{1 - (1-p)e^v} \right)^{r-1} \frac{p(1-p)e^v(1 - (1-p)e^v)^2 + p(1-p)e^{v2}(1 - (1-p)e^v)(1-p)e^v}{(1 - (1-p)e^v)^4} \right] \Big|_{v=0} \\ &= r \left( (r-1) \frac{(1-p)^2}{p^2} + \frac{p(1-p) + 2(1-p)^2}{p^2} \right) = r(1-p) \left( \frac{r-rp+1}{p^2} \right) \end{aligned} \quad (2.7)$$

Unter Verwendung des Satzes von Steiner ergibt sich aus (2.6) und (2.7) die Varianz:

$$\begin{aligned}
 \text{Var}(X) &= \mu_2 - \mu_1^2 \\
 &= r(1-p) \left( \frac{r-rp+1}{p^2} \right) - \left( r \frac{1-p}{p} \right)^2 \\
 &= r(1-p) \left( \frac{r-rp+1}{p^2} - \frac{r-rp}{p^2} \right) \\
 &= r \frac{1-p}{p^2}
 \end{aligned}$$

□

Ebenso wie die verallgemeinerte Poissonverteilung besitzt die negative Binomialverteilung zwei Parameter. Damit erlaubt sie im Vergleich zur Poissonverteilung mehr Flexibilität bei der Modellierung der Varianz. Die Anpassung der Varianz in Abhängigkeit vom Erwartungswert kann wie bei der verallgemeinerten Poissonverteilung über einen Dispersionsparameter erfolgen.

$$\begin{aligned}
 \text{Var}(X) &= \frac{E(X)}{p} \\
 &= \varphi E(X) \quad \text{mit } \varphi = \frac{1}{p}
 \end{aligned} \tag{2.8}$$

Es gilt  $\varphi > 1$ , da der Verteilungsparameter  $p$  nur Werte in  $(0,1)$  annehmen kann. Die Varianz der negativen Binomialverteilung ist daher stets größer als ihr Erwartungswert.

### 2.1.2 Faltungseigenschaft der negativen Binomialverteilung mit ganzzahligem Parameter

Eine weitere Eigenschaft der negativen Binomialverteilung ist deren Additivität.

**Satz 2.4 (Summe zweier negativ binomialverteilter Zufallsvariablen)**  $X \sim NB^*(r_1, p)$  und  $Y \sim NB^*(r_2, p)$  seien zwei unabhängige Zufallsvariablen mit  $r_1, r_2 \in \mathbb{N}$ . Damit ist die Zufallsvariable  $Z := X + Y$  negativ binomialverteilt mit Parametern  $r_1 + r_2$  und  $p$  ( $Z \sim NB^*(r_1 + r_2, p)$ ).

*Beweis:*

Die Zufallsvariablen  $X$  und  $Y$  sind unabhängig. Daher ist die momenterzeugende Funktion der gefalteten Zufallsvariable  $Z$  gleich dem Produkt der momenterzeugenden Funktionen von  $X$  und  $Y$ .

$$\begin{aligned}
 M_Z(v) &= M_{X+Y}(v) = M_X(v)M_Y(v) \\
 &\stackrel{(2.2)}{=} \left( \frac{p}{1 - (1-p)e^v} \right)^{r_1} \left( \frac{p}{1 - (1-p)e^v} \right)^{r_2} \\
 &= \left( \frac{p}{1 - (1-p)e^v} \right)^{r_1+r_2}
 \end{aligned} \tag{2.9}$$

(2.9) ist die momenterzeugende Funktion einer negativ binomialverteilten Zufallsvariable mit Parametern  $r_1 + r_2$  und  $p$ .

□

Durch Induktion folgt weiterhin:

**Korollar 2.5**  $X_i \sim NB^*(r_i, p)$   $i = 1, \dots, n$  seien  $n$  unabhängige Zufallsvariablen mit  $r_i \in \mathbb{N}$ . Dann gilt:

$$\sum_{i=1}^n X_i \sim NB\left(\sum_{i=1}^n r_i, p\right) \quad (2.10)$$

### 2.1.3 Negative Binomialverteilung mit stetigen Parametern

Bisher wurde vorausgesetzt, dass der Parameter  $r$  ganzzahlig ist. Es existiert eine weitere Definition, welche die negative Binomialverteilung auf einen stetigen Parameter  $r$  erweitert.

**Definition 2.6 (Negative Binomialverteilung mit stetigen Parametern)** Eine Zufallsvariable  $X$  mit der Wahrscheinlichkeitsfunktion

$$P_x(\mu, a) = P(X = x) = \begin{cases} \frac{\Gamma(x+a^{-1})}{\Gamma(a^{-1})x!} \left(\frac{\mu}{a^{-1}+\mu}\right)^x \left(\frac{a^{-1}}{a^{-1}+\mu}\right)^{a^{-1}} & x = 0, 1, 2, \dots \\ 0 & \text{sonst} \end{cases} \quad (2.11)$$

heißt negativ binomialverteilt mit Parametern  $\mu \in \mathbb{R}_+$  und  $a \in \mathbb{R}_+$ . Kurz  $X \sim NB(\mu, a)$ .

#### Bemerkungen:

- (i) Nach wie vor handelt es sich hierbei um eine diskrete Verteilung, da  $X$  weiterhin nur Werte aus  $\mathbb{N}$  annimmt.
- (ii) Der Parameter  $a \geq 0$  wird als Dispersionsindex bezeichnet.
- (iii) Übereinstimmung der Definitionen 2.1 und 2.6 erhalten wir, wenn

$$r = a^{-1} \quad (2.12)$$

$$p = \frac{a^{-1}}{a^{-1} + \mu} \quad (2.13)$$

gesetzt wird. Der Kehrwert des Parameters  $a$  aus Definition 2.6 ist damit auf ganzzahlige Werte eingeschränkt.

$$\frac{\Gamma(x+a^{-1})}{\Gamma(a^{-1})x!} \left(\frac{\mu}{a^{-1}+\mu}\right)^x \left(\frac{a^{-1}}{a^{-1}+\mu}\right)^{a^{-1}} = \frac{(x+r-1)!}{(r-1)!x!} \left(\frac{\frac{\mu}{r}}{1+\frac{\mu}{r}}\right)^x p^r \quad \text{für } a^{-1} \in \mathbb{N} \quad (2.14)$$

Lösen wir (2.13) nach  $\mu$  auf, so erhalten wir:

$$\frac{\frac{\mu}{r}}{1+\frac{\mu}{r}} = \frac{\frac{\frac{r}{p}-r}{r}}{1+\frac{\frac{r}{p}-r}{r}} = \frac{\frac{1}{p}-1}{\frac{1}{p}} = 1-p \quad (2.15)$$

Durch Einsetzen von (2.15) in (2.14) folgt die Übereinstimmung von (2.11) aus Definition 2.6 mit (2.1) aus Definition 2.1 für ganzzahlige Werte von  $a^{-1}$ , da

$$\frac{(x+r-1)!}{(r-1)!x!} \left(\frac{\frac{\mu}{r}}{1+\frac{\mu}{r}}\right)^x p^r = \binom{x+r-1}{x} (1-p)^x p^r \quad (2.16)$$

**Satz 2.7 (Erwartungswert und Varianz der negativen Binomialverteilung)** Die Zufallsvariable  $X$  sei negativ binomialverteilt mit stetigen, nicht negativen Parametern  $\mu$  und  $a$ . Erwartungswert und Varianz von  $X$  lauten:

$$E(X) = \mu \quad (2.17)$$

$$\text{Var}(X) = \mu + a\mu^2 \quad (2.18)$$

*Beweis:*

Wir verwenden die Resultate aus Satz 2.3. Für  $a^{-1} \in \mathbb{N}$  gilt:

$$\begin{aligned} E(X) &\stackrel{(2.4)}{=} r \frac{1-p}{p} \\ &\stackrel{(2.12),(2.13)}{=} a^{-1} \frac{1 - \frac{a^{-1}}{a^{-1}+\mu}}{\frac{a^{-1}}{a^{-1}+\mu}} = \frac{1 - \frac{a^{-1}}{a^{-1}+\mu}}{\frac{1}{a^{-1}+\mu}} \\ &= a^{-1} + \mu - a^{-1} = \mu \\ \\ \text{Var}(X) &\stackrel{(2.8)}{=} \frac{E(X)}{p} \\ &\stackrel{(2.17),(2.13)}{=} \frac{\mu}{\frac{a^{-1}}{a^{-1}+\mu}} = \frac{(a^{-1} + \mu)\mu}{a^{-1}} \\ &= (1 + a\mu)\mu = \mu + a\mu^2 \end{aligned}$$

Für beliebiges  $a > 0$  sei auf Sikora (2002, S.39 f.) verwiesen.

□

### 2.1.4 Grenzverteilung der negativen Binomialverteilung

Für die Variante mit stetigem Parameter besitzt die negative Binomialverteilung für  $a = 0$  als Grenzfall die Poissonverteilung.

**Satz 2.8 (Poissonverteilung als Grenzverteilung)** Sei  $X$  eine negativ binomialverteilte Zufallsvariable mit Parametern  $\mu > 0$  und  $a = 0$ . Dann gilt:

$$X \sim \text{Poi}(\mu)$$

*Beweis:*

Es sei zunächst folgende Identität für  $s, t \in \mathbb{N}$  gezeigt:

$$\begin{aligned} \frac{\Gamma(s+t)}{\Gamma(s)} &= \frac{(s+t-1)!}{(s-1)!} \\ &= (s-1+t)(s-2+t) \dots (s-t+t) \\ &= \prod_{k=1}^{t-1} (s+k) \end{aligned} \quad (2.19)$$

Formen wir nun die Wahrscheinlichkeitsfunktion der negativen Binomialverteilung um:

$$\begin{aligned}
P_X(\mu, a) &= \frac{\Gamma(x + a^{-1})}{\Gamma(a^{-1})x!} \left(\frac{\mu}{a^{-1} + \mu}\right)^x \left(\frac{a^{-1}}{a^{-1} + \mu}\right)^{a^{-1}} \\
&\stackrel{r = a^{-1}}{=} \frac{\Gamma(x + r)}{\Gamma(r)x!} \left(\frac{\mu}{r + \mu}\right)^x \left(\frac{r}{r + \mu}\right)^r \\
&\stackrel{(2.19)}{=} \frac{1}{x!} \left(\prod_{k=0}^{x-1} r + k\right) \mu^x \left(\frac{1}{r + \mu}\right)^x \left(\frac{1}{1 + \frac{\mu}{r}}\right)^r \\
&= \frac{\mu^x}{x!} \left(\prod_{k=0}^{x-1} \frac{r + k}{r + \mu}\right) \left(\frac{1}{1 + \frac{\mu}{r}}\right)^r \tag{2.20}
\end{aligned}$$

Für die von  $r$  abhängigen Faktoren in (2.20) gelten folgende Grenzwerte:

$$\lim_{r \rightarrow \infty} \prod_{k=0}^{x-1} \frac{r + k}{r + \mu} = \lim_{r \rightarrow \infty} \prod_{k=0}^{x-1} \frac{1 + \frac{k}{r}}{1 + \frac{\mu}{r}} = 1 \tag{2.21}$$

$$\lim_{r \rightarrow \infty} \left(\frac{1}{1 + \frac{\mu}{r}}\right)^r = \lim_{r \rightarrow \infty} \left(1 + \frac{\mu}{r}\right)^{-r} = e^{-\mu} \tag{2.22}$$

Unter Verwendung von (2.20), (2.21) und (2.22) erhalten wir den Grenzwert von  $P_X(\mu, a)$  für  $a \in \mathbb{Q} \rightarrow 0$ :

$$\lim_{a \rightarrow 0} P_X(\mu, a) = \lim_{r \rightarrow \infty} \frac{\mu^x}{x!} \left(\prod_{k=0}^{x-1} \frac{r + k}{r + \mu}\right) \left(\frac{1}{1 + \frac{\mu}{r}}\right)^r = \frac{\mu^x}{x!} e^{-\mu} \tag{2.23}$$

(2.23) ist die Wahrscheinlichkeitsfunktion der Poissonverteilung mit Parameter  $\mu$ . Da  $\mathbb{Q}$  dicht in  $\mathbb{R}$  liegt, gilt die Konvergenzaussage auch für  $a \in \mathbb{R} \rightarrow 0$  (Königsberger, 1999, S.15).

□

Damit lässt sich die negative Binomialverteilung in der Definition mit diskretem Parameter für große Werte von  $r = a^{-1}$  durch die Poissonverteilung mit Parameter  $\mu$  approximieren.

### 2.1.5 Das Gamma-Poisson-Mischmodell

Ähnlich wie für die verallgemeinerte Poissonverteilung, gibt es auch für die negative Binomialverteilung verschiedene Modelle, mit denen sich diese Verteilung erzeugen lässt.

Die einfachste Methode ist dabei das Gamma-Poisson-Mischmodell. Bei Mischmodellen sind die Verteilungsparameter zufällig. Im Fall der negativen Binomialverteilung ist die Zufallsvariable poissonverteilt mit gammaverteiltem<sup>1</sup> Parameter.

**Satz 2.9 (Gamma-Poisson-Mischmodell)** *Mit*

$$X|\theta \sim Poi(\theta) \quad \text{und} \quad \theta \sim \Gamma\left(q, \frac{q}{\mu}\right)$$

ergibt sich für  $X$  als Mischverteilung die negative Binomialverteilung mit Parametern  $\mu$  und  $\frac{1}{q}$ :

$$X \sim NB\left(\mu, \frac{1}{q}\right) \tag{2.24}$$

*Beweis:* (siehe z.B. Cameron and Trivedi, 1998).

<sup>1</sup>siehe Anhang A.3



## 2.2 Vergleich der verallgemeinerten Poisson- und negativen Binomialverteilung bei Übereinstimmung von Erwartungswert und Varianz

Aus den Abschnitten 1.3 und 2.1.3 sind die Erwartungswerte und Varianzen der verallgemeinerten Poissonverteilung und der negativen Binomialverteilung bekannt.

$$X \sim GP(\theta, \lambda) \quad \Rightarrow \quad E_{GP}X = \frac{\theta}{1 - \lambda} \quad (2.25)$$

$$Var_{GP}X = \frac{\theta}{(1 - \lambda)^3} \quad (2.26)$$

$$Y \sim NB(\mu, a) \quad \Rightarrow \quad E_{NB}Y = \mu \quad (2.27)$$

$$Var_{NB}Y = \mu + a\mu^2 \quad (2.28)$$

Verwenden wir zur Darstellung der Momente der verallgemeinerten Poissonverteilung den Dispersionsparameter  $\varphi := \frac{1}{1 - \lambda}$  aus Abschnitt 1.3.6 und definieren  $\nu := E_{GP}X$ , so erhalten wir

$$E_{GP}X = \nu \quad (2.29)$$

$$Var_{GP}X \stackrel{(1.40)}{=} \varphi^2 \nu \quad (2.30)$$

Für die vier Parameter  $\mu$ ,  $a$ ,  $\nu$  und  $\varphi$  sind mit (2.27)-(2.30) vier Gleichungen gegeben und so können wir einen allgemeinen Zusammenhang zwischen  $\mu$  und  $a$  auf der einen Seite und  $\nu$  und  $\varphi$  auf der anderen Seite herleiten, der zur Gleichheit der ersten beiden zentralen Momente der verallgemeinerten Poisson- und der negativen Binomialverteilung führt.

Dazu setzen wir (2.27) mit (2.29) und (2.28) mit (2.30) gleich:

$$E_{GP}X = E_{NB}Y \quad \Rightarrow \quad \nu = \mu \quad (2.31)$$

$$Var_{GP}X = Var_{NB}Y \quad \Rightarrow \quad \varphi^2 \nu = \mu + a\mu^2 \quad (2.32)$$

Setzen wir nun (2.31) in (2.32) ein und formen unter Verwendung der Positivität von  $\mu$  und der nicht-Negativität von  $\varphi$  um:

$$\begin{aligned} \varphi^2 \mu = \mu + a\mu^2 & \stackrel{\mu \geq 0}{\Rightarrow} \varphi^2 = 1 + a\mu \\ & \stackrel{\varphi \geq 0}{\Rightarrow} \varphi = \sqrt{1 + a\mu} \end{aligned} \quad (2.33)$$

Den Zusammenhang zwischen  $\mu$  und  $a$  sowie den Parametern  $\theta$  und  $\lambda$  der verallgemeinerten Poissonverteilung bei Übereinstimmung von Erwartungswert und Varianz erhalten wir, wenn die Größen  $\nu$  und  $\varphi$  auf  $\theta$  und  $\lambda$  zurücktransformiert werden.

$$\begin{aligned} \varphi & \stackrel{(1.40)}{=} \frac{1}{1 - \lambda} \\ \Rightarrow 1 - \lambda & \stackrel{(2.33)}{=} \frac{1}{\sqrt{1 + a\mu}} \\ \Rightarrow \lambda & = 1 - \frac{1}{\sqrt{1 + a\mu}} \end{aligned} \quad (2.34)$$

$$\begin{aligned} \nu & \stackrel{(1.4)}{=} \frac{\theta}{1 - \lambda} \\ \Rightarrow \theta & \stackrel{(2.31)}{=} \mu(1 - \lambda) \stackrel{(2.34)}{=} \mu \left[ 1 - \left( 1 - \frac{1}{\sqrt{1 + a\mu}} \right) \right] \\ & = \frac{\mu}{\sqrt{1 + a\mu}} \end{aligned} \quad (2.35)$$

Auf Grund der obigen Ausführungen, lässt sich folgender Satz formulieren:

**Satz 2.10** Die Zufallsvariable  $X$  folge der verallgemeinerten Poissonverteilung mit Parametern  $\theta$  und  $\lambda$ , d.h.  $X \sim GP(\theta, \lambda)$ . Eine zweite Zufallsvariable  $Y$  sei negativ binomialverteilt mit Parametern  $\mu$  und  $a$ , d.h.  $Y \sim NB(\mu, a)$ .  $X$  und  $Y$  stimmen in Erwartungswert und Varianz genau dann überein, wenn

$$\theta = \frac{\mu}{\sqrt{1+a\mu}} \quad \text{und} \quad \lambda = 1 - \frac{1}{\sqrt{1+a\mu}}$$

gilt.

Ebenso kann auch das Parameterpaar  $\mu$  und  $a$  durch die Parameter  $\theta$  und  $\lambda$  ausgedrückt werden:

$$\begin{aligned} \mu &= \nu & a &= \frac{\varphi^2 - 1}{\nu} \\ &= \frac{\theta}{1 - \lambda} & &= \frac{\lambda}{\theta(1 - \lambda)} \end{aligned} \quad (2.36)$$

Zur Veranschaulichung der in Erwartungswert und Varianz übereinstimmenden verallgemeinerten Poisson- und negativen Binomialverteilung sind in den Abbildungen 2.1 bis 2.3 links die Wahrscheinlichkeitsfunktionen dieser Verteilungen für entsprechende Parameter abgetragen. In beiden Fällen handelt es sich um diskrete Verteilungen, die Wahrscheinlichkeitsfunktionen sind daher nur punktwise definiert. Um den Funktionsverlauf besser vergleichen zu können, sind die Balken, welche die Wahrscheinlichkeiten der negativen Binomialverteilung repräsentieren, um 0.5 Einheiten nach rechts verschoben. Rechts neben den Wahrscheinlichkeitsfunktionen sind die Abweichungen der einzelnen Eintrittswahrscheinlichkeiten der negativen Binomialverteilung von denen der verallgemeinerten Poissonverteilung ( $P_x^{GP}(\theta, \lambda) - P_x^{NB}(\mu, a)$ ) abgetragen.

Die negative Binomialverteilung wird für kleine Werte des Parameters  $a$  als Approximation der Poissonverteilung verwendet (vgl. Abschnitt 2.1.4). Andererseits stimmt die verallgemeinerte Poissonverteilung mit Parameter  $\lambda = 0$  mit der Poissonverteilung überein. Es ist also zu erwarten, dass für kleine Werte von  $\lambda$  und  $a$  die Wahrscheinlichkeitsfunktionen sehr ähnlich sind. Betrachten wir die obere Zeile in Abbildung 2.1 ( $\lambda \approx 0.09$ ,  $a \approx 0.02$ ), so scheint dies auch der

Erwartungswert	Varianz	GP-Parameter		NB-Parameter		Summe der Abweichungsquadrate
		$\theta$	$\lambda$	$\mu$	$a$	
5	6	4.56	0.09	5.00	0.02	0.00016385
	15	2.89	0.42	5.00	0.25	0.00134641
	150	0.91	0.82	5.00	4.90	0.01851991
10	15	8.16	0.18	10.00	0.03	0.00039089
	150	2.58	0.74	10.00	1.11	0.00164014
	1500	0.82	0.92	10.00	13.78	0.07918916
50	150	28.87	0.42	50.00	0.03	0.00036656
	1500	9.13	0.82	50.00	0.49	0.00189200
	15000	2.87	0.94	50.00	5.65	0.10200490

Tabelle 2.1: Abweichung von GP- und NB-Wahrscheinlichkeitsfunktion

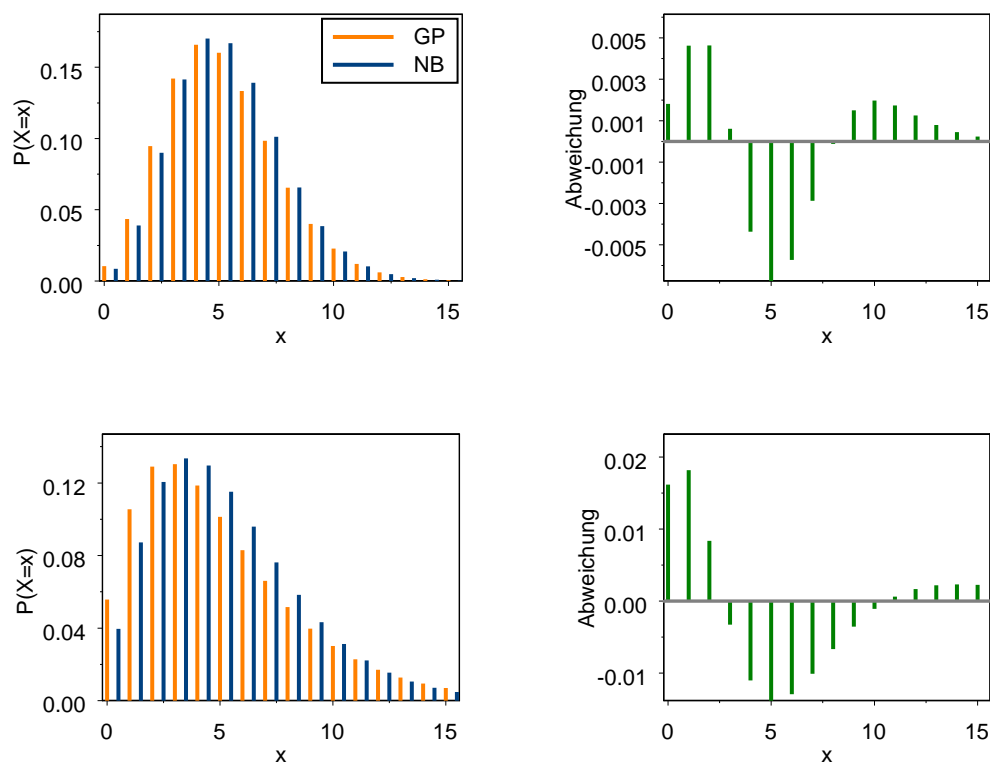


Abbildung 2.1: Verallgemeinerte Poissonverteilung und negative Binomialverteilung mit Erwartungswert 5 und Varianz 6 (oben) bzw. Varianz 15 (unten)

Fall zu sein. Wie wir aus Tabelle 2.1 entnehmen können, liegt für diese beiden Parameterwerte die Summe der Abweichungsquadrate

$$\sum_{x=0}^{1000} (P_x^{GP}(\theta, \lambda) - P_x^{NB}(\mu, a))^2 \quad (2.37)$$

für die Ereignisse  $x \in [0, 1000]$  unter allen betrachteten Konstellationen am niedrigsten.

Wir können der Tabelle 2.1 ebenfalls entnehmen, dass bei konstantem Erwartungswert die Abweichung der Wahrscheinlichkeitsfunktionen mit steigender Varianz zunimmt. In Abbildung 2.2 ist dies deutlich sichtbar. Beträgt die Varianz 15 (obere Zeile), so besitzen beide Wahrscheinlichkeitsfunktionen den selben Modus sowie einen recht ähnlichen Verlauf. Wird die Varianz auf 150 erhöht (untere Zeile), so lautet der Modus der negativen Binomialverteilung 0, während die verallgemeinerte Poissonverteilung den Modus 1 besitzt. Wie an der Skala der Abweichungsschaubilder (rechte Spalte von Abbildung 2.2) erkennbar ist, vergrößert sich gleichzeitig der maximale Unterschied zwischen einzelnen Eintrittswahrscheinlichkeiten.

Wird andererseits bei gleichbleibender Varianz der Erwartungswert erhöht, so nähern sich die verallgemeinerte Poissonverteilung und die negative Binomialverteilung an. Dies lässt sich z.B. bei dem Vergleich der unteren Zeile von Abbildung 2.2 (Erwartungswert 10) mit der oberen Zeile von Abbildung 2.3 (Erwartungswert 50) erkennen. In beiden Fällen beträgt die Varianz 150. Auch die Werte in Tabelle 2.1 weisen eine Annäherung der beiden Verteilungen aus, da die Summe der Abweichungsquadrate bei steigendem Erwartungswert und konstanter Varianz zurück geht.

Generell erkennen wir aus den Abbildungen 2.1-2.3, dass die negative Binomialverteilung meist stärker um ihren Modus konzentriert ist, als die verallgemeinerte Poissonverteilung.

Zusammenfassung:

Für kleine (positive) Werte von  $\lambda$  und  $a$  lassen sich negative Binomialverteilung und verallgemeinerte Poissonverteilung mit gleichem Erwartungswert und gleicher Varianz gegenseitig ersetzen. Wird der Erwartungswert verkleinert bzw. die Varianz erhöht, unterscheiden sich beide Verteilungen in zunehmendem Maße.

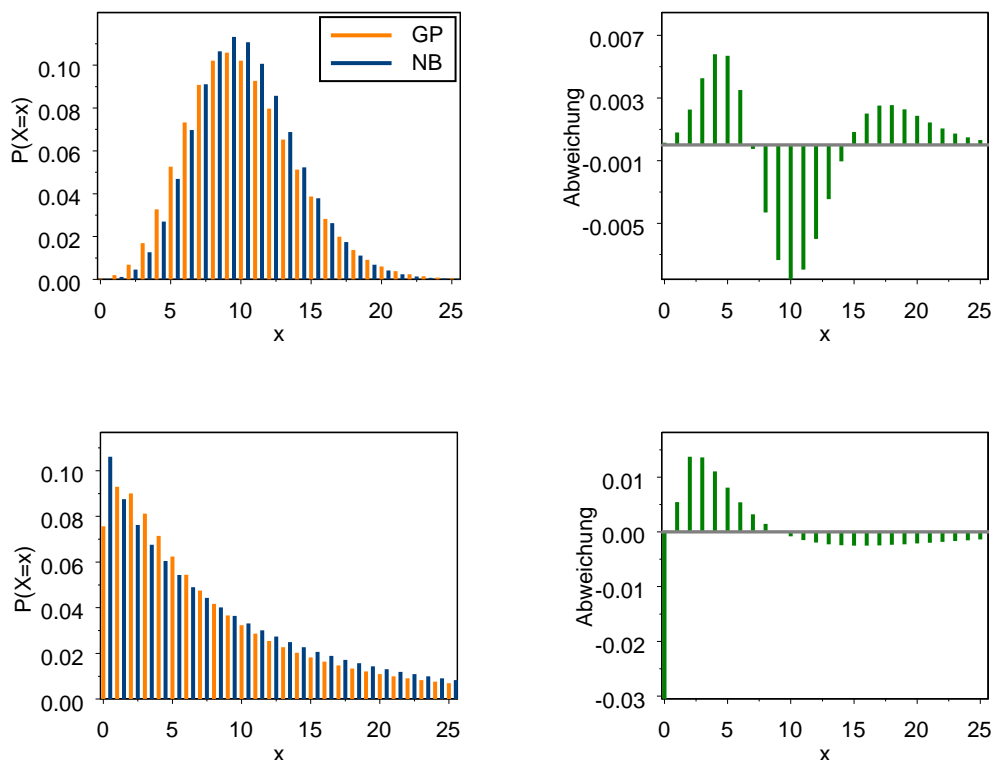


Abbildung 2.2: Verallgemeinerte Poissonverteilung und negative Binomialverteilung mit Erwartungswert 10 und Varianz 15 (oben) bzw. Varianz 150 (unten)

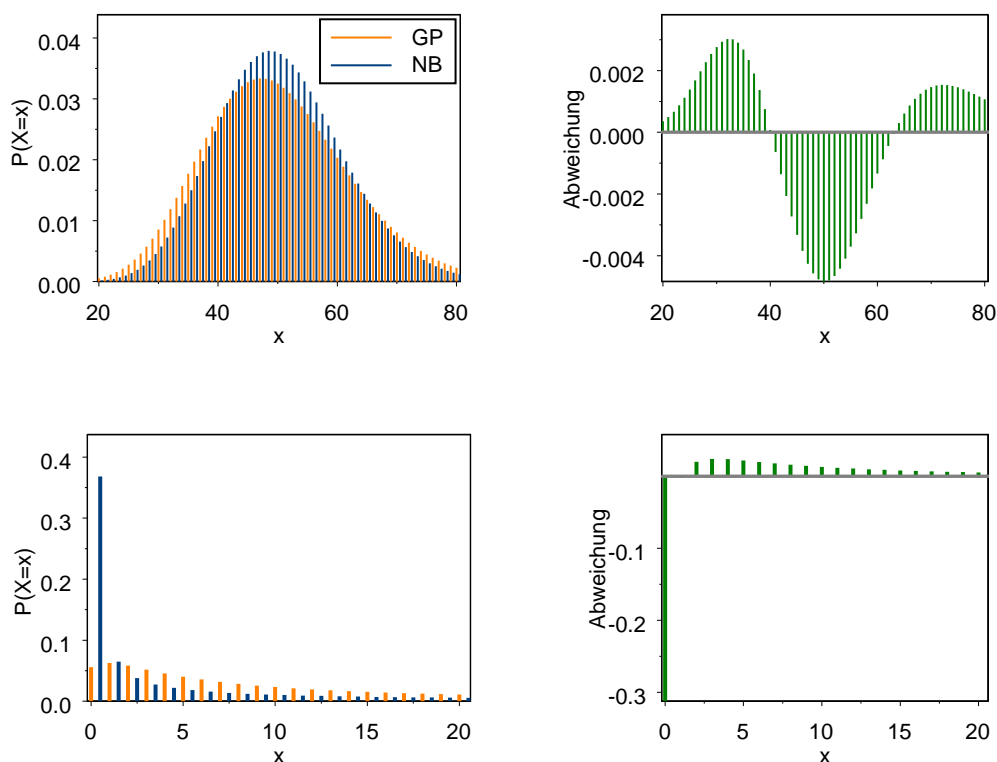


Abbildung 2.3: Verallgemeinerte Poissonverteilung und negative Binomialverteilung mit Erwartungswert 50 und Varianz 150 (oben) bzw. Varianz 15000 (unten)



## Kapitel 3

# Einführung in die Theorie der Regressionsmodelle

Betrachten wir Regressionsmodelle, so liegt eine Situation vor, in der uns eine bestimmte Anzahl an Daten (z.B. Messungen) vorliegt. Es wird angenommen, dass diesen Daten eine gewisse Struktur zu Grunde liegt, die durch das Modell aufgedeckt werden soll. Wir interessieren uns für eine zu untersuchende Größe  $Y_i, i = 1, \dots, n$  aus den Daten, die durch eine gewisse Anzahl von Einflussgrößen  $x_{i1}, \dots, x_{ik}$  erklärt werden soll. Liegen z.B. für 100 verschiedene Tage Messungen von Temperatur, Ozonkonzentration, Windstärke und UV-Strahlung an einem Ort vor, so können wir versuchen die Ozonkonzentration ( $Y_i$ ) durch die jeweils korrespondierenden Messungen der Temperatur ( $x_{i1}$ ), der Windstärke ( $x_{i2}$ ) und der UV-Strahlung ( $x_{i3}$ ) für  $i = 1, \dots, 100$  zu erklären.

Ziel der statistischen Modellbildung mit Hilfe von Regressionsmodellen ist es also eine **Zielvariable**

$$\mathbf{Y} = (Y_1, \dots, Y_n)^t \in \mathbb{R}^n$$

(Response, abhängige Variable) mit Hilfe von **Kovariablen**  $\mathbf{X}_j = (x_{1j}, \dots, x_{nj})^t, j = 1, \dots, k$  (Regressoren, erklärende Variablen) zu modellieren. Diese werden zur sogenannten Design- oder **Modellmatrix**

$$\mathbf{X} = \begin{pmatrix} x_{11} & \dots & x_{1k} \\ \vdots & & \vdots \\ x_{n1} & \dots & x_{nk} \end{pmatrix} \in \mathbb{R}^{n,k}$$

zusammengefasst. Bei der Modellbildung wird zunächst die Modellwahl getroffen. Für das gewählte Regressionsmodell müssen anschließend die darin enthaltenen Regressionsparameter

$$\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)^t \in \mathbb{R}^k$$

geschätzt werden, um letztendlich Responsedaten vorhersagen zu können.

Bei der Modellwahl ist eine Annahme über die Verteilung der beobachteten Responsedaten zu machen. Wesentlich ist bei der Modellwahl auch die Verwendung einer geeigneten Skala für die in das Modell eingehenden Variablen. So kann es unter Umständen nötig sein, anstatt der unskalierten Beobachtungen  $Y_i$ , die Größen  $\log(Y_i)$  oder  $e^{Y_i}$  als Zielgröße zu verwenden. Bei linearen Modellen muss die Skala so gewählt sein, dass die Varianz konstant, die Fehler normalverteilt und der Erwartungswert von  $Y_i | \mathbf{x}_i = Y_i | (x_{i1}, \dots, x_{ik})$  eine Linearkombination der Regressoren ist. Varianzhomogenität und Normalität der Fehler wird bei verallgemeinerten linearen Modellen nicht mehr verlangt. Auch kann die Additivität des Erwartungswertes auf einer transformierten

Skala vorliegen.

Eine entscheidende Rolle bei der Modellwahl kommt der Anzahl der Kovariablen zu, die aufgenommen werden sollen. So geht in das **Nullmodell** keiner der Regressoren ein. In diesem Fall wird keine Beziehung zwischen dem Response und den Regressoren modelliert, da angenommen wird, dass alle Elemente des Response den gleichen Erwartungswert besitzen. Im Gegensatz dazu gehen in das volle oder **gesättigte Modell**<sup>1</sup> so viele Regressoren ein, wie Zielvariablen vorhanden sind. Damit können die beobachteten Daten exakt reproduziert werden, es wird jedoch keine Aussage über die Datenstruktur erreicht. Bei der Modellwahl muss ein Kompromiss zwischen diesen Extremfällen gefunden werden.

Ist das Modell aufgestellt, so gilt es die darin enthaltenen Parameter zu schätzen. Dies geschieht, indem die Anpassungsgüte<sup>2</sup> in Abhängigkeit der Parameter gemessen wird. Dazu wird ein Diskrepanzmaß benötigt. Für das angepasste Modell werden diejenigen Parameterschätzer verwendet, welche das Diskrepanzmaß optimieren. Bei linearen Modellen wird als Diskrepanzmaß die Summe der Fehlerquadrate verwendet. Die Minimierung der Fehlerquadratsumme führt auf den Kleinste-Quadrate-Schätzer. Bei verallgemeinerten linearen Modellen werden Maximum-Likelihood-Schätzer verwendet (s. Abschnitt 3.3).

Für die Vorhersage wird mit den entsprechenden neuen Kovariablendaten eine Modellmatrix  $\mathbf{X}$  nach der Struktur, die durch das Modell vorgegeben ist, bestimmt. Durch Multiplikation dieser Modellmatrix mit dem Schätzer des Parametervektor  $\beta$  erhält man Schätzungen für die neuen, unbekanntes Responseedaten  $\mathbf{Y}$ .

### 3.1 Lineare Modelle

Bei linearen Modellen liegen uns Beobachtungen mit unabhängigen Zielvariablen  $Y_i$ ,  $i = 1, \dots, n$  vor. Gesucht ist eine in den Kovariablen  $\mathbf{X}_j$ ,  $j = 1, \dots, k$  lineare Funktion, die die Erwartungswerte  $\mu_i$  der Zielvariablen modelliert.

Lineare Modelle sind spezifiziert durch:

- (i) *Stochastische Komponente:*

Es wird angenommen, dass die bedingten Responseedaten  $Y_i|\mathbf{x}_i$  unabhängig normalverteilt sind mit individuellem Erwartungswert  $\mu_i$  und gemeinsamer Varianz  $\sigma^2$ :

$$Y_i|\mathbf{x}_i \stackrel{\text{ind.}}{\sim} N(\mu_i, \sigma^2), \quad i = 1, \dots, n \quad (3.1)$$

- (ii) *Systematische Komponente:*

Der Erwartungswert  $\mu_i$  von  $Y_i|\mathbf{x}_i$  wird von den Regressoren  $\mathbf{x}_i = (x_{i1}, \dots, x_{ik})^t$  nur durch eine lineare Funktion bestimmt:

$$\mu_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}, \quad i = 1, \dots, n \quad (3.2)$$

Dabei bezeichnen die Größen  $\beta_0, \dots, \beta_k$  die Regressionsparameter. Diese sind unbekannt und müssen geschätzt werden. Den Parameter  $\beta_0$  bezeichnet man dabei als Intercept. Er modelliert den gemeinsamen Mittelwert der Responseedaten.

---

<sup>1</sup>engl.: saturated model

<sup>2</sup>engl.: goodness of fit



**Lemma 3.1** *Mit Fehlertermen  $\varepsilon_i$  für die*

$$\varepsilon_i \stackrel{\text{ind.}}{\sim} N(0, \sigma^2), \quad i = 1, \dots, n \quad (3.3)$$

*gilt, lässt sich das statistische Modell welches mit (3.1) und (3.2) beschrieben ist, durch die Gleichungen*

$$Y_i = \mathbf{x}_i^t \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \dots, n \quad (3.4)$$

*ausdrücken.*

*Beweis:*

Da für bekannte Kovariablen  $\mathbf{x}_i$  der Term  $\mathbf{x}_i^t \boldsymbol{\beta}$  in (3.4) deterministisch ist, bestimmt die Zufallsgröße  $\varepsilon_i$  die Verteilung von  $Y_i$  gegeben  $\mathbf{x}_i$ . Die bedingten Zielgrößen  $Y_i | \mathbf{x}_i$  sind daher auf Grund der Voraussetzung (3.3) unabhängig normalverteilt.

Für den Erwartungswert gilt:

$$\begin{aligned} E(Y_i | \mathbf{x}_i) &\stackrel{(3.4)}{=} E(\mathbf{x}_i^t \boldsymbol{\beta} + \varepsilon_i | \mathbf{x}_i) \\ &= E(\mathbf{x}_i^t \boldsymbol{\beta} | \mathbf{x}_i) + E(\varepsilon_i | \mathbf{x}_i) \\ &= \mathbf{x}_i^t \boldsymbol{\beta} + E(\varepsilon_i) \stackrel{(3.3)}{=} \mathbf{x}_i^t \boldsymbol{\beta} \\ &\stackrel{(3.2)}{=} \mu_i \end{aligned}$$

Die Varianz lautet:

$$\begin{aligned} \text{Var}(Y_i | \mathbf{x}_i) &\stackrel{(3.4)}{=} \text{Var}(\mathbf{x}_i^t \boldsymbol{\beta} + \varepsilon_i | \mathbf{x}_i) \\ &= \text{Var}(\mathbf{x}_i^t \boldsymbol{\beta} | \mathbf{x}_i) + \text{Var}(\varepsilon_i | \mathbf{x}_i) \\ &= 0 + \text{Var}(\varepsilon_i) \stackrel{(3.3)}{=} \sigma^2 \end{aligned}$$

Zusammenfassend folgt:

$$Y_i | \mathbf{x}_i \stackrel{\text{ind.}}{\sim} N(\mu_i, \sigma^2), \quad i = 1, \dots, n$$

□

Für eine genauere Betrachtung linearer Modelle sei auf Myers (1990) verwiesen.

## 3.2 Verallgemeinerte lineare Modelle

Wiederum sind der

- Responsevektor  $\mathbf{Y} = (Y_1, \dots, Y_n)^t$  mit unabhängigen Einträgen und die
- Kovariablenvektoren  $\mathbf{X}_1, \dots, \mathbf{X}_k$

gegeben. Im Gegensatz zu den linearen Modellen kann der Response bei verallgemeinerten linearen Modellen einer beliebigen Verteilung aus der exponentiellen Familie<sup>3</sup> folgen. Ebenso wie die Verteilungsannahme wird die Voraussetzung der Additivität des Erwartungswert dahingehend gelockert, dass auch Funktionen einer Linearkombination der Regressoren als Erwartungswert zulässig sind. Es existiert jedoch weiterhin nur ein einziger Fehlerterm.

---

<sup>3</sup>s. Abschnitt A.1

Ein verallgemeinertes lineares Modell besteht also aus folgenden Komponenten:

(i) *Stochastische Komponente:*

Als Verteilungsannahme wird vorausgesetzt, dass die Responseverteilung der exponentiellen Familie angehört. Ihr Dichte oder Wahrscheinlichkeitsfunktion muss sich daher in der Form

$$P_x(\theta, \varphi) = \exp \left\{ \frac{\theta x - b(\theta)}{a(\varphi)} + c(x, \varphi) \right\}$$

mit reellwertigen Funktionen  $a$ ,  $b$  und  $c$  darstellen lassen.

(ii) *Systematische Komponente:*

Es wird die strukturelle Annahme getroffen, dass die Vektoren  $\mathbf{x}_i$  die Verteilung von  $\mathbf{Y}$  nur durch eine lineare Funktion  $\boldsymbol{\eta}$  bestimmen.

$$\boldsymbol{\eta} := \beta_0 \mathbf{1}_n + \beta_1 \mathbf{x}_1 + \cdots + \beta_k \mathbf{x}_k = \mathbf{X}\boldsymbol{\beta}$$

$\boldsymbol{\eta}$  wird als **linearer Prädiktor** bezeichnet<sup>4</sup>.

(iii) *Verknüpfungskomponente:*

Es wird angenommen, dass der Erwartungswert  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^t$  eine Funktion des linearen Prädiktors ist:

$$h(\boldsymbol{\eta}) := \boldsymbol{\mu}$$

Die **Responsefunktion**<sup>5</sup>  $h$  verknüpft den linearen Prädiktor  $\boldsymbol{\eta}$  mit dem Erwartungswert  $\boldsymbol{\mu}$ . Damit ist eine Verbindung zwischen der systematischen und der stochastischen Komponente hergestellt.

Meist ist vorausgesetzt, dass  $h$  differenzierbar und monoton ist. In diesem Fall existiert für die Responsefunktion  $h$  eine inverse Funktion  $g$ , die **Linkfunktion** genannt wird.

$$g(\boldsymbol{\mu}) := h^{-1}(\boldsymbol{\mu}) = \boldsymbol{\eta}$$

**Bemerkung:**

In linearen Modellen sind Erwartungswert und linearer Prädiktor identisch, dh.  $\boldsymbol{\eta} = \boldsymbol{\mu}$ .

Eingehend behandelt werden verallgemeinerte lineare Modelle in dem Buch 'Generalized Linear Models' von Peter McCullagh und John Nelder (1989).

### 3.3 Maximum-Likelihood-Schätzung

Bei linearen Modellen wird zur Parameterschätzung der Kleinste-Quadrate-Ansatz gewählt. Dabei wird die Summe der Abstandsquadrate zwischen den beobachteten Daten  $Y_i$  und den angepassten Daten  $\hat{Y}_i = \hat{\mu}_i$  minimiert.  $\hat{\mu}_i$  bezeichnet den Schätzer des Erwartungswertes von  $Y_i | \mathbf{x}_i$ . Bei verallgemeinerten linearen Modellen verwendet man hingegen das Maximum-Likelihood-Konzept. An die aus den Kovariablen bestehende Modellmatrix  $\mathbf{X}$  wird dabei Bedingung eines vollen Spaltenranges gestellt. Ferner muss die dem Modell zugrunde liegende Verteilung vollständig bekannt sein, da die Maximum-Likelihood-Methode modellbasiert ist.

<sup>4</sup>Im Ausdruck von  $\boldsymbol{\eta}$  bezeichnet  $\mathbf{1}_n$  den  $n$ -dimensionalen Vektor  $(1, \dots, 1)^t$

<sup>5</sup>Die Responsefunktion wird auch als Erwartungswertfunktion bezeichnet

### 3.3.1 Maximum-Likelihood-Schätzer

Betrachten wir die gemeinsame Dichte bzw. Verteilungsfunktion  $P(\mathbf{y}|\boldsymbol{\theta})$  nicht als Funktion der Zielgröße  $\mathbf{Y}$  sondern als Funktion des Parameters  $\boldsymbol{\theta}$  für festes  $\mathbf{Y} = \mathbf{y}$ , so erhalten wir die **Likelihoodfunktion**

$$L(\boldsymbol{\theta}|\mathbf{y}) := P(\mathbf{y}|\boldsymbol{\theta}) \quad (3.5)$$

Im diskreten Fall gibt  $P(\mathbf{y}|\boldsymbol{\theta})$  an, wie wahrscheinlich die beobachteten Respondedaten  $\mathbf{Y} = \mathbf{y}$  bei gegebenem Parameter  $\boldsymbol{\theta}$  sind. Daher kann die Likelihoodfunktion  $L(\boldsymbol{\theta}|\mathbf{y})$  für  $\boldsymbol{\theta}$  als Maß dafür aufgefasst werden, wie wahrscheinlich der Parameter  $\boldsymbol{\theta}$  bei den vorliegenden Daten  $\mathbf{Y} = \mathbf{y}$  ist. Die Maximum-Likelihood-Methode besteht nun darin, einen Wert  $\hat{\boldsymbol{\theta}}(\mathbf{y})$  zu finden, welcher die Likelihoodfunktion maximiert und damit für die Erzeugung der beobachteten Daten  $\mathbf{Y} = \mathbf{y}$  am wahrscheinlichsten ist.

**Definition 3.2 (Reguläres Modell)** *Ein parametrisches Verteilungsmodell*

$P_{\boldsymbol{\theta}} = \{P(\mathbf{x}|\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$  wird als regulär bezeichnet, wenn alle  $P_{\boldsymbol{\theta}}$  entweder stetig mit Dichte  $P(\mathbf{x}, \boldsymbol{\theta})$  oder diskret mit Wahrscheinlichkeitsfunktion  $P(\mathbf{x}, \boldsymbol{\theta})$  sind und  $\Theta \subset \mathbb{R}^n$  gilt.

**Definition 3.3 (Maximum-Likelihood-Schätzer)** *Falls in einem regulären parametrischen Modell  $P_{\boldsymbol{\theta}} = \{P(\mathbf{y}|\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$  ein Wert  $\hat{\boldsymbol{\theta}}(\mathbf{y})$  existiert, welcher*

$$\begin{aligned} P(\mathbf{y}|\hat{\boldsymbol{\theta}}(\mathbf{y})) &= \max_{\boldsymbol{\theta} \in \Theta} \{P(\mathbf{y}|\boldsymbol{\theta})\} \\ &= \max_{\boldsymbol{\theta} \in \Theta} \{L(\boldsymbol{\theta}|\mathbf{y})\} \end{aligned} \quad (3.6)$$

erfüllt, so bezeichnen wir diesen Wert als *Maximum-Likelihood-Schätzer (MLE) des Parameters  $\boldsymbol{\theta}$  für die beobachteten Daten  $\mathbf{Y} = \mathbf{y}$ .*

Um das Maximum (3.6) zu bestimmen, muss der Gradient der Likelihoodfunktion Null gesetzt werden. Die Bestimmung der Ableitungen von  $L$  kann jedoch sehr aufwändig sein. Verwenden wir zur Maximierung anstatt  $L$  die **Log-Likelihoodfunktion**

$$\mathcal{L}(\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta}|\mathbf{y}) := \ln [L(\boldsymbol{\theta}|\mathbf{y})] \quad (3.7)$$

so lässt sich die Bestimmung der Ableitungen häufig vereinfachen. Da der Logarithmus eine streng monoton wachsende Transformation ist, gilt:

$$L(\hat{\boldsymbol{\theta}}) \geq L(\boldsymbol{\theta}) \quad \forall \boldsymbol{\theta} \in \Theta \quad \Leftrightarrow \quad \mathcal{L}(\hat{\boldsymbol{\theta}}) \geq \mathcal{L}(\boldsymbol{\theta}) \quad \forall \boldsymbol{\theta} \in \Theta. \quad (3.8)$$

Liegt wie von uns vorausgesetzt Unabhängigkeit der Zielgrößen  $Y_i$  vor, so lässt sich die gemeinsame Dichte von  $\mathbf{Y}|\mathbf{X}$  als Produkt der Einzeldichten schreiben Fahrmeir et al. (2001). Für die Log-Likelihoodfunktion gilt daher:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) &= \ln[L(\boldsymbol{\theta}|\mathbf{y})] = \ln[P(\boldsymbol{\theta}|\mathbf{y})] \\ &= \ln \left[ \prod_{i=1}^n P(y_i|\boldsymbol{\theta}) \right] \\ &= \sum_{i=1}^n \ln [P(y_i|\boldsymbol{\theta})] \end{aligned} \quad (3.9)$$

Da die Summe in (3.9) leichter abzuleiten ist als das Produkt  $\prod_{i=1}^n P(y_i|\boldsymbol{\theta})$ , verwendet man zur Bestimmung des MLE meist die Log-Likelihoodfunktion  $\mathcal{L}$ . Dies ist wegen der Äquivalenz (3.8) möglich.

**Definition 3.4 (Scorefunktion)**  $\mathcal{L}$  bezeichne die Log-Likelihoodfunktion des Parameters  $\boldsymbol{\theta}$  in Abhängigkeit von den Daten  $\mathbf{Y}$ . Dann wird der Gradient

$$s(\boldsymbol{\theta}|\mathbf{y}) := \nabla \mathcal{L}(\boldsymbol{\theta}|\mathbf{y}) = \frac{\partial}{\partial \boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}|\mathbf{y}) \in \mathbb{R}^k \quad (3.10)$$

als Scorefunktion bezeichnet.

Zur Bestimmung des MLE von  $\boldsymbol{\theta}$  müssen wir (3.10) Null setzen. Betrachten wir die dadurch entstehende sog. Scoregleichung  $s(\boldsymbol{\theta}|\mathbf{y}) = 0$  komponentenweise, so erhalten wir ein Gleichungssystem bestehend aus den **Maximum-Likelihood-Gleichungen**

$$s(\boldsymbol{\theta}|\mathbf{y})_j = \frac{\partial}{\partial \theta_j} \mathcal{L}(\boldsymbol{\theta}|\mathbf{y}) = 0, \quad j = 1, \dots, k \quad (3.11)$$

Trotz der Problemvereinfachung durch die Verwendung der Log-Likelihoodfunktion sind die Maximum-Likelihood-Gleichungen im Allgemeinen nicht linear. Das entstehende Gleichungssystem kann daher nicht analytisch gelöst werden, weshalb wir auf numerische, meist iterative Lösungsverfahren zurückgreifen müssen. Das Fisher-Scoring und das Newton-Raphson-Verfahren sind die gängigsten Methoden hierfür (siehe z.B. McCullagh and Nelder, 1989).

**Definition 3.5 (Definitheit symmetrischer Matrizen)**  $\mathbf{A} = (a_{ij})_{i,j=1,\dots,k}$  ist

$$\begin{aligned} \text{positiv definit:} & \Leftrightarrow \mathbf{x}^t \mathbf{A} \mathbf{x} > 0 \quad \forall \mathbf{x} \neq 0 \\ & \Leftrightarrow \lambda_i > 0 \quad \forall \text{ Eigenwerte } \lambda_i, i = 1, \dots, k \\ & \Leftrightarrow \det(\mathbf{A}_i) > 0 \quad \forall \text{ Hauptuntermatrizen } \mathbf{A}_r = (a_{ij})_{i,j=1,\dots,r} \\ \text{negativ definit:} & \Leftrightarrow \mathbf{x}^t \mathbf{A} \mathbf{x} < 0 \quad \forall \mathbf{x} \neq 0 \\ & \Leftrightarrow \lambda_i < 0 \quad \forall \text{ Eigenwerte } \lambda_i, i = 1, \dots, k \\ & \Leftrightarrow \det(\mathbf{A}_1) < 0, \det(\mathbf{A}_2) > 0, \det(\mathbf{A}_3) < 0, \dots \end{aligned}$$

**Definition 3.6 (Beobachtete Informationsmatrix)**  $\mathcal{L}$  bezeichne die Log-Likelihoodfunktion des Parameters  $\boldsymbol{\theta}$  gegeben  $\mathbf{Y} = \mathbf{y}$ . Die Hessematrix

$$I_{\text{obs}}(\boldsymbol{\theta}|\mathbf{y}) := \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta}|\mathbf{y})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^t} = \left( \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta}|\mathbf{y})}{\partial \theta_i \partial \theta_j} \right)_{i,j=1,\dots,k} \in \mathbb{R}^{k \times k} \quad (3.12)$$

wird in diesem Zusammenhang als beobachtete Informationsmatrix<sup>6</sup> bezeichnet.

Ist die beobachtete Informationsmatrix  $I_{\text{obs}}(\boldsymbol{\theta}|\mathbf{y})$  für die Lösung  $\hat{\boldsymbol{\theta}}$  von (3.11) negativ definit, so liegt eine lokale Maximumstelle von  $\mathcal{L}$  vor (siehe z.B. Råde and Westergren, 1997, S.224). Ist  $I_{\text{obs}}(\boldsymbol{\theta}|\mathbf{y})$  zudem für alle Parameterwerte  $\boldsymbol{\theta}$  negativ definit, ist  $\hat{\boldsymbol{\theta}}$  eindeutiger Maximum-Likelihood-Schätzer von  $\boldsymbol{\theta}$ .

<sup>6</sup>auch: empirische Informationsmatrix

### 3.3.2 Eigenschaften von Maximum-Likelihood-Schätzern

**Definition 3.7 (Informationsmatrix)** Der negative Erwartungswert

$$I(\boldsymbol{\theta}) := -E \left( \frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^t} \right) = -E \left( \frac{\partial^2 \mathcal{L}}{\partial \theta_i \partial \theta_j} \right)_{i,j} \quad (3.13)$$

der Hessematrix  $H$  wird als Informationsmatrix<sup>7</sup> bezeichnet.

In der Regel wird von (Log-)Likelihoodfunktionen die Vertauschbarkeit von Differentiation und Integration (im stetigen Fall) bzw. Summation (im diskreten Fall) verlangt.

$$\frac{\partial}{\partial \boldsymbol{\theta}} \sum_{i=1}^n \ln[p(y_i | \mathbf{x}_i, \boldsymbol{\theta})] = \sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\theta}} \ln[p(y_i | \mathbf{x}_i, \boldsymbol{\theta})] \quad (3.14)$$

Diese Regularitätsbedingung ist bei Modellen erfüllt, denen eine Verteilung aus der exponentiellen Familie zu Grunde liegt.

**Satz 3.8** Für Likelihoodfunktionen, die (3.14) erfüllen gilt:

$$E \left[ \frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}} \right] = E[s(\boldsymbol{\theta}, \mathbf{x})] = 0 \quad (3.15)$$

$$E \left[ \frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^t} \right] - E \left( \left[ \frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}} \right] \left[ \frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}} \right]^t \right) = 0 \quad (3.16)$$

$$I(\boldsymbol{\theta}) = E(s(\boldsymbol{\theta}, \mathbf{x})^t s(\boldsymbol{\theta}, \mathbf{x})) = \text{Cov}(s(\boldsymbol{\theta}, \mathbf{x})) \quad (3.17)$$

*Beweis:*

siehe z.B. Casella and Berger (1990)

### 3.3.3 Die Likelihoodfunktion in verallgemeinerten linearen Modellen

Liegt ein verallgemeinertes lineares Modell vor, so gehört die Verteilung von  $Y_i | \mathbf{x}_i$  zur Klasse der exponentiellen Verteilungen. Für die Beobachtung  $i$  lautet die Log-Likelihoodfunktion deshalb

$$\begin{aligned} \mathcal{L}(\theta_i, \varphi | y_i) &= \ln P(y_i | \theta_i, \varphi) \\ &\stackrel{(A.1)}{=} \ln \left[ \exp \left\{ \frac{\theta_i y_i - b(\theta_i)}{a(\varphi)} + c(y_i, \varphi) \right\} \right] \end{aligned} \quad (3.18)$$

$$= \frac{\theta_i y_i - b(\theta_i)}{a(\varphi)} + c(y_i, \varphi) \quad (3.19)$$

Durch Differenzieren von (3.19) nach  $\theta_i$  erhalten wir:

$$\frac{\partial \mathcal{L}}{\partial \theta_i} = \frac{y_i - b'_i(\theta_i)}{a(\varphi)} \quad (3.20)$$

Mit der Identität (3.15) gilt daher:

$$\begin{aligned} E \left[ \frac{y_i - b'_i(\theta_i)}{a(\varphi)} \right] &= 0 \\ \Rightarrow E(y_i) &= b'_i(\theta_i) \end{aligned} \quad (3.21)$$

<sup>7</sup>auch: Informationsnummer oder Fisher-Information

Für den bedingten Response in verallgemeinerten linearen Modellen gilt zusammenfassend:

$$E(\mathbf{Y}|\mathbf{x}) = \mathbf{b}'(\boldsymbol{\theta}) \quad (3.22)$$

$$\text{Var}(\mathbf{Y}|\mathbf{x}) = \mathbf{b}''(\boldsymbol{\theta})a(\varphi) \quad (3.23)$$

Die Größe  $\mathbf{b}''(\boldsymbol{\theta})$  wird in diesem Zusammenhang gelegentlich als **Varianzfunktion** bezeichnet.

**Definition 3.9 (Natürliche Linkfunktion)** Sei  $\theta_i$ ,  $i = 1, \dots, n$  der natürliche Parameter<sup>8</sup> der Verteilung von  $Y_i|\mathbf{x}_i$ ,  $i = 1, \dots, n$ . Gilt für die linearen Prädiktoren

$$\eta_i = \theta_i \quad \forall i \quad (3.24)$$

so wird die zugehörige Linkfunktion  $g$  als natürlicher Link oder natürliche Linkfunktion<sup>9</sup> bezeichnet.

Durch Verwendung der natürlichen Linkfunktion vereinfacht sich die Log-Likelihoodfunktion im Vergleich zur Verwendung einer allgemeinen Linkfunktion (siehe z.B. Pruscha, 1996, S.281).

### 3.4 Messung der Anpassungsqualität

Der Anpassungsprozess eines Modells an die Daten kann als Austausch der Datenwerte  $y_i$  gegen die angepassten Werte

$$\hat{y}_i = \hat{\mu}_i = h(\mathbf{x}_i^t \hat{\boldsymbol{\beta}}) \quad (3.25)$$

betrachtet werden, wobei  $h$  die Responsefunktion des Modells und  $\hat{\boldsymbol{\beta}}$  der ermittelte Schätzer für die Regressionsparameter ist. Für eine Stichprobe mit  $n$  Beobachtungen können Modelle angepasst werden, die bis zu  $n$  Parameter beinhalten.

Das einfachste Modell ist das Nullmodell, welches nur einen Parameter beinhaltet. Dieser Parameter entspricht einem gemeinsamen Erwartungswert (Mittelwert)  $\mu$  für alle  $y_i$ . Damit wird die gesamte Variabilität in den Daten der stochastischen Komponente, d.h. dem Fehlerterm zugewiesen.

Der andere Extremfall liegt in Form des vollen Modells mit  $n$  Parametern vor. Da für jede der  $n$  Beobachtungen ein Parameter vorhanden ist, können die Parameter so gewählt werden, dass die  $\mu_i$  exakt den  $y_i$  entsprechen. Damit ist die gesamte Variabilität in den Daten der systematischen Komponente zugeordnet. Das volle Modell ist an für sich uninteressant, da es keine Vereinfachung der Zusammenhänge beschreibt. Es kann jedoch zur Messung der Anpassungsqualität eines Modelles mit  $k < n$  Parametern verwendet werden.

Da im Allgemeinen die Werte  $\hat{\mu}_i$  nicht den Werten  $y_i$  entsprechen, stellt sich die Frage nach der Diskrepanz zwischen den tatsächlichen und den angepassten Daten. Bei klassischen linearen Modellen wird diese Abweichung mit Hilfe der **Residuen**

$$r_i = y_i - \hat{\mu}_i, \quad i = 1, \dots, n \quad (3.26)$$

gemessen.

---

<sup>8</sup>vgl. Definition A.1

<sup>9</sup>auch: kanonische Linkfunktion

**Lemma 3.10 (Verteilung der Residuen in linearen Modellen)** Die Modellmatrix  $\mathbf{X}$  besitze vollen Rang. Verwenden wir zur Modellanpassung Kleinste-Quadrate- bzw. Maximum-Likelihood-Schätzer, so gilt für die Residuen  $r_i = y_i - \hat{\mu}_i$  in linearen Modellen:

$$r_i \sim N(0, [1 - h_{ii}]\sigma^2), \quad i = 1, \dots, n \quad (3.27)$$

wobei  $h_{ii}$  das  $i$ -te Diagonalelement der sog. Hat-Matrix  $\mathbf{H} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$  ist.

*Beweis:*

Linearen Modellen dient die Identität als Responsefunktion (vgl. strukturelle Annahme (3.2)). Mit  $h(x) = x$  folgt daher:

$$\hat{\boldsymbol{\mu}} \stackrel{(3.25)}{=} \mathbf{X}^t \hat{\boldsymbol{\beta}} \quad (3.28)$$

Andererseits erfüllt jeder Kleinste-Quadrate-Schätzer und damit jeder Maximum-Likelihood-Schätzer<sup>10</sup> die Normalgleichungen

$$(\mathbf{X}^t\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}^t\mathbf{y} \quad (3.29)$$

(siehe z.B. Fahrmeir and Hamerle, 1984, S.87). Da wir einen vollen Rang der Matrix  $\mathbf{X}$  voraussetzen, folgt:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y} \\ \Rightarrow \hat{\boldsymbol{\mu}} &\stackrel{(3.28)}{=} \mathbf{X}^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y} = \mathbf{H}\mathbf{y} \end{aligned}$$

Die Residuen lauten daher:

$$\mathbf{r} = \mathbf{y} - \mathbf{H}\mathbf{y} = (\mathbf{I} - \mathbf{H})\mathbf{y} \quad (3.30)$$

$\mathbf{I}$  bezeichnet dabei die Einheitsmatrix,  $\mathbf{H}$  die Hat-Matrix. Diese ist eine Projektionsmatrix, d.h. sie ist

• idempotent:  $\mathbf{H}^2 = [\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t][\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t] = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t = \mathbf{H} \quad (3.31)$

• symmetrisch:  $\mathbf{H}^t = [\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t]^t = \mathbf{X}[(\mathbf{X}^t\mathbf{X})^{-1}]^t\mathbf{X}^t = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t = \mathbf{H} \quad (3.32)$

Für Erwartungswert und Varianz von  $\mathbf{r}$  gilt:

$$\begin{aligned} E(\mathbf{r}) &= (\mathbf{I} - \mathbf{H})E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} - \mathbf{H}\mathbf{X}\boldsymbol{\beta} = \mathbf{0} \\ \text{Var}(\mathbf{r}) &\stackrel{(3.1)}{=} [\mathbf{I} - \mathbf{H}][\mathbf{I} - \mathbf{H}]^t\sigma^2 \\ &\stackrel{(3.32)}{=} [\mathbf{I} - \mathbf{H}]^2\sigma^2 = [\mathbf{I} - 2\mathbf{H} + \mathbf{H}^2]\sigma^2 \\ &\stackrel{(3.31)}{=} [\mathbf{I} - \mathbf{H}]\sigma^2 \end{aligned}$$

Da  $\mathbf{I} - \mathbf{H}$  deterministisch ist, bestimmt  $\mathbf{y}$  die Verteilung der Residuen  $\mathbf{r} \Rightarrow \mathbf{r} \sim N(\mathbf{0}, [\mathbf{I} - \mathbf{H}]\sigma^2)$ .

□

<sup>10</sup>In linearen Modellen stimmen Kleinste-Quadrate-Schätzer und Maximum-Likelihood-Schätzer überein (siehe z.B. Zhigljavsky, 2003, S.24)

Insbesondere besagt Lemma 3.10, dass die Residuen symmetrisch um Null verteilt sind. Mit Hilfe der **Fehlerquadratsumme**<sup>11</sup>

$$\begin{aligned} RSS &= \sum_{i=1}^n r_i^2 \\ &= \sum_{i=1}^n (y_i - \hat{\mu}_i)^2 \end{aligned} \quad (3.33)$$

lässt sich die Anpassungsqualität des ermittelten Modells quantifizieren. (3.33) ist ein Maß dafür, wie gut die Variabilität der  $y$ -Werte durch die Regressoren erklärt wird.

Bei nicht linearen Modellen hängt der Einfluss der Abweichung von  $y_i$  und  $\hat{\mu}_i$  von dem Wert der Zielgröße  $y_i$  ab. Die Summe der Fehlerquadrate (3.33) kann deshalb hier nicht verwendet werden. Beispiele für geeignete Diskrepanzmaße finden sich z.B. in McCullagh and Nelder (1989) ab Seite 37. Das gängigste Diskrepanzmaß ist die Devianzfunktion.

### 3.4.1 Devianzresiduen

Die Devianzfunktion ist ein Maß dafür, wie gut das gewählte Modell an die vorliegenden Daten angepasst ist. Sie misst die Abweichung zwischen den wahren Daten  $\mathbf{y}$  und den angepassten Daten  $\hat{\mathbf{y}} = \hat{\boldsymbol{\mu}}$

**Definition 3.11 (unskalierte Devianz)** Für den Responsevektor  $\mathbf{y}$  und die durch das vorliegende Regressionsmodell angepassten Daten  $\hat{\boldsymbol{\mu}}$  ist die unskalierte Devianzfunktion definiert als

$$\begin{aligned} D(\mathbf{y}, \hat{\boldsymbol{\mu}}) &= -2 \log \left( \frac{L(\hat{\boldsymbol{\mu}}, \mathbf{y})}{L(\mathbf{y}, \mathbf{y})} \right) \\ &= -2 [\mathcal{L}(\hat{\boldsymbol{\mu}}, \mathbf{y}) - \mathcal{L}(\mathbf{y}, \mathbf{y})] \end{aligned} \quad (3.34)$$

#### Bemerkung:

- In Ausdruck (3.34) ist  $L(\hat{\boldsymbol{\mu}}, \mathbf{y})$  die Likelihoodfunktion für die gewählte Spezifikation.  $L(\mathbf{y}, \mathbf{y})$  ist die Likelihoodfunktion für das volle Modell, das eine exakte Reproduktion ermöglicht und somit die maximal mögliche Anpassung widerspiegelt.
- Da  $L(\mathbf{y}, \mathbf{y})$  nicht von den Parametern  $\boldsymbol{\beta}$  abhängt, entspricht die Maximierung von  $\mathcal{L}(\boldsymbol{\mu}, \mathbf{y})$  der Minimierung der Devianzfunktion  $D(\mathbf{y}, \boldsymbol{\mu})$ .  
 $\Rightarrow$  Der Maximum-Likelihood-Schätzer liefert die niedrigste Devianz.

Jede Beobachtung  $i = 1, \dots, n$  trägt einen gewissen Anteil  $d_i$  zum Devianzmaß  $D$  bei:

$$D = \sum_{i=1}^n d_i$$

---

<sup>11</sup>auch: Residuenquadratsumme



**Definition 3.12 (Devianzresiduen)** Für die Beobachtungen  $i = 1, \dots, n$  und die Größen  $d_i$  definieren wir die Devianzresiduen durch

$$r_i^D = \text{sign}(y_i - \mu_i) \sqrt{d_i} \quad (3.35)$$

$\text{sign}$  bezeichnet dabei die Vorzeichenfunktion

$$\text{sign}(x) = \begin{cases} -1 & x < 0 \\ 0 & x = 0 \\ 1 & x > 0 \end{cases} \quad (3.36)$$

**Bemerkung:**

- $r_i^D$  ist eine von  $y_i - \mu_i$  abhängige Größe
- Für  $r_i^D$  gilt

$$D = \sum_{i=1}^n (r_i^D)^2$$

### 3.4.2 Akaike-Informationskriterium

Vor allem zur Variablenselektion wird die Akaike-Information (siehe z.B. Venables and Ripley, 2003, S.173) verwendet. Sie eignet sich dafür, zu entscheiden, ob gewisse Kovariablen zusätzlich in ein Modell aufgenommen werden sollen.

**Definition 3.13 (Akaike-Information)** Für ein Regressionsmodell mit  $k$  Parametern dessen Log-Likelihoodfunktion  $\mathcal{L}$  berechnet werden kann, definieren wir die Akaike-Information als

$$AIC = -2\mathcal{L}(\hat{\boldsymbol{\mu}}, \mathbf{y}) + 2k \quad (3.37)$$

(AIC = Akaike Information Criterion)

Auf Grund von Definition 3.13 besteht das Akaike-Informationskriterium darin, jenes Modell auszuwählen, welches die Akaike-Information AIC minimiert.  $-2\mathcal{L}(\hat{\boldsymbol{\mu}}, \mathbf{y})$  gibt dabei die Genauigkeit der Anpassung an, während  $2k$  eine Art Strafterm für die im Modell enthaltenen Parameter darstellt. Diese zwei Komponenten sollen es ermöglichen mittels AIC ein gut angepasstes Modell zu finden, bei dem die Anzahl der Parameter möglichst gering ist:

Mit jeder Kovariablen, die in ein bestehendes Modell aufgenommen wird, steigt dessen Genauigkeit. Die damit einhergehende Erhöhung des Wertes der Log-Likelihoodfunktion  $\mathcal{L}$  führt zur Verringerung des ersten Summanden der Akaike-Information. Dieser Verringerung steht jedoch eine Erhöhung des zweiten Summanden entgegen, da mit der Anzahl der Kovariablen auch die Anzahl der Parameter anwächst. Durch die Hinzunahme einer Kovariablen verringert sich der AIC-Wert daher nur dann, wenn die Steigerung der Anpassungsqualität zu einer um mindestens den Wert Eins erhöhten Log-Likelihoodfunktion führt.

Da die Zunahme der Log-Likelihoodfunktion mit steigender Kovariablenanzahl geringer wird, sinkt die Akaike-Information ausgehend vom Nullmodell im Allgemeinen zunächst, um ab einer gewissen Anzahl von Parametern wieder anzuwachsen.

**Definition 3.14 (Eingebettete Modelle)** *Ein Regressionsmodell  $M_1$  heißt eingebettet<sup>12</sup> in ein anderes Regressionsmodell  $M_2$ , falls es vollständig in diesem enthalten ist - d.h. die Parametermenge von  $M_1$  ist eine echte Teilmenge der Parametermenge von  $M_2$ .*

**Bemerkung:**

- Ein in das Modell  $M_2$  eingebettetes Modell  $M_1$  kann dadurch gewonnen werden, dass ein oder mehrere unabhängige Parameter von  $M_2$  restringiert werden.
- Der Vergleich von nicht eingebetteten Regressionsmodellen ist mit der Akaike-Information im Allgemeinen nicht möglich, da sich die Log-Likelihoodfunktionen grundlegend unterscheiden. Ineinander eingebettete Regressionsmodelle können jedoch mit AIC hinsichtlich ihrer Anpassungsqualität verglichen werden.

---

<sup>12</sup>engl.: nested

# Kapitel 4

## Modellierung von Zähldaten

Liegen uns Beobachtungen vor, bei denen die Zielvariable diskrete, nicht negative Werte annimmt, so sprechen wir von **Zähldaten**. Diese können aus der Beobachtung eines Punkt- bzw. Zählprozesses (z.B. des Verzweigungsprozesses in Abschnitt 1.5) entstehen.

### 4.1 Das Poisson-Regressionsmodell

Das Standardmodell um Zähldaten mit Kovariablen zu modellieren ist das Poisson-Regressionsmodell. Es handelt sich hierbei um ein nicht lineares Regressionsmodell, das sich aus der Poisson-Verteilungsannahme für den Response ableitet. Hängt die Intensität  $\theta_i$  der Responseverteilung ausschließlich von den Regressoren  $x_{ij}$  ab, liegt das im folgenden beschriebene Standardmodell der Poissonregression vor. Wird die Intensität von zusätzlichen stochastischen Varianzquellen beeinflusst, so handelt es sich um ein gemischtes Poisson-Regressionsmodell.

#### Modellbeschreibung:

Die zu untersuchenden Daten bestehen aus  $n$  unabhängigen Beobachtungen  $(Y_i, \mathbf{x}_i)$ . Die Zielvariable  $Y_i$  gibt die Anzahl der aufgetretenen Ereignisse wieder, für die wir uns interessieren.  $\mathbf{x}_i$  ist wie zuvor ein Vektor linear unabhängiger Regressoren, von denen vermutet wird, dass sie  $Y_i$  beeinflussen.

Die Spezifikation eines Poisson-Regressionsmodells besteht aus:

(i) *Stochastische Komponente:*

Die bedingten Zielgrößen  $Y_i|\mathbf{x}_i$  seien unabhängig poissonverteilt mit Intensitäten  $\theta_i$ .

(ii) *Systematische Komponente:*

Die Regressoren  $\mathbf{x}_i = (x_{i1}, \dots, x_{ik})^t$  beeinflussen die Verteilung von  $Y_i|\mathbf{x}_i$  nur durch den linearen Prädiktor

$$\eta_i = \mathbf{x}_i^t \boldsymbol{\beta} \tag{4.1}$$

(iii) *Verknüpfungskomponente:*

Der Erwartungswert  $\mu_i = E(Y_i|\mathbf{x}_i)$  ist wegen der Verteilungsannahme (i) strikt positiv. Wählen wir die Identität als Linkfunktion, so kann dies nicht gewährleistet werden. Für  $\mu_i = \eta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$ ,  $i = 1, \dots, n$  können immer Parameter so gewählt

werden, dass die Bedingung  $\mu_i > 0$  für mindestens ein  $i$  verletzt ist. Üblicherweise wird daher der logarithmische Link

$$\eta_i = \ln \mu_i \quad (4.2)$$

gewählt. Dies ist gleichbedeutend mit

$$\mu_i = \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} \quad (4.3)$$

Der Erwartungswert  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^t$  ist daher durch positive multiplikative Effekte bestimmt.

### Bemerkungen:

- Gemäß Satz A.5 gehört die Poissonverteilung zur Klasse der einparametrischen exponentiellen Verteilungen. Die Standardtheorie für verallgemeinerte lineare Modelle ist daher auf Poisson-Regressionsmodelle anwendbar.
- Der natürliche Parameter (vgl. Abschnitt A.1) der Poissonverteilung ist gleich dem Logarithmus der Intensität. Damit ist durch die logarithmische Verknüpfung die kanonische Linkfunktion der Poissonregression bestimmt, da  $\eta_i = \ln \mu_i = \ln \theta_i$  gilt.
- Eine weitere mögliche Linkfunktion ist die Quadratwurzel.

$$\eta_i = \sqrt{\mu_i} \Rightarrow \mu_i = (\mathbf{x}_i^t \boldsymbol{\beta})^2 \geq 0 \quad (4.4)$$

Diese führt jedoch nicht zu einer so einfachen Einflusststruktur der Effekte wie in (4.3).

- Von der Poissonverteilung ist bekannt, dass ihr Erwartungswert mit ihrer Varianz übereinstimmt. Im vorliegenden Regressionsmodell folgt daher für die Varianzfunktion:

$$\text{Var}(Y_i | \mathbf{x}_i) = E(Y_i | \mathbf{x}_i) = \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} \quad (4.5)$$

Die Varianz hängt wie aus (4.5) ersichtlich von den Kovariablen  $\mathbf{x}_i$  ab. Das Regressionsmodell ist also heteroskedastisch, was bedeutet, dass die Zielvariablen  $Y_i$  unterschiedliche Varianz besitzen.

- Ausführlich behandelt wird das Poisson-Regressionsmodell in Cameron and Trivedi (1998).

## 4.2 Abweichende Länge der Beobachtungszeiträume

### 4.2.1 Offset in der Poissonregression

Die Poissonverteilung zählt, wie in Anhang A.2 beschrieben, die Anzahl des Eintritts eines bestimmten Ereignisses während eines vorgegebenen Zeitintervalls. Dieses Zeitintervall bezeichnen wir auch als **Beobachtungszeitraum**. In der bisherigen Parametrisierung des Poisson-Regressionsmodells wurde von einheitlichen Beobachtungszeiträumen ausgegangen. Häufig unterscheiden sich jedoch die Zeiträume, für welche  $Y_i$  die eintretenden Ereignisse zählt, in ihrer Länge  $t_i^*$ . Um die Beobachtungszeiträume hinsichtlich ihrer Länge zu normieren, werden die Intensitäten  $\theta_i$  mit dem Faktor

$$t_i := \frac{1}{t_i^*} \quad (4.6)$$

multipliziert. Damit ist die Zielvariable  $Y_i|\mathbf{x}_i$  im erweiterten Ansatz poissonverteilt mit Intensität

$$\tilde{\theta}_i := t_i\theta_i \quad (4.7)$$

Für  $i = 1, \dots, n$  gilt daher:

$$\begin{aligned} P(Y_i = y_i|\mathbf{x}_i) &\stackrel{(A.2)}{=} \frac{\tilde{\theta}_i^{y_i}}{y_i!} e^{-\tilde{\theta}_i} \\ &\stackrel{(4.7)}{=} \frac{(t_i\theta_i)^{y_i}}{y_i!} e^{-t_i\theta_i}, \quad y_i = 0, 1, 2, \dots \end{aligned}$$

Der Erwartungswert der bedingten Zielvariablen lautet:

$$\begin{aligned} E(Y_i|\mathbf{x}_i) &\stackrel{(A.4)}{=} \tilde{\theta}_i \\ &\stackrel{(4.7)}{=} t_i\theta_i \end{aligned} \quad (4.8)$$

Der Faktor  $t_i$  dient lediglich der Normierung der zu den Beobachtungen  $Y_i$ ,  $i = 1, \dots, n$  gehörenden Beobachtungszeiträumen. Die Längen  $t_i^*$  dieser Zeiträume und damit auch die Faktoren  $t_i$  sind bekannt. Die Regression beeinflusst die Intensität  $\tilde{\theta}_i$  daher nur durch die Größe  $\theta_i$ . Wir verwenden deshalb weiterhin die Verknüpfung

$$\theta_i = \exp\{\mathbf{x}_i^t\boldsymbol{\beta}\}. \quad (4.9)$$

Im Regressionsfall gilt folglich für die Erwartungswertfunktion:

$$\begin{aligned} E(Y_i|\mathbf{x}_i) &\stackrel{(4.8),(4.9)}{=} t_i \exp\{\mathbf{x}_i^t\boldsymbol{\beta}\} \\ &= \exp\{\ln(t_i) + \mathbf{x}_i^t\boldsymbol{\beta}\} \\ &=: \mu_i \end{aligned} \quad (4.10)$$

Mit dieser Gleichung für die Erwartungswertfunktion erhalten wir den linearen Prädiktor des Poisson-Regressionsmodelles bei unterschiedlichen Beobachtungszeiträumen:

$$\begin{aligned} \eta_i &\stackrel{(4.1)}{=} \mathbf{x}_i^t\boldsymbol{\beta} \\ &\stackrel{(4.10)}{=} \ln(\mu_i) - \ln(t_i) \end{aligned} \quad (4.11)$$

Die Größe  $\ln(t_i)$  wird hierbei als **Offset** oder Ausgleichsfaktor bezeichnet. Dieser ist bekannt, falls die Länge  $t_i^*$  des Beobachtungszeitraumes von  $Y_i$  bekannt ist.

### 4.2.2 Modellbeschreibung einer Poissonregression mit Offset

Für die Daten gelten die selben Voraussetzungen wie in Abschnitt 4.1. Die Spezifikation eines Poisson-Regressionsmodells bei Berücksichtigung unterschiedlicher Beobachtungszeiträumen lautet unter Verwendung der Aussagen aus Abschnitt 4.2.1:

(i) *Stochastische Komponente:*

Die bedingten Zielgrößen  $Y_i|\mathbf{x}_i$  seien unabhängig poissonverteilt mit Intensitäten  $\tilde{\theta}_i$ :

$$Y_i|\mathbf{x}_i \stackrel{\text{ind.}}{\sim} Poi(\tilde{\theta}_i), \quad i = 1, \dots, n \quad (4.12)$$

(ii) *Systematische Komponente:*

Die Regressoren  $\mathbf{x}_i = (x_{i1}, \dots, x_{ik})^t$  beeinflussen die Verteilung von  $Y_i|\mathbf{x}_i$  nur durch den linearen Prädiktor

$$\eta_i = \mathbf{x}_i^t \boldsymbol{\beta} \quad (4.13)$$

(iii) *Verknüpfungskomponente:*

Erwartungswert  $\mu_i = E(Y_i|\mathbf{x}_i)$  und linearer Prädiktor  $\eta_i$  sind durch die Funktion

$$\eta_i = \ln \mu_i - \ln t_i \quad (4.14)$$

miteinander verknüpft. Die Ausgleichsfaktoren  $\ln t_i$  sind bekannt.

### 4.2.3 Natürliche Linkfunktion

**Satz 4.1** *In dem durch (4.12) und (4.13) beschriebenen Regressionsmodell mit Offset ist die natürliche Linkfunktion durch (4.14) bestimmt.*

*Beweis:*

Mit (4.8) gilt:

$$P(Y_i = y_i) = \exp\{-t_i \theta_i + y_i \ln(t_i \theta_i) - \ln(y_i!)\} \quad (4.15)$$

Entsprechend (A.6) wählen wir

$$\begin{aligned} a(\varphi) &= \varphi = 1 \\ \text{und } b(\theta_i^*) &= e^{\theta_i^*} \text{ mit } \theta_i^* = \ln(t_i \theta_i) \end{aligned}$$

um (4.15) auf die Form einer Wahrscheinlichkeitsfunktion aus der exponentiellen Familie (A.1) zu bringen. Der natürliche Parameter ist dann durch

$$\theta_i^* = \ln(t_i \theta_i) \quad (4.16)$$

gegeben. Für die Erwartungswertfunktion gilt:

$$\begin{aligned} \mu_i = E(Y_i|\mathbf{x}_i) &\stackrel{(3.22)}{=} b'(\theta_i^*) \\ &= \frac{\partial}{\partial \theta_i^*} e^{\theta_i^*} = e^{\theta_i^*} \end{aligned} \quad (4.17)$$

Setzen wir nun (4.16) ein:

$$\begin{aligned} E(Y_i|\mathbf{x}_i) &= \exp\{\ln(t_i \theta_i)\} = t_i \theta_i \\ &\stackrel{(4.9)}{=} t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} \\ &= \exp\{\mathbf{x}_i^t \boldsymbol{\beta} + \ln t_i\} \end{aligned} \quad (4.18)$$

$$\Rightarrow \ln \mu_i = \mathbf{x}_i^t \boldsymbol{\beta} + \ln t_i \quad (4.19)$$

Für den linearen Prädiktor gilt daher:

$$\begin{aligned} \eta_i &= \ln \mu_i - \ln t_i \\ &\stackrel{(4.17)}{=} \theta_i^* - \ln t_i \end{aligned} \quad (4.20)$$

Falls  $t_i = 1$  für alle Beobachtungen  $i = 1, \dots, n$ , so ist die natürliche Linkfunktion durch  $g(\mu_i) = \ln \mu_i$  gegeben. Sind die Beobachtungszeiträume jedoch unterschiedlich, so lautet die natürliche Linkfunktion

$$g(\mu_i) = \ln \mu_i - \ln t_i$$

□

**Bemerkung:**

Mit der Gleichung (4.18) lässt sich  $\ln t_i$  auch als weitere Kovariable auffassen, für die bekannt ist, dass der Regressionskoeffizient gleich Eins ist.

### 4.3 Missspezifikation

In Abschnitt 4.1 wurden zur Spezifikation des Poisson-Regressionsmodells folgende Annahmen getroffen:

- (i) Annahme unabhängiger Beobachtungen
- (ii) Poisson-Verteilungsannahme (Insbesondere gilt  $E(Y_i|\mathbf{x}_i) = \text{Var}(Y_i|\mathbf{x}_i)$ )
- (iii) strukturelle Annahme  $E(Y_i|\mathbf{x}_i) = \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}$

Als **Missspezifikation** bezeichnet man die Verletzung einer oder mehrerer dieser Annahmen. Bei der Poissonregression besteht eine enge Beziehung zwischen der Erwartungswertfunktion  $E(Y_i|\mathbf{x}_i)$ , der Varianzfunktion  $\text{Var}(Y_i|\mathbf{x}_i)$  und der Verteilung. Die Gleichheit von bedingtem Erwartungswert und Varianz führt zum Verlust eines Freiheitsgrades. Eine Verletzung der Annahme über die Varianzfunktion hat daher auch die Verletzung der Verteilungsannahme zur Folge. Bei Zähldaten wird diese Annahme sehr häufig durch das Auftreten von Überdispersion verletzt.

**Definition 4.2 (Überdispersion im Poisson-Regressionsmodell)** *Gilt für die beobachteten Zähldaten  $(Y_i, \mathbf{x}_i)$*

$$\text{Var}(Y_i|\mathbf{x}_i) > E(Y_i|\mathbf{x}_i) \quad (4.21)$$

so liegt Überdispersion vor.

Missspezifikation führt in vielen Fällen zu Überdispersion. Im folgenden sind mögliche Ursachen für Missspezifikation genannt.

#### 4.3.1 Verknüpfungskomponente

Bei Poissonregression lautet die Regressionsfunktion

$$\begin{aligned} E(Y_i|\mathbf{x}_i) &= \mu_i \\ &\stackrel{(4.3)}{=} \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} = \theta_i \end{aligned} \quad (4.22)$$

Die wahre Intensität sei nun mit  $\theta_i^*$ , die wahre Erwartungswertfunktion sei mit  $E^*$  bezeichnet. Dann gilt:

$$E^*(Y_i|\mathbf{x}_i) = \theta_i^* = f(\mathbf{x}_i, \boldsymbol{\beta}^*)$$

Die Erwartungswertfunktion (4.22) ist unzureichend spezifiziert, falls für beliebiges  $\mathbf{x}_i$  kein  $\boldsymbol{\beta}$  existiert, so dass  $\mu_i = f(\mathbf{x}_i, \boldsymbol{\beta}^*)$  gilt. Mögliche Ursachen hierfür sind (Winkelmann, 2003, S.88ff.):

- Erklärende Variablen wurden nicht berücksichtigt (vgl. Abschnitt 4.3.2).
- Der Prädiktor  $\boldsymbol{\eta}$  ist nicht linear in den Regressionsparametern  $\boldsymbol{\beta}$ .
- Erklärende Variablen gehen in den linearen Prädiktor nicht linear sondern durch eine Transformation  $f(\boldsymbol{x}_i)$  ein.
- Die Linkfunktion ist falsch gewählt. Die wahre Erwartungswertfunktion ist z.B. linear in den Regressoren und nicht log-linear.

### 4.3.2 Unbeobachtete Heterogenität

Können die betrachteten Kovariablen nicht die gesamte Variabilität in den Daten erklären, so liegt **unbeobachtete Heterogenität** vor. Diese Situation lässt sich als Fehler in der Erwartungswertfunktion interpretieren, die in diesem Fall nicht alle bestimmenden Kovariablen enthält. Das wahre Modell sei durch die Erwartungswertfunktion

$$\mu_i^* = \exp\{\boldsymbol{x}_i^t \boldsymbol{\beta} + \boldsymbol{z}_i^t \boldsymbol{\gamma}\} \quad (4.23)$$

gegeben, wobei die Kovariablen  $\boldsymbol{z}_i$  bei der Modellbildung nicht berücksichtigt werden, da sie nicht beobachtet oder weggelassen werden. Falls  $\boldsymbol{x}_i$  und  $\boldsymbol{z}_i$  unabhängig sind, gilt mit  $\mu_i = \exp\{\boldsymbol{x}_i^t \boldsymbol{\beta}\}$  und  $\nu_i = \exp\{\boldsymbol{z}_i^t \boldsymbol{\gamma}\}$ :

$$\mu_i^* = \mu_i \nu_i \quad (4.24)$$

Die Intensität  $\mu_i^*$  besitzt somit zwei Variationsquellen:

- $\mu_i$ : beobachtet, abhängig von  $\boldsymbol{x}_i$
- $\nu_i$ : unbeobachtet, unabhängig von  $\boldsymbol{x}_i$  (da von  $\boldsymbol{z}_i$  verursacht)

Da  $\nu_i$  unbeobachtet ist, muss es wie eine Zufallsvariable behandelt werden, was die Intensität  $\mu_i^*$  ebenfalls zu einer Zufallsvariable werden lässt. Setzen wir  $\varepsilon_i := \ln \nu_i$  so erhalten wir mit  $\varepsilon_i$  einen additiven Fehlerterm auf der logarithmierten Skala:

$$\mu_i^* = \exp\{\boldsymbol{x}_i^t \boldsymbol{\beta} + \varepsilon_i\} \quad (4.25)$$

Wir nehmen im folgenden an, dass der Erwartungswert für die Fehler  $\varepsilon_i$  Null beträgt, was gleichbedeutend mit  $E(\nu_i) = 1$  ist. Ein systematischer Fehler soll also nicht vorliegen. Weiterhin bezeichne  $\sigma_{\nu_i}^2$  die Varianz von  $\nu_i$ . Für den Parameter  $\mu_i^*$  gilt dann:

$$\begin{aligned} E(\mu_i^* | \mu_i) &\stackrel{(4.24)}{=} E(\mu_i \nu_i | \mu_i) \\ &= \mu_i E(\nu_i | \mu_i) = \mu_i E(\nu_i) \\ &= \mu_i \end{aligned} \quad (4.26)$$

$$\begin{aligned} \text{Var}(\mu_i^* | \mu_i) &\stackrel{(4.24)}{=} \text{Var}(\mu_i \nu_i | \mu_i) \\ &= \mu_i^2 \text{Var}(\nu_i | \mu_i) = \mu_i^2 \text{Var}(\nu_i) \\ &= \mu_i^2 \sigma_{\nu_i}^2 \end{aligned} \quad (4.27)$$

Falls  $f(Y_i | \nu_i)$  die bedingte Dichte von  $Y_i$  unter  $\nu_i$  und  $g(\nu_i)$  die Dichte von  $\nu_i$  bezeichnet, gilt für die Dichte der Responseverteilung:

$$f(Y_i) = \int_{-\infty}^{\infty} f(Y_i | \nu_i) g(\nu_i) d\nu_i \quad (4.28)$$



(vgl. Råde and Westergren, 1997, S.413). Da die Verteilung von  $\nu_i$  nicht bekannt ist, kann  $f(Y_i)$  nicht berechnet werden. Sind jedoch Erwartungswert und Varianz von  $\nu_i$  bekannt, können auch die ersten beiden Momente von  $Y_i$  bestimmt werden.

**Satz 4.3**  *$X, Y$  seien Zufallsvariablen. Besitzt  $Y$  einen endlichen Erwartungswert und endliche Varianz, so gilt:*

$$E(Y) = E[E(Y|X)] \quad (4.29)$$

$$\begin{aligned} \text{Var}(Y) &= E[\text{Var}(Y|X)] + \text{Var}[E(Y|X)] \\ &\geq \text{Var}[E(Y|X)] \end{aligned} \quad (4.30)$$

*Beweis:*

Bickel and Doksum (1977, S.6) bzw. Råde and Westergren (1997, S.413)

□

Im Fall von Poissonregression mit unbeobachteter Heterogenität gelten die Annahmen

$$Y_i | \mathbf{x}_i, \nu_i \stackrel{\text{unabh.}}{\sim} \text{Poi}(\theta_i^*) \quad (4.31)$$

$$\theta_i^* = \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} \nu_i \quad (4.32)$$

Die  $\nu_i$  sind unabhängig identisch verteilt mit  $E(\nu_i) = 1$ ,  $\text{Var}(\nu_i) = \sigma_\nu^2$  und unabhängig von  $\mathbf{x}_i$ . Gesucht ist  $Y_i | \mathbf{x}_i$ , da  $Y_i$  und  $\mathbf{x}_i$  beobachtet wurden,  $\nu_i$  aber unbeobachtet bleibt. Mit Satz 4.3 gilt hier:

$$\begin{aligned} \text{Var}(Y_i | \mathbf{x}_i) &= E_{\nu_i}(\theta_i^* | \mathbf{x}_i) + \text{Var}_{\nu_i}(\theta_i^* | \mathbf{x}_i) \\ &\stackrel{(4.26), (4.27)}{=} \mu_i + \sigma_\nu^2 \mu_i^2 \end{aligned} \quad (4.33)$$

Da  $\mu_i = E(Y_i | \mathbf{x}_i)$  gilt, übersteigt die Varianz den Erwartungswert:

$$\text{Var}(Y_i | \mathbf{x}_i) > E(Y_i | \mathbf{x}_i) \quad (4.34)$$

Unbeobachtete Heterogenität dieser Art führt somit zu Überdispersion.

#### **Bemerkung:**

Modelle mit  $E(Y_i | \mathbf{x}_i) = \mu_i \nu_i$  und unabhängig von  $\mu_i$  verteiltem  $\nu_i$  werden auch als Mischungsmodelle bezeichnet. Ein solches ist die negative Binomialverteilung (s. Abschnitt 2.1).

### **4.3.3 Nullenüberschuss**

Nullenüberschuss<sup>1</sup> in Zähldaten liegt vor, wenn die Anzahl der in den Respondedaten zu beobachtenden Nullen die erwartete Anzahl übersteigt. Auf die Behandlung dieses Problems wird in den Kapiteln 6 und 7 näher eingegangen werden. In Kapitel 7 werden kombinierte Modelle mit Hürde Null betrachtet, bei denen für das Ereignis Null und die positiven Ereignisse verschiedene Verteilungen zu Grunde gelegt werden. Die Wahrscheinlichkeit des Ereignisses Null wird durch die sogenannte Hürdenverteilung bestimmt, die Wahrscheinlichkeiten aller übrigen (positiven) Ereignisse durch eine strikt positive Verteilung. Letztere entspricht der bei Null abgeschnittenen

<sup>1</sup>engl.: excess of zeros

sogenannten Ursprungsverteilung für die Zähldaten. Wird als Ursprungsverteilung die Poissonverteilung verwendet, so werden wir in Satz 7.7 folgern:

$$\begin{aligned} \text{Var}(Y_i|\mathbf{x}_i) &= E(Y_i|\mathbf{x}_i) \left( 1 + \frac{1-\varrho}{\varrho} E(Y_i|\mathbf{x}_i) \right) \\ &> E(Y_i|\mathbf{x}_i) \end{aligned}$$

$0 \leq \varrho < 1$  ist dabei eine Normalisierungskonstante für die abgeschnittene Ursprungsverteilung. Bei Zero-Inflated Modells wie sie in Kapitel 6 beschrieben werden, können Nullen nicht nur durch einen Auswahlprozess entstehen, wie er bei kombinierten Modellen durch die Hürdenverteilung bestimmt ist, sondern auch durch die Ursprungsverteilung selbst. Somit existiert keine Einschränkung der Ursprungsverteilung auf echt positive Ereignisse. Wird als Ursprungsverteilung die Poissonverteilung verwendet, so folgt wie wir in Satz 6.7 sehen werden:

$$\begin{aligned} \text{Var}(Y_i|\mathbf{x}_i) &= E(Y_i|\mathbf{x}_i) \left( 1 + \frac{\omega}{1-\omega} E(Y_i|\mathbf{x}_i) \right) \\ &> E(Y_i|\mathbf{x}_i) \end{aligned}$$

wobei  $0 < \omega \leq 1$  den Auswahlprozess steuert.

Nullenüberschuss führt daher wie unbeobachtete Heterogenität zu Überdispersion und liefert somit eine weitere Erklärung dieses Phänomens. Es existieren verschiedene Testverfahren, um poissonverteilte Zähldaten auf Nullenüberschuss zu testen. Es sei hier auf Broeck (1995) und Mullahy (1997) verwiesen.

#### 4.3.4 Messfehler in den Kovariablen

Nehmen wir nun an, dass für die Kovariable  $\mathbf{x}_i$  anstatt dem wahren Wert der Wert  $\mathbf{z}_i := \mathbf{x}_i + \varepsilon$  gemessen wird.  $\varepsilon$  entspricht dann dem Messfehler von dem wir annehmen, dass er

$$\begin{aligned} E(\varepsilon) &= 0 \\ \text{Var}(\varepsilon) &= \Sigma \end{aligned}$$

erfüllt und unabhängig von  $\mathbf{x}_i$  ist. Die bedingte Response-dichte lautet dann:

$$f(Y_i|\mathbf{z}_i) = \int_{-\infty}^{\infty} f(Y_i|\mathbf{x}_i)g(\mathbf{x}_i|\mathbf{z}_i)d\mathbf{x}_i$$

(Råde and Westergren, 1997, S.413). Gilt  $g(\mathbf{x}_i|\mathbf{z}_i) \neq 0$  fast überall, so kann gezeigt werden (Winkelmann, 2003, S.93), dass für  $E(Y_i|\mathbf{x}_i) = \text{Var}(Y_i|\mathbf{x}_i)$

$$E(Y_i|\mathbf{z}_i) < \text{Var}(Y_i|\mathbf{z}_i)$$

gilt. Bei der Poissonregression führen Messfehler der beschriebenen Art daher ebenfalls zu Überdispersion.

### 4.3.5 Abhängigkeiten im Erzeugungsprozess der Daten

Die Poissonverteilung ist dazu geeignet Zählraten zu beschreiben, bei denen aufeinanderfolgende Ereignisse unabhängig voneinander mit derselben Rate eintreten. Die Erzeugung der Daten erfolgt daher durch einen zugrundeliegenden Lévy-Prozess<sup>2</sup>. Treten die gezählten Ereignisse im Zeitverlauf nicht unabhängig voneinander auf, so ist das Poissonmodell unzureichend. Dies ist der Fall, wenn der Prozess nicht gedächtnislos ist - d.h. die Wahrscheinlichkeit eines Ereignisses im Intervall  $(t, t + h]$  ist nicht unabhängig vom Zustand des Prozesses zum Zeitpunkt  $t$ .

Entsprechende Abhängigkeitsstrukturen können z.B. durch Verzweigungsprozesse modelliert werden. In Abschnitt 1.5.2 wurde ein solcher Prozess beschrieben, aus welchem die verallgemeinerte Poissonverteilung entstand. Auch die negative Binomialverteilung lässt sich durch einen entsprechenden Verzweigungsprozess erzeugen (Winkelmann, 2003, S.94).

Eine weitere Möglichkeit zur Modellierung von Abhängigkeitsstrukturen in stochastischen Prozessen wird von Winkelmann (1995) beschrieben. Sind die modellierten Verweildauern des Prozesses dort positiv korreliert, so entsteht Unterdispersion, liegt eine negative Abhängigkeit vor, erhält man Überdispersion. Regressionsmodelle für die Mechanismus, welcher Überdispersion hervorruft, bekannt ist, sind in (McCullagh and Nelder, 1989, S.198 ff.) beschrieben.

### 4.3.6 Selektivität

Stichprobenselektion liegt vor, wenn die Daten so erzeugt werden, dass nicht die zugrundeliegende Zählvariable  $Y^*$  sondern eine **zensierte Größe**  $Y$  beobachtet wird. Eine Größe wird zensiert genannt, wenn z.B. bei der Erhebung Klassen wie "unter x" oder "x und mehr" gebildet werden. Selektivität kann darüber hinaus auftreten, wenn die Zählvariable von einer anderen Variable abhängt (Greene, 2000). In diesem Fall spricht man von endogener Selektivität. Beide Fälle führen zu Misspezifikation.

### 4.3.7 Underreporting

Bei Zählraten wird von **underreporting** gesprochen, wenn nur ein Teil der Gesamtanzahl an Ereignissen gezählt wird. Bezeichnet  $Y^*$  die Anzahl an Ereignissen, so lässt sich die betrachtete Anzahl  $Y$  durch

$$Y := \sum_{i=1}^{Y^*} \delta_i \quad (4.35)$$

mit der Indikatorvariable

$$\delta_i := \begin{cases} 1 & \text{Ereignis } i \text{ wird gezählt} \\ 0 & \text{Ereignis } i \text{ wird nicht gezählt} \end{cases} \quad (4.36)$$

beschreiben. Für die Beobachtungen gilt dann  $Y \leq Y^*$  und die Verteilung von  $Y$  hängt von der gemeinsamen Verteilung von  $Y^*$  und den  $\delta_i$ 's ab.

---

<sup>2</sup>stochastischer Prozess mit unabhängigen, stationären Zuwächsen

## 4.4 Behandlung von Überdispersion in Zähldaten

Die häufigsten Ursachen für Überdispersion sind die eben beschriebenen Missspezifikations-Quellen

- Unbeobachtete Heterogenität
- Nullenüberschuss

Äquivalent zu Definition 4.2 lässt sich Überdispersion im Poissonmodell durch die Gleichung

$$\text{Var}[Y_i|\mathbf{x}_i] = \psi(E[Y_i|\mathbf{x}_i]) \quad (4.37)$$

für ein stetige Funktion  $\psi$  mit  $\psi(x) > x, \forall x > 0$  ausdrücken. Die Funktion  $\psi(x) = \varphi x$  erfüllt für  $\varphi > 1$  diese Anforderung. Es gilt:

$$\text{Var}(Y_i|\mathbf{x}_i) = \varphi E(Y_i|\mathbf{x}_i) \quad (4.38)$$

In diesem Zusammenhang wird die Größe  $\varphi$  als **Dispersionsparameter** bezeichnet. Eine weitere geeignete Funktion ist  $\psi(x) = x(1 + ax)$ ,  $a > 0$ . In diesem Fall gilt:

$$\text{Var}(Y_i|\mathbf{x}_i) = E(Y_i|\mathbf{x}_i) [1 + aE(Y_i|\mathbf{x}_i)] \quad (4.39)$$

Die Größe  $a$  wird hierbei als **Dispersionsindex** bezeichnet<sup>3</sup>. Stellen wir Überdispersion fest, so können wir entweder die Poisson-Verteilungsannahme aufrechterhalten und versuchen  $\varphi$  bzw.  $a$  zu bestimmen oder die Verteilungsannahme abändern.

### 4.4.1 Beibehaltung der Poisson-Verteilungsannahme

Gehen wir weiterhin von poissonverteilten Responsedaten aus, so wird der Erwartungswert wie bisher bestimmt. Die Varianz wird jedoch abweichend zum bisherigen Modell als Funktion des Erwartungswertes und des Dispersionsparameter  $\varphi$  definiert.

Nachdem die Modellparameter  $\boldsymbol{\theta}$  geschätzt sind, wird der Dispersionsparameter an die Daten angepasst. Es findet keine simultane Schätzung von  $\boldsymbol{\theta}$  und  $\varphi$  statt, da der Dispersionsparameter bei dieser Herangehensweise als Störparameter auftritt.

### 4.4.2 Abänderung der Verteilungsannahme

Verwerfen wir die Poisson-Verteilungsannahme, so müssen wir eine andere Verteilung für die Zähldaten wählen. Diese sollte flexibler in Bezug auf ihre Momente sein als die Poissonverteilung - d.h. die Annahmen für die Varianz werden gelockert. Solche Verteilungen können wir gewinnen, wenn wir die Annahmen für den Erzeugungsprozess der Zähldaten lockern. Auf diese Weise entsteht wie in Abschnitt 1.5 beschrieben die verallgemeinerte Poissonverteilung. Es können auch Poisson-Mischmodelle wie die negative Binomialverteilung (vgl. Abschnitt 2.1.5) verwendet werden. Eine weitere Möglichkeit, falls Überdispersion von Nullenüberschuss herrührt, bieten die in den Kapiteln 6 und 7 beschriebenen Zero-Inflated- und Zero-Hurdle-Modelle.

---

<sup>3</sup>Gleichung (4.38) entspricht dem Zusammenhang zwischen Erwartungswert und Varianz bei der verallgemeinerten Poissonverteilung, (4.39) korrespondiert mit der Erwartungswert-Varianz-Struktur der negativen Binomialverteilung.

## Kapitel 5

# Das verallgemeinerte Poisson-Regressionsmodell

Die in Abschnitt 4.1 beschriebene Poissonregression eignet sich sehr gut zur Analyse von Zähl-  
daten, bei denen kein großer Unterschied zwischen empirischer Erwartung und empirischer Va-  
rianz besteht. Weichen diese jedoch signifikant voneinander ab, so wird die Poisson-Verteilungs-  
annahme häufig verworfen. Sowohl für Unter- als auch für Überdispersion wurden daher Re-  
gressionsmodelle entwickelt, welche auf alternativen Verteilungen für nicht negative ganzzahlige  
Zufallsvariablen basieren.

In Kapitel 1 wurde die verallgemeinerte Poissonverteilung vorgestellt. Da diese für nicht negative  
ganzzahlige Zufallsvariablen definiert ist, kann auch sie zur Modellierung von Zähl-  
daten herangezogen werden. Die Besonderheit des resultierenden Regressionsmodells ist, dass die Art der  
Dispersion bzw. deren Vorhandensein keinen Einfluss auf die Anwendbarkeit des Modells hat.  
Im Gegensatz zu anderen Verteilungsalternativen zur Poissonverteilung lässt die verallgemei-  
nerte Poissonverteilung durch die Struktur ihrer Varianz Spielraum sowohl für unter- als auch  
für überdispensierte Daten (siehe Abschnitt 1.3). Wir wollen uns bei der folgenden Betrachtung  
des verallgemeinerten Poisson-Regressionsmodells, wie es von Consul and Famoye (1992)  
vorgeschlagen wird, jedoch auf den häufigeren Fall der Überdispersion beschränken, für den übli-  
cherweise das negative Binomial-Regressionsmodell<sup>1</sup> verwendet wird.

### 5.1 Modellbeschreibung

Die zu untersuchenden Zähl-  
daten bestehen aus  $n$  unabhängigen Beobachtungen  $(Y_i, \mathbf{x}_i)$ . Wir  
spezifizieren das verallgemeinerte Poisson-Regressionsmodell wie folgt:

(i) *Stochastische Komponente:*

Die bedingten Zielvariablen  $Y_i | \mathbf{x}_i$  sind unabhängig verallgemeinert poissonverteilt mit ge-  
meinsamem Dispersionsparameter.

(ii) *Systematische Komponente:*

Die Regressoren  $\mathbf{x}_i = (x_{i1}, \dots, x_{ik})^t$  beeinflussen die Verteilung von  $Y_i | \mathbf{x}_i$  nur durch den  
linearen Prädiktor

$$\eta_i = \mathbf{x}_i^t \boldsymbol{\beta} \tag{5.1}$$

---

<sup>1</sup>siehe hierzu z.B. Lawless (1987)

(iii) *Verknüpfungskomponente:*

Um einen positiven Erwartungswert zu gewährleisten wird der log-lineare Link

$$\begin{aligned}\eta_i &= \ln \mu_i \\ \Leftrightarrow \mu_i &= \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}\end{aligned}\quad (5.2)$$

verwendet.

**Bemerkung:**

Aus obiger Spezifikation ergeben sich die Regressionsfunktionen  $E(Y_i|\mathbf{x}_i)$  und  $Var(Y_i|\mathbf{x}_i)$

- Der bedingte Erwartungswert

$$\begin{aligned}E(Y_i|\mathbf{x}_i) &= \mu(\mathbf{x}_i, \boldsymbol{\beta}) \\ &= \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} = \mu_i\end{aligned}\quad (5.3)$$

ist eine stetige Funktion des Kovariablenvektors  $\mathbf{x}_i$  und den Regressionsparametern  $\boldsymbol{\beta}$ .

- Die bedingte Varianz

$$\begin{aligned}Var(Y_i|\mathbf{x}_i) &= \varphi^2 E(Y_i|\mathbf{x}_i) \\ &= \varphi^2 \mu_i\end{aligned}\quad (5.4)$$

ist eine stetige Funktion des Erwartungswertes  $\mu_i$  und einem Dispersionsparameter  $\varphi$ .

Durch den Dispersionsparameter in (5.4) ist die Beziehung zwischen der Erwartungswertfunktion  $E(Y_i|\mathbf{x}_i)$  und der Varianzfunktion  $Var(Y_i|\mathbf{x}_i)$  im verallgemeinerten Poisson-Regressionsmodell flexibler als im gewöhnlichen Poissonmodell.

## 5.2 Responseverteilung

Betrachten wir die zur Responseverteilung gehörende Wahrscheinlichkeitsfunktion

$$\begin{aligned}P_x(\theta, \lambda) &\stackrel{(1.1)}{=} \frac{\theta(\theta + x\lambda)^{x-1}}{x!} e^{-(\theta+x\lambda)} \\ &= \exp\{\ln(\theta) + (x-1)\ln(\theta + x\lambda) - \ln(x!) - \theta - x\lambda\}\end{aligned}\quad (5.5)$$

so erkennen wir ein zentrales Problem der verallgemeinerten Poissonregression. Da die Parameter  $\theta$  und  $\lambda$  im Term  $\ln(\theta + x\lambda)$  in (5.5) nicht getrennt werden können, kann (1.1) nicht auf die Form (A.1) einer exponentiellen Verteilung gebracht werden. Die verallgemeinerte Poissonverteilung gehört also nicht der exponentiellen Verteilungsfamilie an. Damit ist die wesentliche Voraussetzung über die Responseverteilung für ein verallgemeinertes lineares Modell verletzt. Dies wiederum hat zur Folge, dass sich die Aussagen über Existenz, Eindeutigkeit und asymptotischem Verhalten der Schätzer, die für verallgemeinerte lineare Modelle getroffen wurden, nicht automatisch auf das verallgemeinerte Poisson-Regressionsmodell übertragen lassen.

### 5.2.1 Die eingeschränkte verallgemeinerte Poissonverteilung

In Consul (1989) wird als Variante der verallgemeinerten Poissonverteilung die sog. eingeschränkte verallgemeinerte Poissonverteilung<sup>2</sup> vorgestellt.

**Definition 5.1 (Eingeschränkte verallgemeinerte Poissonverteilung)** Sei  $X$  eine Zufallsvariable mit diskreter Verteilung, welche auf  $\mathbb{N}$  definiert ist.  $P_x(\theta, \alpha)$  bezeichne die Wahrscheinlichkeit  $P(X = x)$ , dass  $X$  den nicht negativen ganzzahligen Wert  $x$  annimmt.  $X$  ist eingeschränkt verallgemeinert poissonverteilt mit Parametern  $\theta$  und  $\alpha$ , falls

$$P_x(\theta, \alpha) = \begin{cases} \frac{\theta^x (1+\alpha x)^{x-1}}{x!} e^{-\theta(1+\alpha x)} & x = 0, 1, 2, \dots \\ 0 & \text{für } x > m, \text{ wenn } \alpha < 0 \end{cases} \quad (5.6)$$

und 0 in allen anderen Fällen (d.h. für  $x \notin \mathbb{N}$ ). Für die Parameter gelten dabei folgende Einschränkungen:

- $\theta > 0$
- $\max\{-\frac{1}{\theta}, -\frac{1}{m}\} \leq \alpha \leq \frac{1}{\theta}$

wobei  $m$  die größte natürliche Zahl ist, sodass  $\theta(1 + m\alpha) > 0$  ist, falls  $\alpha < 0$ .

Wir schreiben  $X \sim \tilde{GP}(\theta, \alpha)$ .

Die eben definierte Verteilung entspricht der verallgemeinerten Poissonverteilung, wenn deren Parameter  $\lambda$  proportional zu  $\theta$  ist:

**Satz 5.2 (Zusammenhang zwischen  $\tilde{GP}$ - und  $GP$ -Verteilung)** Für die Zufallsvariablen  $X$  und  $Y$  gelte:  $X \sim GP(\theta, \lambda)$  und  $Y \sim \tilde{GP}(\theta, \alpha)$ . Verhält sich der Parameter  $\lambda$  in der Verteilung von  $X$  gemäß

$$\lambda = \alpha\theta \quad (5.7)$$

proportional zum Parameter  $\theta$ , so stimmen die Verteilungen der beiden Zufallsvariablen überein.

*Beweis:*

$$\begin{aligned} P(X = \omega) &\stackrel{(1.1)}{=} \frac{\theta(\theta + \omega\lambda)^{\omega-1}}{\omega!} e^{-(\theta + \omega\lambda)} \\ &\stackrel{(5.7)}{=} \frac{\theta(\theta + \omega(\alpha\theta))^{\omega-1}}{\omega!} e^{-(\theta + \omega(\alpha\theta))} \\ &= \frac{\theta^\omega (1 + \alpha\omega)^{\omega-1}}{\omega!} e^{-\theta(1 + \alpha\omega)} \\ &\stackrel{(5.6)}{=} P(Y = \omega), \quad \omega = 0, 1, 2, \dots \end{aligned}$$

Die Wahrscheinlichkeitsfunktionen (1.1) und (5.6) besitzen damit für übereinstimmendes  $m$  den selben Ausdruck. Dividieren wir den Ausdruck für den Zulässigkeitsbereich von  $\lambda$  aus Definition 1.1 durch das positive  $\theta$  so folgt der Zulässigkeitsbereich von  $\alpha$  aus Definition 5.1:

<sup>2</sup>engl.: restricted generalized Poisson distribution

$$\begin{aligned} \max\left\{-1, -\frac{\theta}{m}\right\} &\leq \lambda \leq 1 \\ \Leftrightarrow^{5.7} \max\left\{-1, -\frac{\theta}{m}\right\} &\leq \alpha\theta \leq 1 \\ \Leftrightarrow \max\left\{-\frac{1}{\theta}, -\frac{1}{m}\right\} &\leq \alpha \leq \frac{1}{\theta} \end{aligned}$$

□

**Satz 5.3 (Verteilungsklasse der eingeschränkten verallg. Poissonverteilung)** Die eingeschränkte verallgemeinerte Poissonverteilung gehört für  $\theta \in \Theta$  und festes  $\alpha$  zur Klasse der exponentiellen Verteilungen.

*Beweis:*

$$\begin{aligned} P_x(\theta, \alpha) &\stackrel{(5.6)}{=} \frac{\theta^x (1 + \alpha x)^{x-1}}{x!} e^{-\theta(1 + \alpha x)} \\ &= \exp\left\{x \ln \theta + \ln \left[\frac{(1 + \alpha x)^{x-1}}{x!}\right] - \theta(1 + \alpha x)\right\} \\ &= \exp\left\{x(\ln \theta - \alpha\theta) - \theta + \ln \left[\frac{(1 + \alpha x)^{x-1}}{x!}\right]\right\} \end{aligned} \quad (5.8)$$

Mit den Funktionen

$$\begin{aligned} c(\theta) &= \ln \theta - \alpha\theta & t(x) &= x \\ d(\theta) &= -\theta & s(x) &= \ln \left[\frac{(1 + \alpha x)^{x-1}}{x!}\right] \end{aligned}$$

gilt für die Wahrscheinlichkeitsfunktion der eingeschränkten verallgemeinerten Poissonverteilung

$$P_x(\theta, \alpha) \stackrel{(5.8)}{=} \exp\{c(\theta)t(x) + d(\theta) + s(x)\}$$

□

Der Vorteil der eingeschränkten verallgemeinerten Poissonverteilung im Vergleich zur verallgemeinerten Poissonverteilung liegt darin, dass sie in der Klasse der exponentiellen Verteilungen liegt. Im Regressionskontext führt eine entsprechende Verteilungsannahme daher auf ein verallgemeinertes lineares Modell, was die Anwendbarkeit der dafür vorhandenen Theorie ermöglicht. Wir wollen diesen Ansatz jedoch nicht weiter verfolgen, da  $\alpha$  die unschöne Eigenschaft besitzt über den Zulässigkeitsbereich vom Parameter  $\theta$  abhängig zu sein bzw. dessen Zulässigkeitsbereich einschränkt. Wir können somit nicht beide Größen unabhängig voneinander wählen.

### 5.2.2 Umparametrisierung der verallgemeinerten Poissonverteilung

Um den Zusammenhang zwischen den Modellparametern und den Regressionsfunktionen

$$E(y_i | \mathbf{x}_i) = \mu_i \text{ und } Var(y_i | \mathbf{x}_i) = \varphi^2 \mu_i$$

deutlicher herausstellen, wollen wir für die verallgemeinerte Poissonverteilung ab hier, wie von Consul and Famoye (1992) vorgeschlagen, eine andere Parametrisierung verwenden.



**Definition 5.4 (Verallgemeinerte Poissonverteilung mit  $\varphi$ -Parametrisierung)** Sei  $X$  eine Zufallsvariable mit diskreter Verteilung, welche auf  $\mathbb{N}$  definiert ist.  $P_x(\mu, \varphi)$  bezeichne die Wahrscheinlichkeit  $P(X = x)$ , dass  $X$  den nicht negativen Wert  $x$  annimmt.  $X$  ist verallgemeinert poissonverteilt mit Parametern  $\mu$  und  $\varphi$ , falls

$$P_x(\mu, \varphi) = \begin{cases} \frac{\mu(\mu+(\varphi-1)x)^{x-1}}{x!} \varphi^{-x} e^{-\frac{1}{\varphi}(\mu+(\varphi-1)x)} & x = 0, 1, 2, \dots \\ 0 & \text{für } x > m \text{ wenn } \varphi < 1 \end{cases} \quad (5.9)$$

und 0 in allen anderen Fällen (d.h. für  $x \notin \mathbb{N}$ ) gilt. Für die Parameter gelten dabei folgende Einschränkungen:

- $\mu > 0$
- $\varphi \geq \max\{\frac{1}{2}, 1 - \frac{\mu}{m}\}$

wobei  $m$  die größte natürliche Zahl ist, sodass  $\mu + m(\varphi - 1) > 0$  ist, falls  $\varphi < 1$

Wir schreiben  $X \sim GP^*(\mu, \varphi)$ .

Ist  $\varphi = 1$ , so entspricht die verallgemeinerte Poissonregression der klassischen Poissonregression. Gilt  $\varphi > 1$ , so bildet das Modell (5.9) Überdispersion ab. Für  $\frac{1}{2} \leq \varphi < 1$  stellt es Unterdispersion dar, falls  $\mu > \frac{m}{2}$  ist.

**Satz 5.5 (Übereinstimmung der Definitionen 1.1 und 5.4)** Die Zufallsvariable  $X$  sei  $GP(\theta, \lambda)$ -verteilt, die Zufallsvariable  $Y$  sei  $GP^*(\mu, \varphi)$ -verteilt. Falls

$$\mu = \theta\varphi \quad \text{und} \quad \varphi = \frac{1}{1 - \lambda}$$

gilt, stimmen die Verteilungen von  $X$  und  $Y$  überein.

*Beweis:*

Für die Parameter  $\theta$ ,  $\lambda$ ,  $\mu$  und  $\varphi$  besteht folgender Zusammenhang:

$$\begin{aligned} \mu = \theta\varphi & & \varphi = \frac{1}{1 - \lambda} \\ \Leftrightarrow \theta = \frac{\mu}{\varphi} & & \lambda = 1 - \frac{1}{\varphi} \end{aligned} \quad (5.10)$$

Verwenden wir (5.10) nun um die Wahrscheinlichkeitsfunktion von  $X$  umzuformen:

$$\begin{aligned} P(X = \omega) & \stackrel{(1.1)}{=} \frac{\theta(\theta + \omega\lambda)^{\omega-1}}{\omega!} e^{-(\theta + \omega\lambda)} \\ & \stackrel{(5.10)}{=} \frac{\frac{\mu}{\varphi}(\frac{\mu}{\varphi} + (1 - \frac{1}{\varphi})\omega)^{\omega-1}}{\omega!} e^{-(\frac{\mu}{\varphi} + (1 - \frac{1}{\varphi})\omega)} \\ & = \frac{\mu(\frac{1}{\varphi}(\mu + (\varphi - 1)\omega))^{\omega-1}}{\varphi \omega!} e^{-\frac{1}{\varphi}(\mu + (\varphi - 1)\omega)} \\ & = \frac{\mu(\mu + (\varphi - 1)\omega)^{\omega-1}}{\omega!} \varphi^{-\omega+1-1} e^{-\frac{1}{\varphi}(\mu + (\varphi - 1)\omega)} \\ & = \frac{\mu(\mu + (\varphi - 1)\omega)^{\omega-1}}{\omega!} \varphi^{-\omega} e^{-\frac{1}{\varphi}(\mu + (\varphi - 1)\omega)} \\ & \stackrel{(5.9)}{=} P(Y = \omega), \quad \omega = 0, 1, 2, \dots \end{aligned}$$

Damit stimmen die Wahrscheinlichkeitsfunktionen (1.1) und (5.9) überein. Für die Parameterrestriktionen gilt:

$$\max\{-1, -\frac{\theta}{m}\} \leq \lambda \leq 1 \quad \stackrel{(5.10)}{\Leftrightarrow} \quad \max\{-1, -\frac{\mu}{m\varphi}\} \leq 1 - \frac{1}{\varphi} \leq 1$$

Fall 1 ( $-1 > -\frac{\mu}{m\varphi}$ ):

$$\begin{aligned} -1 \leq 1 - \frac{1}{\varphi} \leq 1 &\Leftrightarrow -2 \leq -\frac{1}{\varphi} \leq 0 \\ &\Rightarrow \varphi \geq \frac{1}{2} \end{aligned}$$

Fall 2 ( $-1 \leq -\frac{\mu}{m\varphi}$ ):

$$\begin{aligned} \frac{\mu}{m\varphi} \leq 1 - \frac{1}{\varphi} \leq 1 &\Leftrightarrow -\frac{\mu}{m\varphi} - 1 \leq -\frac{1}{\varphi} \leq 0 \\ &\Rightarrow -\frac{\mu}{m} - \varphi \leq -1 \\ &\Rightarrow \varphi \geq 1 - \frac{\mu}{m} \end{aligned}$$

Zusammenfassend folgt für den Parameter  $\varphi$  der Zulässigkeitsbereich

$$\varphi \geq \max\{\frac{1}{2}, 1 - \frac{\mu}{m}\} \quad (5.11)$$

Da  $\varphi$  nur positive Werte annimmt, kann der Zulässigkeitsbereich von  $\mu$  wie folgt festgelegt werden:

$$\begin{aligned} \theta > 0 &\stackrel{(5.10)}{\Leftrightarrow} \frac{\mu}{\varphi} > 0 \\ &\stackrel{(5.11)}{\Rightarrow} \mu > 0 \end{aligned} \quad (5.12)$$

Auch die Zulässigkeitsbereiche der Parameter für die Verteilungen von X und Y stimmen überein. □

**Korollar 5.6** Für die verallgemeinert poissonverteilte Zufallsvariable  $X \sim GP^*(\mu, \varphi)$  gilt:

$$E(X) = \mu \quad (5.13)$$

$$Var(X) = \varphi^2 \mu \quad (5.14)$$

*Beweis:*

Aus Satz 5.5 folgern wir:

$$X \sim GP^*(\mu, \varphi) \Leftrightarrow X \sim GP(\frac{\mu}{\varphi}, 1 - \frac{1}{\varphi})$$

Wir können also in den in Abschnitt 1.3 bestimmten Ausdrücken für Erwartungswert und Varianz die Parameter  $\theta$  und  $\lambda$  entsprechend (5.10) ersetzen und erhalten die Ausdrücke für den

Erwartungswert und die Varianz von  $X \sim GP^*(\mu, \varphi)$ :

$$\begin{aligned} E(X) &\stackrel{(1.4)}{=} \frac{\frac{\mu}{\varphi}}{1 - (1 - \frac{1}{\varphi})} \\ &= \frac{\mu}{\varphi} \varphi = \mu \\ \text{Var}(X) &\stackrel{(1.36)}{=} \frac{\frac{\mu}{\varphi}}{(1 - (1 - \frac{1}{\varphi}))^3} \\ &= \frac{\mu}{\varphi} \varphi^3 = \varphi^2 \mu \end{aligned}$$

□

Wir können nun erkennen, dass der Ausdruck für den Erwartungswert (5.13) mit dem Ausdruck für die Regressionsfunktion (5.3) übereinstimmt. Entsprechendes gilt für die Varianz und die Regressionsfunktion (5.4).

## 5.3 Maximum-Likelihood-Schätzung im verallgemeinerten Poissonmodell

Bei der verallgemeinerten Poissonregression treten unter Verwendung der Definition 5.4 die Modellparameter  $\mu$ ,  $\varphi$  und die Regressionsparameter  $\boldsymbol{\beta}$  auf. Diese sind über die Regressionsfunktionen (5.3) und (5.4) miteinander verknüpft. Um Schätzer für die Modellparameter zu erhalten, müssen die Parameter  $\boldsymbol{\beta}$  und  $\varphi$  geschätzt werden.

### 5.3.1 Die Likelihoodfunktion im verallgemeinerten Poissonmodell

**Satz 5.7 (Log-Likelihoodfunktion im verallgemeinerten Poissonmodell)** *Wählen wir die Spezifikation aus Abschnitt 5.1 und die Definition 5.4 für die verallgemeinerte Poissonverteilung, so gilt für die Log-Likelihoodfunktion des resultierenden Regressionsmodells:*

$$\begin{aligned} \mathcal{L}(\varphi, \boldsymbol{\beta}) &\propto \sum_{i=1}^n \mathbf{x}_i^t \boldsymbol{\beta} + (y_i - 1) \ln \left[ e^{\mathbf{x}_i^t \boldsymbol{\beta}} + (\varphi - 1) y_i \right] \\ &\quad - y_i \ln \varphi - \frac{1}{\varphi} \left[ e^{\mathbf{x}_i^t \boldsymbol{\beta}} + (\varphi - 1) y_i \right] \end{aligned} \quad (5.15)$$

*Beweis:*

Entsprechend Abschnitt 3.3 erhalten wir im verallgemeinerten Poisson-Regressionsmodell wegen der Unabhängigkeit der Beobachtungen die Likelihoodfunktion:

$$\begin{aligned} L(\varphi, \boldsymbol{\beta}) &= L(\boldsymbol{\beta}, \varphi | \mathbf{y}) \\ &= P(\mathbf{Y} = \mathbf{y} | \varphi, \boldsymbol{\beta}) = \prod_{i=1}^n P(Y_i = y_i | \boldsymbol{\beta}, \varphi) \\ &\stackrel{(5.9)}{=} \prod_{i=1}^n \frac{\mu_i [\mu_i + (\varphi - 1) y_i]^{y_i - 1}}{y_i!} \varphi^{-y_i} \exp \left\{ -\frac{\mu_i + (\varphi - 1) y_i}{\varphi} \right\} \end{aligned} \quad (5.16)$$

und damit die Log-Likelihoodfunktion

$$\begin{aligned}\mathcal{L}(\varphi, \boldsymbol{\beta}) &= \sum_{i=1}^n \ln \mu_i + (y_i - 1) \ln[\mu_i + (\varphi - 1)y_i] \\ &\quad - y_i \ln \varphi - \frac{1}{\varphi} [\mu_i + (\varphi - 1)y_i] - \ln(y_i!)\end{aligned}\quad (5.17)$$

Wird die Verknüpfung (5.2) gewählt, lautet Gleichung (5.17):

$$\begin{aligned}\mathcal{L}(\varphi, \boldsymbol{\beta}) &= \sum_{i=1}^n \mathbf{x}_i^t \boldsymbol{\beta} + (y_i - 1) \ln [e^{\mathbf{x}_i^t \boldsymbol{\beta}} + (\varphi - 1)y_i] \\ &\quad - y_i \ln \varphi - \frac{1}{\varphi} [e^{\mathbf{x}_i^t \boldsymbol{\beta}} + (\varphi - 1)y_i] - \ln(y_i!)\end{aligned}\quad (5.18)$$

□

### Eigenschaften der Log-Likelihoodfunktion

Maximum-Likelihood-Schätzer für die Parameter  $\boldsymbol{\beta}$  und  $\varphi$  erhalten wir durch Maximierung der Log-Likelihoodfunktion. Unstetigkeiten in der Zielfunktion können jedoch zu Komplikationen bei der (numerischen) Maximierung führen. Wird ein Optimierungsalgorithmus verwendet, so kann dieser unter Umständen nicht korrekt arbeiten. Dies ist insbesondere der Fall, wenn zur Maximierung Gradienten erforderlich sind. Die Log-Likelihoodfunktion (5.15) ist sowohl für  $\varphi$  als auch für  $\boldsymbol{\beta}$  stetig und somit unter diesem Aspekt unproblematisch.

Von entscheidender Bedeutung für die Schätzbarkeit von  $\boldsymbol{\beta}$  und  $\varphi$  ist die Anzahl und die Lage deren Maximalstellen. Wir wollen daher die Log-Likelihoodfunktion (5.15) etwas genauer betrachten. Für variablen Dispersionsparameter  $\varphi$  lässt sich diese Funktion allerdings nur für den einfachen Fall eines ein-dimensionalen Regressionsparameters  $\beta$  grafisch veranschaulichen. Dennoch können wir mit Hilfe der sich ergebenden Log-Likelihoodfunktion auf die Gutartigkeit des Maximierungsproblems höherdimensionaler Schätzprobleme schließen. Betrachten wir im folgenden also die Log-Likelihoodfunktion

$$\begin{aligned}\mathcal{L}(\varphi, \beta) &\stackrel{(5.15)}{=} \sum_{i=1}^n \beta x_i + (y_i - 1) \ln [e^{\beta x_i} + (\varphi - 1)y_i] \\ &\quad - y_i \ln \varphi - \frac{1}{\varphi} [e^{\beta x_i} + (\varphi - 1)y_i] - \ln(y_i!)\end{aligned}\quad (5.19)$$

Zur Visualisierung dieser Funktion für variierende  $\varphi$  und  $\beta$  benötigen wir einen Kovariablenvektor  $\mathbf{x}$  und einen gemäß der Verknüpfungsvorschrift (5.2) generierten, verallgemeinert poissonverteilten Responsevektor  $\mathbf{y}$ . Wählen wir daher als Kovariablenvektor eine Sequenz von -1 bis 1 der Länge 1000 mit äquidistanten Einträgen:

$$\mathbf{x} = \begin{pmatrix} -1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^{1000}$$

Der wahre Wert  $\beta^*$  des Regressionsparameters sei Eins, für den wahren Dispersionsparameter gelte  $\varphi^* = 1$  ( $\varphi^* = 2$ ,  $\varphi^* = 10$ ). Unter Verwendung dieser Daten erzeugen wir mit der S-Plus Funktion `response.gpois` (siehe Anhang C.2.1) die Zielgröße  $\mathbf{y}$ . Wird mit den Größen  $\mathbf{x}$  und  $\mathbf{y}$

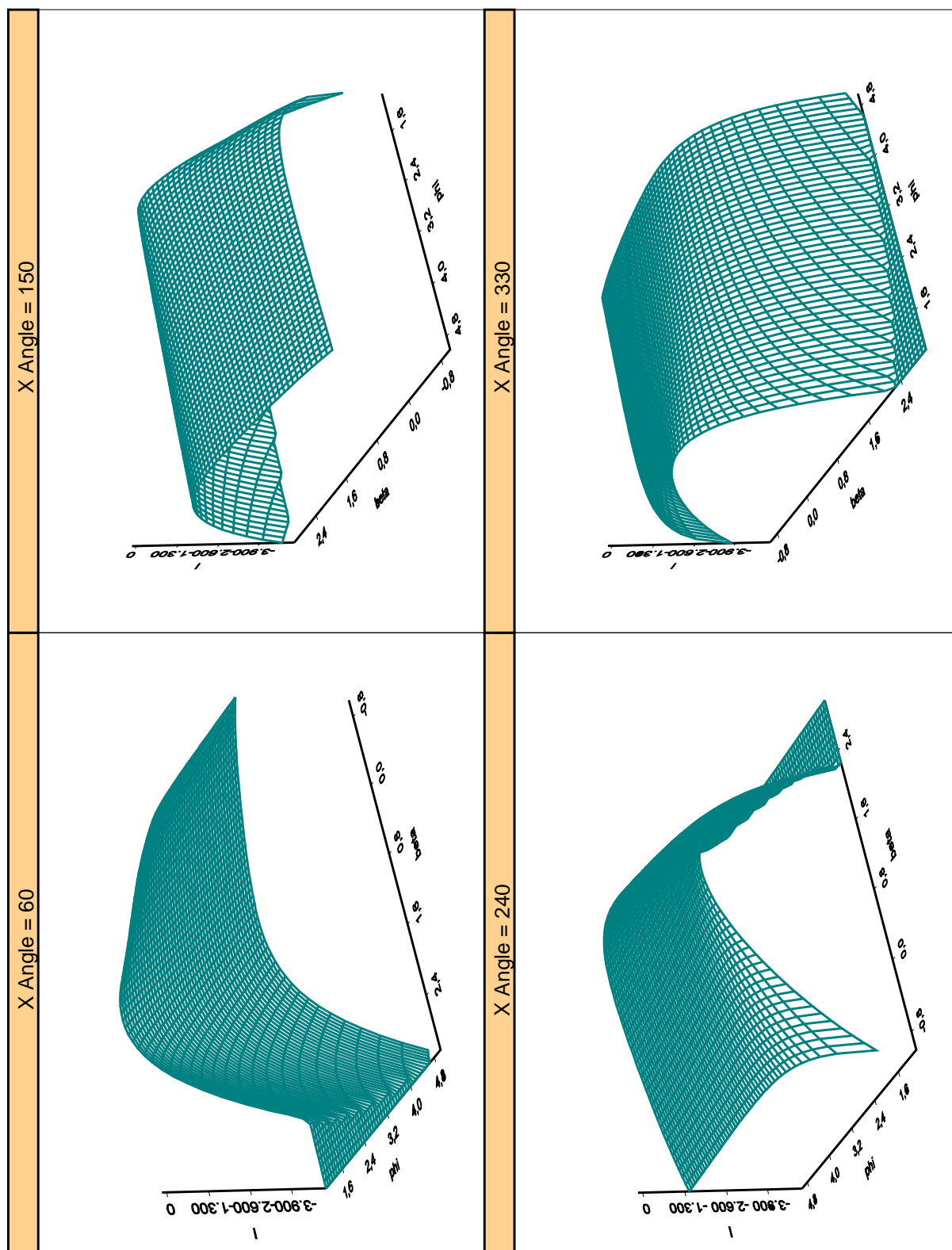


Abbildung 5.1: Log-Likelihoodfunktion des verallgemeinerten Poisson-Regressionsmodells mit wahren Parametern  $\beta^* = 1$  und  $\varphi^* = 1$

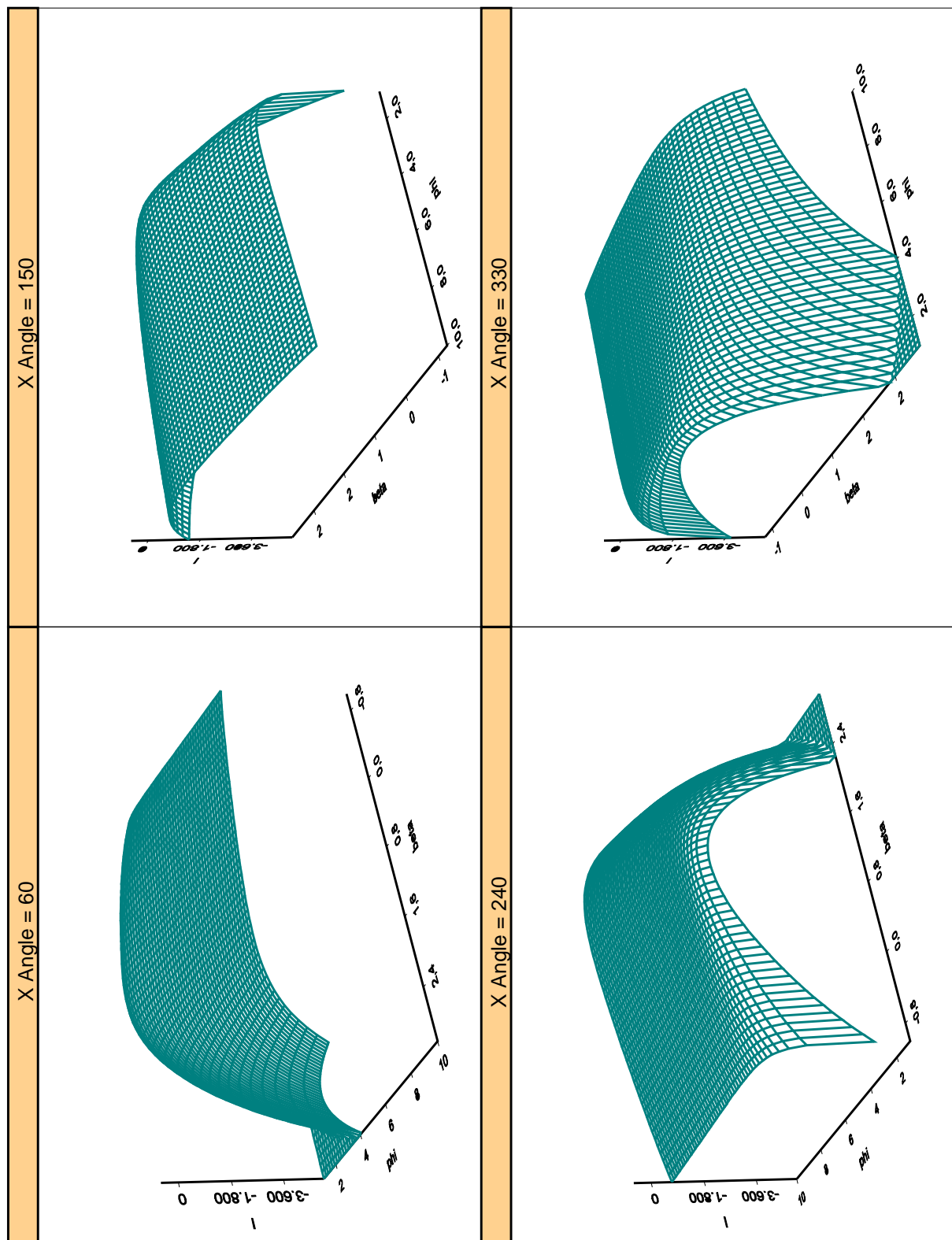


Abbildung 5.2: Log-Likelihoodfunktion des verallgemeinerten Poisson-Regressionsmodells mit wahren Parametern  $\beta^* = 1$  und  $\varphi^* = 2$

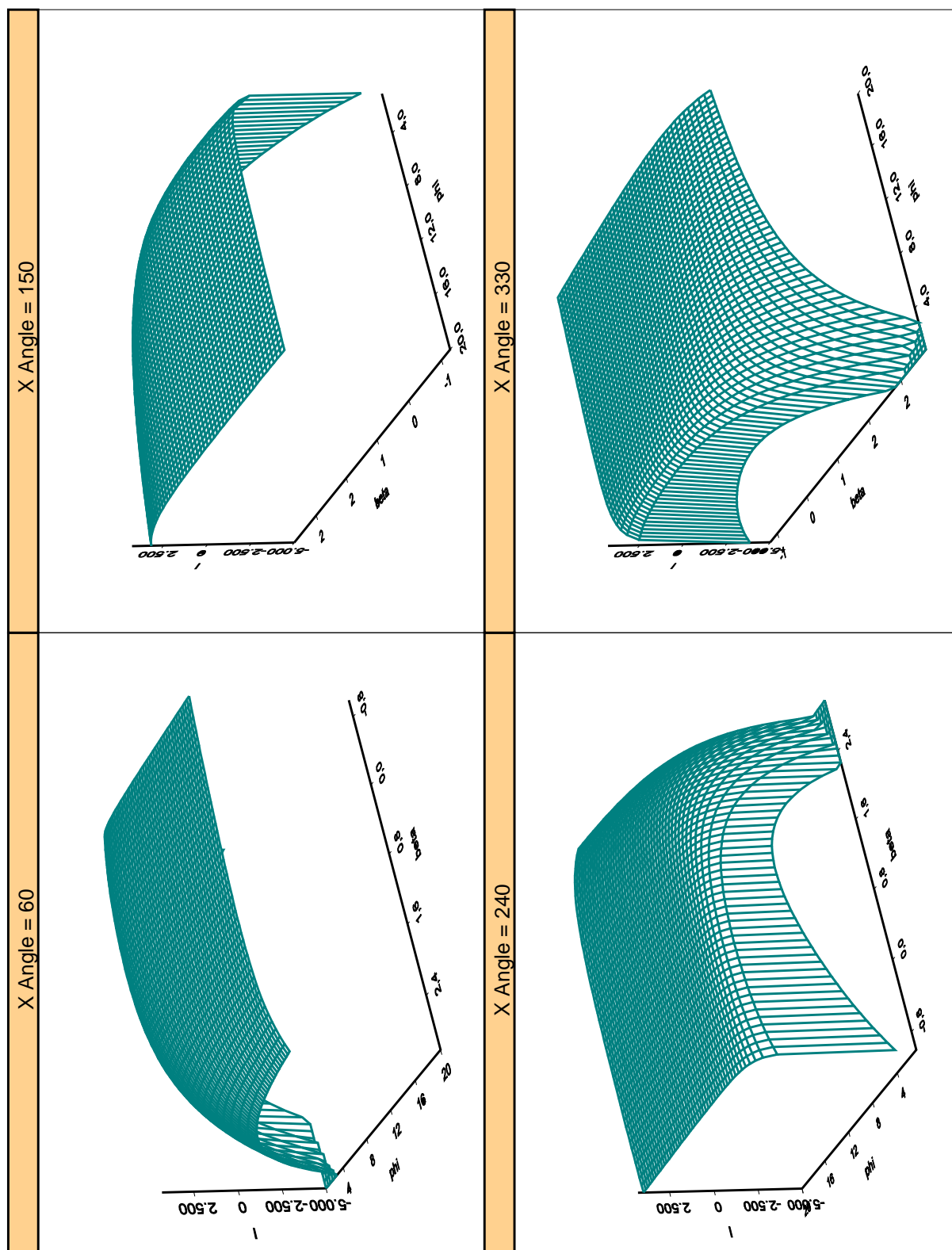


Abbildung 5.3: Log-Likelihoodfunktion des verallgemeinerten Poisson-Regressionsmodells mit wahren Parametern  $\beta^* = 1$  und  $\varphi^* = 10$

für variierende Parameterwerte  $1 < \varphi \leq 10$ ,  $0 \leq \beta \leq 2$  die Log-Likelihoodfunktion (5.19) ausgewertet, so erhalten wir die zu den Abbildungen 5.1 - 5.6 gehörenden Daten.

Bei der Maximierung der Log-Likelihoodfunktion ist Unimodalität wünschenswert, damit bei numerischer Optimierung der verwendete Algorithmus wegen der Existenz eines Maximums zum einen konvergiert und andererseits nicht in einem lokalen Maximum hängen bleibt. Multimodalität in (5.19) dürfte sich auf Log-Likelihoodfunktionen von Regressionsansätzen mit mehrdimensionalen Regressionsparametern übertragen. In der Umgebung der wahren Parameterwerte zeigt  $\mathcal{L}(\varphi, \beta)$  in den Abbildungen 5.1 bis 5.3 einen unimodalen Verlauf.

Führen wir in der Abbildung 5.1 für festes  $\beta$  Schnitte durch, so scheint bzgl.  $\varphi$  Konvexität vorzuliegen (Abbildung 5.4 links). Da Abbildung 5.1 die Log-Likelihoodfunktion für poissonverteiltes  $\mathbf{y}$  darstellt, erreicht  $\mathcal{L}$  sein Maximum bzgl.  $\varphi$  wie zu erwarten ist für den Grenzwert  $\varphi = 1$ . Sind die Zähldaten  $\mathbf{y}$  mit Überdispersion behaftet, so erhalten wir für  $\beta = 1$  die Schnitte auf der rechten Seite der Abbildungen 5.5 und 5.6. Der Graph geht nun für steigendes  $\varphi$  von einem konkaven Bereich in der Umgebung des wahren Dispersionsparameters in einen konvexen Verlauf über. In  $\beta$ -Richtung scheint  $\mathcal{L}(\varphi, \beta)$  konkav zu sein (vgl. Abbildungen 5.4 bis 5.6 (links)). Daher ist ein eindeutiges Maximum zu erwarten.

Ausschlaggebend für die Konvergenzgeschwindigkeit und die Genauigkeit der gefundenen Lösung bei numerischer Maximierung ist die Norm der Gradienten. Diese geben an, wie steil die Funktion verläuft. Je flacher der Verlauf ist, desto mehr Iterationen wird der Optimierungsalgorithmus benötigen. Auch ist in diesem Fall anzunehmen, dass die gefundene Lösung ungenau sein wird, da sich der Zielfunktionswert trotz einer gewissen Entfernung zum Maximum nur noch gering ändert und der Algorithmus daher abbricht.

In großer Entfernung zu den wahren Parameterwerten verläuft die Likelihoodfunktion sehr steil. Nähern wir uns  $\beta^*$  und  $\varphi^*$ , so flacht der Verlauf ab. Für schlecht gewählte Startwerte verbessern sich die Schätzer durch die ersten Optimierungsschritte sehr stark. Je mehr sich die Schätzwerte  $\beta^*$  und  $\varphi^*$  nähern, desto schwieriger scheint es jedoch zu sein eine nennenswerte Genauigkeitssteigerung zu erreichen.

Maximum-Likelihood-Schätzer von  $\varphi$  und  $\beta$  erfüllen die in Abschnitt (3.3) definierten Maximum-Likelihood-Gleichungen. Um diese bilden zu können, benötigen wir die Scorefunktion:

### 5.3.2 Die Scorefunktion im verallgemeinerten Poisson-Regressionsmodell

**Satz 5.8 (Scorefunktion im verallgemeinerten Poisson-Regressionsmodell)** Für das Regressionsmodell mit der Spezifikation aus Abschnitt (5.1) lautet die Scorefunktion:

$$s(\varphi, \beta | \mathbf{y}) = \begin{pmatrix} s_0(\varphi, \beta | \mathbf{y}) \\ s_1(\varphi, \beta | \mathbf{y}) \\ \vdots \\ s_k(\varphi, \beta | \mathbf{y}) \end{pmatrix} \quad (5.20)$$

mit den Einträgen

$$s_0(\varphi, \beta | \mathbf{y}) = \sum_{i=1}^n \left( \frac{y_i(y_i - 1)}{e^{\mathbf{x}_i^t \beta} + (\varphi - 1)y_i} - \frac{y_i}{\varphi} + \frac{e^{\mathbf{x}_i^t \beta} - y_i}{\varphi^2} \right) \quad (5.21)$$

$$s_r(\varphi, \beta | \mathbf{y}) = \sum_{i=1}^n x_{ir} \left( 1 + \frac{e^{\mathbf{x}_i^t \beta} (y_i - 1)}{e^{\mathbf{x}_i^t \beta} + (\varphi - 1)y_i} - \frac{e^{\mathbf{x}_i^t \beta}}{\varphi} \right), \quad r = 1, \dots, k \quad (5.22)$$



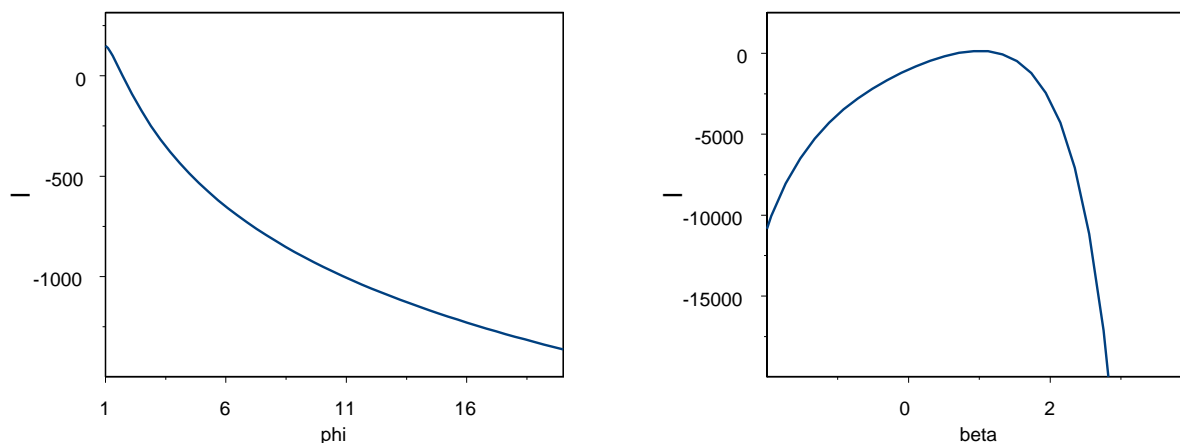


Abbildung 5.4: Schnitt bei  $\beta = 1$  (links) und  $\varphi = 1$  (rechts) durch die Log-Likelihoodfunktion des verallgemeinerten Poisson-Regressionsmodells mit den wahren Parametern  $\beta^* = 1$  und  $\varphi^* = 1$

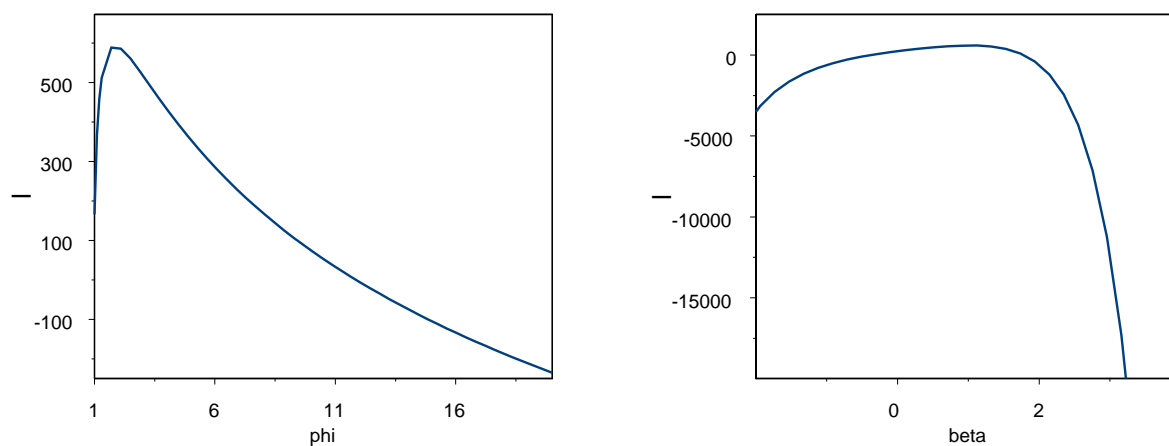


Abbildung 5.5: Schnitt bei  $\beta = 1$  (links) und  $\varphi = 2$  (rechts) durch die Log-Likelihoodfunktion des verallgemeinerten Poisson-Regressionsmodells mit den wahren Parametern  $\beta^* = 1$  und  $\varphi^* = 2$

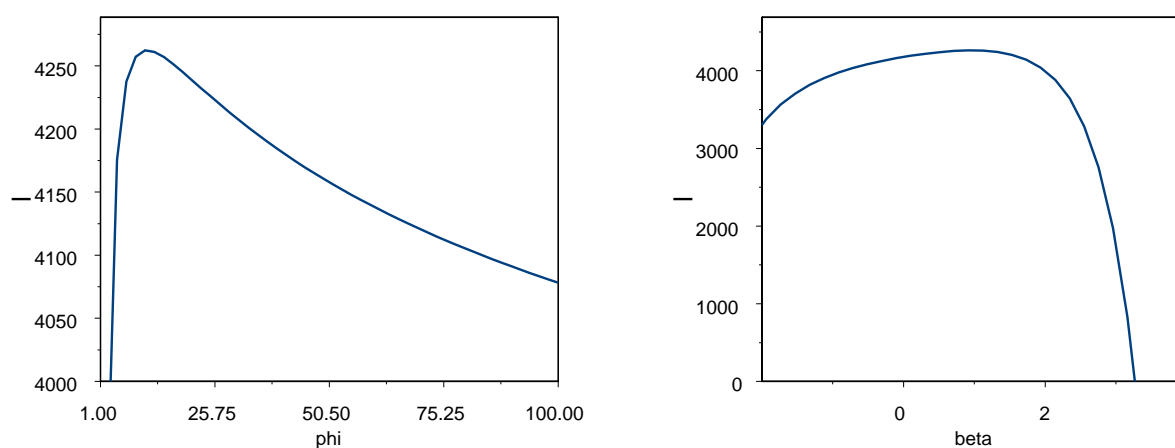


Abbildung 5.6: Schnitt bei  $\beta = 1$  (links) und  $\varphi = 10$  (rechts) durch die Log-Likelihoodfunktion des verallgemeinerten Poisson-Regressionsmodells mit den wahren Parametern  $\beta^* = 1$  und  $\varphi^* = 10$

*Beweis:*

Betrachten wir zunächst den linearen Prädiktor  $\eta_i = \mathbf{x}_i^t \boldsymbol{\beta}$ . Für diesen gilt:

$$\frac{\partial \eta_i}{\partial \varphi} = 0 \quad \text{und} \quad \frac{\partial \eta_i}{\partial \beta_r} = x_{ir} \quad (5.23)$$

Verwenden wir die Verknüpfung (5.2) so folgt für die Erwartungswerte  $\mu_i$ :

$$\frac{\partial \mu_i}{\partial \varphi} = \frac{\partial e^{\eta_i}}{\partial \varphi} = 0 \quad \text{und} \quad \frac{\partial \mu_i}{\partial \beta_r} = \frac{\partial e^{\eta_i}}{\partial \beta_r} = e^{\eta_i} \frac{\partial \eta_i}{\partial \beta_r} = \mu_i x_{ir} \quad (5.24)$$

Die Scorefunktion ist der Gradient der Log-Likelihoodfunktion. Für die Log-Likelihoodfunktion (5.15) gilt daher:

$$s(\varphi, \boldsymbol{\beta} | \mathbf{y}) = \left( \frac{\partial \mathcal{L}}{\partial \varphi}, \frac{\partial \mathcal{L}}{\partial \beta_1}, \dots, \frac{\partial \mathcal{L}}{\partial \beta_k} \right)^t$$

Bilden wir die partiellen Ableitungen in obigem Ausdruck:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \varphi} &\stackrel{(5.18)}{=} \frac{\partial}{\partial \varphi} \left( \sum_{i=1}^n \ln \mu_i + (y_i - 1) \ln [\mu_i + (\varphi - 1)y_i] - y_i \ln \varphi - \frac{1}{\varphi} [\mu_i + (\varphi - 1)y_i] - \ln(y_i!) \right) \\ &= \sum_{i=1}^n \frac{\partial \ln \mu_i}{\partial \varphi} + (y_i - 1) \frac{\partial [\mu_i + (\varphi - 1)y_i]}{\mu_i + (\varphi - 1)y_i} - \frac{y_i}{\varphi} - \frac{\partial \ln(y_i!)}{\partial \varphi} \\ &\quad - \left( -\frac{1}{\varphi^2} [\mu_i + (\varphi - 1)y_i] + \frac{1}{\varphi} \frac{\partial [\mu_i + (\varphi - 1)y_i]}{\partial \varphi} \right) \\ &\stackrel{(5.24)}{=} \sum_{i=1}^n \frac{y_i(y_i - 1)}{\mu_i + (\varphi - 1)y_i} - \frac{y_i}{\varphi} + \frac{\mu_i + (\varphi - 1)y_i}{\varphi^2} - \frac{y_i}{\varphi} \\ &= \sum_{i=1}^n \frac{y_i(y_i - 1)}{\mu_i + (\varphi - 1)y_i} - \frac{y_i}{\varphi} + \frac{\mu_i - y_i}{\varphi^2} \quad (5.25) \\ &\stackrel{(5.2)}{=} \sum_{i=1}^n \left( \frac{y_i(y_i - 1)}{e^{\mathbf{x}_i^t \boldsymbol{\beta}} + (\varphi - 1)y_i} - \frac{y_i}{\varphi} + \frac{e^{\mathbf{x}_i^t \boldsymbol{\beta}} - y_i}{\varphi^2} \right) \end{aligned}$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \beta_r} &\stackrel{(5.18)}{=} \frac{\partial}{\partial \beta_r} \left( \sum_{i=1}^n \ln \mu_i + (y_i - 1) \ln [\mu_i + (\varphi - 1)y_i] - y_i \ln \varphi - \frac{1}{\varphi} [\mu_i + (\varphi - 1)y_i] - \ln(y_i!) \right) \\ &= \sum_{i=1}^n \frac{1}{\mu_i} \frac{\partial \mu_i}{\partial \beta_r} + \frac{y_i - 1}{\mu_i + (\varphi - 1)y_i} \frac{\partial \mu_i}{\partial \beta_r} - \frac{1}{\varphi} \frac{\partial \mu_i}{\partial \beta_r} \\ &= \sum_{i=1}^n \left( \frac{1}{\mu_i} + \frac{y_i - 1}{\mu_i + (\varphi - 1)y_i} - \frac{1}{\varphi} \right) \frac{\partial \mu_i}{\partial \beta_r} \quad (5.26) \\ &\stackrel{(5.24)}{=} \sum_{i=1}^n x_{ir} \left( 1 + \frac{\mu_i(y_i - 1)}{\mu_i + (\varphi - 1)y_i} - \frac{\mu_i}{\varphi} \right) \\ &\stackrel{(5.2)}{=} \sum_{i=1}^n x_{ir} \left( 1 + \frac{e^{\mathbf{x}_i^t \boldsymbol{\beta}}(y_i - 1)}{e^{\mathbf{x}_i^t \boldsymbol{\beta}} + (\varphi - 1)y_i} - \frac{e^{\mathbf{x}_i^t \boldsymbol{\beta}}}{\varphi} \right), \quad r = 1, \dots, k \end{aligned}$$

□

Die Score-Gleichungen lauten also:

$$\sum_{i=1}^n \left( \frac{y_i(y_i - 1)}{e^{\mathbf{x}_i^t \boldsymbol{\beta}} + (\varphi - 1)y_i} - \frac{y_i}{\varphi} + \frac{e^{\mathbf{x}_i^t \boldsymbol{\beta}} - y_i}{\varphi^2} \right) = 0 \quad (5.27)$$

$$\sum_{i=1}^n x_{ir} \left( 1 + \frac{e^{\mathbf{x}_i^t \boldsymbol{\beta}}(y_i - 1)}{e^{\mathbf{x}_i^t \boldsymbol{\beta}} + (\varphi - 1)y_i} - \frac{e^{\mathbf{x}_i^t \boldsymbol{\beta}}}{\varphi} \right) = 0, \quad r = 1, \dots, k \quad (5.28)$$

**Satz 5.9 (Beobachtete Fisher-Information im verallgemeinerten Poissonmodell)** *Im betrachteten Regressionsmodell lautet die Hessematrix (beobachtete Fisher-Information)*

$$H(\varphi, \boldsymbol{\beta}) = I_{obs} = \begin{pmatrix} h_{00} & h_{01} & \dots & h_{0k} \\ h_{10} & h_{11} & \dots & h_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ h_{k0} & h_{k1} & \dots & h_{kk} \end{pmatrix} \quad (5.29)$$

mit den Einträgen

$$\begin{aligned} h_{00} &= \frac{\partial^2 \mathcal{L}}{\partial \varphi^2} = - \sum_{i=1}^n \frac{y_i^2(y_i - 1)}{[e^{\mathbf{x}_i^t \boldsymbol{\beta}} + (\varphi - 1)y_i]^2} + 2 \frac{e^{\mathbf{x}_i^t \boldsymbol{\beta}} - y_i}{\varphi^3} - \frac{y_i}{\varphi^2} \\ h_{0r} = h_{r0} &= \frac{\partial^2 \mathcal{L}}{\partial \beta_r \partial \varphi} = \sum_{i=1}^n x_{ir} e^{\mathbf{x}_i^t \boldsymbol{\beta}} \left( \frac{1}{\varphi^2} - \frac{y_i(y_i - 1)}{[e^{\mathbf{x}_i^t \boldsymbol{\beta}} + (\varphi - 1)y_i]^2} \right), \quad r = 1, \dots, k \\ h_{rs} &= \frac{\partial^2 \mathcal{L}}{\partial \beta_r \partial \beta_s} = - \sum_{i=1}^n x_{ir} x_{is} e^{\mathbf{x}_i^t \boldsymbol{\beta}} \left( \frac{1}{\varphi} - \frac{y_i(y_i - 1)(\varphi - 1)}{[e^{\mathbf{x}_i^t \boldsymbol{\beta}} + (\varphi - 1)y_i]^2} \right), \quad r, s = 1, \dots, k \end{aligned}$$

*Beweis:*

Betrachten wir wiederum zunächst die Ableitungen von  $\mu_i$ . Mit (5.24) gilt:

$$\frac{\partial^2 \mu_i}{\partial \beta_r \partial \varphi} = 0 \quad \text{und} \quad \frac{\partial^2 \mu_i}{\partial \beta_r \partial \beta_s} = \frac{\partial}{\partial \beta_s} \mu_i x_{ir} = \mu_i x_{ir} x_{is} \quad (5.30)$$

Die zweiten partiellen Ableitungen der Log-Likelihoodfunktion (5.15) lauten:

$$\begin{aligned} \frac{\partial^2 \mathcal{L}}{\partial \varphi^2} &\stackrel{(5.25)}{=} \frac{\partial}{\partial \varphi} \sum_{i=1}^n \frac{y_i(y_i - 1)}{\mu_i + (\varphi - 1)y_i} - \frac{y_i}{\varphi} + \frac{\mu_i - y_i}{\varphi^2} \\ &= - \sum_{i=1}^n \frac{y_i(y_i - 1)}{(\mu_i + (\varphi - 1)y_i)^2} \left( \frac{\partial \mu_i}{\partial \varphi} + y_i \right) - \frac{y_i}{\varphi^2} + \frac{2(\mu_i - y_i) - \frac{\partial \mu_i}{\partial \varphi}}{\varphi^3} \\ &\stackrel{(5.24)}{=} - \sum_{i=1}^n \frac{y_i^2(y_i - 1)}{[\mu_i + (\varphi - 1)y_i]^2} + 2 \frac{\mu_i - y_i}{\varphi^3} - \frac{y_i}{\varphi^2} \\ &\stackrel{(5.2)}{=} - \sum_{i=1}^n \frac{y_i^2(y_i - 1)}{[e^{\mathbf{x}_i^t \boldsymbol{\beta}} + (\varphi - 1)y_i]^2} + 2 \frac{e^{\mathbf{x}_i^t \boldsymbol{\beta}} - y_i}{\varphi^3} - \frac{y_i}{\varphi^2} \end{aligned} \quad (5.31)$$

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}}{\partial \beta_r \partial \varphi} &\stackrel{(5.26)}{=} \frac{\partial}{\partial \varphi} \sum_{i=1}^n \left( \frac{1}{\mu_i} + \frac{y_i - 1}{\mu_i + (\varphi - 1)y_i} - \frac{1}{\varphi} \right) \frac{\partial \mu_i}{\partial \beta_r} \\
&= \sum_{i=1}^n \left[ \left( \frac{\partial \frac{1}{\mu_i}}{\partial \varphi} - \frac{(y_i - 1)(\frac{\partial \mu_i}{\partial \varphi} + y_i)}{(\mu_i + (\varphi - 1)y_i)^2} + \frac{1}{\varphi^2} \right) \frac{\partial \mu_i}{\partial \beta_r} \right. \\
&\quad \left. + \left( \frac{1}{\mu_i} + \frac{y_i - 1}{\mu_i + (\varphi - 1)y_i} - \frac{1}{\varphi} \right) \frac{\partial^2 \mu_i}{\partial \beta_r \partial \varphi} \right] \\
&\stackrel{(5.30)}{=} \sum_{i=1}^n x_{ir} \mu_i \left( \frac{1}{\varphi^2} - \frac{y_i(y_i - 1)}{[\mu_i + (\varphi - 1)y_i]^2} \right), \\
&\stackrel{(5.2)}{=} \sum_{i=1}^n x_{ir} e^{\mathbf{x}_i^t \boldsymbol{\beta}} \left( \frac{1}{\varphi^2} - \frac{y_i(y_i - 1)}{[e^{\mathbf{x}_i^t \boldsymbol{\beta}} + (\varphi - 1)y_i]^2} \right), \quad r = 1, \dots, k
\end{aligned} \tag{5.32}$$

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}}{\partial \beta_r \partial \beta_s} &\stackrel{(5.26)}{=} \frac{\partial}{\partial \beta_s} \sum_{i=1}^n \left( \frac{1}{\mu_i} + \frac{y_i - 1}{\mu_i + (\varphi - 1)y_i} - \frac{1}{\varphi} \right) \frac{\partial \mu_i}{\partial \beta_r} \\
&= \sum_{i=1}^n \left[ \left( \frac{1}{\mu_i} + \frac{y_i - 1}{\mu_i + (\varphi - 1)y_i} - \frac{1}{\varphi} \right) \frac{\partial^2 \mu_i}{\partial \beta_r \partial \beta_s} \right. \\
&\quad \left. - \left( \frac{1}{\mu_i^2} + \frac{y_i - 1}{[\mu_i + (\varphi - 1)y_i]^2} \right) \frac{\partial \mu_i}{\partial \beta_s} \frac{\partial \mu_i}{\partial \beta_r} \right] \\
&\stackrel{(5.30)}{=} - \sum_{i=1}^n x_{ir} x_{is} \mu_i \left( \frac{1}{\varphi} - \frac{y_i(y_i - 1)(\varphi - 1)}{[\mu_i + (\varphi - 1)y_i]^2} \right) \\
&\stackrel{(5.2)}{=} - \sum_{i=1}^n x_{ir} x_{is} e^{\mathbf{x}_i^t \boldsymbol{\beta}} \left( \frac{1}{\varphi} - \frac{y_i(y_i - 1)(\varphi - 1)}{[e^{\mathbf{x}_i^t \boldsymbol{\beta}} + (\varphi - 1)y_i]^2} \right), \quad r, s = 1, \dots, k
\end{aligned} \tag{5.33}$$

□

### 5.3.3 Fisher-Information

Für  $Y \sim GP^*(\mu, \varphi)$  gilt:

$$\begin{aligned}
E \left( \frac{Y(Y-1)}{[\mu + (\varphi - 1)Y]^2} \right) &= \sum_{y=0}^{\infty} \frac{y(y-1)}{[\mu + (\varphi - 1)y]^2} \\
&\quad \frac{\mu[\mu + (\varphi - 1)y]^{y-1}}{y!} \varphi^{-y} e^{-\frac{1}{\varphi}[\mu + (\varphi - 1)y]}
\end{aligned} \tag{5.34}$$

Die Summanden für  $y = 0$  und  $y = 1$  sind wegen des Terms  $y(y - 1)$  gleich Null. Daher können wir die Summation bei  $y = 2$  starten:

$$\begin{aligned}
E\left(\frac{Y(Y-1)}{[\mu+(\varphi-1)Y]^2}\right) &= \sum_{y=2}^{\infty} \frac{y(y-1)}{[\mu+(\varphi-1)y]^2} \frac{\mu[\mu+(\varphi-1)y]^{y-1}}{y!} \varphi^{-y} e^{-\frac{1}{\varphi}[\mu+(\varphi-1)y]} \\
&= \sum_{y=2}^{\infty} \frac{\mu[\mu+(\varphi-1)y]^{y-3}}{(y-2)!} \varphi^{-y} e^{-\frac{1}{\varphi}[\mu+(\varphi-1)y]} \\
&= \frac{\mu}{\varphi^2} \sum_{y=2}^{\infty} \frac{[\mu+2(\varphi-1)+(\varphi-1)(y-2)]^{y-3}}{(y-2)!} \\
&\quad \varphi^{-(y-2)} e^{-\frac{1}{\varphi}[\mu+2(\varphi-1)+(\varphi-1)(y-2)]} \\
&\stackrel{(*)}{=} \frac{\mu}{\varphi^2} \sum_{\tilde{y}=0}^{\infty} \frac{(\tilde{\mu}+(\varphi-1)\tilde{y})^{\tilde{y}-1}}{\tilde{y}!} \varphi^{-\tilde{y}} e^{-\frac{1}{\varphi}[\tilde{\mu}+(\varphi-1)\tilde{y}]} \\
&= \frac{\mu}{\varphi^2 \tilde{\mu}} \sum_{\tilde{y}=0}^{\infty} \frac{\tilde{\mu}[\tilde{\mu}+(\varphi-1)\tilde{y}]^{\tilde{y}-1}}{\tilde{y}!} \varphi^{-\tilde{y}} e^{-\frac{1}{\varphi}[\tilde{\mu}+(\varphi-1)\tilde{y}]} \quad (5.35) \\
&\stackrel{(5.9)}{=} \frac{\mu}{\varphi^2 \tilde{\mu}} = \frac{\mu}{\varphi^2(\mu-2+2\varphi)} \quad (5.36)
\end{aligned}$$

(\*) Wir definieren  $\tilde{\mu} := \mu + 2(\varphi - 1)$  und  $\tilde{y} := y - 2$ .

Der Ausdruck  $y^2(y-1)$  ist ebenfalls für  $y=0$  und  $y=1$  gleich Null. Es gilt daher:

$$\begin{aligned}
E\left(\frac{Y^2(Y-1)}{[\mu+(\varphi-1)Y]^2}\right) &= \sum_{y=2}^{\infty} \frac{y^2(y-1)}{[\mu+(\varphi-1)y]^2} \frac{\mu[\mu+(\varphi-1)y]^{y-1}}{y!} \varphi^{-y} e^{-\frac{1}{\varphi}[\mu+(\varphi-1)y]} \\
&\stackrel{(5.35)}{=} \frac{\mu}{\varphi^2 \tilde{\mu}} \sum_{\tilde{y}=0}^{\infty} y \frac{\tilde{\mu}[\tilde{\mu}+(\varphi-1)\tilde{y}]^{\tilde{y}-1}}{\tilde{y}!} \varphi^{-\tilde{y}} e^{-\frac{1}{\varphi}[\tilde{\mu}+(\varphi-1)\tilde{y}]} \\
&= \frac{\mu}{\varphi^2 \tilde{\mu}} \sum_{\tilde{y}=0}^{\infty} (\tilde{y}+2) \frac{\tilde{\mu}[\tilde{\mu}+(\varphi-1)\tilde{y}]^{\tilde{y}-1}}{\tilde{y}!} \varphi^{-\tilde{y}} e^{-\frac{1}{\varphi}[\tilde{\mu}+(\varphi-1)\tilde{y}]} \\
&= \frac{\mu}{\varphi^2 \tilde{\mu}} (\tilde{\mu}+2) = \frac{\mu(\mu-2+2\varphi+2)}{\varphi^2(\mu-2+2\varphi)} = \frac{\mu(\mu+2\varphi)}{\varphi^2(\mu-2+2\varphi)} \quad (5.37)
\end{aligned}$$

Die Fisher-Information lautet:

$$I(\boldsymbol{\beta}, \varphi) = -E \begin{pmatrix} \frac{\partial^2 \mathcal{L}}{\partial \varphi^2} & \frac{\partial^2 \mathcal{L}}{\partial \varphi \partial \boldsymbol{\beta}^t} \\ \frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\beta} \partial \varphi} & \frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^t} \end{pmatrix}$$

mit den Einträgen:

$$\begin{aligned}
-E\left(\frac{\partial^2 \mathcal{L}}{\partial \varphi^2}\right) &\stackrel{(5.31)}{=} E\left(\sum_{i=1}^n \frac{y_i^2(y_i-1)}{[\mu_i+(\varphi-1)y_i]^2} + 2\frac{\mu_i-y_i}{\varphi^3} - \frac{y_i}{\varphi^2}\right) \\
&= \sum_{i=1}^n \left[ E\left(\frac{y_i^2(y_i-1)}{[\mu_i+(\varphi-1)y_i]^2}\right) + 2\frac{\mu_i-E(y_i)}{\varphi^3} - \frac{E(y_i)}{\varphi^2} \right] \\
&\stackrel{(5.37)}{=} \sum_{i=1}^n \left( \frac{\mu_i(\mu_i+2\varphi)}{\varphi^2(\mu_i-2+2\varphi)} + 2\frac{\mu_i-\mu_i}{\varphi^3} - \frac{\mu_i}{\varphi^2} \right) \\
&= \sum_{i=1}^n \frac{2\mu_i}{\varphi^2(\mu_i-2+2\varphi)} \quad (5.38)
\end{aligned}$$

$$\begin{aligned}
-E \left( \frac{\partial^2 \mathcal{L}}{\partial \beta \partial \varphi} \right) &\stackrel{(5.32)}{=} -E \left( \sum_{i=1}^n \mu_i \left( \frac{1}{\varphi^2} - \frac{y_i(y_i - 1)}{[\mu_i + (\varphi - 1)y_i]^2} \right) \mathbf{x}_i \right) \\
&= - \sum_{i=1}^n \mu_i \left( \frac{1}{\varphi^2} - E \left( \frac{y_i(y_i - 1)}{[\mu_i + (\varphi - 1)y_i]^2} \right) \right) \mathbf{x}_i \\
&\stackrel{(5.36)}{=} - \sum_{i=1}^n \mu_i \left( \frac{1}{\varphi^2} - \frac{\mu_i}{\varphi^2(\mu_i - 2 + 2\varphi)} \right) \mathbf{x}_i \\
&= - \sum_{i=1}^n \frac{2\mu_i(\varphi - 1)}{\varphi^2(\mu_i - 2 + 2\varphi)} \mathbf{x}_i \tag{5.39}
\end{aligned}$$

$$\begin{aligned}
-E \left( \frac{\partial^2 \mathcal{L}}{\partial \beta \partial \beta^t} \right) &\stackrel{(5.33)}{=} E \left( \sum_{i=1}^n \mu_i \left( \frac{1}{\varphi} - \frac{y_i(y_i - 1)(\varphi - 1)}{[\mu_i + (\varphi - 1)y_i]^2} \right) \mathbf{x}_i \mathbf{x}_i^t \right) \\
&= \sum_{i=1}^n \left( \frac{\mu_i}{\varphi} - \mu_i(\varphi - 1) E \left( \frac{y_i(y_i - 1)}{[\mu_i + (\varphi - 1)y_i]^2} \right) \right) \mathbf{x}_i \mathbf{x}_i^t \\
&\stackrel{(5.36)}{=} \sum_{i=1}^n \left( \frac{\mu_i}{\varphi} - \mu_i(\varphi - 1) \left( \frac{\mu_i}{\varphi^2[\mu_i - 2 + 2\varphi]} \right) \right) \mathbf{x}_i \mathbf{x}_i^t \\
&= \sum_{i=1}^n \frac{\mu_i(\mu_i - 2\varphi + 2\varphi^2)}{\varphi^2(\mu_i - 2 + 2\varphi)} \mathbf{x}_i \mathbf{x}_i^t \tag{5.40}
\end{aligned}$$

## 5.4 Offset in der verallgemeinerten Poissonregression

Wir wollen im Folgenden das verallgemeinerte Poisson-Regressionsmodell, wie zuvor schon das Poisson-Regressionsmodell, für den Fall von Beobachtungszeiträumen unterschiedlicher Länge erweitern. Wie in Abschnitt 4.2.1 bezeichne  $t_i^*$  die Länge des mit  $Y_i$  korrespondierenden Beobachtungszeitraumes. Die Intensitäten  $\theta_i$  der verallgemeinerten Poissonverteilung werden daher zur Normierung der Beobachtungszeiten mit  $t_i = (t_i^*)^{-1}$  multipliziert:

$$\tilde{\theta}_i := t_i \theta_i \tag{5.41}$$

Für ein  $\lambda \in (0, 1)$  sind die Zielgrößen  $Y_i | \mathbf{x}_i$  im erweiterten Ansatz  $GP(\tilde{\theta}_i, \lambda)$ -verteilt:

$$P(Y_i = y_i | \mathbf{x}_i) \stackrel{(1.1)}{=} \frac{\tilde{\theta}_i (\tilde{\theta}_i + y_i \lambda)^{y_i - 1}}{y_i!} e^{-(\tilde{\theta}_i + y_i \lambda)} \tag{5.42}$$

Setzen wir

$$\begin{aligned}
\tilde{\mu}_i &= \frac{\tilde{\theta}_i}{1 - \lambda} = t_i \frac{\theta_i}{1 - \lambda} = t_i \mu_i \quad \text{und} \\
\varphi &= \frac{1}{1 - \lambda}
\end{aligned} \tag{5.43}$$

lässt sich (5.42) auf die Form der Wahrscheinlichkeitsfunktion (5.9) bringen:

$$P(Y_i = y_i | \mathbf{x}_i) = \frac{\tilde{\mu}_i [\tilde{\mu}_i + (\varphi - 1)y_i]^{y_i - 1}}{y_i!} \varphi^{-y_i} \exp \left\{ -\frac{1}{\varphi} [\tilde{\mu}_i + (\varphi - 1)y_i] \right\} \tag{5.44}$$

(vgl. Abschnitt 5.4). Für die bedingte Zielgröße gilt daher:  $Y_i|\mathbf{x}_i \sim GP^*(\tilde{\mu}_i, \varphi)$ . Dem entsprechend lautet der Erwartungswert von  $Y_i|\mathbf{x}_i$ :

$$E(Y_i|\mathbf{x}_i) = \tilde{\mu}_i \stackrel{(5.43)}{=} t_i \mu_i \quad (5.45)$$

Verwenden wir den linearen Prädiktor

$$\eta_i = \mathbf{x}_i^t \boldsymbol{\beta} \quad (5.46)$$

und die Verknüpfung

$$\eta_i = \ln \mu_i \quad (5.47)$$

so erhalten wir die Regressionsfunktionen

$$\begin{aligned} E(Y_i|\mathbf{x}_i) &= \tilde{\mu}_i = t_i \mu_i = t_i e^{\eta_i} \\ &= t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} \\ &= \exp\{\mathbf{x}_i^t \boldsymbol{\beta} + \ln t_i\} \end{aligned} \quad (5.48)$$

$$\begin{aligned} Var(Y_i|\mathbf{x}_i) &= \varphi^2 E(Y_i|\mathbf{x}_i) \\ &= \varphi^2 \exp\{\mathbf{x}_i^t \boldsymbol{\beta} + \ln t_i\} \end{aligned} \quad (5.49)$$

Folglich gilt:

$$\eta_i = \ln \tilde{\mu}_i - \ln t_i \quad (5.50)$$

**Satz 5.10 (Log-Likelihoodfunktion für abweichende Beobachtungszeiträume)** *Im verallgemeinerten Poisson-Regressionsmodell gilt für die Log-Likelihoodfunktion unter Berücksichtigung abweichender Längen der Beobachtungszeiträume:*

$$\begin{aligned} \mathcal{L}(\varphi, \boldsymbol{\beta}) &\propto \sum_{i=1}^n \mathbf{x}_i^t \boldsymbol{\beta} + (y_i - 1) \ln [t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} + (\varphi - 1)y_i] \\ &\quad - y_i \ln \varphi - \frac{1}{\varphi} [t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} + (\varphi - 1)y_i] \end{aligned} \quad (5.51)$$

*Beweis:*

Für die Likelihoodfunktion erhalten wir mit (5.44) und der Unabhängigkeit der Responseedaten:

$$\begin{aligned} L(\varphi, \boldsymbol{\beta}) &= \prod_{i=1}^n P(Y_i = y_i|\mathbf{x}_i) \\ &\stackrel{(5.44)}{=} \prod_{i=1}^n \frac{\tilde{\mu}_i [\tilde{\mu}_i + (\varphi - 1)y_i]^{y_i - 1}}{y_i!} \varphi^{-y_i} \exp\left\{-\frac{1}{\varphi} [\tilde{\mu}_i + (\varphi - 1)y_i]\right\} \end{aligned} \quad (5.52)$$

Logarithmieren ergibt die Log-Likelihoodfunktion

$$\begin{aligned}
\mathcal{L}(\varphi, \boldsymbol{\beta}) &= \sum_{i=1}^n \ln \tilde{\mu}_i + (y_i - 1) \ln[\tilde{\mu}_i + (\varphi - 1)y_i] \\
&\quad - \ln(y_i!) - y_i \ln \varphi - \frac{1}{\varphi} [\tilde{\mu}_i + (\varphi - 1)y_i] \\
&\stackrel{(5.43)}{=} \sum_{i=1}^n \ln t_i + \ln \mu_i + (y_i - 1) \ln[t_i \mu_i (\varphi - 1)y_i] \\
&\quad - \ln(y_i!) - y_i \ln \varphi - \frac{1}{\varphi} [t_i \mu_i + (\varphi - 1)y_i] \\
&\stackrel{(5.46)}{=} \sum_{i=1}^n \ln t_i + \mathbf{x}_i^t \boldsymbol{\beta} + (y_i - 1) \ln[t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} (\varphi - 1)y_i] \\
&\quad - \ln(y_i!) - y_i \ln \varphi - \frac{1}{\varphi} [t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} + (\varphi - 1)y_i] \tag{5.53}
\end{aligned}$$

□

## 5.5 Schätzprozeduren

### 5.5.1 Dispersionsparameter als Störparameter

Der Vorteil der verallgemeinerten Poissonregression gegenüber der klassischen Poissonregression besteht darin, dass durch die veränderte Verteilungsannahme Überdispersion in das Modell integriert werden kann. Der Dispersionsparameter  $\varphi$  ist bei der verallgemeinerten Poissonregression ein Modellparameter, der gemeinsam mit den Regressionsparametern durch Maximierung der Likelihoodfunktion (5.16) geschätzt werden kann.

Wie in Abschnitt 4.4 erwähnt, kann  $\varphi$  aber auch als Störparameter aufgefasst werden. In diesem Fall dient  $\varphi$  als Justierungsgröße, um Erwartungswert- und Varianzfunktion nach der Schätzung der Regressionsparameter  $\boldsymbol{\beta}$  auf die Form (4.38) zu bringen. Es findet keine simultane Schätzung von Dispersions- und Regressionsparametern statt.

Die Regressionsparameter werden wie üblich mit der Maximum-Likelihood-Methode geschätzt. Da  $\varphi$  als fest angesehen wird und lediglich der Justierung der sich aus den Maximum-Likelihood-Schätzern ergebenden Regressionsfunktionen dient, wird vor der Maximierung in der Likelihoodfunktion  $\varphi = 1$  gesetzt. Wir erhalten die Likelihoodfunktion des klassischen Poisson-Regressionsmodells.

### 5.5.2 Momentenschätzer für $\varphi$

Sind Schätzer der Regressionsparameter bekannt, so lässt sich auch  $\varphi$  schätzen. Breslow (1984) schlägt vor, den Dispersionsparameter über die Gleichsetzung der aus dem Modell abgeleiteten  $\chi^2$ -Statistik mit ihren Freiheitsgraden zu schätzen.

$$\chi_{n-k}^2 = n - k \tag{5.54}$$

Lösen wir diese Gleichung nach dem Dispersionsparameter auf, erhalten wir einen Schätzer für diesen Parameter.

Im verallgemeinerten Poisson-Regressionsmodell gilt:

$$\begin{aligned}
E(y_i | \mathbf{x}_i) &= \mu_i \\
Var(y_i | \mathbf{x}_i) &= \varphi^2 \mu_i
\end{aligned}$$



Mit den standardisierten Größen

$$\frac{y_i - E(y_i)}{\sqrt{\text{Var}(y_i)}} = \frac{y_i - \mu_i}{\sqrt{\varphi^2 \mu_i}}, \quad i = 1, \dots, n \quad (5.55)$$

folgt die  $\chi^2$ -Statistik

$$\chi_{n-k}^2 = \sum_{i=1}^n \left( \frac{y_i - \mu_i}{\varphi \sqrt{\mu_i}} \right)^2 \quad (5.56)$$

Die Freiheitsgrade entsprechen der Anzahl der Beobachtungen  $n$  abzüglich der im Modell verwendeten Regressionsparameter  $k$ . Gemäß dem beschriebenen Ansatz erhalten wir also für die verallgemeinerte Poissonregression:

$$\sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\varphi^2 \mu_i} = n - k \quad (5.57)$$

Lösen wir (5.57) nach  $\varphi$  auf, so erhalten wir einen Ausdruck in Abhängigkeit der unbekanntenen  $\mu_i$ . Setzen wir für die  $\mu_i$  deren Schätzer  $\hat{\mu}_i = \exp\{\mathbf{x}_i^t \hat{\boldsymbol{\beta}}\}$  ein, erhalten wir den Momentenschätzer

$$\hat{\varphi} = \sqrt{\frac{1}{n-k} \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}} \quad (5.58)$$

### 5.5.3 Zweistufige Schätzprozedur

Im Gegensatz zur Likelihoodfunktion der Poissonregression tritt in der Likelihoodfunktion der verallgemeinerten Poissonregression der Parameter  $\varphi$  auf. Somit ist es möglich, die Maximum-Likelihood-Schätzer  $\hat{\boldsymbol{\beta}}$  zu verbessern, nachdem der Momentenschätzer  $\hat{\varphi}$  gebildet wurde. Consul and Famoye (1992) beschreiben dazu folgende Vorgehensweise:

Gehen wir zunächst von Poissonregression aus ( $\hat{\varphi}_{(0)} = 1$ ), so erhalten wir durch Maximieren der Likelihoodfunktion den Maximum-Likelihood-Schätzer  $\hat{\boldsymbol{\beta}}_{(0)}$ . Haben wir  $\hat{\boldsymbol{\beta}}_{(0)}$  bestimmt so können wir wie im vorangegangenen Abschnitt beschrieben den Momentenschätzer  $\hat{\varphi}_{(1)}$  berechnen. Mit diesem wiederum ist es möglich durch erneute Maximierung der Likelihoodfunktion einen verbesserten Schätzer  $\hat{\boldsymbol{\beta}}_{(1)}$  für die Regressionsparameter zu erhalten. Diese zweistufige Verbesserung der Schätzer ( $\hat{\varphi}_{(i)}, \hat{\boldsymbol{\beta}}_{(i)}$ ) wird solange fortgesetzt, bis die Paare ( $\hat{\varphi}_{(i)}, \hat{\boldsymbol{\beta}}_{(i)}$ ) konvergieren.

#### Algorithmus:

- Eingabe:
  - (i) Responsevektor  $\mathbf{Y}$
  - (ii) Designmatrix  $\mathbf{X}$
  - (iii) Toleranzgrenze  $\tau$
- Initialisierung:
  - (i)  $\hat{\varphi}_{(0)} = 1$
  - (ii) Bestimme Maximum-Likelihood-Schätzer  $\hat{\boldsymbol{\beta}}_{(0)}$  für  $\hat{\varphi}_{(0)}$

- (iii)  $i = 1$
- Iteration:
  - (i) Bestimme Momentenschätzer  $\hat{\varphi}_{(i)}$
  - (ii) Bestimme Maximum-Likelihood-Schätzer  $\hat{\boldsymbol{\beta}}_{(i)}$  für  $\hat{\varphi}_{(i)}$
  - (iii) Stop, falls  $\|(\hat{\varphi}_{(i)}, \hat{\boldsymbol{\beta}}_{(i)}) - (\hat{\varphi}_{(i-1)}, \hat{\boldsymbol{\beta}}_{(i-1)})\| < \tau$
  - (iv)  $i = i + 1$
- Ausgabe:
  - $(\hat{\varphi}_{(i)}, \hat{\boldsymbol{\beta}}_{(i)})$

Dieser Ansatz hat zudem den Vorteil, dass wir erkennen können, ob die Verwendung des verallgemeinerten Poissonmodells überhaupt erforderlich ist oder nicht auch die Schätzung mittels Poissonregression zufriedenstellend ist. Gilt  $\hat{\varphi}_{(1)} \simeq 1$ , so können wir folgern, dass Überdispersion nicht vorliegt. In diesem Fall lohnt sich der Übergang vom Poissonmodell zum verallgemeinerten Poissonmodell nicht. Weicht  $\hat{\varphi}_{(1)}$  jedoch signifikant von Eins ab, so führen wir die eben beschriebene zweistufige Schätzprozedur aus.

#### 5.5.4 Simultane Schätzung von $\varphi$ und $\boldsymbol{\beta}$

Der Dispersionsparameter  $\varphi$  wird bei der verallgemeinerten Poissonregression als Modellparameter betrachtet. Die zugehörige Likelihoodfunktion (5.16) ermöglicht eine simultane Schätzung von  $\varphi$  und  $\boldsymbol{\beta}$  mit Hilfe der Maximum-Likelihood-Methode.

Soll die Likelihoodfunktion numerisch maximiert werden, so werden Startwerte  $\varphi_{(0)}$  und  $\boldsymbol{\beta}_{(0)}$  benötigt. Um solche zu erhalten gehen wir wie oben zunächst von gewöhnlicher Poissonregression aus. Somit gilt wiederum  $\varphi_{(0)} = 1$ . Die Werte  $\boldsymbol{\beta}_{(0)}$  erhalten wir über die vorhandene Schätzroutine `glm` mit der Option `family=poisson`. Ausgehend von  $(1, \boldsymbol{\beta}_{(0)})$  wird die Likelihoodfunktion (5.16) über  $(\varphi, \boldsymbol{\beta})$  maximiert.

Um den Maximierungsalgorithmus zu beschleunigen wird in der Schätzroutine `mle.gpois` als Startwert  $\varphi_{(0)}$  nicht Eins sondern der Momentenschätzer von  $\varphi$  für  $\boldsymbol{\beta}_{(0)}$  verwendet (vgl. Anhang C.2.2).

#### Algorithmus:

- Eingabe:
  - (i) Responsevektor  $\mathbf{Y}$
  - (ii) Designmatrix  $\mathbf{X}$
- Initialisierung:
  - (i) Bestimme Maximum-Likelihood-Schätzer  $\hat{\boldsymbol{\beta}}_{(0)}$  für die klassische Poissonregression
  - (ii) Bestimme mit  $\hat{\boldsymbol{\beta}}_{(0)}$  den Momentenschätzer  $\hat{\varphi}_{(0)}$
- Maximierung:
  - Maximiere ausgehend von dem Startvektor  $(\hat{\varphi}_{(0)}, \hat{\boldsymbol{\beta}}_{(0)})$  die Log-Likelihoodfunktion der verallgemeinerten Poissonregression.
- Ausgabe:
  - Maximum-Likelihood-Schätzer  $(\hat{\varphi}, \hat{\boldsymbol{\beta}})$

## 5.6 Messung der Anpassungsqualität

### 5.6.1 Unskalierte Devianzresiduen

Betrachten wir  $\varphi$  als Störparameter, so können die Devianzresiduen zur Messung der Anpassungsqualität verwendet werden.

**Satz 5.11 (Unskalierte Devianzresiduen in der verallgemeinerten Poissonregression)**

Liegt verallgemeinerte Poissonregression vor, so lauten die Devianzresiduen:

$$r_i^D = \text{sign}(y_i - \hat{\mu}_i) \sqrt{2 \left[ -\ln \hat{\mu}_i - (y_i - 1) \ln \left( \frac{\hat{\mu}_i + (\varphi - 1)y_i}{\varphi} \right) + y_i \ln y_i + \frac{\hat{\mu}_i - y_i}{\varphi} \right]} \quad (5.59)$$

mit  $\hat{\mu}_i = \exp\{\mathbf{x}_i^t \hat{\boldsymbol{\beta}}\}$  und den Maximum-Likelihood-Schätzern  $\hat{\boldsymbol{\beta}}$ .

*Beweis:*

Bestimmen wir zunächst die Likelihoodfunktionen  $L(\boldsymbol{\mu}, \mathbf{y})$  und  $L(\mathbf{y}, \mathbf{y})$ :

$$\begin{aligned} L(\boldsymbol{\mu}, \mathbf{y}) &= L(\varphi, \boldsymbol{\beta}) \\ &\stackrel{(5.16)}{=} \prod_{i=1}^n \frac{\mu_i [\mu_i + (\varphi - 1)y_i]^{y_i - 1}}{y_i!} \varphi^{-y_i} \exp \left\{ -\frac{\mu_i + (\varphi - 1)y_i}{\varphi} \right\} \\ L(\mathbf{y}, \mathbf{y}) &\stackrel{(5.16)}{=} \prod_{i=1}^n \frac{y_i [y_i + (\varphi - 1)y_i]^{y_i - 1}}{y_i!} \varphi^{-y_i} \exp \left\{ -\frac{y_i + (\varphi - 1)y_i}{\varphi} \right\} \\ &= \prod_{i=1}^n \frac{y_i^{y_i}}{y_i!} \varphi^{y_i - 1} \varphi^{-y_i} e^{-y_i} \\ &= \prod_{i=1}^n \frac{y_i^{y_i}}{y_i!} \varphi^{-1} e^{-y_i} \end{aligned} \quad (5.60)$$

Unter Verwendung der Definition 3.11 erhalten wir die zugehörige Devianzfunktion:

$$\begin{aligned} D(\mathbf{y}, \hat{\boldsymbol{\mu}}) &\stackrel{(5.60), (5.61)}{=} -2 \ln \prod_{i=1}^n \frac{\hat{\mu}_i [\hat{\mu}_i + (\varphi - 1)y_i]^{y_i - 1} \varphi^{-y_i} \exp \left\{ -\frac{\hat{\mu}_i + (\varphi - 1)y_i}{\varphi} \right\}}{y_i^{y_i} \varphi^{-1} \exp \{-y_i\}} \\ &= -2 \ln \prod_{i=1}^n \hat{\mu}_i [\hat{\mu}_i + (\varphi - 1)y_i]^{y_i - 1} y_i^{-y_i} \varphi^{1 - y_i} \exp \left\{ -\frac{\hat{\mu}_i + \varphi y_i - y_i}{\varphi} + \frac{y_i \varphi}{\varphi} \right\} \\ &= -2 \sum_{i=1}^n \left[ \ln \hat{\mu}_i + (y_i - 1) \ln [\hat{\mu}_i + (\varphi - 1)y_i] \right. \\ &\quad \left. - y_i \ln y_i + (1 - y_i) \ln \varphi - \frac{\hat{\mu}_i - y_i}{\varphi} \right] \\ &= \sum_{i=1}^n 2 \left[ -\ln \hat{\mu}_i - (y_i - 1) \ln \left( \frac{\hat{\mu}_i + (\varphi - 1)y_i}{\varphi} \right) + y_i \ln y_i + \frac{\hat{\mu}_i - y_i}{\varphi} \right] \end{aligned} \quad (5.62)$$

Bezeichnen wir die Summanden in (5.62) mit  $d_i$  so gilt:

$$D(\mathbf{y}, \hat{\boldsymbol{\mu}}) = \sum_{i=1}^n d_i \quad (5.63)$$

Setzen wir schließlich den Ausdruck für  $d_i$  in die Definition 3.12 ein, erhalten wir die Devianzresiduen der verallgemeinerten Poissonregression.

□

**Bemerkung:**

Bei der Verwendung der Devianzresiduen ist zu beachten, dass allein eine Erhöhung des Dispersionsparameters  $\varphi$  ohne die Verbesserung der angepassten Werte  $\hat{\boldsymbol{\mu}}$  zu einer Reduzierung des Devianzmaßes führen kann.

**5.6.2 Akaike-Informationskriterium**

Zur Variablenselektion sowie zum Vergleich mit Modellen, in welche das verallgemeinerte Poissonmodell eingebettet ist bzw. zum Vergleich mit Modellen die in das verallgemeinerte Poissonmodell eingebettet sind, können wir das Akaike-Informationskriterium verwenden. So ist es mit AIC möglich, die Anpassungsqualität eines verallgemeinerten Poissonmodells mit der Anpassungsqualität des entsprechenden klassischen Poissonmodells zu vergleichen.

**Satz 5.12 (Akaike-Information für das verallg. Poisson-Regressionsmodell)** *Die Akaike-Information des in Abschnitt 5.1 definierten Regressionsmodells lautet:*

$$\begin{aligned}
 AIC &= 2k - 2 \sum_{i=1}^n \mathbf{x}_i^t \boldsymbol{\beta} + (y_i - 1) \ln \left[ e^{\mathbf{x}_i^t \boldsymbol{\beta}} + (\varphi - 1)y_i \right] \\
 &\quad - y_i \ln \varphi - \frac{1}{\varphi} \left[ e^{\mathbf{x}_i^t \boldsymbol{\beta}} + (\varphi - 1)y_i \right] - \ln(y_i!) \quad (5.64)
 \end{aligned}$$

*Beweis:*

Einsetzen von (5.17) und (5.2) in (3.37)

□

# Kapitel 6

## Zero-Inflated-Regressionsmodelle

### 6.1 Zero-Inflated-Verteilungen

Mit Hilfe von ZI-Verteilungen<sup>1</sup> lässt sich Nullenüberschuss bei Zähldaten modellieren. Einer gewöhnlichen Zählvariable  $X^*$  wird dabei eine binäre Variable  $\Omega \in \{0, 1\}$  vorgeschaltet. Die resultierende Zählvariable  $X$  kann als Produkt dieser (unabhängigen) Größen betrachtet werden. Beobachtungen von  $X$  entstehen dem entsprechend durch die Verbindung zweier unabhängiger Prozesse:

(i) **Ursprungsprozess** zur Erzeugung der Zähldaten  $X^* = x^*$

(ii) **Auswahlprozess**<sup>2</sup> durch die **Auswahlvariable**  $\Omega$

In ZI-Modellen gibt es daher zwei Arten von Nullen. Ein Ereignis Null entstammt entweder dem Auswahlprozess ( $\Omega = 0$ ) oder dem Ursprungsprozess ( $\Omega = 1, X^* = 0$ ). Auf Grund dieser klassifizierenden Eigenschaft für  $X = 0$  wird  $\Omega$  auch als Klassifizierungsvariable bezeichnet.

**Definition 6.1 (ZI-Zufallsvariable)** Sei  $X^* \in \mathbb{N}$  eine Zufallsvariable. Weiterhin seien  $X^*$  und die Größe  $\Omega \in \{0, 1\}$  unabhängig. Die ZI-Zufallsvariable  $X \in \mathbb{N}$  wird definiert als:

$$X := \Omega X^* = \begin{cases} 0 & \Omega = 0 \\ X^* & \Omega = 1 \end{cases} \quad (6.1)$$

Ist der Zustand von  $\Omega$  bekannt, so gilt für die bedingte Wahrscheinlichkeit von  $X$ :

$$\begin{aligned} P(X = x|\Omega) &= \begin{cases} P(X^* = x) & \Omega = 1, \forall x \\ 0 & \Omega = 0, x > 0 \\ 1 & \Omega = 0, x = 0 \end{cases} \\ &= P(X^* = x)[1 - \delta(\Omega)] + \delta(x)\delta(\Omega) \end{aligned} \quad (6.2)$$

Dabei bezeichnet  $\delta$  die Indikatorfunktion

$$\delta(x) = \begin{cases} 1 & x = 0 \\ 0 & \text{sonst} \end{cases} \quad (6.3)$$

---

<sup>1</sup>ZI-Verteilungen (zero inflated distributions) werden gelegentlich auch als Verteilungen mit Nullen bezeichnet.

<sup>2</sup>engl.: selection process

**Satz 6.2 (Wahrscheinlichkeitsfunktion der ZI-Verteilung)** Ist  $\omega \in [0, 1]$  die Wahrscheinlichkeit für  $\Omega = 0$  und  $P^*$  die Wahrscheinlichkeitsfunktion der Verteilung von  $X^*$ , so gilt für die Wahrscheinlichkeitsfunktion der Verteilung von  $X$ :

$$\begin{aligned} P_x = P(X = x) &= \begin{cases} \omega + (1 - \omega)P_x^* & x = 0 \\ (1 - \omega)P_x^* & x = 1, 2, \dots \end{cases} \\ &= \omega \delta(x) + (1 - \omega) P_x^* \quad \forall x \in \mathbb{N} \end{aligned} \quad (6.4)$$

wobei  $\delta$  die Indikatorfunktion (6.3) bezeichnet.

*Beweis:*

Mit  $P(\Omega = 0) = \omega$  gilt:

$$\begin{aligned} P(X = x) &\stackrel{(6.2)}{=} P(\Omega = 0)\{P(X^* = x)[1 - \delta(0)] + \delta(x)\delta(0)\} \\ &+ P(\Omega \neq 0)\{P(X^* = x)[1 - \delta(1)] + \delta(x)\delta(1)\} \\ &= P(\Omega = 0)\{P(X^* = x)[1 - 1] + \delta(x)1\} \\ &+ P(\Omega \neq 0)\{P(X^* = x)[1 - 0] + \delta(x)0\} \\ &= P(\Omega = 0)\delta(x) + P(\Omega \neq 0)P(X^* = x) \\ &= \omega\delta(x) + (1 - \omega)P_x^* \end{aligned}$$

□

**Bemerkung:**

Wir bezeichnen  $\omega$  in diesem Zusammenhang als **Auswahlparameter**. Gibt  $\omega$  wie im vorangehenden Satz die Eintrittswahrscheinlichkeit von  $\Omega = 0$  an, so kann die Auswahlvariable  $\Omega$  als bernoulliverteilt mit Erfolgswahrscheinlichkeit  $1 - \omega$  angesehen werden:

$$\Omega \sim \text{Bernoulli}(1 - \omega) \quad (6.5)$$

Gupta et al. (1996) interpretieren die Größe  $\omega$  in (6.4) nicht als Eintrittswahrscheinlichkeit, sondern als Justierungsfaktor. Falls  $0 < \omega < 1$  gilt, enthält das Modell zusätzliche Nullen, gilt  $\omega < 0$  sind weniger Nullen vorhanden. Wird dieser Ansatz verwendet, so spricht man von einer **ZA-Verteilung**<sup>3</sup>. Mit ZA-Verteilungen lassen sich auch Zähldaten modellieren, die weniger Nullen aufweisen, als durch das Ursprungsmodell zu erwarten ist. ZI-Verteilungsmodelle sind daher ein Spezialfall von ZA-Modellen.

Um  $P(X = 0) \geq 0$  sicherzustellen, muss  $\omega$  nach unten beschränkt werden:

$$\begin{aligned} P(X = 0) = \omega + (1 - \omega)P(X = 0) &\geq 0 \\ \Leftrightarrow \omega - \omega P(X = 0) &\geq -P(X = 0) \\ \Leftrightarrow \omega &\geq -\frac{P(X = 0)}{1 - P(X = 0)} \end{aligned}$$

In Satz 6.4 werden einige Eigenschaften von ZA-Modellen gezeigt werden, welche für  $0 < \omega < 1$  auch für ZI-Modelle gelten.

**Definition 6.3 (Stochastische Dominanz)** Sei  $X$  eine Zufallsvariable mit Verteilungsfunktion  $F_X$  und  $Y$  eine Zufallsvariable mit Verteilungsfunktion  $F_Y$ .  $Y \stackrel{st}{\leq} X$ , d.h.  $X$  dominiert  $Y$  stochastisch, falls

$$F_X(k) \geq F_Y(k) \quad (6.6)$$

für alle Werte  $k$  gilt.

---

<sup>3</sup>engl.: zero adjusted distribution

Gupta et al. (1996) nennen einige Eigenschaften von ZA-Modellen, die wir in folgendem Satz zusammenfassen.

**Satz 6.4 (Eigenschaften von ZA-Modellen)** Die ZA-Zufallsvariable  $X$  folge der Verteilung mit Wahrscheinlichkeitsfunktion (6.4).

(i) Für die Zählvariablen  $X$  und  $X^*$  gilt:

$$\begin{aligned} X &\stackrel{st}{\leq} X^* && \text{für } \omega \geq 0 \\ X &\stackrel{st}{\geq} X^* && \text{für } \omega \leq 0 \end{aligned}$$

(ii) Für die Ausfallraten  $r(x) = \frac{P(X=x)}{P(X \geq x)}$  und  $r^*(x) = \frac{P(X^*=x)}{P(X^* \geq x)}$  gilt:

$$\begin{aligned} r(0) &\geq r^*(0) && \text{für } \omega \geq 0 \\ r(0) &\leq r^*(0) && \text{für } \omega \leq 0 \end{aligned}$$

(iii) Für die Quotienten der Wahrscheinlichkeiten aufeinanderfolgender positiver Ereignisse  $k$  und  $k+1$  gilt Gleichheit zwischen Ursprungs- und ZA-Verteilung:

$$\frac{P(X = k)}{P(X = k + 1)} = \frac{P(X^* = k)}{P(X^* = k + 1)} \quad \forall k = 1, 2, \dots$$

Die Differenz zwischen entsprechenden Quotienten für die Ereignisse 0 und 1 ist eine wachsende Funktion bzgl.  $\omega$ :

$$\frac{P(X = 0)}{P(X = 1)} - \frac{P(X^* = 0)}{P(X^* = 1)} = \left( \frac{\omega}{1 - \omega} \right) \frac{1}{P(X^* = 1)}$$

(iv) Der relative Fehler

$$e(x) = \left| \frac{F(x) - F^*(x)}{F^*(x)} \right| \tag{6.7}$$

ist eine fallende Funktion bzgl.  $x$  ( $F$  und  $F^*$  bezeichnen die Verteilungsfunktionen für  $X$  und  $X^*$ ).

*Beweis:*

(i) Sei  $F(x) = P(X \leq x)$  die Verteilungsfunktion von  $X$  und  $F^*(x) = P(X^* \leq x)$  die Verteilungsfunktion der Zählvariable  $X^*$ . Dann gilt:

$$\begin{aligned} F(x) &= \sum_{k=0}^x P(X = k) \stackrel{(6.4)}{=} \omega + (1 - \omega)P(X^* = 0) + \sum_{k=1}^x (1 - \omega)P(X^* = k) \\ &= \omega + (1 - \omega) \sum_{k=0}^x P(X^* = k) = \omega + (1 - \omega)F^*(x) \\ &= F^*(x) + \omega [1 - F^*(x)] \end{aligned}$$

Für  $\omega \geq 0$  folgt daher:

$$F(x) = F^*(x) + \omega [1 - F^*(x)] \geq F^*(x)$$

für  $\omega < 0$  hingegen:

$$F(x) = F^*(x) + \omega [1 - F^*(x)] < F^*(x)$$

(ii) Ist  $x \geq 1$  so erhalten wir Gleichheit der Ausfallraten:

$$\begin{aligned}
 r(x) &= \frac{P(X = x)}{P(X \geq x)} \\
 &\stackrel{(6.4)}{=} \frac{(1 - \omega)P(X^* = x)}{\sum_{k=x}^{\infty} P(X = k)} = \frac{(1 - \omega)P(X^* = x)}{(1 - \omega) \sum_{k=x}^{\infty} P(X^* = k)} \\
 &= \frac{P(X^* = x)}{P(X^* \geq x)} = r^*(x)
 \end{aligned}$$

Für  $x = 0$  gilt:

$$\begin{aligned}
 r(0) &= \frac{P(X = 0)}{P(X \geq 0)} = \frac{\omega + (1 - \omega)P(X^* = 0)}{1} \\
 &= P(X^* = 0) + \omega [1 - P(X^* = 0)] \\
 &= r^*(0) + \omega [1 - r^*(0)]
 \end{aligned}$$

Damit unterscheiden sich die Ausfallraten nur im Ursprung und es gilt:

$$\begin{aligned}
 \omega \geq 0 &\Rightarrow r(0) \geq r^*(0) \\
 \omega < 0 &\Rightarrow r(0) < r^*(0)
 \end{aligned}$$

(iii)

$$\begin{aligned}
 \frac{P(X = x)}{P(X = x + 1)} &= \frac{(1 - \omega)P(X^* = x)}{(1 - \omega)P(X^* = x + 1)} \\
 &= \frac{P(X^* = x)}{P(X^* = x + 1)}, \quad x = 1, 2, \dots \\
 \frac{P(X = 0)}{P(X = 1)} &\stackrel{(6.4)}{=} \frac{\omega + (1 - \omega)P(X^* = 0)}{(1 - \omega)P(X^* = 1)} \\
 &= \frac{\omega}{(1 - \omega)P(X^* = 1)} + \frac{P(X^* = 0)}{P(X^* = 1)} \\
 \Rightarrow \frac{P(X = 0)}{P(X = 1)} - \frac{P(X^* = 0)}{P(X^* = 1)} &= \left( \frac{\omega}{1 - \omega} \right) \frac{1}{P(X^* = 1)}
 \end{aligned}$$

(iv) Der Ausdruck des relativen Fehlers lässt sich vereinfachen:

$$\begin{aligned}
 |F(x) - F^*(x)| &= \left| \sum_{k=0}^x P(X = k) - \sum_{k=0}^x P(X^* = k) \right| \\
 &= \left| P(X = 0) + \sum_{k=1}^x P(X = k) - P(X^* = 0) - \sum_{k=1}^x P(X^* = k) \right| \quad (6.8)
 \end{aligned}$$



$$\begin{aligned}
&\stackrel{(6.4)}{=} \left| \omega + (1 - \omega)P(X^* = 0) - P(X^* = 0) + \sum_{k=1}^x (1 - \omega)P(X^* = k) - \sum_{k=1}^x P(X^* = k) \right| \\
&= \left| \omega - \omega P(X^* = 0) - \omega \sum_{k=1}^x P(X^* = k) \right| \\
&= \left| \omega - \omega \sum_{k=0}^x P(X^* = k) \right| \\
&= |\omega - \omega F^*(x)| \tag{6.9}
\end{aligned}$$

Die Verteilungsfunktion  $F^*$  nimmt Werte zwischen 0 und 1 an, daher gilt mit (6.9):

$$|F(x) - F^*(x)| = |\omega| [1 - F^*(x)] \tag{6.10}$$

Bilden wir nun die Differenz zweier aufeinander folgenden relativen Fehlern.

$$\begin{aligned}
e(x+1) - e(x) &\stackrel{(6.10)}{=} |\omega| \frac{1 - F^*(x+1)}{F^*(x+1)} - |\omega| \frac{1 - F^*(x)}{F^*(x)} \\
&= |\omega| \left[ \frac{F^*(x) - F^*(x)F^*(x+1) - F^*(x+1) + F^*(x+1)F^*(x)}{F^*(x)F^*(x+1)} \right] \\
&= |\omega| \left[ \frac{P(X^* \leq x) - P(X^* \leq x+1)}{P(X^* \leq x)P(X^* \leq x+1)} \right] \\
&= |\omega| \left[ -\frac{P(X^* = x+1)}{P(X^* \leq x)P(X^* \leq x+1)} \right] < 0
\end{aligned}$$

Die Ungleichung gilt wegen der Positivität der Wahrscheinlichkeiten für  $x \geq 0$ .

□

### Bemerkung:

Eigenschaft (iv) besagt, dass die Justierung durch  $\omega$  einen erheblichen Einfluss bei kleinen Werten von  $X$  besitzt. Für größere Werte nimmt der Einfluss ab.

Gilt die Verteilungsannahme (6.5) und ist die Wahrscheinlichkeit  $\omega$  bekannt, so können wir weitere Aussagen über die ZI-Zählgröße  $X$  treffen.

**Satz 6.5 (Erwartungswert und Varianz von ZI-Verteilungen)** *Die ZI-Zufallsvariable  $X$  folge einer Verteilung mit Wahrscheinlichkeitsfunktion (6.4).  $\omega$  sei die Eintrittswahrscheinlichkeit des Ereignisses  $\Omega = 0$ . Dann gilt:*

$$E(X) = (1 - \omega)E(X^*) \tag{6.11}$$

$$\text{Var}(X) = (1 - \omega)\text{Var}(X^*) + \omega(1 - \omega)[E(X^*)]^2 \tag{6.12}$$

*Beweis:*

Da  $X^*$  und  $\Omega$  unabhängig sind, gilt für den Erwartungswert von  $X$ :

$$\begin{aligned} E(X) &= E_{\Omega} [E(\Omega X^* | \Omega)] = E_{\Omega} [\Omega E(X^* | \Omega)] \\ &= E_{\Omega} [\Omega E(X^*)] = E_{\Omega}(\Omega) E_{\Omega} [E(X^*)] \\ &= (1 - \omega) E(X^*) \end{aligned} \quad (6.13)$$

Bilden wir nun die Varianz von  $\Omega$ :

$$\begin{aligned} E(\Omega) &= 0P(\Omega = 0) + 1P(\Omega = 1) \\ &= 1 - \omega \end{aligned} \quad (6.14)$$

$$\begin{aligned} E(\Omega^2) &= 0^2P(\Omega = 0) + 1^2P(\Omega = 1) \\ &= 1 - \omega \end{aligned} \quad (6.15)$$

$$\begin{aligned} \Rightarrow \text{Var}(\Omega) &= E(\Omega^2) - [E(\Omega)]^2 = (1 - \omega) - (1 - \omega)^2 \\ &= 1 - \omega - 1 + 2\omega - \omega^2 \\ &= \omega(1 - \omega) \end{aligned} \quad (6.16)$$

Mit Ausdruck (6.16) läßt sich die Varianz von  $X$  für unabhängige  $X^*$  und  $\Omega$  bestimmen:

$$\begin{aligned} \text{Var}(X) &= E_{\Omega} [\text{Var}(\Omega X^* | \Omega)] + \text{Var}_{\Omega} [E(\Omega X^* | \Omega)] \\ &= E_{\Omega} [\Omega^2 \text{Var}(X^* | \Omega)] + \text{Var}_{\Omega} [\Omega E(X^* | \Omega)] \\ &= E_{\Omega}(\Omega^2) \text{Var}(X^*) + [E(X^*)]^2 \text{Var}_{\Omega}(\Omega) \\ &\stackrel{(6.14), (6.16)}{=} (1 - \omega) \text{Var}(X^*) + \omega(1 - \omega) [E(X^*)]^2 \end{aligned} \quad (6.17)$$

□

## 6.2 Zero-Inflated-Poissonregression (ZIP-Regression)

### 6.2.1 Zero-Inflated-Poissonverteilung (ZIP-Verteilung)

Als Ursprungsverteilung sind bei ZI-Modellen beliebige nicht negative, diskrete Verteilungen zugelassen. Häufig wird die Poissonverteilung verwendet. Unter den Annahmen:

- $X^* \sim \text{Poi}(\theta)$
- $\Omega \sim \text{Bernoulli}(1 - \omega)$

erhalten wir bei Verwendung des Ansatzes (6.1) das ZIP-Modell<sup>4</sup>.

**Definition 6.6 (ZIP-Verteilung)** Eine diskrete Zufallsvariable  $X$  mit der Wahrscheinlichkeitsfunktion

$$P_x(\theta, \omega) = P(X = x) = \begin{cases} \omega + (1 - \omega)e^{-\theta} & x = 0 \\ (1 - \omega) \frac{\theta^x}{x!} e^{-\theta} & x = 1, 2, \dots \end{cases} \quad (6.18)$$

heißt ZIP-verteilt mit Parametern  $\theta \in \mathbb{R}_+$  und  $\omega \in [0, 1]$ . Kurz  $X \sim \text{ZIP}(\theta, \omega)$ .

<sup>4</sup>ZIP: zero inflated Poisson

(6.18) erhalten wir, wenn in den allgemeinen Ausdruck (6.4) für ZI-Wahrscheinlichkeitsfunktionen die Wahrscheinlichkeitsfunktion der Poissonverteilung mit Parameter  $\theta$  als  $P^*$  eingesetzt wird.

**Satz 6.7 (Erwartungswert und Varianz der ZIP-Verteilung)** Für die Zufallsvariable  $X \sim ZIP(\theta, \omega)$  gilt:

$$E(X) = (1 - \omega)\theta \quad (6.19)$$

$$Var(X) = E(X) \left( 1 + \frac{\omega}{1 - \omega} E(X) \right) \quad (6.20)$$

*Beweis:*

$$\begin{aligned} E(X) &\stackrel{(6.11)}{=} (1 - \omega)E(X^*) \\ &\stackrel{(A.4)}{=} (1 - \omega)\theta \\ Var(X) &\stackrel{(6.12)}{=} (1 - \omega)Var(X^*) + \omega(1 - \omega)[E(X^*)]^2 \\ &\stackrel{(A.4)}{=} (1 - \omega)\theta + \omega(1 - \omega)\theta^2 \\ &= (1 - \omega)\theta(1 + \omega\theta) \\ &= E(X) \left( 1 + \frac{\omega}{1 - \omega} E(X) \right) \end{aligned}$$

□

**Bemerkung:**

Aus (6.20) lässt sich erkennen, dass sich ZIP-Modelle zur Beschreibung von Überdispersion eignen. Der Ausdruck  $\frac{\omega}{1 - \omega}$  übernimmt dabei die Rolle eines Dispersionsindex (vgl. Abschnitt 4.4). Die Struktur der Varianz der ZIP-Verteilung entspricht daher der Varianzstruktur der negativen Binomialverteilung.

### 6.2.2 Spezifikation eines ZIP-Regressionsmodells mit individuellen Auswahlparametern

Wenden wir uns nun dem Regressionsfall zu. Für die unabhängigen Beobachtungen  $(Y_i | \mathbf{x}_i)$ ,  $i = 1, \dots, n$  gelte

$$\begin{aligned} P(Y_i = y_i | \mathbf{x}_i) &\stackrel{(6.18)}{=} \begin{cases} \omega_i + (1 - \omega_i)e^{-\theta_i} & y_i = 0 \\ (1 - \omega_i) \frac{1}{y_i!} \theta_i^{y_i} e^{-\theta_i} & y_i = 1, 2, \dots \end{cases} \\ &= \omega_i \delta_i + (1 - \omega_i) \frac{\theta_i^{y_i}}{y_i!} e^{-\theta_i} \quad \forall y_i > 0 \end{aligned} \quad (6.21)$$

mit  $\delta_i = 1 - \min\{1, y_i\}$ . In Mullahy (1997) wird der Parameter  $\omega_i$  als konstant angesehen ( $\omega_i = \omega \forall i = 1, \dots, n$ ). Lambert (1992) erweitert dieses Modell, indem sie  $\omega_i$  durch einen Logit-Link spezifiziert und somit von Kovariablen abhängig lässt. Für  $\theta$  wählt sie einen log-linearen Ansatz. Damit lautet die Spezifikation der Verteilungsparameter:

$$\text{logit}(\omega_i) = \log \left( \frac{\omega_i}{1 - \omega_i} \right) = \mathbf{a}_i^t \boldsymbol{\alpha} \quad (6.22)$$

$$\log(\theta_i) = \mathbf{b}_i^t \boldsymbol{\beta} \quad (6.23)$$

wobei  $\mathbf{a}_i^t$  die  $i$ -te Zeile einer Modellmatrix  $\mathbf{A} \in \mathbb{R}^{n, k_a}$  und  $\mathbf{b}_i^t$  die  $i$ -te Zeile einer Modellmatrix  $\mathbf{B} \in \mathbb{R}^{n, k_b}$  ist. Sind die Parameter  $\boldsymbol{\theta}$  und  $\boldsymbol{\omega}$  unabhängig voneinander, dann besitzt das ZIP-Regressionsmodell maximal  $k_a + k_b$  Regressionsparameter und damit mehr als das gewöhnliche Poisson-Regressionsmodell. Falls  $\boldsymbol{\omega}$  nicht von Kovariablen abhängt, besitzt das Modell  $k + 1$  Parameter. Werden  $\boldsymbol{\theta}$  und  $\boldsymbol{\omega}$  von den selben Kovariablen beeinflusst, so sind beide Parameter nicht mehr unabhängig voneinander. Es lässt sich ein Zusammenhang  $\boldsymbol{\omega} = f(\boldsymbol{\theta})$  beschreiben. Für die Behandlung von ZIP-Regressionsmodellen mit abhängigen Parametern sei auf Lambert (1992) verwiesen. Wir setzen im Folgenden die Unabhängigkeit von  $\boldsymbol{\theta}$  und  $\boldsymbol{\omega}$  voraus. Mit der Verknüpfung (6.22) gilt für den Parameter  $\omega_i$ :

$$\omega_i = \frac{e^{\mathbf{a}_i^t \boldsymbol{\alpha}}}{1 + e^{\mathbf{a}_i^t \boldsymbol{\alpha}}} \quad (6.24)$$

Dies garantiert  $0 \leq \omega_i \leq 1$ , da

$$\begin{aligned} \frac{e^{\mathbf{a}_i^t \boldsymbol{\alpha}}}{1 + e^{\mathbf{a}_i^t \boldsymbol{\alpha}}} &\rightarrow 0 && \text{für } \mathbf{a}_i^t \boldsymbol{\alpha} \rightarrow -\infty \\ \frac{e^{\mathbf{a}_i^t \boldsymbol{\alpha}}}{1 + e^{\mathbf{a}_i^t \boldsymbol{\alpha}}} &\rightarrow 1 && \text{für } \mathbf{a}_i^t \boldsymbol{\alpha} \rightarrow \infty \end{aligned}$$

Aus (6.23) folgt:

$$\theta_i = e^{\mathbf{b}_i^t \boldsymbol{\beta}} \quad (6.25)$$

Wie bereits in Abschnitt 4.1 gesehen, liegt auch  $e^{\mathbf{b}_i^t \boldsymbol{\beta}}$  für beliebige  $\mathbf{b}_i$  und  $\boldsymbol{\beta}$  im Zulässigkeitsbereich des korrespondierenden Parameters  $\theta_i$

### 6.2.3 Die Likelihoodfunktion im ZIP-Regressionsmodell mit individuellen Auswahlparametern

Unter Verwendung der Indikatorfunktion  $\delta_i = 1 - \min\{1, y_i\}$  lautet die Likelihoodfunktion für  $\mathbf{Y} | \mathbf{X} \sim ZIP(\boldsymbol{\theta}, \boldsymbol{\omega})$ :

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta} | \mathbf{y}) \stackrel{(6.21)}{=} \prod_{i=1}^n \left[ \omega_i + (1 - \omega_i) e^{-\theta_i} \right]^{\delta_i} \left[ (1 - \omega_i) e^{-\theta_i} \theta_i^{y_i} (y_i!)^{-1} \right]^{1 - \delta_i} \quad (6.26)$$

$$\begin{aligned} \Rightarrow \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta} | \mathbf{y}) &= \sum_{i=1}^n \delta_i \ln \left[ \omega_i + (1 - \omega_i) e^{-\theta_i} \right] \\ &\quad + (1 - \delta_i) [\ln(1 - \omega_i) - \theta_i + y_i \ln \theta_i - \ln(y_i!)] \end{aligned} \quad (6.27)$$

Zur Klassifizierung der Realisierungen  $y_i$  werden zwei disjunkten Indexmengen definiert:

$$I \subseteq \{1, \dots, n\} \quad \text{mit} \quad y_i = 0 \quad \forall i \in I \quad (6.28)$$

$$J \subseteq \{1, \dots, n\} \quad \text{mit} \quad y_j > 0 \quad \forall j \in J \quad (6.29)$$

Damit lässt sich (6.27) in zwei Summen aufspalten:

$$\begin{aligned}\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta} | \mathbf{y}) &= \sum_{i \in I} \ln \left[ \omega_i + (1 - \omega_i) e^{-\theta_i} \right] \\ &+ \sum_{j \in J} \left[ \ln(1 - \omega_j) - \theta_j + y_j \ln \theta_j - \ln(y_j!) \right]\end{aligned}\quad (6.30)$$

Wir setzen (6.25) und (6.24) in (6.30) ein und formen um:

$$\begin{aligned}\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta} | \mathbf{y}) &= \sum_{i \in I} \ln \left[ \frac{e^{\mathbf{a}_i^t \boldsymbol{\alpha}}}{1 + e^{\mathbf{a}_i^t \boldsymbol{\alpha}}} + \left( 1 - \frac{e^{\mathbf{a}_i^t \boldsymbol{\alpha}}}{1 + e^{\mathbf{a}_i^t \boldsymbol{\alpha}}} \right) \exp\{-e^{\mathbf{b}_i^t \boldsymbol{\beta}}\} \right] \\ &+ \sum_{j \in J} \left[ \ln \left( 1 - \frac{e^{\mathbf{a}_j^t \boldsymbol{\alpha}}}{1 + e^{\mathbf{a}_j^t \boldsymbol{\alpha}}} \right) - e^{\mathbf{b}_j^t \boldsymbol{\beta}} + y_j \mathbf{b}_j^t \boldsymbol{\beta} - \ln(y_j!) \right] \\ &= \sum_{i \in I} \ln \left( \frac{e^{\mathbf{a}_i^t \boldsymbol{\alpha}}}{1 + e^{\mathbf{a}_i^t \boldsymbol{\alpha}}} + \frac{1}{1 + e^{\mathbf{a}_i^t \boldsymbol{\beta}}} \exp\{-e^{\mathbf{b}_i^t \boldsymbol{\beta}}\} \right) \\ &+ \sum_{j \in J} \left[ \ln \left( \frac{1}{1 + e^{\mathbf{a}_j^t \boldsymbol{\alpha}}} \right) - e^{\mathbf{b}_j^t \boldsymbol{\beta}} + y_j \mathbf{b}_j^t \boldsymbol{\beta} - \ln(y_j!) \right] \\ &= \sum_{i \in I} \left[ \ln \left( e^{\mathbf{a}_i^t \boldsymbol{\alpha}} + \exp\{-e^{\mathbf{b}_i^t \boldsymbol{\beta}}\} \right) - \ln \left( 1 + e^{\mathbf{a}_i^t \boldsymbol{\alpha}} \right) \right] \\ &+ \sum_{j \in J} \left[ y_j \mathbf{b}_j^t \boldsymbol{\beta} - \ln \left( 1 + e^{\mathbf{a}_j^t \boldsymbol{\alpha}} \right) - e^{\mathbf{b}_j^t \boldsymbol{\beta}} - \ln(y_j!) \right] \\ &= \sum_{i \in I} \ln \left( e^{\mathbf{a}_i^t \boldsymbol{\alpha}} + \exp\{-e^{\mathbf{b}_i^t \boldsymbol{\beta}}\} \right) - \sum_{i=1}^n \ln \left( 1 + e^{\mathbf{a}_i^t \boldsymbol{\alpha}} \right) \\ &+ \sum_{j \in J} \left[ y_j \mathbf{b}_j^t \boldsymbol{\beta} - e^{\mathbf{b}_j^t \boldsymbol{\beta}} - \ln(y_j!) \right]\end{aligned}\quad (6.31)$$

Wir erhalten aus (6.31):

$$\begin{aligned}\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta} | \mathbf{y}) &\propto \sum_{i \in I} \ln \left( e^{\mathbf{a}_i^t \boldsymbol{\alpha}} + \exp\{-e^{\mathbf{b}_i^t \boldsymbol{\beta}}\} \right) \\ &+ \sum_{j \in J} \left[ y_j \mathbf{b}_j^t \boldsymbol{\beta} - e^{\mathbf{b}_j^t \boldsymbol{\beta}} \right] \\ &- \sum_{i=1}^n \ln \left( 1 + e^{\mathbf{a}_i^t \boldsymbol{\alpha}} \right)\end{aligned}\quad (6.32)$$

**Bemerkung:**

Für konstantes  $\omega$  lautet die log-Likelihoodfunktion des ZIP-Modells:

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta}, \omega | \mathbf{y}) &\stackrel{(6.30)}{=} \sum_{i \in I} \ln \left( \omega + (1 - \omega) \exp\{-e^{\mathbf{b}_i^t \boldsymbol{\beta}}\} \right) \\ &+ \sum_{j \in J} \left[ y_j \mathbf{b}_j^t \boldsymbol{\beta} - e^{\mathbf{b}_j^t \boldsymbol{\beta}} + \ln(1 - \omega) - \ln(y_j!) \right]\end{aligned}\quad (6.33)$$

### 6.3 Verallgemeinerte Zero-Inflated-Poissonregression mit gemeinsamem Auswahlparameter (ZIGP $\mu$ -Regression)

Äquivalent zu der Vorgehensweise für das gängige ZIP-Regressionsmodell mit konstantem Auswahlparameter  $\omega$ , wollen wir nun ein ZI-Regressionsmodell herleiten, welchem die verallgemeinerte Poissonverteilung zu Grunde liegt. Dazu definieren wir zunächst entsprechend der ZIP-Verteilung eine ZI-Verteilung, bei der die Zufallsvariable  $X^*$  verallgemeinert poissonverteilt ist.

#### 6.3.1 Verallgemeinerte Zero-Inflated-Poissonverteilung (ZIGP-Verteilung)

Werden in Ansatz (6.1) die Verteilungsannahmen

- $X^* \sim GP^*(\mu, \varphi)$
- $\Omega \sim \text{Bernoulli}(1 - \omega)$

getroffen, so erhalten wir das ZIGP-Verteilungsmodell.

**Definition 6.8 (ZIGP-Verteilung)** Eine Zufallsvariable  $X$  mit der Wahrscheinlichkeitsfunktion

$$P_x(\mu, \varphi, \omega) = P(X = x) = \begin{cases} \omega + (1 - \omega)e^{-\frac{\mu}{\varphi}} & x = 0 \\ (1 - \omega) \frac{\mu[\mu + (\varphi - 1)x]^{x-1}}{x!} \varphi^{-x} e^{-\frac{1}{\varphi}[\mu + (\varphi - 1)x]} & x = 1, 2, \dots \end{cases} \quad (6.34)$$

heißt ZIGP-verteilt<sup>5</sup> mit Parametern  $\mu$ ,  $\varphi$  und  $\omega$ . Kurz  $X \sim \text{ZIGP}(\mu, \varphi, \omega)$ .

**Satz 6.9 (Erwartungswert und Varianz der ZIGP-Verteilung)** Für die Zufallsvariable  $X \sim \text{ZIGP}(\mu, \varphi, \omega)$  gilt:

$$E(X) = (1 - \omega)\mu \quad (6.35)$$

$$\begin{aligned} \text{Var}(X) &= (1 - \omega)\mu(\varphi^2 + \omega\mu) \\ &= E(X) \left( \varphi^2 + \frac{\omega}{1 - \omega} E(X) \right) \end{aligned} \quad (6.36)$$

*Beweis:*

$$\begin{aligned} E(X) &\stackrel{(6.11)}{=} (1 - \omega)E(X^*) \stackrel{(5.13)}{=} (1 - \omega)\mu \\ \text{Var}(X) &\stackrel{(6.12)}{=} (1 - \omega)\text{Var}(X^*) + \omega(1 - \omega)[E(X^*)]^2 \\ &\stackrel{(5.14)}{=} (1 - \omega)\varphi^2\mu + \omega(1 - \omega)\mu^2 \\ &= \varphi^2(1 - \omega)\mu + \frac{\omega}{1 - \omega}(1 - \omega)^2\mu^2 \\ &\stackrel{(6.35)}{=} \varphi^2 E(X) + \frac{\omega}{1 - \omega} [E(X)]^2 \\ &= E(X) \left( \varphi^2 + \frac{\omega}{1 - \omega} E(X) \right) \end{aligned} \quad (6.37)$$

<sup>5</sup>ZIGP: zero inflated generalized Poisson

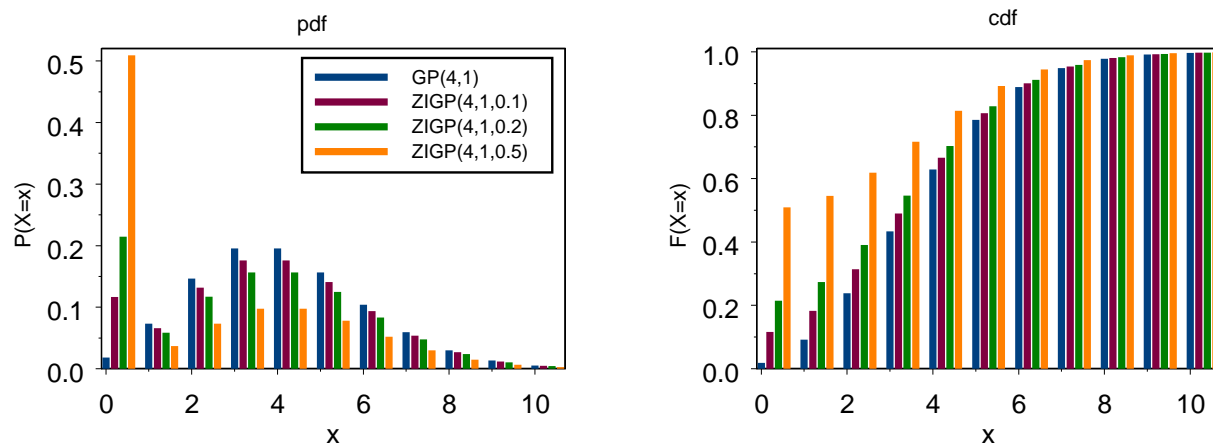


Abbildung 6.1: Wahrscheinlichkeitsfunktionen (links) und Verteilungsfunktionen (rechts) von ZIGP-Verteilungen mit  $\mu = 4$ ,  $\varphi = 1$  und  $\omega = 0, 0.1, 0.2$  und  $0.5$

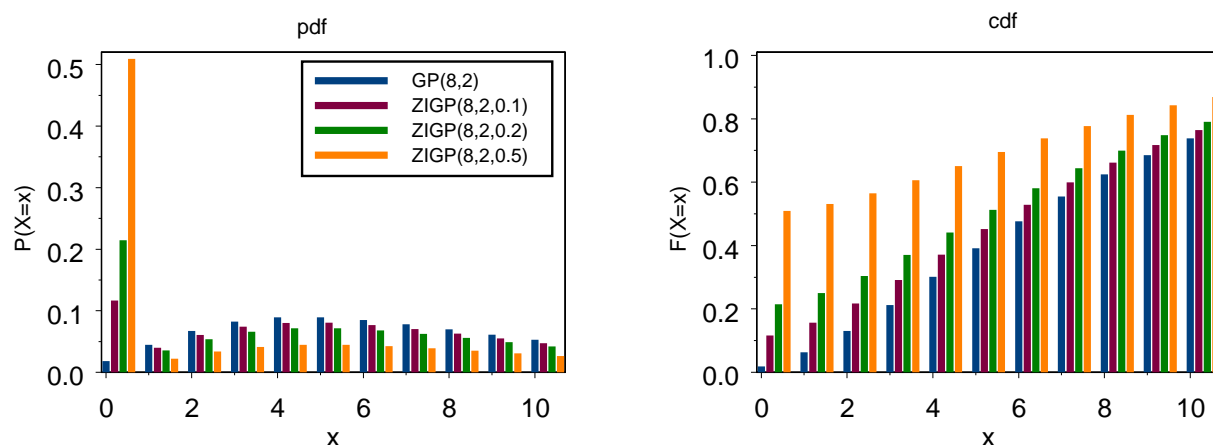


Abbildung 6.2: Wahrscheinlichkeitsfunktionen (links) und Verteilungsfunktionen (rechts) von ZIGP-Verteilungen mit  $\mu = 8$ ,  $\varphi = 2$  und  $\omega = 0, 0.1, 0.2$  und  $0.5$

□

**Bemerkung:**

An Hand von Gleichung (6.37) erkennen wir im Vergleich zur verallgemeinerten Poissonverteilung eine weitere Quelle für Überdispersion in Gestalt des Faktors

$$\frac{\omega}{1 - \omega} \tag{6.38}$$

Dieser gibt den Einfluss des Auswahlprozesses auf die ZIGP-Verteilung wider. Der Wert von (6.38) liegt im Intervall  $[0, \infty)$ .

### 6.3.2 Vergleich von verallgemeinerter Poisson- und ZIGP-Verteilung

Auf der linken Seite der Abbildungen 6.1 und 6.2 sind die Wahrscheinlichkeitsfunktionen von verallgemeinert poisson- und ZIGP-verteilten Zufallsvariablen für verschiedene Parameter zu sehen. In Abbildung 6.1 sind die Wahrscheinlichkeitsfunktionen von  $GP^*(4, 1)$ ,  $ZIGP(4, 1, 0.1)$ ,  $ZIGP(4, 1, 0.2)$  und  $ZIGP(4, 1, 0.5)$  abgetragen. Zur besseren Unterscheidbarkeit der unterschiedlichen Wahrscheinlichkeiten für einen Wert  $x \in \mathbb{N}$  sind die Eintrittswahrscheinlichkeiten für  $X \sim ZIGP(4, 1, 0.1)$  um 0.2 nach rechts versetzt abgetragen. Die Wahrscheinlichkeiten von  $X \sim ZIGP(4, 1, 0.2)$  sind um 0.4, die von  $X \sim ZIGP(4, 1, 0.5)$  um 0.6 nach rechts versetzt abgetragen. Entsprechendes gilt für die Wahrscheinlichkeiten von  $GP^*(8, 2)$ ,  $ZIGP(8, 2, 0.1)$ ,  $ZIGP(8, 2, 0.2)$  und  $ZIGP(8, 2, 0.5)$  in Abbildung 6.2. Rechts neben den Wahrscheinlichkeitsfunktionen sind in der selben Anordnung für die selben Parameterkonstellationen die Verteilungsfunktionen dargestellt.

Zunächst lässt sich erkennen, dass die Wahrscheinlichkeit des Ereignisses  $X = 0$  bei ZIGP-Verteilungen höher ist als bei der verallgemeinerten Poissonverteilung. Mit steigendem Auswahlparameter  $\omega$  steigt die Wahrscheinlichkeit von  $X = 0$  an, die Wahrscheinlichkeiten der positiven Ereignisse nehmen ab. Die Form der Wahrscheinlichkeitsfunktion für die positiven Ereignisse bleibt beim Übergang von verallgemeinerter Poissonverteilung zur ZIGP-Verteilung jedoch im wesentlichen erhalten.

Für beide Parameterkonstellationen  $(\mu, \varphi) = (4, 1)$  und  $(\mu, \varphi) = (8, 2)$  gilt  $\theta = \frac{\mu}{\varphi} = 4$ . Vergleichen wir die Wahrscheinlichkeit des Ereignisses  $X = 0$  in Abbildung 6.1 mit der entsprechenden Wahrscheinlichkeit in Abbildung 6.2 für den selben Parameter  $\omega$ , so stellen wir fest, dass diese übereinstimmen. Dies ist der Fall, da der Erwartungswert  $\mu$  und der Dispersionsparameter  $\varphi$  der Ursprungsverteilung im selben Maße ansteigen:

$$P(X = 0) \stackrel{(6.34)}{=} \omega + (1 - \omega)e^{-\frac{\mu}{\varphi}} = \omega + (1 - \omega)e^{-\frac{2\mu}{2\varphi}} \quad (6.39)$$

$P(X = 0)$  verändert sich daher bei gleichbleibendem  $\omega$  nicht,  $P(X = x)$ ,  $x = 1, 2, \dots$  fällt in Abbildung 6.2 auf Grund des höheren Dispersionsparameters  $\varphi$  jedoch geringer aus als in Abbildung 6.1. Verwenden wir die Parametrisierung (1.1) der verallgemeinerten Poissonverteilung, so bedeutet dies, dass bei gleichbleibenden Parametern  $\theta = \frac{\mu}{\varphi}$  und  $\omega$  die Erhöhung des Dispersionsparameters  $\lambda = 1 - \frac{1}{\varphi}$  keinen Einfluss auf die Wahrscheinlichkeit des Ereignisses  $X = 0$  hat, während  $P(X = x)$ ,  $x = 1, 2, \dots$  sinkt.

Aus den Abbildungen der Verteilungsfunktionen ist zunächst die in Satz 6.4 gezeigte stochastische Dominanz der ZIGP-Verteilungen über deren Ursprungsverteilung, d.h. über die verallgemeinerte Poissonverteilung zu erkennen. Desweiteren erkennen wir, dass ZIGP-Verteilungen mit höherem Auswahlparameter  $\omega$  über entsprechende ZIGP-Verteilungen mit geringerem  $\omega$  stochastisch dominieren. Wir erkennen ebenfalls, dass sich die Verteilungsfunktionen von verallgemeinerten Poisson- und ZIGP-Verteilung für kleine Werte  $x$  stark, für große Werte jedoch kaum noch unterscheiden (vgl. Satz 6.4 iv). Der relative Fehler (6.7) nimmt für steigende Werte  $x$  umso schneller ab, je geringer der Dispersionsparameter  $\varphi$  ist. Der gleiche Zusammenhang scheint zwischen dem relativen Fehler und dem Auswahlparameter  $\omega$  vorzuliegen.



### 6.3.3 Spezifikation des ZIGP-Regressionsmodells mit gemeinsamem Auswahlparameter

Wie in vorangegangenen Abschnitten liegen uns  $n$  unabhängige Beobachtungen  $(Y_i, \mathbf{x}_i)$  sowie die Faktoren  $t_i$  vor. Die Zielvariable  $Y_i$  ist eine Zählgröße,  $\mathbf{x}_i$  ist ein Vektor linear unabhängiger Regressoren,  $t_i$  dient der Normierung des Beobachtungszeitraums.

#### Definition 6.10 (ZIGP $\mu$ -Regressionsmodell)

(i) *Stochastische Komponente:*

Die bedingten Zielvariablen  $Y_i|\mathbf{x}_i$  seien unabhängig ZIGP-verteilt mit gemeinsamem Auswahlparameter  $\omega$  und Dispersionsparameter  $\varphi$ :

$$Y_i|\mathbf{x}_i \sim \text{ZIGP}(\mu_i, \varphi, \omega)$$

(ii) *Systematische Komponente:*

Die Regressoren  $\mathbf{x}_i = (x_{i1}, \dots, x_{ik})^t$  beeinflussen die Verteilung von  $Y_i|\mathbf{x}_i$  nur durch den linearen Prädiktor

$$\eta_i = \mathbf{x}_i^t \boldsymbol{\beta}$$

(iii) *Verknüpfungskomponente:*

Unter Einbeziehung des Offsets  $\ln t_i$  wird der Link

$$\eta_i = \ln \mu_i - \ln t_i \Leftrightarrow \mu_i = t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} \quad (6.40)$$

verwendet.

#### Bemerkung:

Definition 6.10 führt auf die Regressionsfunktionen

$$\begin{aligned} E(Y_i|\mathbf{x}_i) &\stackrel{(6.11)}{=} (1 - \omega)\mu_i \\ &\stackrel{(6.40)}{=} (1 - \omega)t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} \end{aligned} \quad (6.41)$$

$$\begin{aligned} \text{Var}(Y_i|\mathbf{x}_i) &\stackrel{(6.12)}{=} (1 - \omega)\varphi^2\mu_i + \omega(1 - \omega)\mu_i^2 \\ &\stackrel{(6.40)}{=} (1 - \omega)t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} [\varphi^2 + \omega t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}] \\ &\stackrel{(6.41)}{=} E(Y_i|\mathbf{x}_i) \left[ \varphi^2 + \frac{\omega}{1 - \omega} E(Y_i|\mathbf{x}_i) \right] \end{aligned}$$

### 6.3.4 Die Likelihoodfunktion im ZIGP-Regressionsmodell mit gemeinsamem Auswahlparameter

**Satz 6.11 (Log-Likelihoodfunktion im ZIGP $\mu$ -Modell)** Wird die Spezifikation aus Definition 6.10 verwendet, so gilt für die Log-Likelihoodfunktion:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\beta}, \varphi, \omega) &\propto n \ln(1 - \omega) + \sum_{i \in I} \ln \left[ \frac{\omega}{1 - \omega} + \exp \left\{ -\frac{t_i}{\varphi} \mathbf{x}_i^t \boldsymbol{\beta} \right\} \right] \\ &\quad + \sum_{j \in J} \mathbf{x}_j^t \boldsymbol{\beta} + (y_j - 1) \ln \left[ t_j e^{\mathbf{x}_j^t \boldsymbol{\beta}} + (\varphi - 1) y_j \right] \\ &\quad - y_j \ln \varphi - \frac{1}{\varphi} \left[ t_j e^{\mathbf{x}_j^t \boldsymbol{\beta}} + (\varphi - 1) y_j \right] \end{aligned}$$

mit  $I \subseteq \{1, \dots, n\}$  so, dass  $y_i = 0 \ \forall i \in I$  und  $J \subseteq \{1, \dots, n\}$  so, dass  $y_j > 0 \ \forall j \in J$

*Beweis:*

Die Likelihoodfunktion für  $\mathbf{Y}$  lautet bei Verwendung der Indikatorfunktion (6.3)

$$\begin{aligned}
L(\boldsymbol{\beta}, \varphi, \omega | \mathbf{y}) &\stackrel{(6.34)}{=} \prod_{i=1}^n \left[ \omega + (1 - \omega) e^{-\frac{\mu_i}{\varphi}} \right]^{\delta_i} \\
&\times \left[ (1 - \omega) \frac{\mu_i [\mu_i + (\varphi - 1) y_i]^{y_i - 1}}{y_i!} \varphi^{-y_i} e^{-\frac{1}{\varphi} [\mu_i + (\varphi - 1) y_i]} \right]^{1 - \delta_i} \\
\Rightarrow \mathcal{L}(\boldsymbol{\beta}, \varphi, \omega | \mathbf{y}) &= \sum_{i=1}^n \delta_i \ln \left[ \omega + (1 - \omega) e^{-\frac{\mu_i}{\varphi}} \right] \\
&+ (1 - \delta_i) \ln \left[ (1 - \omega) \frac{\mu_i [\mu_i + (\varphi - 1) y_i]^{y_i - 1}}{y_i!} \varphi^{-y_i} e^{-\frac{1}{\varphi} [\mu_i + (\varphi - 1) y_i]} \right] \quad (6.42)
\end{aligned}$$

Mit Hilfe der Teilmengen (6.28) und (6.29) spalten wir die Likelihoodfunktion (6.42) in zwei Summen auf:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\beta}, \varphi, \omega | \mathbf{y}) &= \sum_{i \in I} \ln \left[ \omega + (1 - \omega) e^{-\frac{\mu_i}{\varphi}} \right] \\
&+ \sum_{j \in J} \left\langle \ln(1 - \omega) \ln \mu_j + (y_j - 1) \ln [\mu_j + (\varphi - 1) y_j] - \ln(y_j!) \right. \\
&\quad \left. - y_j \ln \varphi - \frac{1}{\varphi} [\mu_j + (\varphi - 1) y_j] \right\rangle \quad (6.43)
\end{aligned}$$

Verwenden wir die Verknüpfung (6.40) so folgt

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\beta}, \varphi, \omega | \mathbf{y}) &= \sum_{i \in I} \ln \left[ \omega + (1 - \omega) \exp \left\{ -\frac{t_i}{\varphi} \mathbf{x}_i^t \boldsymbol{\beta} \right\} \right] \\
&+ \sum_{j \in J} \left\langle \ln(1 - \omega) + \ln \left( t_j e^{\mathbf{x}_j^t \boldsymbol{\beta}} \right) + (y_j - 1) \ln \left[ t_j e^{\mathbf{x}_j^t \boldsymbol{\beta}} + (\varphi - 1) y_j \right] \right. \\
&\quad \left. - \ln(y_j!) - y_j \ln \varphi - \frac{1}{\varphi} \left[ t_j e^{\mathbf{x}_j^t \boldsymbol{\beta}} + (\varphi - 1) y_j \right] \right\rangle \\
&= \sum_{i \in I} \left\langle \ln \left[ \frac{\omega}{1 - \omega} + \exp \left\{ -\frac{t_i}{\varphi} \mathbf{x}_i^t \boldsymbol{\beta} \right\} \right] + \ln(1 - \omega) \right\rangle \\
&+ \sum_{j \in J} \left\langle \ln t_j + \mathbf{x}_j^t \boldsymbol{\beta} + (y_j - 1) \ln \left[ t_j e^{\mathbf{x}_j^t \boldsymbol{\beta}} + (\varphi - 1) y_j \right] - \ln(y_j!) \right. \\
&\quad \left. - y_j \ln \varphi - \frac{1}{\varphi} \left[ t_j e^{\mathbf{x}_j^t \boldsymbol{\beta}} + (\varphi - 1) y_j \right] + \ln(1 - \omega) \right\rangle \\
&= n \ln(1 - \omega) + \sum_{i \in I} \ln \left[ \frac{\omega}{1 - \omega} + \exp \left\{ -\frac{t_i}{\varphi} \mathbf{x}_i^t \boldsymbol{\beta} \right\} \right] \\
&+ \sum_{j \in J} \left\langle \ln t_j + \mathbf{x}_j^t \boldsymbol{\beta} + (y_j - 1) \ln \left[ t_j e^{\mathbf{x}_j^t \boldsymbol{\beta}} + (\varphi - 1) y_j \right] - \ln(y_j!) \right. \\
&\quad \left. - y_j \ln \varphi - \frac{1}{\varphi} \left[ t_j e^{\mathbf{x}_j^t \boldsymbol{\beta}} + (\varphi - 1) y_j \right] \right\rangle \quad (6.44)
\end{aligned}$$

□

### 6.3.5 Messung der Anpassungsqualität des ZIGP-Regressionsmodells mit gemeinsamem Auswahlparameter

**Satz 6.12 (Akaike-Information für das ZIGP $\mu$ -Regressionsmodell)** Die Akaike-Information des Regressionsmodells aus Definition 6.10 lautet:

$$\begin{aligned} AIC &= 2k - n \ln(1 - \omega) - \sum_{i \in I} \ln \left[ \frac{\omega}{1 - \omega} + \exp \left\{ -\frac{t_i}{\varphi} e^{\mathbf{x}_i^t \boldsymbol{\beta}} \right\} \right] \\ &\quad - \sum_{j \in J} \left\langle \ln t_j + \mathbf{x}_j \boldsymbol{\beta} + (y_j - 1) \ln \left[ t_j e^{\mathbf{x}_j^t \boldsymbol{\beta}} + (\varphi - 1) y_j \right] \right. \\ &\quad \left. - y_j \ln \varphi - \frac{1}{\varphi} \left[ t_j e^{\mathbf{x}_j^t \boldsymbol{\beta}} + (\varphi - 1) y_j \right] - \ln(y_j!) \right\rangle \end{aligned}$$

*Beweis:*

Einsetzen von (6.44) in (3.37)

□

Mit Hilfe der Akaike-Information können wir ZIGP $\mu$ -Modelle mit unterschiedlichen Kovariablen vergleichen. Darüber hinaus lässt sich die Anpassungsqualität eines ZIGP $\mu$ -Modells über AIC auch mit der Anpassungsqualität eines entsprechenden verallgemeinerten Poissonmodells oder gewöhnlichen Poissonmodells vergleichen. Dies ist möglich, da das ZIGP $\mu$ - in das GP-Regressionsmodell und dieses wiederum in das Poisson-Regressionsmodell eingebettet ist:

$$ZIGP(\mu, \varphi, \omega) \stackrel{\omega \equiv 0}{=} GP^*(\mu, \varphi) \stackrel{\varphi \equiv 1}{=} Poi(\mu)$$

$$\Rightarrow \text{Poissonregression} \subset \text{verallgemeinerte Poissonregression} \subset \text{ZIGP}\mu\text{-Regression}$$

## 6.4 Verallgemeinerte Zero-Inflated-Poissonregression mit individuellen Auswahlparametern (ZIGP $\mu\omega$ -Regression)

Wir wandeln das ZIGP-Regressionsmodell 1 dahin gehend ab, dass nun auch für die Auswahlvariable  $\Omega$  Regression zugelassen wird. Die Auswahlparameter  $\omega_i$  hängen daher im Folgenden ebenfalls, wie von Lambert (1992) für die ZIP-Regression vorgeschlagen, von Regressoren ab. Dazu sind uns  $n$  unabhängige Beobachtungen  $(Y_i, \mathbf{a}_i, \mathbf{b}_i)$  und die Faktoren  $t_i$  gegeben. Der Vektor  $\mathbf{a}_i$  besteht aus den Regressoren welche mit dem Parameter  $\omega_i$  korrespondieren,  $\mu_i$  wird durch die Regressoren  $\mathbf{b}_i$  beeinflusst. Mit  $t_i$  lassen sich die Beobachtungszeiträume normieren.

**Definition 6.13 (ZIGP $\mu\omega$ -Regressionsmodell)**

(i) *Stochastische Komponente:*

Die bedingten Zielvariablen  $Y_i | \mathbf{a}_i, \mathbf{b}_i$  seien unabhängig ZIGP-verteilt mit übereinstimmendem Dispersionsparameter  $\varphi$ :

$$Y_i | \mathbf{a}_i, \mathbf{b}_i \sim ZIGP(\mu_i, \varphi, \omega_i)$$

(ii) *Systematische Komponente:*

Die Regressoren  $\mathbf{a}_i = (a_{i1}, \dots, a_{il})^t$  beeinflussen die Verteilung von  $Y_i | \mathbf{a}_i, \mathbf{b}_i$  nur durch den linearen Prädiktor

$$\nu_i = \mathbf{a}_i^t \boldsymbol{\alpha} \tag{6.45}$$

Die Regressoren  $\mathbf{b}_i = (b_{i1}, \dots, b_{ik})^t$  beeinflussen die Verteilung von  $Y_i | \mathbf{a}_i, \mathbf{b}_i$  nur durch den linearen Prädiktor

$$\eta_i = \mathbf{b}_i^t \boldsymbol{\beta} \quad (6.46)$$

(iii) Verknüpfungskomponente:

Unter Einbeziehung des Offsets  $\ln t_i$  werden die Linkfunktionen

$$\nu_i = \text{logit } \omega_i \Leftrightarrow \omega_i = \frac{\exp\{\mathbf{a}_i^t \boldsymbol{\alpha}\}}{1 + \exp\{\mathbf{a}_i^t \boldsymbol{\alpha}\}} \quad (6.47)$$

$$\eta_i = \ln \mu_i - \ln t_i \Leftrightarrow \mu_i = t_i \exp\{\mathbf{b}_i^t \boldsymbol{\beta}\} \quad (6.48)$$

verwendet.

### Bemerkung:

Definition 6.13 führt auf die Regressionsfunktionen

$$E(Y_i | \mathbf{a}_i, \mathbf{b}_i) \stackrel{(6.11)}{=} (1 - \omega_i) \mu_i = \frac{t_i \exp\{\mathbf{b}_i^t \boldsymbol{\beta}\}}{1 + \exp\{\mathbf{a}_i^t \boldsymbol{\alpha}\}} \quad (6.49)$$

$$(6.50)$$

$$\text{Var}(Y_i | \mathbf{a}_i, \mathbf{b}_i) \stackrel{(6.12)}{=} (1 - \omega_i) \mu_i (\varphi^2 + \omega_i \mu_i) \quad (6.51)$$

$$= \frac{t_i \exp\{\mathbf{b}_i^t \boldsymbol{\beta}\}}{1 + \exp\{\mathbf{a}_i^t \boldsymbol{\alpha}\}} \left[ \varphi^2 + \frac{t_i \exp\{\mathbf{b}_i^t \boldsymbol{\beta}\} \exp\{\mathbf{a}_i^t \boldsymbol{\alpha}\}}{1 + \exp\{\mathbf{a}_i^t \boldsymbol{\alpha}\}} \right] \quad (6.52)$$

**Satz 6.14 (Log-Likelihoodfunktion im ZIGP $\mu\omega$ -Regressionsmodell)** Wird die Spezifikation aus Definition 6.13 verwendet, so gilt für die Log-Likelihoodfunktion:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \varphi | \mathbf{y}) &\propto \sum_{i \in I} \ln \left[ e^{\mathbf{a}_i^t \boldsymbol{\alpha}} + \exp \left\{ -\frac{t_i}{\varphi} e^{\mathbf{b}_i^t \boldsymbol{\beta}} \right\} \right] \\ &+ \sum_{j \in J} \left\langle \mathbf{b}_j^t \boldsymbol{\beta} + (y_j - 1) \ln \left[ t_j e^{\mathbf{b}_j^t \boldsymbol{\beta}} + (\varphi - 1) y_j \right] \right. \\ &\quad \left. - y_j \ln \varphi - \frac{1}{\varphi} \left[ t_j e^{\mathbf{b}_j^t \boldsymbol{\beta}} + (\varphi - 1) y_j \right] \right\rangle \\ &- \sum_{i=1}^n \ln (1 + e^{\mathbf{a}_i^t \boldsymbol{\alpha}}) \end{aligned} \quad (6.53)$$

mit  $I \subseteq \{1, \dots, n\}$  sodass  $y_i = 0 \ \forall i \in I$  und  $J \subseteq \{1, \dots, n\}$  sodass  $y_j > 0 \ \forall j \in J$

*Beweis:*

Entsprechend (6.43) gilt für die Log-Likelihoodfunktion des ZIGP $\mu\omega$ -Regressionsmodells

$$\begin{aligned} \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \varphi | \mathbf{y}) &= \sum_{i \in I} \ln \left[ \omega_i + (1 - \omega_i) e^{-\frac{\mu_i}{\varphi}} \right] \\ &+ \sum_{j \in J} \left\langle \ln(1 - \omega_j) + \ln \mu_j + (y_j - 1) \ln [\mu_j + (\varphi - 1) y_j] - \ln(y_j!) \right. \\ &\quad \left. - y_j \ln \varphi - \frac{1}{\varphi} [\mu_j + (\varphi - 1) y_j] \right\rangle \end{aligned} \quad (6.54)$$

Verwenden wir in (6.54) die Verknüpfungen (6.47) und (6.48):

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \varphi | \mathbf{y}) &= \sum_{i \in I} \ln \left[ \frac{\exp\{\mathbf{a}_i^t \boldsymbol{\alpha}\}}{1 + \exp\{\mathbf{a}_i^t \boldsymbol{\alpha}\}} + \left(1 - \frac{\exp\{\mathbf{a}_i^t \boldsymbol{\alpha}\}}{1 + \exp\{\mathbf{a}_i^t \boldsymbol{\alpha}\}}\right) \exp\left\{-\frac{t_i}{\varphi} e^{\mathbf{b}_i^t \boldsymbol{\beta}}\right\}\right] \\
&+ \sum_{j \in J} \left\langle \ln \left(1 - \frac{\exp\{\mathbf{a}_j^t \boldsymbol{\alpha}\}}{1 + \exp\{\mathbf{a}_j^t \boldsymbol{\alpha}\}}\right) + \ln(t_j e^{\mathbf{b}_j^t \boldsymbol{\beta}}) + (y_j - 1) \ln[t_j e^{\mathbf{b}_j^t \boldsymbol{\beta}} + (\varphi - 1)y_j] \right. \\
&\quad \left. - \ln(y_j!) - y_j \ln \varphi - \frac{1}{\varphi} [t_j e^{\mathbf{b}_j^t \boldsymbol{\beta}} + (\varphi - 1)y_j] \right\rangle \\
&= \sum_{i \in I} \ln \left[ e^{\mathbf{a}_i^t \boldsymbol{\alpha}} + \exp\left\{-\frac{t_i}{\varphi} e^{\mathbf{b}_i^t \boldsymbol{\beta}}\right\}\right] - \ln(1 + e^{\mathbf{a}_i^t \boldsymbol{\alpha}}) \\
&+ \sum_{j \in J} \left\langle \ln t_j + \mathbf{b}_j \boldsymbol{\beta} + (y_j - 1) \ln[t_j e^{\mathbf{b}_j^t \boldsymbol{\beta}} + (\varphi - 1)y_j] - \ln(y_j!) \right. \\
&\quad \left. - y_j \ln \varphi - \frac{1}{\varphi} [t_j e^{\mathbf{b}_j^t \boldsymbol{\beta}} + (\varphi - 1)y_j] - \ln(1 + e^{\mathbf{a}_j^t \boldsymbol{\alpha}}) \right\rangle \\
&= \sum_{i \in I} \ln \left[ e^{\mathbf{a}_i^t \boldsymbol{\alpha}} + \exp\left\{-\frac{t_i}{\varphi} e^{\mathbf{b}_i^t \boldsymbol{\beta}}\right\}\right] \\
&+ \sum_{j \in J} \left\langle \ln t_j + \mathbf{b}_j \boldsymbol{\beta} + (y_j - 1) \ln[t_j e^{\mathbf{b}_j^t \boldsymbol{\beta}} + (\varphi - 1)y_j] - \ln(y_j!) \right. \\
&\quad \left. - y_j \ln \varphi - \frac{1}{\varphi} [t_j e^{\mathbf{b}_j^t \boldsymbol{\beta}} + (\varphi - 1)y_j] - \sum_{i=1}^n \ln(1 + e^{\mathbf{a}_i^t \boldsymbol{\alpha}}) \right\rangle
\end{aligned}$$

□

## 6.5 Verallgemeinerte Zero-Inflated-Poissonregression mit individuellen Auswahlparametern aber konstantem Parameter $\mu$ (ZIGP $\omega$ -Regression)

Die  $n$  unabhängigen Beobachtungen  $(Y_i, \mathbf{x}_i)$  bestehen aus der Zielgröße  $Y_i$  und den Regressoren  $\mathbf{x}_i$ . Zusätzlich sind uns die Normierungsfaktoren  $t_i$ ,  $i = 1, \dots, n$  für die Beobachtungszeiträume gegeben.

Wir wollen Regression nun ausschließlich auf die Auswahlvariable  $\Omega$  anwenden. Die Verteilungsparameter  $\mu$  und  $\varphi$  des Ursprungsprozesses seien daher konstant. Damit sind die Zufallsvariablen  $Y_i^*$ , welche durch den Ursprungsprozess erzeugt werden, unabhängig identisch verteilt.

### Definition 6.15 (ZIGP $\omega$ -Regressionsmodell)

(i) *Stochastische Komponente:*

Die bedingten Zielvariablen  $Y_i | \mathbf{x}_i$  seien unabhängig ZIGP-verteilt mit übereinstimmendem Dispersionsparameter  $\varphi$  und übereinstimmendem Erwartungswert  $\mu$  der Ursprungsverteilung:

$$Y_i | \mathbf{x}_i \sim \text{ZIGP}(\mu, \varphi, \omega_i)$$

(ii) *Systematische Komponente:*

Die Regressoren  $\mathbf{x}_i = (x_{i1}, \dots, x_{in})^t$  beeinflussen die Verteilung von  $Y_i | \mathbf{x}_i$  nur durch den linearen Prädiktor

$$\nu_i = \mathbf{x}_i^t \boldsymbol{\alpha}$$

(iii) *Verknüpfungskomponente:*

Unter Einbeziehung des Offsets  $\ln t_i$  wird die Linkfunktion

$$\nu_i = \text{logit } \omega_i \Leftrightarrow \omega_i = \frac{\exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\}}{1 + \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\}} \quad (6.55)$$

verwendet.

**Bemerkung:**

Definition 6.15 führt bei Berücksichtigung des Normierungsfaktors  $t_i$  auf die Regressionsfunktionen

$$E(Y_i | \mathbf{x}_i) = (1 - \omega_i)t_i\mu = \frac{t_i\mu}{1 + \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\}} \quad (6.56)$$

$$(6.57)$$

$$\text{Var}(Y_i | \mathbf{x}_i) = (1 - \omega_i)t_i\mu(\varphi^2 + \omega_i t_i\mu) \quad (6.58)$$

$$= \frac{t_i\mu}{1 + \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\}} \left[ \varphi^2 + \frac{\exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\} t_i\mu}{1 + \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\}} \right] \quad (6.59)$$

**Satz 6.16 (Log-Likelihoodfunktion im ZIGP $\omega$ -Regressionsmodell)** Wird die Spezifikation aus Definition 6.15 verwendet, so gilt für die Log-Likelihoodfunktion:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\alpha}, \mu, \varphi | \mathbf{y}) &\propto \sum_{i \in I} \ln \left[ e^{\mathbf{x}_i^t \boldsymbol{\alpha}} + e^{-\frac{t_i \mu}{\varphi}} \right] - \sum_{i=1}^n \ln \left( 1 + e^{\mathbf{x}_i^t \boldsymbol{\alpha}} \right) \\ &+ \sum_{j \in J} \left\langle \ln \mu + (y_j - 1) \ln [t_j \mu + (\varphi - 1)y_j] \right. \\ &\quad \left. - y_j \ln \varphi - \frac{1}{\varphi} [t_j \mu + (\varphi - 1)y_j] \right\rangle \end{aligned} \quad (6.60)$$

mit  $I \subseteq \{1, \dots, n\}$  sodass  $y_i = 0 \quad \forall i \in I$  und  $J \subseteq \{1, \dots, n\}$  sodass  $y_j > 0 \quad \forall j \in J$

*Beweis:*

Entsprechend (6.43) gilt für die Log-Likelihoodfunktion des ZIGP $\omega$ -Regressionsmodells

$$\begin{aligned} \mathcal{L}(\boldsymbol{\alpha}, \mu, \varphi | \mathbf{y}) &= \sum_{i \in I} \ln \left[ \omega_i + (1 - \omega_i) e^{-\frac{t_i \mu}{\varphi}} \right] \\ &+ \sum_{j \in J} \left\langle \ln(1 - \omega_j) + \ln(t_j \mu) + (y_j - 1) \ln [t_j \mu + (\varphi - 1)y_j] - \ln(y_j!) \right. \\ &\quad \left. - y_j \ln \varphi - \frac{1}{\varphi} [t_j \mu + (\varphi - 1)y_j] \right\rangle \end{aligned} \quad (6.61)$$

Verwenden wir in (6.54) die Verknüpfungen (6.47) und (6.48):

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\alpha}, \mu, \varphi | \mathbf{y}) &= \sum_{i \in I} \ln \left[ \frac{\exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\}}{1 + \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\}} + \left(1 - \frac{\exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\}}{1 + \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\}}\right) e^{-t_i \frac{\mu}{\varphi}} \right] \\
&+ \sum_{j \in J} \left\langle \ln \left(1 - \frac{\exp\{\mathbf{x}_j^t \boldsymbol{\alpha}\}}{1 + \exp\{\mathbf{x}_j^t \boldsymbol{\alpha}\}}\right) + \ln(t_j \mu) + (y_j - 1) \ln[t_j \mu + (\varphi - 1)y_j] \right. \\
&\quad \left. - \ln(y_j!) - y_j \ln \varphi - \frac{1}{\varphi} [t_j \mu + (\varphi - 1)y_j] \right\rangle \\
&= \sum_{i \in I} \ln \left[ e^{\mathbf{x}_i^t \boldsymbol{\alpha}} + e^{-t_i \frac{\mu}{\varphi}} \right] - \ln(1 + e^{\mathbf{x}_i^t \boldsymbol{\alpha}}) \\
&+ \sum_{j \in J} \left\langle \ln t_j + \ln \mu + (y_j - 1) \ln[t_j \mu + (\varphi - 1)y_j] \right. \\
&\quad \left. - \ln(y_j!) - y_j \ln \varphi - \frac{1}{\varphi} [t_j \mu + (\varphi - 1)y_j] - \ln(1 + e^{\mathbf{x}_i^t \boldsymbol{\alpha}}) \right\rangle \\
&= \sum_{i \in I} \ln \left[ e^{\mathbf{x}_i^t \boldsymbol{\alpha}} + e^{-t_i \frac{\mu}{\varphi}} \right] \\
&+ \sum_{j \in J} \left\langle \ln t_j + \ln \mu + (y_j - 1) \ln[t_j \mu + (\varphi - 1)y_j] \right. \\
&\quad \left. - \ln(y_j!) - y_j \ln \varphi - \frac{1}{\varphi} [t_j \mu + (\varphi - 1)y_j] \right\rangle \\
&- \sum_{i=1}^n \ln(1 + e^{\mathbf{x}_i^t \boldsymbol{\alpha}})
\end{aligned}$$

□





# Kapitel 7

## Hürden-Regressionsmodelle

### 7.1 Kombinierte Verteilungen

Als kombiniertes Modell<sup>1</sup> wird ein Verteilungsansatz bezeichnet, welcher eine Verteilung für Ereignisse unterhalb einer gesetzten Hürde mit einer Verteilung für Ereignisse oberhalb dieser Hürde kombiniert. Damit ist eine systematische Unterscheidung zwischen Ereignissen unter und über der festgelegten Hürde möglich. Man spricht daher auch von **Hürdenmodellen**<sup>2</sup>.

**Definition 7.1 (Wahrscheinlichkeitsfunktion eines Hürdenmodells)** Sei  $P_1(x)$  die Wahrscheinlichkeitsfunktion einer nicht negativen, ganzzahligen Verteilung.  $P_2(x)$  sei eine Wahrscheinlichkeitsfunktion für ganzzahlige Zufallsvariablen  $X > k$ . Die Wahrscheinlichkeitsfunktion des kombinierten Modells mit Hürde  $k$  lautet:

$$P(x) = P(X = x) = \begin{cases} P_1(x) & x \leq k \\ \xi_k P_2(x) & x > k \end{cases} \quad (7.1)$$

mit  $\xi_k = 1 - \sum_{i=1}^k P_1(X = i)$ .

Die Größe  $\xi_k$  in Definition 7.1 dient der Normalisierung von  $P(x)$ . Wir bezeichnen die zu  $P_1(x)$  gehörende Verteilung als **Hürdenverteilung**, die zu  $P_2(x)$  gehörende Verteilung als **Ursprungsverteilung**<sup>3</sup> des kombinierten Modells. Kombinierte Modelle werden z.B. von Winkelmann (2003) und Mullahy (1986) behandelt.

### 7.2 Kombinierte Regressionsmodelle mit Hürde Null

Bei kombinierten Modellen wird als Hürde meist der Wert Null gewählt. Im folgenden beschränken wir uns auf den Fall dieser sogenannten **Null-Hürdenmodelle**. Wir werden sehen, dass sie eine weitere Möglichkeit bieten Nullenüberschuss zu modellieren.

#### 7.2.1 Verteilung in Null-Hürdenmodellen

Als Ursprungsverteilung kann bei Null-Hürdenmodellen gemäß Definition 7.1 jede positive oder passend verschobene Wahrscheinlichkeitsverteilung verwendet werden.  $P_1$  korrespondiert mit einer binären Auswahlvariable zwischen den Ereignissen  $X = 0$  und  $X > 0$ .

---

<sup>1</sup>engl.: two-part model

<sup>2</sup>engl.: hurdle model

<sup>3</sup>engl.: parent distribution

Dem Ansatz von Mullahy (1986) für Null-Hürdenmodelle hingegen liegt sowohl als Hürden- als auch als Ursprungsverteilung eine nicht negative, ganzzahlige Wahrscheinlichkeitsverteilung zu Grunde. Die Ursprungsverteilung muss dabei jedoch auf positive Werte eingeschränkt werden, was die Verwendung eines weiteren Normalisierungsfaktors erforderlich macht. Diese Einschränkung bezieht sich nur auf den Definitionsbereich der Ursprungsverteilung, nicht auf die zu betrachtende Population. Für diesen Ansatz können wir die Wahrscheinlichkeitsfunktion wie folgt definieren:

**Definition 7.2 (Wahrscheinlichkeitsfunktion eines Null-Hürdenmodells)** Seien  $P_1$  und  $P_2$  zwei Wahrscheinlichkeitsfunktionen für nicht negative, ganzzahlige Zufallsvariable. Die Wahrscheinlichkeit des Null-Hürdenmodells lautet:

$$P(x) = P(X = x) = \begin{cases} P_1(x) & x = 0 \\ \varrho P_2(x) & x = 1, 2, \dots \end{cases} \quad (7.2)$$

$$\text{mit } \varrho = \frac{1 - P_1(0)}{1 - P_2(0)}$$

Der Zähler von  $\varrho$  gibt die Wahrscheinlichkeit an, die Hürde zu überspringen, während der Nenner zur Normalisierung der abgeschnittenen Ursprungsverteilung  $P_2$  dient.

**Bemerkungen:**

- Das kombinierte Modell (7.2) reduziert sich zum Ursprungsmodell, falls  $\varrho = 1$  gilt. Dies ist der Fall, wenn  $P_1(0)$  und  $P_2(0)$  übereinstimmen. Gilt  $\varrho < 1$ , erhalten wir im Vergleich zum Ursprungsmodell Nullenüberschuss, während für  $\varrho > 1$  die Wahrscheinlichkeit des Ereignisses Null geringer ist.
- Kombinierte Modelle können durch die Hinzunahme weiterer Hürden erweitert werden. Sind  $P_1$ ,  $P_2$  und  $P_3$  beliebige Wahrscheinlichkeitsfunktionen nicht-negativer ganzzahliger Zufallsvariablen, so lautet die Wahrscheinlichkeitsfunktion des **Doppel-Hürdenmodells** mit Hürden bei Null und Eins:

$$P(x) = P(X = x) = \begin{cases} P_1(x) & x = 0 \\ \frac{1 - P_1(0)}{1 - P_2(0)} P_2(x) & x = 1 \\ \left[ 1 - P_1(0) - \frac{1 - P_1(0)}{1 - P_2(0)} P_2(1) \right] \frac{P_3(x)}{1 - P_3(0) - P_3(1)} & x = 2, 3, \dots \end{cases} \quad (7.3)$$

(Winkelmann, 2003, S.141). Das zu (7.3) gehörende kombinierte Modell stellt eine Verallgemeinerung des Modells (7.2) dar. Äquivalent zu dieser Erweiterung lassen sich auch Modelle mit mehr als zwei Hürden konstruieren.

**Satz 7.3 (Erwartungswert und Varianz eines Null-Hürdenmodells)** Die Zufallsvariable  $X$  folge einer kombinierten Verteilung mit Wahrscheinlichkeitsfunktion gemäß (7.2).  $E_2(X)$  sei der Erwartungswert,  $Var_2(X)$  die Varianz der Ursprungsverteilung  $P_2$ . Dann gilt für Erwartungswert  $E(X)$  und Varianz  $Var(X)$  des kombinierten Modells:

$$E(X) = \varrho E_2(X) \quad (7.4)$$

$$Var(X) = \varrho [Var_2(X) + (1 - \varrho)E_2(X)^2] \quad (7.5)$$

*Beweis:*

$$\begin{aligned}
E(X) &= \sum_{x=0}^{\infty} xP(x) = 0 P_1(0) + \sum_{x=1}^{\infty} x \varrho P_2(x) \\
&= \varrho \sum_{x=1}^{\infty} x P_2(x) = \varrho E_2(X) \\
\text{Var}(X) &= E(X^2) - E(X)^2 = \sum_{x=1}^{\infty} x^2 \varrho P_2(x) - \left( \sum_{x=1}^{\infty} x \varrho P_2(x) \right)^2 \\
&= \varrho E_2(X^2) - \varrho^2 E_2(X)^2 \\
&= \varrho [E_2(X^2) - E_2(X)^2 + E_2(X)^2 - \varrho E_2(X)^2] \\
&= \varrho [\text{Var}_2(X) + (1 - \varrho) E_2(X)^2]
\end{aligned}$$

□

Der Erwartungswert des kombinierten Modells weicht durch den Faktor  $\varrho$  vom Erwartungswert des Ursprungsmodells ab. Ist die Wahrscheinlichkeit, dass die Hürde überwunden wird größer als die Summe der Wahrscheinlichkeiten positiver Ereignisse im Ursprungsmodell, so gilt  $\varrho > 1$ . Der Erwartungswert der kombinierten Verteilung übersteigt in diesem Fall den Erwartungswert der Ursprungsverteilung. Falls die Wahrscheinlichkeit dass die Hürde übersprungen wird größer ist als die Wahrscheinlichkeit für das Ereignis Null im Ursprungsmodell, so gilt  $\varrho < 1$ . Nullenüberschuss kann daher auch als Eigenschaft des Erwartungswertes aufgefasst werden.

### 7.2.2 Zusammenhang zwischen Dispersion und Null-Hürdenmodellen

Dispersion kann bei Poissonverteilung äquivalent zu (4.38) durch den Varianz-Erwartungswert Quotienten ausgedrückt werden. Dieser entspricht dem Dispersionsparameter in (4.38).

Im Null-Hürdenmodell (7.2) gilt:

$$\begin{aligned}
\frac{\text{Var}(X)}{E(X)} &\stackrel{(7.4),(7.5)}{=} \frac{\varrho [\text{Var}_2(X) + (1 - \varrho) E_2(X)^2]}{\varrho E_2(X)} \\
&= \frac{E_2(X^2) - E_2(X)^2 + E_2(x)^2 - \varrho E_2(X)^2}{E_2(X)} \\
&= \frac{\sum_{x=1}^{\infty} x^2 P_2(x) - \varrho [\sum_{x=1}^{\infty} x P_2(x)]^2}{\sum_{x=1}^{\infty} x P_2(x)} \tag{7.6}
\end{aligned}$$

Für  $\varrho = 1$  reduziert sich (7.6) zum Varianz-Erwartungswert Quotienten des Ursprungsmodells.

$$\begin{aligned}
\frac{\text{Var}(X)}{E(X)} &= \frac{\sum_{x=1}^{\infty} x^2 P_2(x) - [\sum_{x=1}^{\infty} x P_2(x)]^2}{\sum_{x=1}^{\infty} x P_2(x)} \\
&= \frac{\text{Var}_2(X)}{E_2(X)}
\end{aligned}$$

Ist  $P_2$  die Wahrscheinlichkeitsfunktion einer Poissonverteilung, so bedeutet dies Äquidispersion, da in diesem Fall  $\frac{\text{Var}(X)}{E(X)} = 1$  gilt. Bei  $\varrho \neq 1$  liegt ein echtes kombiniertes Modell vor. Gilt  $0 < \varrho < 1$ , so erhalten wir Überdispersion, gilt  $1 < \varrho < c$  erhalten wir Unterdispersion.

Im Allgemeinen existiert für  $\varrho$  eine obere Schranke  $c$ . Diese wird benötigt, damit die Varianz positiv ist. Im Fall der Poissonverteilung mit Intensität  $\theta$  gilt für die Varianz:

$$\text{Var}(X) \stackrel{(7.5)}{=} \varrho(\theta + (1 - \varrho)\theta^2) = \varrho\theta(1 + (1 - \varrho)\theta) \quad (7.7)$$

Die Nullstellen von (7.7) bzgl.  $\varrho$  sind  $\varrho_1 = 0$  und  $\varrho_2 = \frac{\theta+1}{\theta}$ . Bei kombinierten Poissonmodellen mit Hürde 0 tritt Unterdispersion daher für  $1 < \varrho < \frac{\theta+1}{\theta}$  auf. Folglich wird Unterdispersion für  $\theta \rightarrow \infty$  unmöglich. Dies spiegelt die Tatsache wieder, dass Unterdispersion in diesen Modellen bedeutet, dass das Auftreten von Nullen seltener ist als durch die Poissonverteilung zu erwarten wäre. Steigt der Erwartungswert  $E(X) = \theta$  an, so nimmt die Wahrscheinlichkeit von 0 und mit ihr der Spielraum für Unterdispersion ab.

### 7.2.3 Maximum-Likelihood-Schätzung in Null-Hürdenmodellen

**Satz 7.4 (Likelihoodfunktion des Null-Hürdenmodells)** Die Likelihoodfunktion des Null-Hürdenmodells (7.2) lässt sich in zur Hürde gehörende Terme  $Q_1^{(i)}$  und der Likelihoodfunktion des abgeschnittenen Ursprungsmodells entsprechende Terme  $Q_2^{(i)}$  aufspalten:

$$L = \prod_{i=1}^n Q_1^{(i)} Q_2^{(i)} \quad (7.8)$$

Dabei lauten diese Terme:

$$Q_1^{(i)} := P_1(0)^{\delta_i} [1 - P_1(0)]^{1-\delta_i} \quad (7.9)$$

$$Q_2^{(i)} := \left[ \frac{P_2(y_i)}{1 - P_2(0)} \right]^{1-\delta_i} \quad (7.10)$$

wobei  $\delta_i = 1 - \min\{y_i, 1\}$  gilt.

*Beweis:*

Der Vektor  $\mathbf{y} = (y_i)_{i=1, \dots, n}$  enthalte Realisierungen der unabhängigen Zufallsgrößen  $Y_i$ , die einer Verteilung mit Wahrscheinlichkeitsfunktion (7.2) folgen. Entsprechend der üblichen Definition lautet die Likelihoodfunktion:

$$L(\mathbf{y}) = \prod_{i=1}^n P(y_i) \quad (7.11)$$

Falls  $y_i = 0$  ist, entspricht der Term in (7.11) mit Index  $i$  dem Wert  $P_1(0)$ . Für positives  $y_i$  kommt der dazu korrespondierende Wert  $\varrho P_2(y_i)$  zum Zug. Dies kann mit der Auswahlvariable  $\delta_i$  beschrieben werden, die gleich 1 ist für  $y_i = 0$  und ansonsten gleich 0.

$$\begin{aligned} \prod_{i=1}^n P(y_i) &= \prod_{i=1}^n P_1(0)^{\delta_i} [\varrho P_2(y_i)]^{1-\delta_i} \\ &= \prod_{i=1}^n P_1(0)^{\delta_i} \left[ \frac{1 - P_1(0)}{1 - P_2(0)} P_2(y_i) \right]^{1-\delta_i} \\ &= \prod_{i=1}^n P_1(0)^{\delta_i} [1 - P_1(0)]^{1-\delta_i} \left[ \frac{P_2(y_i)}{1 - P_2(0)} \right]^{1-\delta_i} = \prod_{i=1}^n Q_1^{(i)} Q_2^{(i)} \end{aligned}$$

□

**Satz 7.5 (Maximum-Likelihood-Schätzung im Null-Hürdenmodell)** Die Maximierung der Likelihoodfunktion (7.8) ist äquivalent dazu die Funktionen

$$\sum_{i=1}^n \delta_i \ln[P_1(0)] + (1 - \delta_i) \ln[1 - P_1(0)] \quad (7.12)$$

$$\sum_{y_i > 0} \ln[P_2(y_i)] - \ln[1 - P_2(0)] \quad (7.13)$$

unabhängig voneinander zu maximieren.

*Beweis:*

Durch logarithmieren von (7.8) erhalten wir die Log-Likelihoodfunktion:

$$\begin{aligned} \mathcal{L}(\mathbf{y}) &= \ln L(\mathbf{y}) = \ln \prod_{i=1}^n Q_1^{(i)} Q_2^{(i)} \\ &= \sum_{i=1}^n \left( \ln Q_1^{(i)} + \ln Q_2^{(i)} \right) \\ &\stackrel{(7.9),(7.10)}{=} \sum_{i=1}^n \left( \delta_i \ln P_1(0) + (1 - \delta_i) \ln[1 - P_1(0)] + (1 - \delta_i) \ln \left[ \frac{P_2(y_i)}{1 - P_2(0)} \right] \right) \\ &= \sum_{i=1}^n \left( \delta_i \ln P_1(0) + (1 - \delta_i) \ln[1 - P_1(0)] \right) \\ &+ \sum_{y_i > 0} \left( \ln P_2(y_i) - \ln[1 - P_2(0)] \right) \end{aligned} \quad (7.14)$$

□

**Bemerkung:**

Die erste Summe in (7.14) gehört zu der binären Hürdenverteilung. Zur Maximierung dieses Ausdruckes werden alle Beobachtungen  $y_i$  herangezogen. Bei der Maximierung der zweiten Summe in (7.14), die mit der abgeschnittenen Ursprungsverteilung korrespondiert, wird nur die Untermenge der positiven Beobachtungen verwendet.

## 7.3 Zero-Hurdle-Poissonregression (ZHP-Regression)

### 7.3.1 Zero-Hurdle-Poissonverteilung (ZHP-Verteilung)

Bei kombinierten Poissonmodellen wie sie von Mullahy (1986) vorgeschlagen werden, bilden Poissonverteilungen sowohl die Hürden- als auch die Ursprungsverteilung. Solch ein Verteilungsmodell bezeichnen wir als **ZHP<sup>4</sup>-Verteilung**.

**Satz 7.6 (Wahrscheinlichkeitsfunktion der ZHP-Verteilung)** Die Wahrscheinlichkeitsfunktion des kombinierten Poissonmodells mit Hürdenverteilung  $Poi(\theta_1)$ , Ursprungsverteilung  $Poi(\theta_2)$  und Hürde Null lautet:

$$P_x(\theta_1, \theta_2) = P(X = x) = \begin{cases} e^{-\theta_1} & x = 0 \\ \varrho \frac{\theta_2^x}{x!} e^{-\theta_2} & x = 1, 2, \dots \end{cases} \quad (7.15)$$

mit  $\varrho = \frac{1 - e^{-\theta_1}}{1 - e^{-\theta_2}}$ .

---

<sup>4</sup>ZHP: zero hurdle Poisson

*Beweis:*

Hürden- und Ursprungsverteilung lauten:

$$P_1(x) = \frac{\theta_1^x}{x!} e^{-\theta_1}, \quad P_2(x) = \frac{\theta_2^x}{x!} e^{-\theta_2}$$

Für die Konstante  $\varrho$  gilt dann gemäß Definition 7.2:

$$\varrho = \frac{1 - P_1(0)}{1 - P_2(0)} = \frac{1 - \frac{\theta_1^0}{0!} e^{-\theta_1}}{1 - \frac{\theta_2^0}{0!} e^{-\theta_2}} = \frac{1 - e^{-\theta_1}}{1 - e^{-\theta_2}}$$

Durch Einsetzen von  $P_1(0)$  und  $P_2(x)$  in (7.2) erhalten wir (7.15)

□

**Satz 7.7 (Erwartungswert und Varianz bei der ZHP-Verteilung)** Die Zufallsvariable  $X$  folge einer kombinierten Verteilung mit Wahrscheinlichkeitsfunktion (7.15). Dann gilt:

$$E(X) = \varrho \theta_2 \quad (7.16)$$

$$\text{Var}(X) = E(X) + \frac{1 - \varrho}{\varrho} E(X)^2 \quad (7.17)$$

*Beweis:*

$$\begin{aligned} E(X) &\stackrel{(7.4)}{=} \varrho E_2(X) = \varrho \theta_2 \\ \text{Var}(X) &\stackrel{(7.5)}{=} \varrho [\text{Var}_2(X) + (1 - \varrho) E_2(X)^2] \\ &= \varrho [\theta_2 + (1 - \varrho) \theta_2^2] = \varrho \theta_2 + \frac{1 - \varrho}{\varrho} \varrho^2 \theta_2^2 \\ &\stackrel{(7.16)}{=} E(X) + \frac{1 - \varrho}{\varrho} E(X)^2 \end{aligned}$$

□

### 7.3.2 Spezifikation des ZHP-Regressionsmodells

(i) *Stochastische Komponente:*

Für die unabhängigen Beobachtungen  $(Y_i, \mathbf{x}_i)$ ,  $i = 1, \dots, n$  gelte

$$P_{y_i}(\gamma_i, \theta_i) = P(Y_i = y_i | \mathbf{x}_i) = \begin{cases} e^{-\gamma_i} & y_i = 0 \\ \varrho_i \frac{\theta_i^{y_i}}{y_i!} e^{-\theta_i} & y_i = 1, 2, \dots \end{cases}$$

$$\text{mit } \varrho_i = \frac{1 - e^{-\gamma_i}}{1 - e^{-\theta_i}}$$

(ii) *Systematische Komponente:*

Die Regressoren  $\mathbf{x}_i$  beeinflussen die Verteilung von  $Y_i | \mathbf{x}_i$  nur durch die linearen Prädiktoren

$$\begin{aligned} \nu_i &= \mathbf{x}_i^t \boldsymbol{\alpha} \\ \eta_i &= \mathbf{x}_i^t \boldsymbol{\beta} \end{aligned}$$

(iii) *Verknüpfungskomponente:*

Die Positivität der Intensitäten  $\gamma_i$  und  $\theta_i$  wird durch die Verknüpfungen

$$\nu_i = \ln \gamma_i - \ln t_i \Leftrightarrow \gamma_i = \exp\{\mathbf{x}_i^t \boldsymbol{\alpha} + \ln t_i\} \quad (7.18)$$

$$\eta_i = \ln \theta_i - \ln t_i \Leftrightarrow \theta_i = \exp\{\mathbf{x}_i^t \boldsymbol{\beta} + \ln t_i\} \quad (7.19)$$

gewährleistet.  $\ln t_i$  bezeichnet dabei den Offset.

### 7.3.3 Maximum-Likelihood-Schätzung im ZHP-Regressionsmodell

**Lemma 7.8** Für das ZHP-Regressionsmodell mit obiger Spezifikation lautet die Likelihoodfunktion

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{i=1}^n \exp\{-t_i \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\} \delta_i\} (1 - \exp\{-t_i \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\}\})^{1-\delta_i} \times \left[ \frac{(t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\})^{y_i} \exp\{-t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}\}}{y_i! (1 - \exp\{-t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}\})} \right]^{1-\delta_i} \quad (7.20)$$

*Beweis:*

Verwenden wir die Aussage  $L = \prod_{i=1}^n Q_1^{(i)} Q_2^{(i)}$  von Satz 7.4 mit den entsprechenden Ausdrücken (7.9) für  $Q_1^{(i)}$  und (7.10) für  $Q_2^{(i)}$  erhalten wir:

$$\begin{aligned} L(\boldsymbol{\alpha}, \boldsymbol{\beta}) &\stackrel{(7.8)}{=} \prod_{i=1}^n \left[ \frac{\gamma_i^0}{0!} e^{-\gamma_i} \right]^{\delta_i} \left[ 1 - \frac{\gamma_i^0}{0!} e^{-\gamma_i} \right]^{1-\delta_i} \left[ \frac{\theta_i^{y_i}}{y_i!} e^{-\theta_i} \right]^{1-\delta_i} \left[ 1 - \frac{\theta_i^0}{0!} e^{-\theta_i} \right]^{-(1-\delta_i)} \\ &= \prod_{i=1}^n e^{-\gamma_i \delta_i} (1 - e^{-\gamma_i})^{1-\delta_i} (1 - e^{-\theta_i})^{\delta_i - 1} \left[ \frac{\theta_i^{y_i}}{y_i!} e^{-\theta_i} \right]^{1-\delta_i} \\ &= \prod_{i=1}^n e^{-\gamma_i \delta_i} (1 - e^{-\gamma_i})^{1-\delta_i} \left[ \frac{\theta_i^{y_i} e^{-\theta_i}}{y_i! (1 - e^{-\theta_i})} \right]^{1-\delta_i} \\ &\stackrel{(7.18), (7.19)}{=} \prod_{i=1}^n \exp\{-\exp\{\mathbf{x}_i^t \boldsymbol{\alpha} + \ln t_i\} \delta_i\} (1 - \exp\{-\exp\{\mathbf{x}_i^t \boldsymbol{\alpha} + \ln t_i\}\})^{1-\delta_i} \\ &\quad \times \left[ \frac{(\exp\{\mathbf{x}_i^t \boldsymbol{\beta} + \ln t_i\})^{y_i} \exp\{-\exp\{\mathbf{x}_i^t \boldsymbol{\beta} + \ln t_i\}\}}{y_i! (1 - \exp\{-\exp\{\mathbf{x}_i^t \boldsymbol{\beta} + \ln t_i\}\})} \right]^{1-\delta_i} \end{aligned}$$

□

**Lemma 7.9 (Maximum-Likelihood-Schätzung)** Die Maximierung der Likelihoodfunktion (7.20) ist äquivalent dazu die Funktionen

$$\sum_{i=1}^n -\delta_i t_i \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\} + (1 - \delta_i) \ln (1 - \exp\{-t_i \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\}\}) \quad (7.21)$$

$$\sum_{y_i > 0} y_i (\mathbf{x}_i^t \boldsymbol{\beta} + \ln t_i) - t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} - \ln (1 - \exp\{-t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}\}) \quad (7.22)$$

unabhängig voneinander zu maximieren.

*Beweis:*

Wir verwenden für den Beweis die Aussagen von Satz 7.5 und setzen in die Funktionen (7.12) und (7.13) die entsprechenden Werte der Wahrscheinlichkeitsfunktion (7.15) der ZHP-Verteilung

ein:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}) &\stackrel{(7.14)}{=} \sum_{i=1}^n \left( \delta_i \ln P_1(0) + (1 - \delta_i) \ln[1 - P_1(0)] \right) + \sum_{y_i > 0} \left( \ln P_2(y_i) - \ln[1 - P_2(0)] \right) \\
&\stackrel{(7.15)}{=} \sum_{i=1}^n \left( \delta_i \ln e^{-\gamma_i} + (1 - \delta_i) \ln[1 - e^{-\gamma_i}] \right) \\
&\quad + \sum_{y_i > 0} \left( y_i \ln \theta_i - \ln(y_i!) - \theta_i - \ln[1 - e^{-\theta_i}] \right) \\
&\stackrel{(7.18), (7.19)}{=} \sum_{i=1}^n \left( -\delta_i t_i \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\} + (1 - \delta_i) \ln(1 - \exp\{-t_i \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\}\}) \right) \\
&\quad + \sum_{y_i > 0} \left( y_i (\mathbf{x}_i^t \boldsymbol{\beta} + \ln t_i) - \ln(y_i!) - t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} \right. \\
&\quad \quad \left. - \ln(1 - \exp\{-t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}\}) \right)
\end{aligned}$$

□

## 7.4 Verallgemeinerte Zero-Hurdle-Poissonregression (ZHGP-Regression)

Entsprechend dem ZHP-Modell das eine Abwandlung der Poissonverteilung darstellt, wollen wir nun eine abgewandelte verallgemeinerte Poissonverteilung konstruieren. Wir erhalten das **ZHGP<sup>5</sup>-Modell**.

### 7.4.1 Verallgemeinerte Zero-Hurdle-Poissonverteilung (ZHGP-Verteilung)

**Satz 7.10 (Wahrscheinlichkeitsfunktion der ZHGP-Verteilung)** *Die Wahrscheinlichkeitsfunktion des kombinierten verallgemeinerten Poissonmodells mit Hürdenverteilung  $GP(\theta_1, \lambda_1)$ , Ursprungsverteilung  $GP(\theta_2, \lambda_2)$  und Hürde Null lautet:*

$$P_x(\boldsymbol{\theta}, \lambda) = P(X = x) = \begin{cases} e^{-\theta_1} & x = 0 \\ \varrho \frac{\theta_2(\theta_2 + \lambda x)^{x-1}}{x!} e^{-(\theta_2 + \lambda x)} & x = 1, 2, \dots \end{cases} \quad (7.23)$$

mit  $\varrho = \frac{1 - e^{-\theta_1}}{1 - e^{-\theta_2}}$ ,  $\boldsymbol{\theta} = (\theta_1, \theta_2)^t$  und  $\lambda = \lambda_2$ .

Für die Zufallsvariable  $X$  schreiben wir  $X \sim ZHGP(\boldsymbol{\theta}, \lambda)$ .

*Beweis:*

Für die Hürden- und die Ursprungsverteilung gilt:

$$\begin{aligned}
P_1(x) &= \frac{\theta_1(\theta_1 + \lambda_1 x)^{x-1}}{x!} e^{-(\theta_1 + \lambda_1 x)} \\
P_2(x) &= \frac{\theta_2(\theta_2 + \lambda_2 x)^{x-1}}{x!} e^{-(\theta_2 + \lambda_2 x)}
\end{aligned}$$

<sup>5</sup>ZHGP: zero hurdle generalized Poisson



Den Parameter  $\varrho$  erhalten wir wie in Definition 7.2:

$$\begin{aligned}\varrho &= \frac{1 - P_1(0)}{1 - P_2(0)} = \frac{1 - \frac{\theta_1 \theta_1^{-1}}{0!} e^{-\theta_1}}{1 - \frac{\theta_2 \theta_2^{-1}}{0!} e^{-\theta_2}} \\ &= \frac{1 - e^{-\theta_1}}{1 - e^{-\theta_2}}\end{aligned}$$

Setzen wir dieses Resultat sowie  $P_1(0)$  und  $P_2(x)$  in (7.2) ein, so erhalten wir (7.23).

□

### Bemerkungen:

- Von den Parametern der Hürdenverteilung übernimmt die ZHGP-Verteilung nur die Intensität  $\theta_1$ . Der Dispersionsparameter  $\lambda_1$  hat hingegen keinen Einfluss auf die ZHGP-Verteilung. Deren Dispersionsparameter  $\lambda$  entspricht dem Parameter  $\lambda_2$  der Ursprungsverteilung.
- Übersteigt die Intensität der Ursprungsverteilung die Intensität der Hürdenverteilung ( $\theta_1 < \theta_2$ ), so gilt  $\varrho < 1$ . Im Vergleich zum Ursprungsmodell liegt daher Nullenüberschuss vor.

### Satz 7.11 (Erwartungswert und Varianz bei ZHGP-Verteilung)

$$E(X) = \varrho \frac{\theta_2}{1 - \lambda} \quad (7.24)$$

$$Var(X) = \varphi [\varphi + (1 - \varrho)\theta_2] E(X) \quad (7.25)$$

mit dem Dispersionsparameter  $\varphi = \frac{1}{1 - \lambda}$ .

*Beweis:*

$$\begin{aligned}E(X) &\stackrel{(7.4)}{=} \varrho E_2(X) \stackrel{(1.4)}{=} \varrho \frac{\theta_2}{1 - \lambda_2} \\ Var(X) &\stackrel{(7.5)}{=} \varrho (Var_2(X) + (1 - \varrho) E_2(X)^2) \\ &\stackrel{(1.36)}{=} \varrho \left( \frac{\theta_2}{(1 - \lambda_2)^3} + (1 - \varrho) \frac{\theta_2^2}{(1 - \theta_2)^2} \right) \\ &= \frac{\varrho \theta_2}{(1 - \lambda_2)^2} \left( \frac{1}{1 - \lambda_2} + (1 - \varrho) \theta_2 \right) \\ &= \frac{1}{1 - \lambda_2} \left( \frac{1}{1 - \lambda_2} + (1 - \varrho) \theta_2 \right) \varrho \frac{\theta_2}{1 - \lambda_2} \\ &\stackrel{(7.24)}{=} \varphi [\varphi + (1 - \varrho) \theta_2] E(X)\end{aligned}$$

□

### Bemerkung:

Über den Dispersionsparameter  $\varphi$  der Ursprungsverteilung hinaus besitzen ZHGP-Modelle eine weitere Quelle für Überdispersion in Form des Ausdrucks  $(1 - \varrho)\theta_2$ . Dieser ist charakteristisch für ein Hürdenmodell.

### 7.4.2 Vergleich von verallgemeinerter Poisson- und ZHGP-Verteilung

In Abbildung 7.1 sind Wahrscheinlichkeits- und Verteilungsfunktion für die Verteilungen  $GP(4, 0)$ ,  $ZHGP(3, 4, 0)$ ,  $ZHGP(2, 4, 0)$  und  $ZHGP(1, 4, 0)$  in der genannten Reihenfolge<sup>6</sup> abgetragen. In Abbildung 7.2 befinden sich entsprechende Abbildungen für  $GP(4, 0.5)$ ,  $ZHGP(3, 4, 0.5)$ ,  $ZHGP(2, 4, 0.5)$  und  $ZHGP(1, 4, 0.5)$ .

Wie bei den ZIGP-Verteilungen (vgl. Abschnitt 6.3.2) hat der Dispersionsparameter  $\lambda$  keinen Einfluss auf  $P(X = 0)$ , während die Wahrscheinlichkeiten für positive Ereignisse auf steigendes  $\lambda$  entsprechend den Wahrscheinlichkeiten bei der verallgemeinerten Poissonverteilung reagieren. Bei sinkenden Intensitäten  $\theta_1$  steigt die Wahrscheinlichkeit des Ereignisses Null. Insbesondere ist diese für  $\theta_1 < \theta_2$  höher als bei der  $GP(\theta_2, \lambda)$ -Verteilung. In diesem Fall sind die Wahrscheinlichkeiten positiver Ereignisse bei ZHGP-Verteilung geringer als bei der verallgemeinerten Poissonverteilung. Die Kurve für positive Ereignisse nimmt bei ZHGP-Verteilung im wesentlichen den selben Verlauf wie bei der verallgemeinerten Poissonverteilung, da die entsprechenden Einzelwahrscheinlichkeiten der Ursprungsverteilung lediglich mit dem Faktor  $\varrho < 1$  multipliziert werden.

Für den relativen Fehler der Verteilungsfunktionen und die stochastische Dominanz der ZHGP-Verteilungen mit  $\theta_1 < \theta_2$  über die verallgemeinerte Poissonverteilung gilt entsprechendes zu den Aussagen über die ZIGP-Verteilung aus Abschnitt 6.3.2.

### 7.4.3 Spezifikation des ZHGP-Regressionsmodells

(i) *Stochastische Komponente:*

Für die unabhängigen Beobachtungen  $(Y_i, \mathbf{x}_i)$ ,  $i = 1, \dots, n$  gelte

$$P_{y_i}(\gamma_i, \theta_i, \lambda) = P(Y_i = y_i | \mathbf{x}_i) = \begin{cases} e^{-\gamma_i} & y_i = 0 \\ \varrho_i \frac{\theta_i(\theta_i + \lambda y_i)^{y_i - 1}}{y_i!} e^{-(\theta_i + \lambda y_i)} & y_i = 1, 2, \dots \end{cases}$$

mit  $\varrho_i = \frac{1 - e^{-\gamma_i}}{1 - e^{-\theta_i}}$ .

(ii) *Systematische Komponente:*

Die Regressoren  $\mathbf{x}_i$  beeinflussen die Verteilung von  $Y_i | \mathbf{x}_i$  nur durch die linearen Prädiktoren

$$\begin{aligned} \nu_i &= \mathbf{x}_i^t \boldsymbol{\alpha} \\ \eta_i &= \mathbf{x}_i^t \boldsymbol{\beta} \end{aligned}$$

(iii) *Verknüpfungskomponente:*

Die Positivität der Intensitäten  $\gamma_i$  und  $\theta_i$  wird durch die Verknüpfungen

$$\nu_i = \ln \gamma_i - \ln t_i \Leftrightarrow \gamma_i = \exp\{\mathbf{x}_i^t \boldsymbol{\alpha} + \ln t_i\} \quad (7.26)$$

$$\eta_i = \ln \theta_i - \ln t_i \Leftrightarrow \theta_i = \exp\{\mathbf{x}_i^t \boldsymbol{\beta} + \ln t_i\} \quad (7.27)$$

gewährleistet.  $\ln t_i$  bezeichnet dabei den Offset.

<sup>6</sup>Die ZHGP-Größen sind äquivalent zu der Vorgehensweise bei Abbildung 6.1 versetzt dargestellt

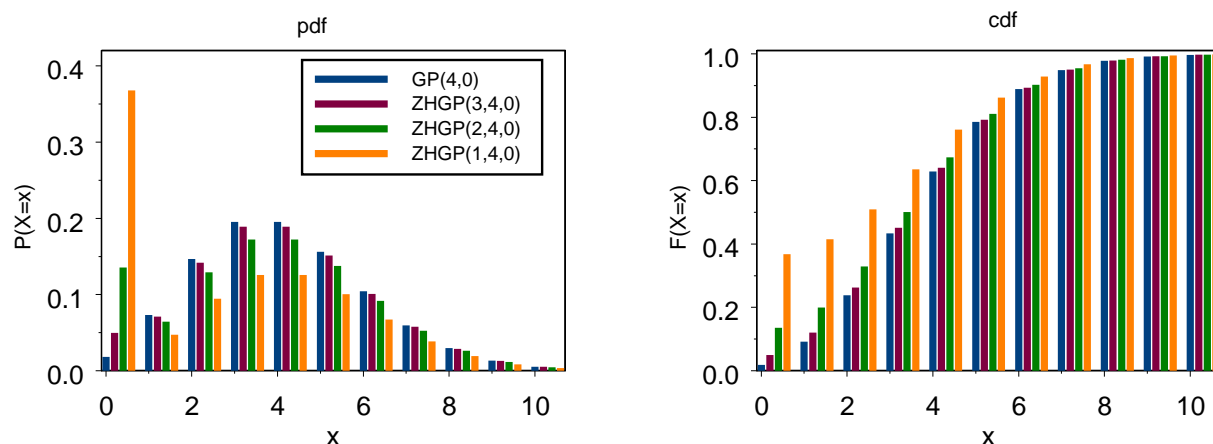


Abbildung 7.1: Wahrscheinlichkeitsfunktionen (links) und Verteilungsfunktionen (rechts) von ZHGP-Verteilungen mit  $\lambda = 0$ ,  $\theta_1 = 4$  und  $\theta_2 = 4, 3, 2$  und  $1$

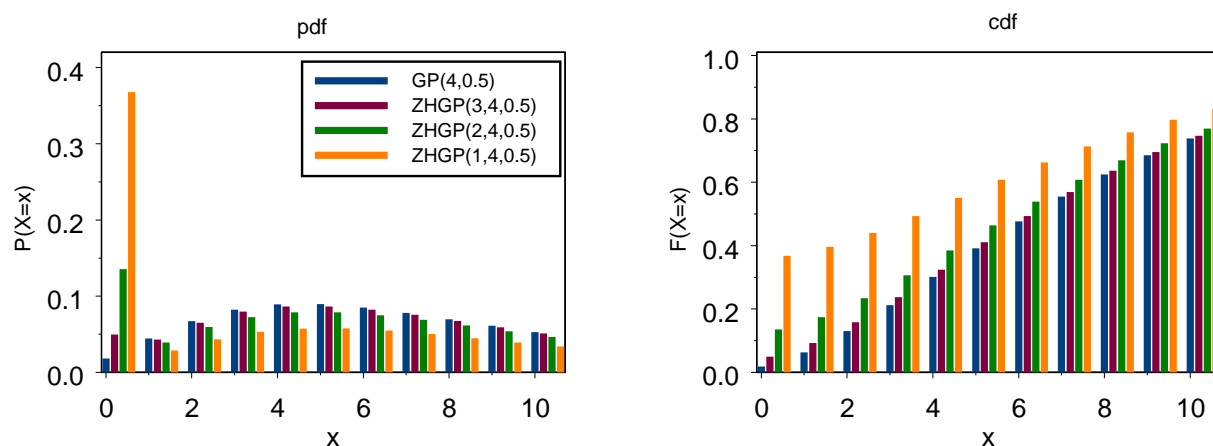


Abbildung 7.2: Wahrscheinlichkeitsfunktionen (links) und Verteilungsfunktionen (rechts) von ZHGP-Verteilungen mit  $\lambda = 0.5$ ,  $\theta_1 = 4$  und  $\theta_2 = 4, 3, 2$  und  $1$

**Bemerkung:**

Verwenden wir Spezifikation 7.4.3, so lautet die Erwartungswertfunktion:

$$\begin{aligned}
 E(Y_i | \mathbf{x}_i) &\stackrel{(7.24)}{=} \frac{1 - e^{-\gamma_i}}{1 - e^{-\theta_i}} \frac{\theta_i}{1 - \lambda} \\
 &\stackrel{(7.26), (7.27)}{=} \frac{1 - \exp\{-t_i \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\}\}}{1 - \exp\{-t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}\}} \frac{t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}}{1 - \lambda}
 \end{aligned} \tag{7.28}$$

#### 7.4.4 Maximum-Likelihood-Schätzung im ZHGP-Reggressionsmodell

**Lemma 7.12** Für das ZHGP-Reggressionsmodell mit obiger Spezifikation lautet die Likelihoodfunktion

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda) = \prod_{i=1}^n \exp\{-t_i \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\} \delta_i\} [1 - \exp\{-t_i \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\}\}]^{1-\delta_i} \\ \times \left[ \frac{t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} (t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} + y_i \lambda)^{y_i-1}}{y_i! (1 - \exp\{-t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}\})} \exp\{-t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} + y_i \lambda\} \right]^{1-\delta_i} \quad (7.29)$$

*Beweis:*

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda) \stackrel{(7.8)}{=} \prod_{i=1}^n [e^{-\gamma_i}]^{\delta_i} [1 - e^{-\gamma_i}]^{1-\delta_i} \\ \times \left[ \frac{\theta_i (\theta_i + y_i \lambda)^{y_i-1}}{y_i!} e^{-(\theta_i + y_i \lambda)} \right]^{1-\delta_i} [1 - e^{-\theta_i}]^{-(1-\delta_i)} \\ = \prod_{i=1}^n e^{-\gamma_i \delta_i} [1 - e^{-\gamma_i}]^{1-\delta_i} \left[ \frac{\theta_i (\theta_i + y_i \lambda)^{y_i-1}}{y_i! (1 - e^{-\theta_i})} e^{-(\theta_i + y_i \lambda)} \right]^{1-\delta_i} \\ \stackrel{(7.26), (7.27)}{=} \prod_{i=1}^n \exp\{-t_i \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\} \delta_i\} [1 - \exp\{-t_i \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\}\}]^{1-\delta_i} \\ \times \left[ \frac{t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} (t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} + y_i \lambda)^{y_i-1}}{y_i! (1 - \exp\{-t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}\})} \exp\{y_i \lambda - t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}\} \right]^{1-\delta_i}$$

□

**Lemma 7.13 (Maximum-Likelihood-Schätzung)** Die Maximierung der Likelihoodfunktion (7.29) ist äquivalent dazu die Funktionen

$$\sum_{i=1}^n -\delta_i t_i \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\} + (1 - \delta_i) \ln (1 - \exp\{-t_i \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\}\}) \quad (7.30)$$

$$\sum_{y_i > 0} \mathbf{x}_i^t \boldsymbol{\beta} + \ln t_i + (y_i - 1) \ln (t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} + \lambda y_i) - \ln(y_i!) \\ - t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} - \lambda y_i - \ln (1 - \exp\{-t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}\}) \quad (7.31)$$

unabhängig voneinander zu maximieren.

*Beweis:*

Für den Beweis verwenden wir erneut die Aussagen von Satz 7.5 und setzen in die Funktionen (7.12) und (7.13) die entsprechenden Werte der Wahrscheinlichkeitsfunktion (7.23) der ZHP-

Verteilung ein:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda) &\stackrel{(7.14)}{=} \sum_{i=1}^n \delta_i \ln P_1(0) + (1 - \delta_i) \ln[1 - P_1(0)] + \sum_{y_i > 0} \ln P_2(y_i) - \ln[1 - P_2(0)] \\
&\stackrel{(7.23)}{=} \sum_{i=1}^n \delta_i \ln e^{-\gamma_i} + (1 - \delta_i) \ln(1 - e^{-\gamma_i}) \\
&\quad + \sum_{y_i > 0} \ln \left( \frac{\theta_i (\theta_i + \lambda y_i)^{y_i - 1}}{y_i!} e^{-(\theta_i + \lambda y_i)} \right) - \ln(1 - e^{-\theta_i}) \\
&= \sum_{i=1}^n \delta_i \ln e^{-\gamma_i} + (1 - \delta_i) \ln(1 - e^{-\gamma_i}) \\
&\quad + \sum_{y_i > 0} \ln \theta_i + (y_i - 1) \ln(\theta_i + \lambda y_i) - \ln(y_i!) - \theta_i - \lambda y_i - \ln(1 - e^{-\theta_i}) \\
&\stackrel{(7.26), (7.27)}{=} \sum_{i=1}^n -\delta_i t_i \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\} + (1 - \delta_i) \ln(1 - \exp\{-t_i \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\}\}) \\
&\quad + \sum_{y_i > 0} \mathbf{x}_i^t \boldsymbol{\beta} + \ln t_i + (y_i - 1) \ln(t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} + \lambda y_i) - \ln(y_i!) \\
&\quad - t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} - \lambda y_i - \ln(1 - \exp\{-t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}\}) \tag{7.32}
\end{aligned}$$

□

#### 7.4.5 Messung der Anpassungsqualität des ZHGP-Regressionsmodells

**Satz 7.14 (Akaike-Information für das ZHGP-Regressionsmodell)** Die Akaike-Information des ZHGP-Regressionsmodells lautet:

$$\begin{aligned}
AIC &= 2k - \sum_{i=1}^n \langle -\delta_i t_i \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\} + (1 - \delta_i) \ln(1 - \exp\{-t_i \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\}\}) \rangle \\
&\quad - \sum_{y_i > 0} \langle \mathbf{x}_i^t \boldsymbol{\beta} + \ln t_i + (y_i - 1) \ln(t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} + \lambda y_i) - \ln(y_i!) \\
&\quad - t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} - \lambda y_i - \ln(1 - \exp\{-t_i \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}\}) \rangle
\end{aligned}$$

*Beweis:*

Einsetzen von (7.32) in (3.37)

□

Mit Hilfe der Akaike-Information können wir ZHGP-Modelle mit unterschiedlichen Kovariablen vergleichen. Darüber hinaus lässt sich die Anpassungsqualität eines ZHGP-Regressionsmodells über AIC auch mit der Anpassungsqualität eines entsprechenden GP-Regressionsmodells oder eines gewöhnlichen Poisson-Regressionsmodells vergleichen. Dies ist möglich, da das ZHGP in das verallgemeinerte Poissonmodell und dieses wiederum in das gewöhnliche Poissonmodell eingebettet ist:

$$ZHGP(\theta_1, \theta_2, \lambda) \stackrel{\theta = \theta_1 = \theta_2}{=} GP(\theta, \lambda) \stackrel{\lambda \equiv 0}{=} Poi(\theta)$$

⇒ Poissonregression  $\subset$  verallgemeinerte Poissonregression  $\subset$  ZHGP-Regression



## Kapitel 8

# Simulationsstudie zu den Maximum-Likelihood-Schätzern

Im Folgenden wollen wir die Maximum-Likelihood-Schätzer betrachten, die wir für das

- GP-Regressionsmodell (vgl. Abschnitt 5.1)
- ZIGP- $\mu$ -Regressionsmodell (vgl. Definition 6.10)
- ZIGP- $\mu\omega$ -Regressionsmodell (vgl. Definition 6.13)
- ZIGP- $\omega$ -Regressionsmodell (vgl. Definition 6.15)
- ZHGP-Regressionsmodell (vgl. Abschnitt 7.4.3)

aus den vorangegangenen Kapiteln erhalten. Dabei wollen wir untersuchen, wie sich der

- Stichprobenumfang  $n$
- Dispersionsparameter  $\varphi$
- Auswahlparameter  $\omega$  / Parameterbereich  $[\omega_{\min}, \omega_{\max}]$
- Parameterbereich  $[\mu_{\min}, \mu_{\max}]$

auf die Güte der Schätzer auswirkt. Um diese bewerten zu können, werden wir zunächst zwei grundlegende Eigenschaften von Schätzern herleiten.

### 8.1 Eigenschaften von Schätzern

Möchten wir die Güte eines Schätzers messen, so erwarten wir zunächst, dass dieser tendenziell korrekte Werte liefert. Er soll weder systematisch unter- noch überschätzen. Systematische Unter- oder Überschätzung eines Parameterschätzers  $\hat{\theta}$  wird durch dessen Verzerrung erfasst.

**Definition 8.1 (Verzerrung und Erwartungstreue eines Schätzers)** *Betrachten wir einen Parameter  $\theta$ , so gibt die Größe*

$$b(\theta, \hat{\theta}) = E(\hat{\theta}) - \theta \quad (8.1)$$

die Verzerrung<sup>1</sup> des Schätzers  $\hat{\theta}$  an. Stimmen Erwartungswert des Schätzers und der Wert des zu schätzenden Parameters überein, d.h.

$$b(\theta, \hat{\theta}) = 0 \quad \Leftrightarrow \quad E(\hat{\theta}) = \theta$$

so bezeichnet man den Schätzer  $\hat{\theta}$  als unverzerrt oder erwartungstreu.

Für die Genauigkeit eines Schätzers ist jedoch nicht nur der Schätzwert selbst ausschlaggebend. Vielmehr muss auch die Varianz des Schätzers berücksichtigt werden, damit die Schätzwerte nicht zu stark um  $E(\hat{\theta})$  schwanken. Von einem guten Schätzer erwarten wir, dass die doppelte Standardabweichung den Betrag seines Erwartungswertes nicht übersteigt. Folglich sollte

$$\frac{|E(\hat{\theta})|}{\sqrt{\text{Var}(\hat{\theta})}} > 2 \quad (8.2)$$

erfüllt sein. Ein Maß für die Schätzgüte, welches sowohl die Genauigkeit als auch die Präzision des Schätzers mit einbezieht ist die mittlere quadratische Abweichung.

**Definition 8.2 (Mittlere quadratische Abweichung)** *Die mittlere quadratische Abweichung<sup>2</sup> ist bestimmt durch*

$$MSE = R(\theta, \hat{\theta}) = E([\hat{\theta} - \theta]^2) \quad (8.3)$$

Die mittlere quadratische Abweichung gibt damit an, welche Abweichung zwischen dem Schätzwert  $\hat{\theta}$  und dem wahren Wert  $\theta$  zu erwarten ist. Formen wir den Ausdruck (8.3) um, so erkennen wir, dass sich die mittlere quadratische Abweichung als Summe der Varianz und der quadrierten Verzerrung von  $\hat{\theta}$  darstellen lässt:

$$\begin{aligned} R(\hat{\theta}, \theta) &\stackrel{(8.3)}{=} E([\hat{\theta} - \theta]^2) = E([\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta]^2) \\ &= E([\hat{\theta} - E(\hat{\theta})]^2) + 2E([\hat{\theta} - E(\hat{\theta})][E(\hat{\theta}) - \theta]) + E([E(\hat{\theta}) - \theta]^2) \\ &= E([\hat{\theta} - E(\hat{\theta})]^2) + 2E(\hat{\theta}E(\hat{\theta}) - \hat{\theta}\theta - E(\hat{\theta})^2 + E(\hat{\theta})\theta) + E([E(\hat{\theta}) - \theta]^2) \\ &= E([\hat{\theta} - E(\hat{\theta})]^2) + 2[E(\hat{\theta})E(\hat{\theta}) - \theta E(\hat{\theta}) - E(\hat{\theta})^2 + \theta E(\hat{\theta})] + E([E(\hat{\theta}) - \theta]^2) \\ &= E([\hat{\theta} - E(\hat{\theta})]^2) + E([E(\hat{\theta}) - \theta]^2) \\ &\stackrel{(8.1)}{=} \text{Var}(\hat{\theta}) + b^2(\hat{\theta}, \theta) \end{aligned} \quad (8.4)$$

Aus der Darstellung (8.4) für die mittlere quadratische Abweichung erkennen wir, dass die Varianz eines Schätzers als Maß für dessen Güte nur für unverzerrte Schätzer sinnvoll ist.

Wollen wir die Güte der Maximum-Likelihood-Schätzer aus den vorangegangenen Kapiteln messen, scheitert die Anwendung der Maße (8.1) und (8.3) daran, dass wir die Verteilung der betrachteten Schätzer nicht kennen. Wir können deren Erwartungswert und deren Varianz nicht explizit bestimmen. Aus diesem Grunde werden wir versuchen durch die wiederholte Schätzung der Regressionsparameter für  $n$  unabhängig identisch verteilte Responsedaten das asymptotische

<sup>1</sup>engl.: Bias

<sup>2</sup>engl.: mean squared error (MSE)



Verhalten der Schätzer abzuleiten.

Ein asymptotisch erwartungstreu und konsistenter<sup>3</sup> Schätzer des Erwartungswertes ist das arithmetische Mittel (siehe z.B. Fahrmeir et al., 2001, S.368). Die Verzerrung lautet daher:

$$\hat{b}(\theta) = \bar{\theta} - \theta = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_i - \theta \quad (8.5)$$

Für die Verzerrung  $b$  des Schätzers  $\hat{\theta}$  bezeichne  $p_\alpha$ ,  $0 < \alpha < 1$  das zugehörige, auf den unabhängig identisch verteilten Größen  $b_i = \hat{\theta}_i - \theta$ ,  $i = 1, \dots, n$  basierende, empirische  $\alpha$ -Quantil (Fahrmeir et al., 2001, S.62). Damit ist durch

$$KI_{n,1-\alpha}^b = [p_{\frac{\alpha}{2}}, p_{1-\frac{\alpha}{2}}] \quad (8.6)$$

ein empirisches zweiseitiges  $100(1 - \alpha)\%$ -Konfidenzintervall<sup>4</sup> für  $b$  gegeben. Vom betrachteten Schätzer  $\hat{\theta}$  erwarten wir, dass der Wert Null zumindest im empirischen 95%-Konfidenzintervall der Verzerrung  $b$  enthalten ist:

$$0 \in KI_{n,0.95}^b = [p_{0.025}, p_{0.975}] \quad (8.7)$$

Die Varianz der Verzerrung bestimmen wir wie folgt:

$$\begin{aligned} \text{Var}(\hat{b}) &= \text{Var}(\bar{\theta} - \theta) = \text{Var}(\bar{\theta}) \\ &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n \hat{\theta}_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(\hat{\theta}_i) \end{aligned} \quad (8.8)$$

Die Schätzer  $\hat{\theta}_i$  sind auf Grund der Vorgehensweise bei der Responseerzeugung unabhängig identisch verteilt. Der Ausdruck (8.8) lässt sich deshalb vereinfachen:

$$\frac{1}{n^2} \sum_{i=1}^n \text{Var}(\hat{\theta}_i) = \frac{1}{n} \text{Var}(\hat{\theta}) \quad (8.9)$$

Anstatt  $\text{Var}(\hat{\theta})$  verwenden wir die erwartungstreu und konsistente Stichprobenvarianz  $s^2$  (siehe z.B. Fahrmeir et al., 2001, S.369). Aus (8.9) folgt daher:

$$\text{Var}(\hat{b}) = \frac{1}{n} s^2 = \frac{1}{n} \left[ \frac{1}{n-1} \sum_{i=1}^n (\hat{\theta}_i - \bar{\theta})^2 \right] \quad (8.10)$$

Bei Verwendung der Stichprobenvarianz  $s^2$  lautet die mittlere quadratische Abweichung:

$$\begin{aligned} \hat{R}(\theta) &= s^2 + \hat{b}^2(\theta) \\ &= \frac{1}{n-1} \sum_{i=1}^n (\hat{\theta}_i - \bar{\theta})^2 + \hat{b}^2(\theta) \end{aligned} \quad (8.11)$$

<sup>3</sup>Eine Schätzfolge  $\hat{\theta}_1, \dots, \hat{\theta}_n$  heißt (im quadratischen Mittel) konsistent, wenn  $\text{MSE} \rightarrow 0$  für  $n \rightarrow \infty$  gilt.

<sup>4</sup>Gelegentlich wird das empirische Konfidenzintervall auch als Realisierung des entsprechenden (theoretischen) Konfidenzintervalls bezeichnet (Kredler, 1998, S.183).

Berechnen wir auch für die mittlere quadratische Abweichung die Varianz:

$$\begin{aligned}
\text{Var}(\hat{R}) &\stackrel{(8.3)}{=} \text{Var}[E(\hat{\theta} - \theta)^2] = \text{Var}\left[\frac{1}{n} \sum_{i=1}^n (\hat{\theta}_i - \theta)^2\right] \\
&= \frac{1}{n^2} \sum_{i=1}^n \text{Var}[(\hat{\theta}_i - \theta)^2] = \frac{1}{n^2} n \text{Var}[(\hat{\theta} - \theta)^2] \\
&= \frac{1}{n} \text{Var}[\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2] = \frac{1}{n} \text{Var}[\hat{\theta}^2 - 2\hat{\theta}\theta] \\
&= \frac{1}{n} \left\{ E[(\hat{\theta}^2 - 2\hat{\theta}\theta)^2] - [E(\hat{\theta}^2 - 2\hat{\theta}\theta)]^2 \right\} \\
&= \frac{1}{n} \left\{ E(\hat{\theta}^4) - 4E(\hat{\theta}^3)\theta + 4E(\hat{\theta}^2)\theta^2 \right. \\
&\quad \left. - E(\hat{\theta}^2)^2 + 4E(\hat{\theta}^2)E(\hat{\theta})\theta - 4E(\hat{\theta})^2\theta^2 \right\} \tag{8.12}
\end{aligned}$$

Die Momente  $\mu_k = E(\hat{\theta}^k)$  in Gleichung (8.12) ersetzen wir durch deren Schätzer  $m_k = \frac{1}{n} \sum_{i=1}^n \theta_i^k$  (Råde and Westergren, 1997, S.471). Für die Varianz der mittleren quadratischen Abweichung gilt daher:

$$\text{Var}(\hat{R}) = \frac{1}{n} (m_4 - 4\theta m_3 + 4\theta^2 m_2 - m_2^2 + 4\theta m_1 m_2 - 4\theta^2 m_2) \tag{8.13}$$

Da im Allgemeinen  $\hat{b}(\theta) \neq 0$  gilt, wollen wir als weiteres Kriterium für die Güte der Schätzer den relativen Fehler

$$\text{Err}(\theta) = \frac{E(\hat{\theta}) - \theta}{\theta} \tag{8.14}$$

beziehungsweise den geschätzten relativen Fehler

$$\widehat{\text{Err}}(\theta) = \frac{\bar{\theta} - \theta}{\theta} \tag{8.15}$$

betrachten. Dieser berücksichtigt, wie stark die Abweichung im Verhältnis zum wahren Wert ist. Die Varianz lautet hier:

$$\begin{aligned}
\text{Var}(\widehat{\text{Err}}) &= \text{Var}\left(\frac{\frac{1}{n} \sum_{i=1}^n \hat{\theta}_i - \theta}{\theta}\right) = \frac{1}{\theta^2} \text{Var}\left(\frac{1}{n} \sum_{i=1}^n \hat{\theta}_i\right) \\
&= \frac{1}{\theta^2} \frac{1}{n} \text{Var}(\hat{\theta}) = \frac{1}{n\theta^2} \left[ \frac{1}{n-1} \sum_{i=1}^n (\hat{\theta}_i - \theta)^2 \right] \tag{8.16}
\end{aligned}$$

Damit wir bei wiederholter Schätzung eines Parameters  $\theta \neq 0$  für  $n$  unabhängig identisch verteilte Responsedaten nicht zufällige Schwankungen um Null beobachten, wollen wir diesen Wert aus dem empirischen 95%-Konfidenzintervall der ermittelten Schätzer ausschließen.  $q_\alpha$  bezeichne das empirische  $\alpha$ -Quantil der Schätzer  $\hat{\theta}_i$ ,  $i = 1, \dots, n$ . Wir betrachten den Schätzer  $\hat{\theta}$  als signifikant verschieden von Null, falls

$$0 \notin KI_{n,0.95}^\theta = [q_{0.025}, q_{0.975}] \tag{8.17}$$

erfüllt ist.

Im folgenden seien die Kriterien, welche wir bei der Bewertung der Schätzgüte betrachten wollen kurz zusammengefasst:

1. Signifikanter Unterschied des Schätzers von Null (falls  $\theta \neq 0$ )
2. Verzerrung
  - a) Bias  $\approx 0$
  - b)  $0 \in KI_{n,0.95}^b$
  - c) Err  $\approx 0$
3. Präzision
  - a) MSE  $\approx 0$
  - b)  $\frac{|\hat{\theta}|}{s} > 2$

## 8.2 Verallgemeinerte Poissonregression

### 8.2.1 Datensimulation

Mit der S-Plus Funktion `response.gpois` (Anhang C.2.1) erzeugen wir für unterschiedlichen Stichprobenumfang  $n$  sowie verschiedene Parameterwerte  $\varphi$  und  $\beta$  100 unabhängig identisch verallgemeinert poissonverteilte Responsevektoren  $\mathbf{Y}$ . Als Modellmatrix dient uns dabei

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \quad (8.18)$$

mit den Elementen  $x_i := -1 + \frac{2}{n-1} * (i - 1)$ ,  $i = 1, \dots, n$ . Der lineare Prädiktor (5.1) der verallgemeinerten Poissonregression ( $\eta_i = \mathbf{x}_i^t \beta$ ) enthält somit einen Intercept- und einen Regressions-term. Mit der üblichen log-linearen Verknüpfung (5.2) folgt für die Parameter  $\mu_i = \exp\{\mathbf{x}_i^t \beta\}$ ,  $i = 1, \dots, n$ :

$$\begin{aligned} \mu_i \in M &:= [\mu_{\min}, \mu_{\max}] \\ &= \left[ \min_{1 \leq i \leq n} \{e^{\beta_0 + \beta_1 x_i}\}, \max_{1 \leq i \leq n} \{e^{\beta_0 + \beta_1 x_i}\} \right] \end{aligned}$$

Um die Güte der Schätzer bei unterschiedlichem Parameterbereich  $M$  vergleichen zu können, wählen wir:

$$\beta' = \begin{pmatrix} -1 \\ 2 \end{pmatrix} \quad \text{und} \quad \beta'' = \begin{pmatrix} -1 \\ 3 \end{pmatrix}$$

Wir erhalten die Bereiche

$$\begin{aligned} M' &= [e^{-3}, e^1] \approx [0.05, 2.72] \\ M'' &= [e^{-4}, e^2] \approx [0.02, 7.39] \end{aligned}$$

$[\mu_{\min}, \mu_{\max}]$	$\varphi$	Parameter	Schätzer					
			Bias		MSE		relativer Fehler	
[0.05, 2.72]	1.25	$\varphi = 1.25$	-0.03	(0.02)	0.06	(0.02)	-0.02	(0.02)
		$\beta_0 = -1$	-0.28	(0.09)	0.96	(7.75)	0.28	(0.09)
		$\beta_1 = 2$	0.19	(0.11)	1.29	(13.77)	0.09	(0.06)
	5.00	$\varphi = 5.00$	0.83	(0.71)	50.70	(12388.36)	0.17	(0.14)
		$\beta_0 = -1$	-4.95	(0.93)	111.63	(92105.43)	4.95	(0.93)
		$\beta_1 = 2$	2.16	(0.48)	27.95	(4927.41)	1.08	(0.24)
[0.02, 7.39]	1.25	$\varphi = 1.25$	-0.09	(0.02)	0.06	(0.02)	-0.07	(0.02)
		$\beta_0 = -1$	-0.27	(0.06)	0.44	(0.71)	0.27	(0.06)
		$\beta_1 = 3$	0.27	(0.07)	0.56	(1.47)	0.09	(0.02)
	5.00	$\varphi = 5.00$	-0.94	(0.45)	20.77	(3446.96)	0.19	(0.09)
		$\beta_0 = -1$	-1.88	(0.35)	16.00	(2009.87)	1.88	(0.35)
		$\beta_1 = 3$	1.55	(0.37)	15.90	(1852.93)	0.52	(0.12)

Tabelle 8.1: Bias, MSE und relativer Fehler der Maximum-Likelihood-Schätzer im verallgemeinerten Poisson-Regressionsmodell bei Stichprobenumfang  $n=20$

$[\mu_{\min}, \mu_{\max}]$	$\varphi$	Parameter	Schätzer					
			Bias		MSE		relativer Fehler	
[0.05, 2.72]	1.25	$\varphi = 1.25$	-0.04	(0.01)	0.02	(0.01)	-0.03	(0.01)
		$\beta_0 = -1$	-0.04	(0.03)	0.07	(0.02)	0.04	(0.03)
		$\beta_1 = 2$	0.07	(0.04)	0.16	(0.04)	0.03	(0.02)
	5.00	$\varphi = 5.00$	-0.57	(0.23)	5.83	(87.56)	-0.11	(0.05)
		$\beta_0 = -1$	-0.45	(0.08)	0.85	(1.93)	0.45	(0.08)
		$\beta_1 = 2$	0.19	(0.08)	0.68	(1.72)	0.10	(0.04)
[0.02, 7.39]	1.25	$\varphi = 1.25$	-0.04	(0.01)	0.01	(0.004)	-0.03	(0.01)
		$\beta_0 = -1$	-0.04	(0.03)	0.09	(0.03)	0.04	(0.03)
		$\beta_1 = 3$	0.03	(0.04)	0.16	(0.10)	0.01	(0.01)
	5.00	$\varphi = 5.00$	-0.06	(0.20)	3.88	(56.78)	-0.01	(0.04)
		$\beta_0 = -1$	-0.19	(0.07)	0.46	(0.55)	0.19	(0.07)
		$\beta_1 = 3$	0.14	(0.08)	0.63	(1.05)	0.05	(0.03)

Tabelle 8.2: Bias, MSE und relativer Fehler der Maximum-Likelihood-Schätzer im verallgemeinerten Poisson-Regressionsmodell bei Stichprobenumfang  $n=100$

Ebenso wollen wir den Dispersionsgrad variieren. Dazu wählen wir:

$$\begin{aligned}\varphi' &= 1.25 & (\Leftrightarrow \lambda = 0.2) \\ \varphi'' &= 2.00 & (\Leftrightarrow \lambda = 0.5) \\ \varphi''' &= 5.00 & (\Leftrightarrow \lambda = 0.8)\end{aligned}$$

Für jede Kombination aus  $M$  und  $\varphi$  erzeugen wir 100 Responsevektoren mit Stichprobenumfang

$$\begin{aligned}n' &= 20 \\ n'' &= 100\end{aligned}$$

Für die Modellmatrix  $\mathbf{X}$  und die Responsevektoren  $\mathbf{Y}$  bestimmen wir mit der S-Plus Funktion `mle.gpois` (Anhang C.2.2) die Maximum-Likelihood-Schätzer der Parameter  $\varphi$ ,  $\beta_0$  und  $\beta_1$ .

### 8.2.2 Güte der Maximum-Likelihood-Schätzer

Führen wir oben beschriebene Simulation durch, erhalten wir die Werte in den Tabellen 8.1 und 8.2. Für die einzelnen Parameterkonstellationen sind jeweils die Verzerrung, die mittlere quadratische Abweichung und der relative Fehler der Schätzer  $\hat{\varphi}$ ,  $\hat{\beta}_0$  und  $\hat{\beta}_1$  angegeben. In Abschnitt 8.1 wurde für diese drei Maße die Varianz bestimmt. Ziehen wir die Quadratwurzel der jeweiligen Varianz, so erhalten wir die Standardabweichung die in den Tabellen 8.1 und 8.2 hinter dem betreffenden Wert von Verzerrung, mittlerer quadratischer Abweichung und relativem Fehler in Klammer zu finden ist.

Zunächst erkennen wir, dass der Schätzer  $\hat{\varphi}$  meist unterschätzt. Der Regressionsparameter  $\beta_0$  wird immer unter-,  $\beta_1$  hingegen überschätzt. Wir erkennen ebenfalls, dass der relative Fehler und dessen Standardabweichung für den Schätzer  $\hat{\beta}_0$  immer am höchsten, für  $\hat{\varphi}$  hingegen meist am geringsten ist. Alle Schätzer erfüllen das Erwartungswert-Varianz-Kriterium (8.2), nachdem der Erwartungswert größer als die doppelte Standardabweichung ist. Desweiteren sind alle Schätzer bis auf  $\hat{\beta}_0$  und  $\hat{\beta}_1$  bei  $n=20/M = [0.05, 2.72]/\varphi = 5$  gemäß (8.17) signifikant von Null verschieden. Für  $\hat{\beta}_0$  erhalten wir bei der genannten Parameterkonstellation das empirische 95%-Konfidenzintervall  $[-36.04, 0.27]$ , für  $\hat{\beta}_1$  erhalten wir das 95%-Konfidenzintervall  $[-0.02, 19.12]$ . Betrachten wir nun im einzelnen den Einfluss des Stichprobenumfangs  $n$ , des Dispersionsparameters  $\varphi$  und des Parameterbereichs  $M$  auf die Schätzgüte<sup>5</sup>.

- Stichprobenumfang  $n$ :  
Vergleichen wir für  $n=20$  und  $n=100$  die Schätzwerte, denen der selbe  $\varphi$ -Wert und der selbe  $\mu$ -Bereich zu Grunde liegt. Die Verzerrung für  $n=100$  ist erwartungsgemäß meist geringer als bei entsprechenden Schätzern für  $n=20$ . Die Präzision ist bei  $n=100$  stets höher als bei  $n=20$ . Am stärksten ist dies an den Werten von  $\hat{\beta}_0$  und  $\hat{\beta}_1$  zu beobachten. Am geringsten scheint der Einfluss des Stichprobenumfangs auf  $\hat{\varphi}$  zu sein. Die mittlere quadratische Abweichung aller Schätzer nimmt bei allen Parameterkonstellationen mit steigendem Stichprobenumfang deutlich ab. Dies entspricht einer erhöhten Schätzgüte.
- Parameterbereich  $M$ :  
Mit dem Übergang von  $M = [0.05, 2.72]$  auf  $M = [0.02, 7.39]$  steigt die Verzerrung von  $\hat{\varphi}$  im Durchschnitt an, was sich mit der größeren Schwankungsbreite von  $\mu$  erklären lässt.

<sup>5</sup>Wird in den folgenden Betrachtungen der Begriff "Schätzgüte" verwendet, so wird auf die mittlere quadratische Abweichung Bezug genommen. Eine hohe Schätzgüte ist äquivalent zu einer niedrigen mittleren quadratischen Abweichung des betreffenden Schätzers.

Wir müssen hier jedoch zwischen kleinem und großem Stichprobenumfang unterscheiden. Bei  $n=20$  erhöht sich die Verzerrung mit der Vergrößerung des  $\mu$ -Bereichs, während sie bei  $n=100$  abnimmt. Die Verzerrung von  $\hat{\beta}_0$  und  $\hat{\beta}_1$  reduziert sich meist, bei kleinem Stichprobenumfang etwas stärker als bei großem. Die mittlere quadratische Abweichung der Schätzer bei  $M = [0.02, 7.39]$  ist in den meisten Fällen geringer als die mittlere quadratische Abweichung der Schätzer bei  $M = [0.05, 2.72]$ .

- Dispersionsparameter  $\varphi$ :  
Eindeutiger als bei Stichprobenumfang  $n$  und Parameterbereich  $M$  scheint der Einfluss des Dispersionsparameters  $\varphi$  auf die Schätzgüte zu sein. Am stärksten wirkt sich der Anstieg von  $\varphi$  auf dessen Schätzer  $\hat{\varphi}$  aus. Verzerrung und vor allem die Standardabweichung liegen bei  $\varphi = 5$  deutlich höher als bei  $\varphi = 1.25$ . Der Anstieg von Verzerrung und Standardabweichung ist auch bei  $\hat{\beta}_0$  und  $\hat{\beta}_1$  zu beobachten, wenn auch das Ausmaß des Anstiegs, insbesondere der Standardabweichung, nicht so groß ist wie bei  $\hat{\varphi}$ . Der Einfluss von  $\varphi$  auf Verzerrung und Präzision der Schätzer schwindet mit steigendem Stichprobenumfang und sich ausweitendem Parameterbereich  $M$ .

Zusammenfassung:

Bei geringem bis mittlerem Dispersionsgrad erhalten wir stets gute Schätzungen für  $\varphi$ . Sind die Daten jedoch stark überdispersioniert, so reicht selbst ein Stichprobenumfang von 100 Beobachtungen nicht mehr aus um für  $\hat{\varphi}$  eine geringe mittlere quadratische Abweichung zu erhalten.

Die Parameter  $\beta_0$  und  $\beta_1$  werden bei kleinem Dispersionsgrad schon bei Stichprobenumfang  $n=20$  zufriedenstellend geschätzt. Bei hohem Dispersionsgrad erhalten wir nur noch bei Stichprobenumfang  $n=100$  akzeptable Schätzungen.

### 8.2.3 Grafische Überprüfung der asymptotischen Verteilung der Maximum-Likelihood-Schätzer

Für verallgemeinerte lineare Modelle ist bekannt, dass die Maximum-Likelihood-Schätzer asymptotisch normalverteilt sind (siehe z.B. Fahrmeir and Hamerle, 1984, S.58). Da das verallgemeinerte Poisson-Regressionsmodell nicht zur Klasse der verallgemeinerten linearen Modelle gehört (vgl. Abschnitt 5.2), können wir hier jedoch nicht auf diese allgemeine Aussage zurückgreifen. An Hand der berechneten Schätzer wollen wir grafisch überprüfen, ob asymptotische Normalität der Maximum-Likelihood-Schätzer vorliegt.

Zur grafischen Überprüfung der Normalverteilungsannahme eignen sich Histogramm und QQ-Normal-Plot. Im Histogramm sind die (relativen) Häufigkeiten der Ereignisse (meist in Klassen zusammengefasst) abgebildet. Das resultierende Schaubild sollte der Glockenkurve einer Normalverteilungsdichte gleichen. Beim QQ-Normal-Plot werden die Quantile der zu untersuchenden Daten, in unserem Fall die der Schätzwerte, gegen die Quantile der Standard-Normalverteilung abgetragen. Die so gebildeten Punkte sollten im Schaubild auf der ersten Winkelhalbierenden liegen.

In den Abbildungen 8.1-8.6 ist links jeweils das Histogramm und rechts der QQ-Normal-Plot zu einem Parameter zu finden. Über die Histogramme ist dabei die geschätzte Dichtekurve gelegt, wie sie durch die vorhandene S-Plus Grafikroutine berechnet wird.

Für einen Stichprobenumfang von 20 Beobachtungen, den Dispersionsparameter  $\varphi'$  und den Parameterbereich  $M''$  (vgl. Abbildung 8.1) scheint vor allem die Verteilung von  $\hat{\varphi}$  von der Normalverteilung abzuweichen. Das Histogramm lässt viel mehr auf Poisson- oder Exponentialverteilung schließen. Die Linkslastigkeit dürfte allerdings vor allem darauf zurückzuführen sein, dass die Schätzzoutine `mle.gpois` auf Grund der Beschränkung auf Überdispersion als untere

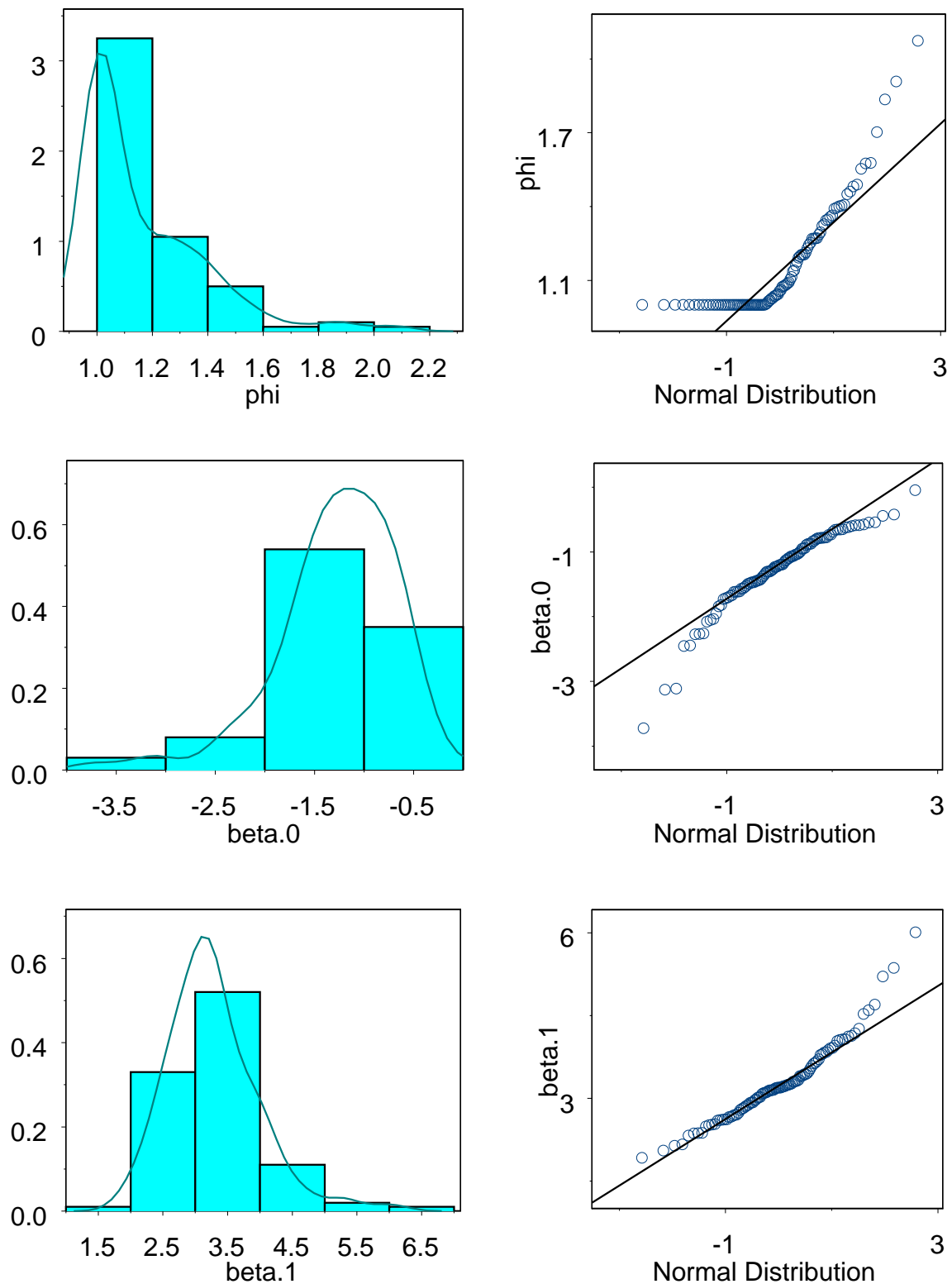


Abbildung 8.1: Histogramm und QQ-Normal-Plot der Maximum-Likelihood-Schätzer im verallgemeinerten Poisson-Regressionsmodell bei  $n=20$  /  $\varphi=1.25$  /  $M = [0.02, 7.39]$

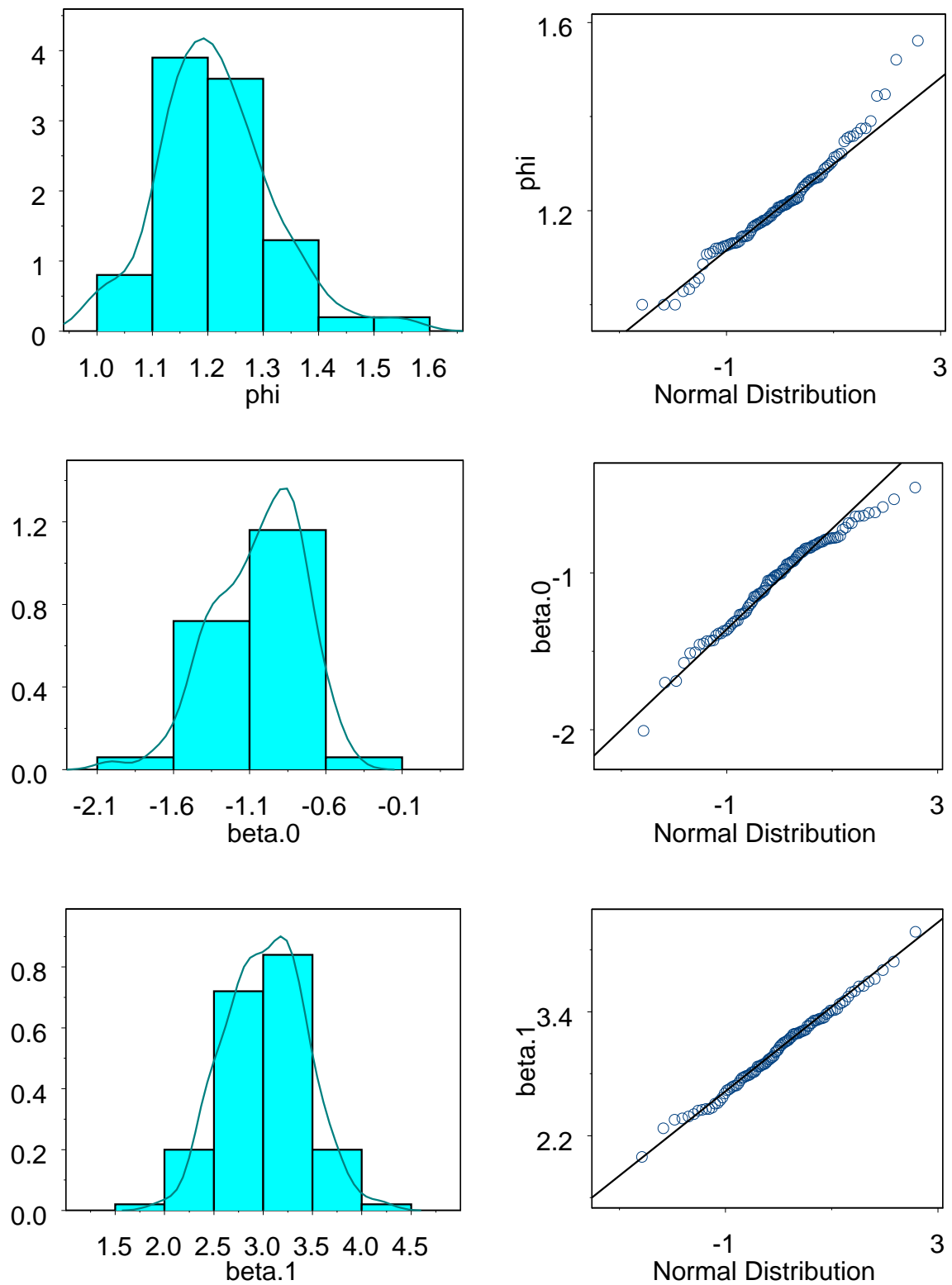


Abbildung 8.2: Histogramm und QQ-Normal-Plot der Maximum-Likelihood-Schätzer im verallgemeinerten Poisson-Regressionsmodell bei  $n=100$  /  $\varphi=1.25$  /  $M = [0.02, 7.39]$



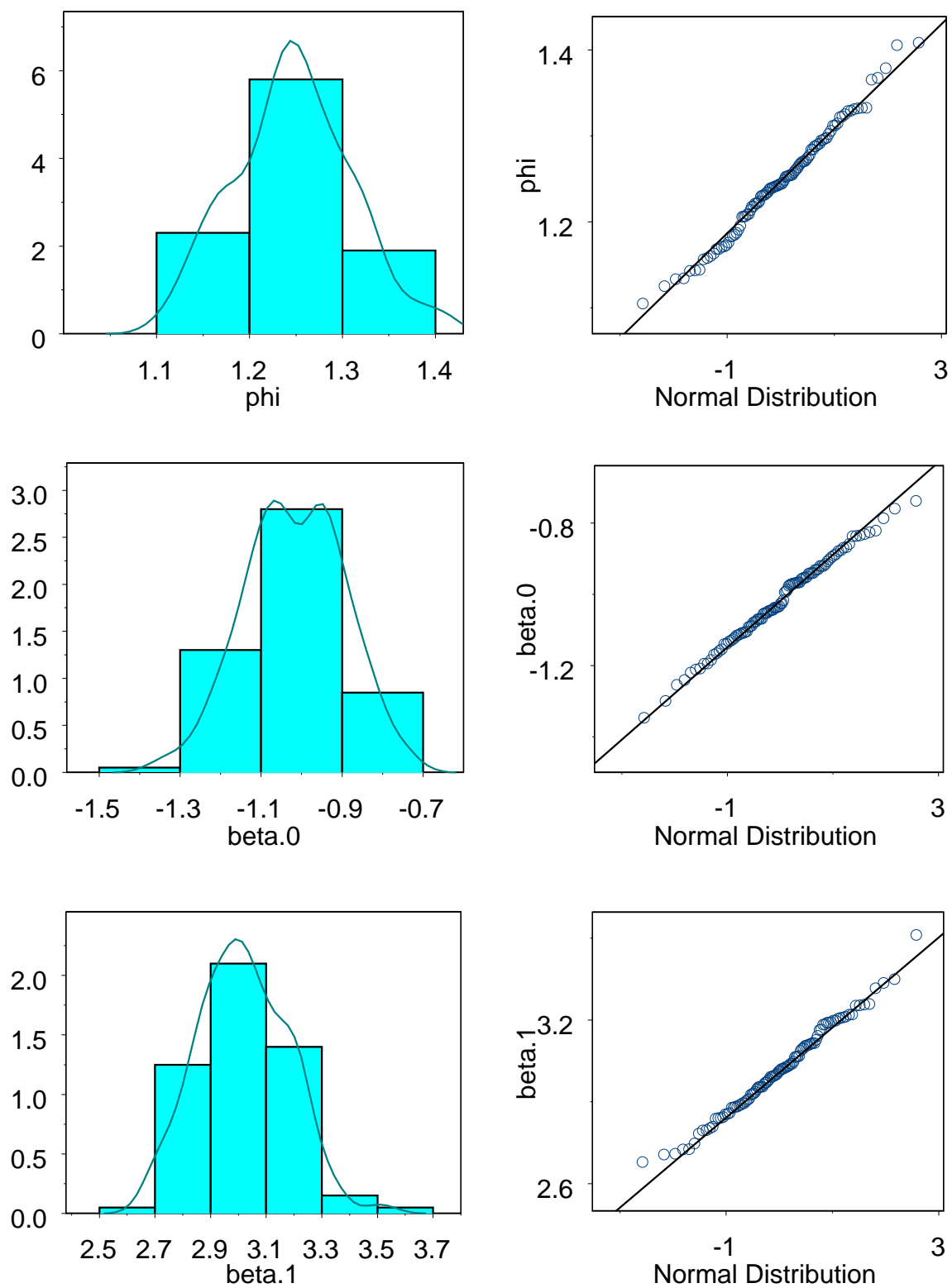


Abbildung 8.3: Histogramm und QQ-Normal-Plot der Maximum-Likelihood-Schätzer im verallgemeinerten Poisson-Regressionsmodell bei  $n=500$  /  $\varphi=1.25$  /  $M = [0.02, 7.39]$

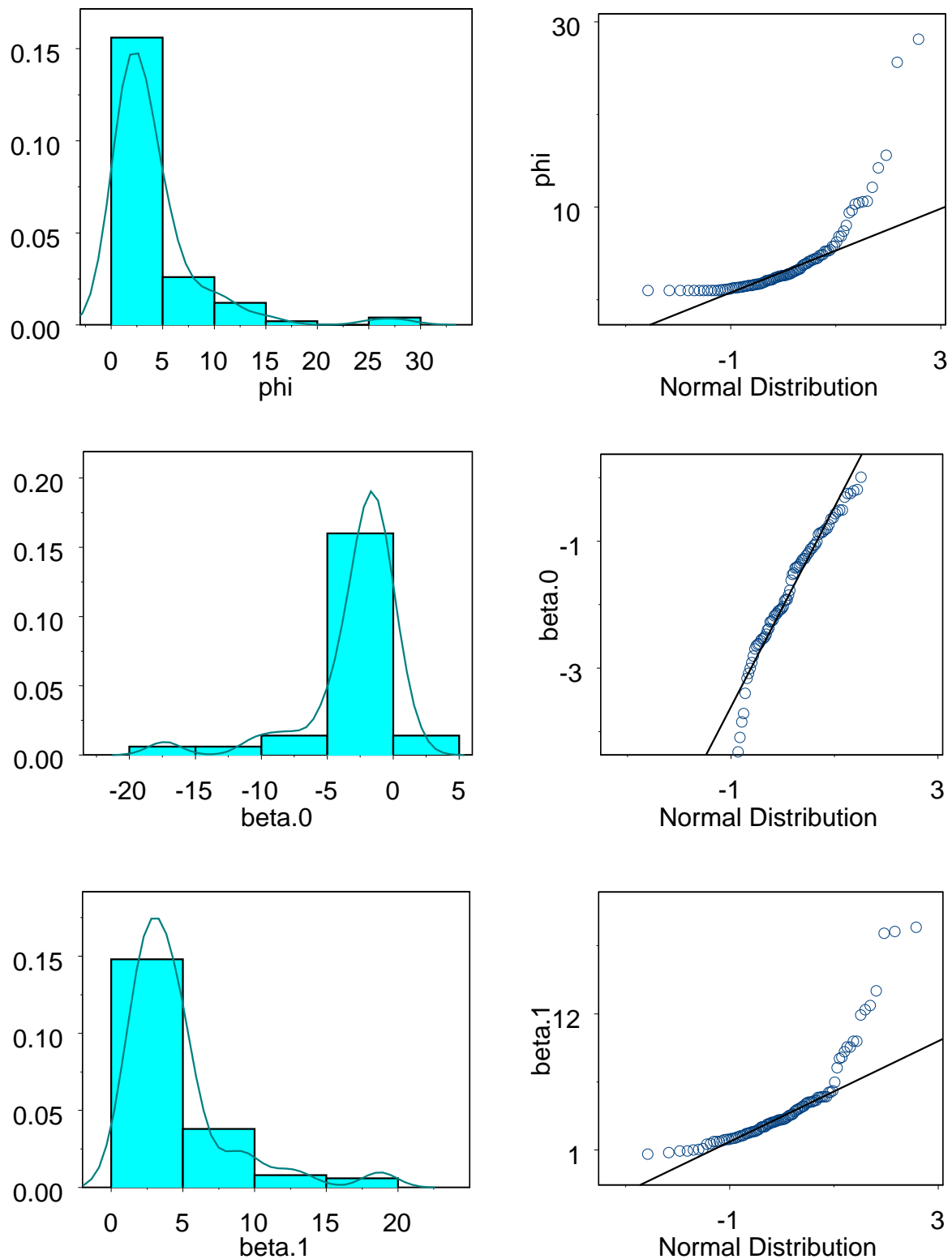


Abbildung 8.4: Histogramm und QQ-Normal-Plot der Maximum-Likelihood-Schätzer im verallgemeinerten Poisson-Regressionsmodell bei  $n=20$  /  $\varphi=5.00$  /  $M = [0.02, 7.39]$

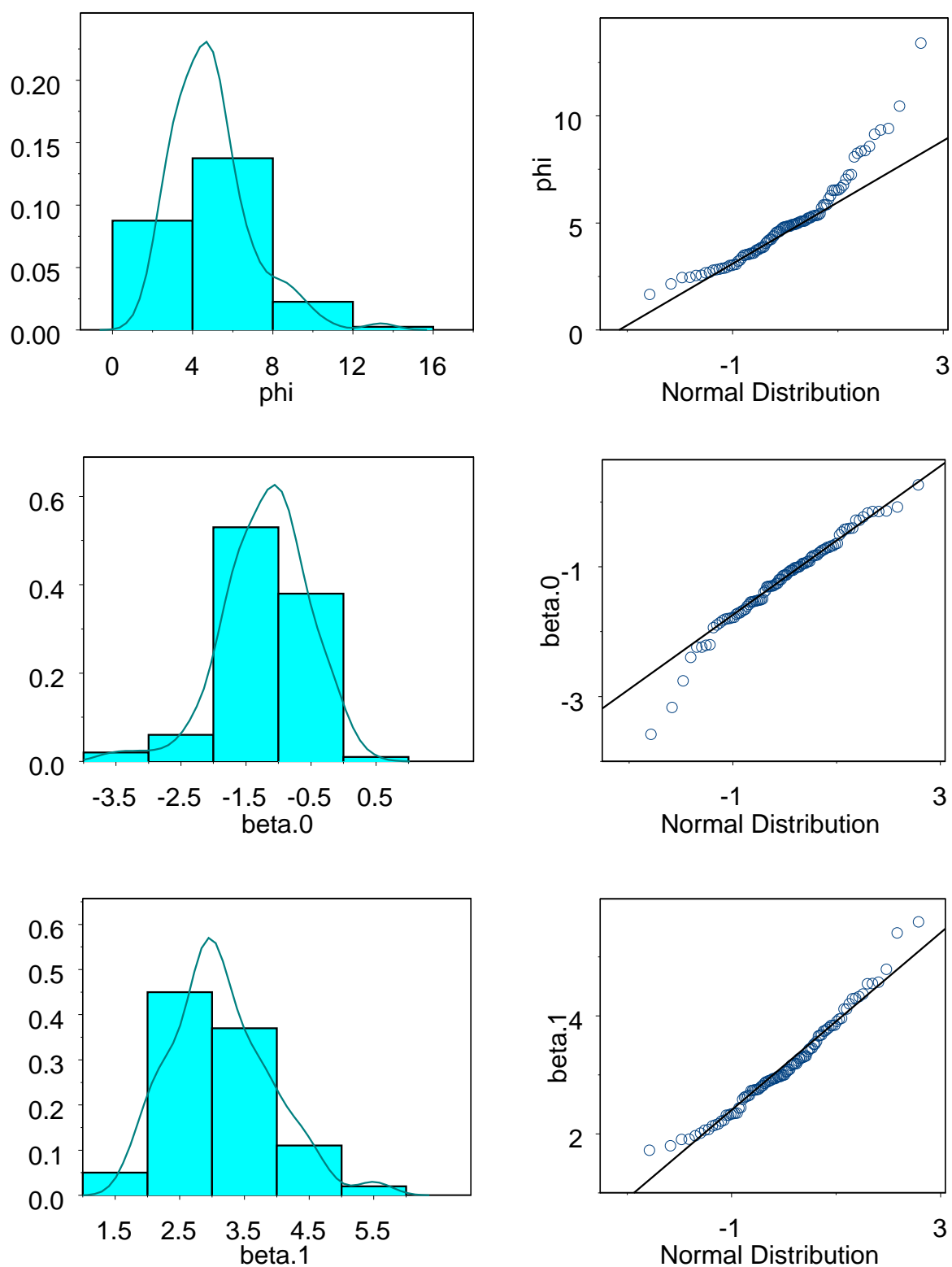


Abbildung 8.5: Histogramm und QQ-Normal-Plot der Maximum-Likelihood-Schätzer im verallgemeinerten Poisson-Regressionsmodell bei  $n=100$  /  $\varphi=5.00$  /  $M = [0.02, 7.39]$

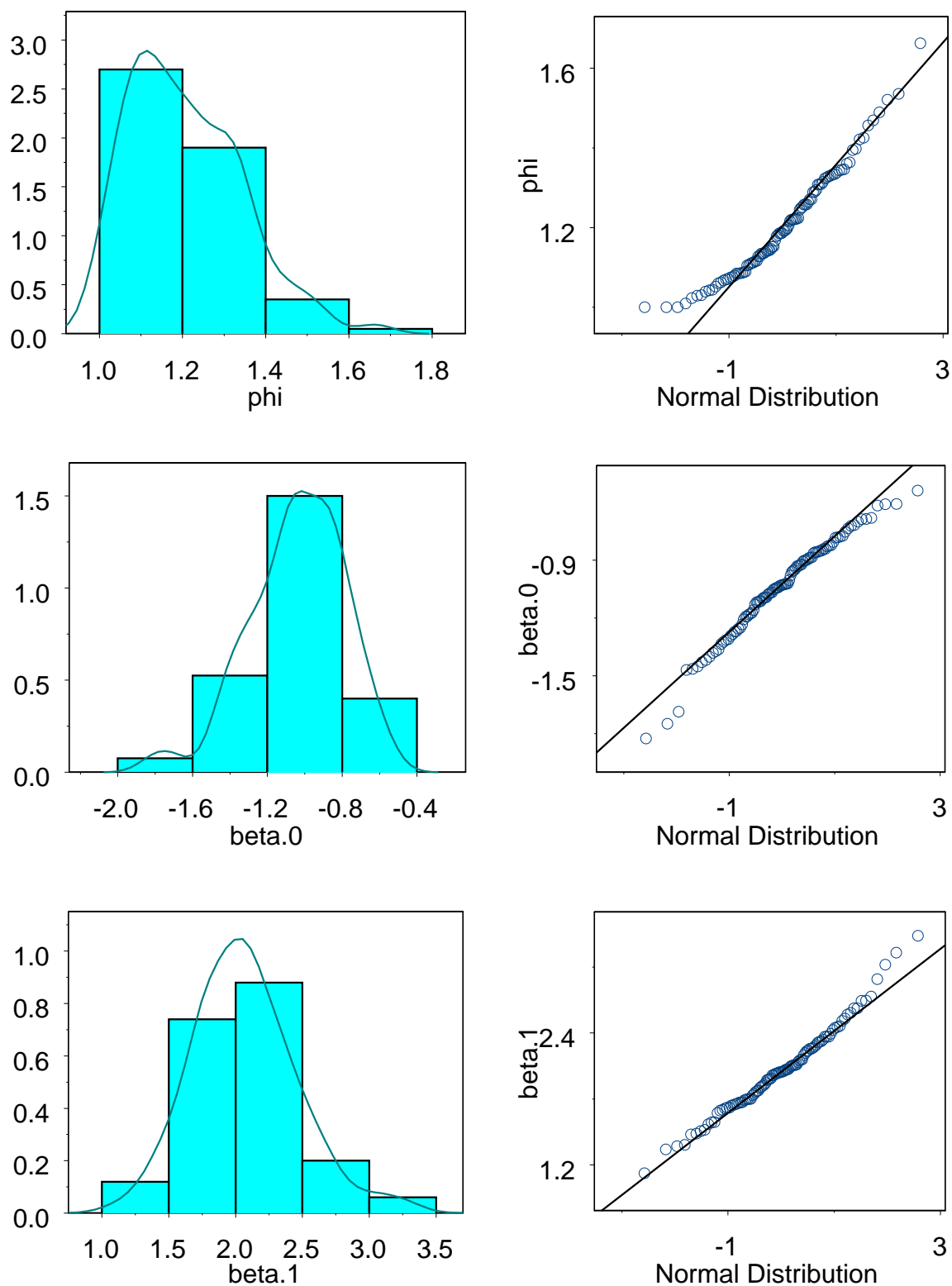


Abbildung 8.6: Histogramm und QQ-Normal-Plot der Maximum-Likelihood-Schätzer im verallgemeinerten Poisson-Regressionsmodell bei  $n=100$  /  $\varphi=1.25$  /  $M = [0.05, 2.72]$

Grenze für  $\hat{\varphi}$  Eins verwendet. Die Quantile der 100 Schätzer  $\hat{\varphi}$  weichen deutlich von denen der Normalverteilung ab. Recht eng legen sich hingegen die Quantile der Schätzer  $\hat{\beta}_1$  bereits für  $n=20$  an die Gerade der Normalverteilungsquantile.

Erhöhen wir nun den Stichprobenumfang auf  $n=100$ . Wir erkennen eine deutliche Annäherung der Schätzerquantile an die Normalverteilungsquantile (vgl. Abbildung 8.2). Auch an den Histogrammen für  $\hat{\varphi}$  und  $\hat{\beta}_0$  können wir eine größere Ähnlichkeit mit der Normalverteilung erkennen. Dieser Trend verfestigt sich bei einer weiteren Erhöhung des Stichprobenumfangs auf  $n=500$  (vgl. Abbildung 8.3). Wiederum entnehmen wir den QQ-Normal-Plots eine Annäherung der Quantile. Dies ist vor allem an dem Maßstab der horizontalen Achsen ablesbar. Die Histogramme nähern sich ebenfalls weiter der Form einer Normalverteilung an, wenn auch die geschätzte Dichte von  $\hat{\beta}_0$  Multimodalität aufweist.

Legen wir nun den erzeugten Responseedaten einen höheren Dispersionsgrad zu Grunde. In Abbildung 8.4 erkennen wir, dass die Verteilungen der Schätzer bei wahren Dispersionsparameter  $\varphi = 5$  und einem Stichprobenumfang von 20 Beobachtung wesentlich weiter von der Normalität entfernt sind, als bei  $\varphi = 1.25$ , gleichem Stichprobenumfang und gleichem Parameterbereich  $M''$ . Dies lässt sich vor allem am Maßstab der QQ-Normal-Plots ablesen. Erhöhen wir den Stichprobenumfang, so erkennen wir auch hier eine Annäherung an die Normalverteilung. Betrachten wir die QQ-Normal-Plots von  $\hat{\beta}_0$  und  $\hat{\beta}_1$  bei  $\varphi = 1.25/n = 20$  (Abbildung 8.1) und  $\varphi = 5.00/n = 100$  (Abbildung 8.5), so weisen diese ungefähr den selben Verlauf aus. Für höheren Dispersionsgrad scheint also ein höherer Stichprobenumfang notwendig zu sein, um einen vergleichbaren Ähnlichkeitsgrad zur Normalverteilung zu erhalten, als bei niedrigerem Dispersionsgrad.

Nur unwesentlich beeinflusst der Parameterbereich  $M$  die Anzahl der Stichproben, die nötig sind, um eine bestimmte Annäherung der Schätzerverteilungen an die Normalverteilung zu erreichen. Vergleichen wir die Verteilung von  $\hat{\varphi}$  bei  $n=100/\varphi = 1.25/M'$  (Abbildung 8.6) und bei  $n=100/\varphi = 1.25/M''$  (Abbildung 8.2). Histogramm und geschätzte Dichtefunktion weisen bei größerem  $\mu$ -Bereich eine höhere Ähnlichkeit mit der Normalverteilung aus, als bei  $M'$ . Die QQ-Normal-Plots sind in beiden Fällen vergleichbar. Die Verteilung von  $\hat{\beta}_0$  scheint hingegen bei  $M'$  näher an der Normalverteilung zu liegen, als bei  $M''$ . Dies lässt sich an Histogramm und Dichtefunktion, vor allem aber am QQ-Normal-Plot erkennen. Bei der Verteilung von  $\hat{\beta}_1$  lässt sich kein allzu großer Unterschied bei verschiedenen  $\mu$ -Bereichen ausmachen.

## 8.3 ZIGP $\mu$ -Regression

### 8.3.1 Datensimulation

Zur Simulation ZIGP-verteilter Responseedaten verwenden wir die S-Plus Funktion `response.zigp1` (Anhang C.3.1). Dabei folgen wir der selben Vorgehensweise, wie in Abschnitt 8.2.1. Die Responsefunktion im ZIGP $\mu$ -Modell lautet

$$h(\eta_i) \stackrel{(6.41)}{=} (1 - \omega) \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\} \quad (8.19)$$

Den zusätzlichen Auswahlparameter  $\omega$  wollen wir ebenfalls variieren. Dazu wählen wir:

$$\begin{aligned} \omega' &= 0.2 \\ \omega'' &= 0.5 \\ \omega''' &= 0.8 \end{aligned}$$

Die Maximum-Likelihood-Schätzer der Parameter  $\varphi$ ,  $\omega$ ,  $\beta_0$  und  $\beta_1$  für die Modellmatrix (8.18) und die erzeugten Responsevektoren bestimmen wir mit der S-Plus Funktion `mle.zigp1` (Anhang C.3.2).

$\varphi$	$\omega$	Parameter	Schätzer					
			Bias		MSE		relativer Fehler	
1.25	0.2	$\varphi = 1.25$	-0.08	(0.02)	0.07	(0.02)	-0.06	(0.02)
		$\omega = 0.2$	-0.04	(0.02)	0.03	(0.001)	-0.20	(0.08)
		$\beta_0 = -1$	-0.18	(0.08)	0.66	(1.16)	0.18	(0.08)
		$\beta_1 = -3$	0.13	(0.09)	0.86	(2.20)	0.04	(0.03)
	0.5	$\varphi = 1.25$	-0.12	(0.02)	0.06	(0.01)	-0.09	(0.02)
		$\omega = 0.5$	-0.004	(0.008)	0.006	(0.002)	-0.01	(0.01)
		$\beta_0 = -1$	-0.32	(0.10)	1.11	(6.60)	0.32	(0.10)
		$\beta_1 = 3$	0.39	(0.13)	1.78	(25.94)	0.13	(0.04)
	0.8	$\varphi = 1.25$	-0.001	(0.04)	0.17	(0.01)	-0.001	(0.03)
		$\omega = 0.8$	-0.26	(0.04)	0.21	(0.07)	-0.32	(0.05)
		$\beta_0 = -1$	-29.07	(5.88)	4297.58	( $10^8$ )	29.07	(5.88)
		$\beta_1 = 3$	35.11	(7.89)	7462.20	( $10^8$ )	11.70	(2.63)
5.00	0.2	$\varphi = 5.00$	-2.87	(0.23)	13.40	(69.64)	-0.57	(0.05)
		$\omega = 0.2$	0.09	(0.03)	0.11	(0.02)	0.47	(0.16)
		$\beta_0 = -1$	-26.83	(6.49)	4931.94	( $10^8$ )	26.83	(6.49)
		$\beta_1 = 3$	29.33	(7.16)	5986.90	( $10^8$ )	9.78	(2.39)
	0.8	$\varphi = 5.00$	-3.13	(0.17)	12.73	(66.77)	-0.63	(0.03)
		$\omega = 0.8$	-0.14	(0.04)	0.18	(0.06)	-0.18	(0.05)
		$\beta_0 = -1$	-48.16	(7.53)	7996.16	( $10^8$ )	48.16	(7.53)
		$\beta_1 = 3$	50.86	(9.29)	11218.13	( $10^8$ )	16.95	(3.10)

Tabelle 8.3: Bias, MSE und relativer Fehler der Maximum-Likelihood-Schätzer im ZIGP $\mu$ -Regressionsmodell für den  $\mu$ -Bereich [0.02, 7.39] bei Stichprobenumfang n=20

$\varphi$	$\omega$	Parameter	Schätzer					
			Bias		MSE		relativer Fehler	
1.25	0.2	$\varphi = 1.25$	-0.05	(0.02)	0.03	(0.01)	-0.04	(0.01)
		$\omega = 0.2$	-0.01	(0.01)	0.01	(0.001)	-0.02	(0.05)
		$\beta_0 = -1$	-0.03	(0.03)	0.11	(0.03)	0.03	(0.03)
		$\beta_1 = 3$	0.03	(0.04)	0.17	(0.10)	0.01	(0.01)
	0.8	$\varphi = 1.25$	-0.03	(0.04)	0.19	(1005.36)	-0.02	(0.04)
		$\omega = 0.8$	-0.03	(0.01)	0.02	(0.04)	-0.04	(0.02)
		$\beta_0 = -1$	-0.21	(0.09)	0.84	(1.87)	0.21	(0.09)
		$\beta_1 = 3$	0.24	(0.10)	1.15	(4.61)	0.08	(0.03)
5.00	0.2	$\varphi = 5.00$	-0.48	(0.25)	6.61	(255.32)	-0.10	(0.05)
		$\omega = 0.2$	0.05	(0.02)	0.05	(0.01)	0.23	(0.11)
		$\beta_0 = -1$	-0.29	(0.09)	0.95	(2.99)	0.29	(0.09)
		$\beta_1 = 3$	0.47	(0.10)	1.17	(5.30)	0.16	(0.03)
	0.5	$\varphi = 5.00$	-0.96	(0.29)	9.62	(63.17)	-0.19	(0.06)
		$\omega = 0.8$	-0.01	(0.03)	0.08	(0.05)	-0.03	(0.06)
		$\beta_0 = -1$	-0.62	(0.16)	3.00	(66.53)	0.62	(0.16)
		$\beta_1 = 3$	0.62	(0.17)	3.36	(78.42)	0.21	(0.06)
	0.8	$\varphi = 5.00$	-2.54	(0.22)	11.22	(1108.20)	-0.51	(0.04)
		$\omega = 0.8$	-0.13	(0.03)	0.13	(0.004)	-0.17	(0.04)
		$\beta_0 = -1$	-55.11	(18.87)	38644.76	( $10^{10}$ )	55.11	(18.87)
		$\beta_1 = 3$	71.27	(23.83)	61857.52	( $10^{10}$ )	23.76	(7.94)

Tabelle 8.4: Bias, MSE und relativer Fehler der Maximum-Likelihood-Schätzer im ZIGP $\mu$ -Regressionsmodell für den  $\mu$ -Bereich [0.02, 7.39] bei Stichprobenumfang n=100

### 8.3.2 Güte der Maximum-Likelihood-Schätzer

Die Tabellen 8.3 und 8.4 enthalten die Auswertung der ermittelten Schätzer für einige Parameterkonstellationen. Für alle Parameterkonstellationen wird der Parameter  $\beta_1$  überschätzt, während  $\varphi$  und  $\beta_0$  unterschätzt werden. Vergleichen wir die Präzision der Schätzer, so finden wir bis auf zwei Parameterkonstellationen stets die selbe Rangfolge der Schätzer vor. Die höchste Präzision weist  $\hat{\omega}$  auf, gefolgt von  $\hat{\varphi}$  und  $\hat{\beta}_0$ . Ein ähnliches Bild zeigt der Vergleich der mittleren quadratischen Abweichungen. Alle Schätzer erfüllen das Erwartungswert-Varianz-Kriterium (8.2).

Betrachten wir die Auswirkungen der Einflussgrößen Stichprobenumfang  $n$ , Dispersionsparameter  $\varphi$ , Auswahlparameter  $\omega$  und Parameterbereich  $M$  auf die Schätzgüte erneut im einzelnen:

- Stichprobenumfang  $n$ :

Der Schätzer  $\hat{\varphi}$  weist bei  $n=100$  meist eine geringere Verzerrung auf als bei  $n=20$ . Dem entgegen kann aus den vorliegenden Daten keine eindeutige Veränderung der Präzision dieses Schätzers abgelesen werden. Mittlere quadratische Abweichung und relativer Fehler fallen für  $\hat{\varphi}$  bei großem Stichprobenumfang meist geringer aus als bei kleinem. Bei  $\varphi = 5/\omega = 0.8$  ist eine Ausweitung des Stichprobenumfangs über 100 nötig um eine akzeptabel kleine Verzerrung zu erhalten.

Sowohl die Verzerrung als auch die Präzision, die mittlere quadratische Abweichung und der relative Fehler des Schätzers  $\hat{\omega}$  verbessern sich bei der Erhöhung des Stichprobenumfangs.

Sind Dispersions- und Auswahlparameter klein, so sind Verzerrung, Standardabweichung, mittlere quadratische Abweichung und relativer Fehler der Schätzer  $\hat{\beta}_0$  und  $\hat{\beta}_1$  sowohl bei  $n=20$  als auch bei  $n=100$  gering. Dennoch verbessern sich hier diese Größen mit steigendem Stichprobenumfang weiter. Eine erhebliche Reduktion erfährt die Verzerrung, vor allem aber die Standardabweichung dieser Schätzer, wenn sich die Werte des Dispersions- und Auswahlparameters deutlich unterscheiden, d.h. für die Kombinationen  $\varphi = 1.25/\omega = 0.8$  und  $\varphi = 5/\omega = 0.2$ . Sind beide Parameter hoch, führen unsere Beispieldaten jedoch zu einer stärkeren Verzerrung und einer geringeren Präzision bei einem höheren Stichprobenumfang. Bis auf diese Parameterkonstellation sinkt mit steigendem  $n$  die mittlere quadratische Abweichung. Insgesamt zeigen die Schätzer  $\hat{\beta}_0$  und  $\hat{\beta}_1$  die stärkste Reaktion auf die Veränderung des Stichprobenumfangs.

- Parameterbereich  $M$ :

Auf die Veränderung des Parameterbereichs reagieren die Schätzer im ZIGP $\mu$ -Modell im wesentlichen wie die Schätzer im verallgemeinerten Poissonmodell. Wird  $M$  verkleinert, so steigt die mittlere quadratische Abweichung bis auf wenige Ausnahmen an. Damit die Tabellen 8.3 und 8.4 übersichtlich bleiben, sind lediglich die Werte bei  $M = [0.02, 7.39]$  angegeben, da die Auswirkung des verkleinerten Parameterbereichs geringer ist als die Auswirkung eines veränderten Dispersionsparameters, Stichprobenumfangs und Auswahlparameters.

- Dispersionsparameter  $\varphi$ :

Eine Erhöhung des Dispersionsgrades führt bei allen Schätzern zu einer stärkeren Verzerrung. Das Ausmaß indem die Verzerrung der Schätzer  $\hat{\beta}_0$  und  $\hat{\beta}_1$  zunimmt, scheint deutlich vom Stichprobenumfang und dem Auswahlparameter abzuhängen. Für  $\varphi = 1.25$  liegen die Schätzwerte  $\hat{\beta}_0$  und  $\hat{\beta}_1$  bei  $\omega = 0.2/n = 20$  und  $\omega = 0.8/n = 100$  nahe an den wahren Werten. In beiden Fällen erhöht sich die Verzerrung jedoch mit anwachsendem  $\varphi$  stark. Bei hohem Dispersionsgrad ist der jeweilige Stichprobenumfang nicht mehr ausreichend

um angemessene Schätzwerte zu erhalten.

Entsprechend der Verzerrung verhält sich auch die Standardabweichung und folglich die mittlere quadratische Abweichung der Schätzer. Mit steigendem Dispersionsgrad sinkt daher die Schätzgüte. Am stärksten ist dies bei  $\hat{\beta}_0$  und  $\hat{\beta}_1$  zu beobachten, am schwächsten bei  $\hat{\omega}$ . Der relative Fehler steigt, vergleichbar der Verzerrung, bis auf eine Ausnahme mit dem Dispersionsparameter an.

- Auswahlparameter  $\omega$ :

Sehr deutlich ist die Abhängigkeit der Schätzgüte von  $\hat{\beta}_0$  und  $\hat{\beta}_1$  von dem Auswahlparameter  $\omega$  zu erkennen. Ist der Anteil der Zero-Inflation hoch ( $\omega = 0.8$ ), so stehen der  $\mu$ -Regression vor allem bei  $n=20$  zu wenig Daten zur Verfügung. Dem entsprechend verschlechtern sich Verzerrung, Präzision, mittlere quadratische Abweichung und relativer Fehler der korrespondierenden Schätzer bei Erhöhung des Auswahlparameters.

Die Verzerrung der Schätzer  $\hat{\varphi}$  und  $\hat{\omega}$  steigt bei Ausweitung der Zero-Inflation meist, während die  $\omega$  keinen Einfluss auf die Standardabweichung dieser Schätzer zu haben scheint. In den meisten Fällen steigt die mittlere quadratische Abweichung und der relative Fehler von  $\hat{\varphi}$  und  $\hat{\omega}$ , wenn wir den wahren Wert von  $\omega$  erhöhen.

Zusammenfassung:

Die Güte des Schätzers  $\hat{\omega}$  ist stets hoch. Der Schätzer  $\hat{\varphi}$  hingegen ist bei stark überdispensierten Daten kaum brauchbar. Um akzeptable Schätzer  $\hat{\beta}_0$  und  $\hat{\beta}_1$  bei hohem Dispersionsgrad und/oder hohem Auswahlparameter zu erhalten, ist ein großer Stichprobenumfang erforderlich.

## 8.4 ZIGP $\mu\omega$ -Regression

### 8.4.1 Datensimulation

Im Gegensatz zur ZIGP $\mu$ -Regression darf der Auswahlparameter  $\omega$  bei ZIGP $\mu\omega$ -Regression von Kovariablen abhängen. Es tritt daher auch für  $\omega$  Regression auf. Die Linkfunktionen in diesem Modell lauten:

$$\begin{aligned}\omega_i &\stackrel{(6.47)}{=} \frac{\exp\{\mathbf{a}_i^t \boldsymbol{\alpha}\}}{1 + \exp\{\mathbf{a}_i^t \boldsymbol{\alpha}\}} \\ \mu_i &\stackrel{(6.48)}{=} \exp\{\mathbf{b}_i^t \boldsymbol{\beta}\}\end{aligned}$$

Für die Responsefunktion gilt:

$$E(Y_i | \mathbf{a}_i, \mathbf{b}_i) \stackrel{(6.49)}{=} \frac{t_i \exp\{\mathbf{b}_i^t \boldsymbol{\beta}\}}{1 + \exp\{\mathbf{a}_i^t \boldsymbol{\alpha}\}}$$

### Design A

Zunächst wollen wir der  $\omega$ -Regression dieselben Kovariablen (8.18) zur Verfügung stellen, wie der  $\mu$ -Regression. Der lineare Prädiktor (6.45) enthält deshalb ebenfalls einen Intercept- und einen Regressionsterm. Ähnlich wie für die Parameter  $\mu_i$ , wollen wir auch für  $\omega_i$  unterschiedliche Parameterbereiche betrachten. Mit der Verknüpfung (6.47) folgt für  $\omega_i$ ,  $i = 1, \dots, n$ :

$$\begin{aligned}\omega_i \in \Omega &= [\omega_{\min}, \omega_{\max}] \\ &= \left[ \min_{1 \leq i \leq n} \left\{ \frac{\exp\{\alpha_0 + \alpha_1 x_i\}}{1 - \exp\{\alpha_0 + \alpha_1 x_i\}} \right\}, \max_{1 \leq i \leq n} \left\{ \frac{\exp\{\alpha_0 + \alpha_1 x_i\}}{1 - \exp\{\alpha_0 + \alpha_1 x_i\}} \right\} \right]\end{aligned}$$



Wählen wir

$$\boldsymbol{\alpha}' = \begin{pmatrix} -2 \\ 0.5 \end{pmatrix}, \quad \boldsymbol{\alpha}'' = \begin{pmatrix} -2 \\ 2 \end{pmatrix}, \quad \boldsymbol{\alpha}''' = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

erhalten wir die Bereiche

$$\begin{aligned} \Omega' &\approx [0.08, 0.18] \\ \Omega'' &\approx [0.02, 0.50] \\ \Omega''' &\approx [0.05, 0.73] \end{aligned}$$

Damit entspricht  $\omega_{\max}$  jeweils ungefähr den Werten  $\omega'$ ,  $\omega''$  und  $\omega'''$  aus Abschnitt 8.3.1. Die Verwendung der Modellmatrix  $\mathbf{X}$  (8.18) sowohl für die  $\omega$ - als auch für die  $\mu$ -Regression, wollen wir als Design A bezeichnen. Bei der Wahl des Dispersionsgrades und des Parameterbereichs  $M$  gehen wir wiederum entsprechend Abschnitt 8.2.1 vor. Als Stichprobenumfang wählen wir  $n=100$ .

### Orthogonale Kovariablen

Um die Auswirkung unterschiedlicher Kovariablen für  $\omega$ - und  $\mu$ -Regression zu beobachten, führen wir eine weitere Modellmatrix  $\tilde{\mathbf{X}}$  ein. Wir wählen die Spalten von  $\tilde{\mathbf{X}}$  so, dass wir orthogonale Kovariablen erhalten.  $\tilde{\mathbf{X}}_i$  bzw.  $\mathbf{X}_i$  bezeichne dabei die  $i$ -te Spalte der entsprechenden Matrix. Für

$$\tilde{\mathbf{X}} = \begin{pmatrix} o_1 & o_1 \tilde{x}_1 \\ \vdots & \vdots \\ o_{\frac{n}{2}} & o_{\frac{n}{2}} \tilde{x}_{\frac{n}{2}} \\ -o_{\frac{n+2}{2}} & -o_{\frac{n+2}{2}} \tilde{x}_{\frac{n+2}{2}} \\ \vdots & \vdots \\ -o_n & -o_n \tilde{x}_n \end{pmatrix} \quad (8.20)$$

mit

$$\tilde{x}_i = \left| \frac{n-1}{2(i-1) - (n-1)} \right|, \quad i = 1, \dots, n$$

$n$  gerade und

$$o_i = \begin{cases} +1 & \text{für } i \bmod 2 = 1 \\ -1 & \text{für } i \bmod 2 = 0 \end{cases}$$

gilt

$$\begin{aligned} \tilde{\mathbf{X}}_i^t \mathbf{X}_j &= 0 \quad i, j \in \{1, 2\}, \\ \tilde{\mathbf{X}}_1^t \tilde{\mathbf{X}}_2 &= 0 \quad \text{und} \\ \mathbf{X}_1^t \mathbf{X}_2 &= 0 \end{aligned}$$

Für  $\tilde{\mathbf{X}}_1^t \mathbf{X}_1$  ist dies leicht ersichtlich. Unter Verwendung der Gleichung

$$x_{(n+1)-i} = \frac{2((n+1) - i - 1) - (n-1)}{n-1} = \frac{(n-1) - 2(i-1)}{n-1} = -x_i$$

Design	$\omega$ -Regression			$\mu$ -Regression		
	Modellmatrix	Regressionsparameter	$\Omega$	Modellmatrix	Regressionsparameter	$M$
A	$\mathbf{X}$	$\boldsymbol{\alpha}' = \begin{pmatrix} -2 \\ 0.5 \end{pmatrix}$ $\boldsymbol{\alpha}'' = \begin{pmatrix} -2 \\ 2 \end{pmatrix}$ $\boldsymbol{\alpha}''' = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$	[0.08, 0.18] [0.02, 0.50] [0.05, 0.73]	$\mathbf{X}$	$\boldsymbol{\beta}' = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$ $\boldsymbol{\beta}'' = \begin{pmatrix} -1 \\ 3 \end{pmatrix}$	[0.05, 2.72] [0.02, 7.39]
B	$\tilde{\mathbf{X}}$	$\tilde{\boldsymbol{\alpha}} = \begin{pmatrix} 0.01 \\ 0.01 \end{pmatrix}$	[0.27, 0.73]	$\mathbf{X}$	$\boldsymbol{\beta}' = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$ $\boldsymbol{\beta}'' = \begin{pmatrix} -1 \\ 3 \end{pmatrix}$	[0.05, 2.72] [0.02, 7.39]
C	$\mathbf{X}$	$\boldsymbol{\alpha}' = \begin{pmatrix} -2 \\ 0.5 \end{pmatrix}$ $\boldsymbol{\alpha}'' = \begin{pmatrix} -2 \\ 2 \end{pmatrix}$ $\boldsymbol{\alpha}''' = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$	[0.08, 0.18] [0.02, 0.50] [0.05, 0.73]	$\tilde{\mathbf{X}}$	$\tilde{\boldsymbol{\beta}} = \begin{pmatrix} -1 \\ 0.01 \end{pmatrix}$	[0.26, 7.32]

Tabelle 8.5: Übersicht über die ZIGP $\mu\omega$ -Designs, die darin verwendeten Regressionsparameter und die resultierenden Parameterbereiche  $\Omega$  und  $M$

können wir bei geradem  $n$   $\mathbf{X}_1^t \mathbf{X}_2 = 0$  und  $\tilde{\mathbf{X}}_1^t \mathbf{X}_2 = 0$  folgern. Gleiches gilt für  $\mathbf{X}_1^t \tilde{\mathbf{X}}_2$  und  $\tilde{\mathbf{X}}_1^t \tilde{\mathbf{X}}_2$  wegen

$$\tilde{x}_{(n+1)-i} = \left| \frac{n-1}{2((n+1)-i-1)-(n-1)} \right| = \left| \frac{n-1}{(n-1)-2(i-1)} \right| = \tilde{x}_i$$

Desweiteren erhalten wir für gerades  $n$ :

$$\begin{aligned} \tilde{\mathbf{X}}_2^t \mathbf{X}_2 &= \sum_{i=1}^n \tilde{x}_i x_i = \sum_{i=1}^n o_i \left| \frac{n-1}{2(i-1)-(n-1)} \right| \left( \frac{2(i-1)-(n-1)}{n-1} \right) \\ &= \sum_{i=1}^{\frac{n}{2}} -o_i - \sum_{i=\frac{n}{2}+1}^n o_i = 0 \end{aligned}$$

**Design B**

Verwenden wir  $\tilde{\mathbf{X}}$  als Modellmatrix für die  $\omega$ -Regression, so erhalten mit dem Regressionsparameter

$$\tilde{\boldsymbol{\alpha}} = \begin{pmatrix} 0.01 \\ 0.01 \end{pmatrix} \quad (8.21)$$

den Parameterbereich

$$\tilde{\Omega} \approx [0.27, 0.73] \quad (8.22)$$

für  $\omega_i$ . Die  $\mu$ -Regression verhalte sich wie in Design A, ebenso Stichprobenumfang und Dispersionsgrad. Wir bezeichnen diese Anordnung als Design B.

**Design C**

Bei Design C sind die Rollen von  $\tilde{\mathbf{X}}$  und  $\mathbf{X}$  im Vergleich zu Design B vertauscht. Um einen, zu  $M'''$  vergleichbaren Parameterbereich zu erhalten, wählen wir hier als Regressionsparameter  $\tilde{\boldsymbol{\beta}}$ :

$$\tilde{\boldsymbol{\beta}} = \begin{pmatrix} -1 \\ 0.01 \end{pmatrix} \quad (8.23)$$

Für  $\mu_i$  erhalten wir damit den Parameterbereich

$$\tilde{M} \approx [0.26, 7.32] \quad (8.24)$$

Stichprobenumfang, Dispersionsgrad und Regressionsparameter  $\alpha$  wählen wir entsprechend Design A. Für alle drei Designs und die jeweiligen Parameterkonstellationen, erzeugen wir 100 Responsevektoren  $\mathbf{Y}$  mit der S-Plus Funktion `response.zigp2` (Anhang C.3.3). Für diese bestimmen wir Maximum-Likelihood-Schätzer der Parameter  $\varphi$ ,  $\alpha_0$ ,  $\alpha_1$ ,  $\beta_0$  und  $\beta_1$  mit der S-Plus Funktion `mle.zigp2` (Anhang C.3.4).

**8.4.2 Güte der Maximum-Likelihood-Schätzer**

Die aus der Simulation resultierenden Schätzer lassen sich in drei Kategorien unterscheiden. Während die Schätzer  $\hat{\varphi}$  und  $\hat{\beta}_0$  stets zufriedenstellend sind, unterscheidet sich der Schätzer  $\hat{\beta}_1$  in einigen wenigen Fällen nicht signifikant von Null. Die Schätzer  $\hat{\alpha}_0$  und  $\hat{\alpha}_1$  sind hingegen in den meisten Fällen nicht signifikant von Null verschieden. Für diese scheint ein Identifizierbarkeitsproblem vorzuliegen. Häufig kann das Erwartungswert-Varianz-Kriterium (8.2) von  $\hat{\alpha}_0$  und  $\hat{\alpha}_1$  nicht erfüllt werden.

Die Verzerrung (vgl. Tabelle 8.6) zeigt, dass für die vorliegenden Daten  $\varphi$ ,  $\alpha_0$  und  $\beta_0$  stets unterschätzt,  $\beta_1$  aber überschätzt wird. Meist ist die Standardabweichung, die mittlere quadratische Abweichung und der relative Fehler des Schätzers  $\hat{\varphi}$  am geringsten, während diese Größen bei  $\hat{\alpha}_0$  und  $\hat{\alpha}_1$  mit Abstand am schlechtesten sind. Auf veränderte Parameter reagieren die Schätzer wie folgt:

- Dispersionsgrad  $\varphi$ :  
Erhöhen wir den Dispersionsgrad, erfahren  $\hat{\varphi}$ ,  $\hat{\beta}_0$  und  $\hat{\beta}_1$  eine eindeutige Verschlechterung. Über das Verhalten von  $\hat{\alpha}_0$  und  $\hat{\alpha}_1$  lassen die vorliegenden Daten keine Aussage zu.

Design	$\varphi$	$\Omega$	Parameter	Schätzer					
				Bias		MSE		relativer Fehler	
A	1.25	[0.08, 0.18]	$\varphi = 1.25$	-0.06	(0.02)	0.03	(0.01)	-0.05	(0.01)
			$\alpha_0 = -2$	-437.97	(421.21)	$10^7$	( $10^{15}$ )	218.99	(210.61)
			$\alpha_1 = 0.5$	-797.43	(758.65)	$10^8$	( $10^{16}$ )	-1594.86	(1517.31)
			$\beta_0 = -1$	0.13	(0.05)	0.28	(0.01)	-0.13	(0.05)
			$\beta_1 = 3$	-0.16	(0.06)	0.44	(0.28)	-0.05	(0.02)
		[0.02, 0.50]	$\varphi = 1.25$	-0.07	(0.02)	0.04	(0.01)	-0.06	(0.01)
			$\alpha_0 = -2$	-1.95	(0.70)	52.43	(29154.27)	0.98	(0.35)
			$\alpha_1 = 2$	2.22	(0.81)	70.38	(54139.72)	1.11	(0.40)
			$\beta_0 = -1$	0.07	(0.05)	0.25	(0.08)	-0.07	(0.05)
	[0.05, 0.73]	$\varphi = 1.25$	-0.04	(0.02)	0.04	(0.01)	-0.04	(0.01)	
		$\alpha_0 = -1$	-0.74	(0.41)	17.37	(2605.60)	0.74	(0.41)	
		$\alpha_1 = 2$	1.04	(0.51)	26.77	(6576.73)	0.52	(0.25)	
$\beta_0 = -1$		0.06	(0.06)	0.33	(0.27)	-0.06	(0.06)		
2.00	[0.08, 0.18]	$\varphi = 2$	-0.16	(0.04)	0.16	(0.06)	-0.08	(0.02)	
		$\alpha_0 = -2$	-127.87	(45.13)	$10^5$	( $10^{11}$ )	63.93	(22.57)	
		$\alpha_1 = 0.5$	-151.50	(111.13)	$10^6$	( $10^{13}$ )	-303.00	(222.26)	
		$\beta_0 = -1$	0.38	(0.08)	0.74	(1.07)	-0.38	(0.08)	
		$\beta_1 = 3$	-0.49	(0.09)	1.09	(2.58)	-0.16	(0.03)	
	[0.02, 0.50]	$\varphi = 2$	-0.15	(0.06)	0.34	(0.36)	-0.08	(0.03)	
		$\alpha_0 = -2$	-57.33	(36.32)	$10^5$	( $10^{11}$ )	28.66	(18.16)	
		$\alpha_1 = 2$	20.68	(48.69)	$10^5$	( $10^{12}$ )	10.34	(24.34)	
		$\beta_0 = -1$	0.35	(0.08)	0.75	(1.01)	-0.35	(0.08)	
B	1.25	[0.27, 7.32]	$\varphi = 1.25$	-0.03	(0.02)	0.04	(0.01)	-0.03	(0.02)
			$\alpha_0 = 0.01$	0.06	(0.08)	0.63	(2.50)	5.84	(7.90)
			$\alpha_1 = 0.01$	0.06	(0.05)	0.28	(0.90)	6.09	(5.30)
			$\beta_0 = -1$	-0.11	(0.04)	0.17	(0.03)	0.11	(0.04)
			$\beta_1 = 3$	0.11	(0.06)	0.35	(0.16)	0.04	(0.02)
			$\varphi = 1.25$	-0.04	(0.01)	0.02	(0.01)	-0.03	(0.01)
	2.00	[0.27, 7.32]	$\alpha_0 = -2$	-65.34	(37.47)	$10^5$	( $10^{11}$ )	32.67	(18.74)
			$\alpha_1 = 0.5$	23.24	(42.32)	$10^5$	( $10^{11}$ )	46.49	(84.65)
			$\beta_0 = -1$	0.01	(0.002)	0.01	(0.003)	-0.01	(0.01)
			$\beta_1 = 0.01$	-0.003	(0.003)	0.0001	(0.0004)	-0.09	(0.08)
			$\varphi = 2$	-0.15	(0.06)	0.34	(0.36)	-0.08	(0.03)
			$\alpha_0 = -2$	-57.33	(36.32)	$10^5$	( $10^{11}$ )	28.66	(18.16)
C	1.25	[0.08, 0.18]	$\alpha_1 = 2$	20.68	(48.69)	$10^5$	( $10^{12}$ )	10.34	(24.34)
			$\beta_0 = -1$	0.35	(0.08)	0.75	(1.01)	-0.35	(0.08)
			$\beta_1 = 3$	-0.47	(0.11)	1.44	(5.22)	-0.16	(0.04)
			$\varphi = 1.25$	-0.03	(0.02)	0.04	(0.01)	-0.03	(0.02)
			$\alpha_0 = 0.01$	0.06	(0.08)	0.63	(2.50)	5.84	(7.90)
			$\alpha_1 = 0.01$	0.06	(0.05)	0.28	(0.90)	6.09	(5.30)

Tabelle 8.6: Bias, MSE und relativer Fehler der Maximum-Likelihood-Schätzer im ZIGP $_{\mu\omega}$ -Regressionsmodell bei Stichprobenumfang n=100

- Parameterbereich  $M$ :  
Die Reduzierung des  $\mu$ -Bereichs von  $[0.02, 7.39]$  auf  $[0.05, 2.72]$  führt für  $\hat{\varphi}$  in den meisten Fällen zu einer Erhöhung, für die übrigen Parameter zu einem Rückgang von Verzerrung, Standardabweichung, mittlerer quadratischer Abweichung und relativen Fehlern. In Tabelle 8.6 sind aus Gründen der Übersichtlichkeit nur Werte für den Parameterbereich  $[0.02, 7.39]$  zu finde.
- Parameterbereich  $\Omega$ :  
Wird der  $\omega$ -Bereich  $[0.08, 0.18]$  verwendet, so stehen der  $\omega$ -Regression auf Grund des geringen Nullenüberschusses verhältnismäßig wenig Daten zur Verfügung. Die korrespondierenden Schätzer  $\hat{\alpha}_0$  und  $\hat{\alpha}_1$  können nicht identifiziert werden. Wechseln wir den  $\omega$ -Bereich  $[0.08, 0.18]$  gegen  $[0.02, 0.50]$  und dann gegen  $[0.05, 0.73]$  aus, dann steigt mit der Zero-Inflation auch die Anzahl der Beobachtungen für die  $\omega$ -Regression. Wir beobachten für diesen Fall eine eindeutige Verbesserung aller Kriterien für die Schätzgüte von  $\hat{\alpha}_0$  und  $\hat{\alpha}_1$ . Steigt die Zero-Inflation, so muss die  $\mu$ -Regression im Gegenzug mit weniger Daten auskommen. Dies führt in den Simulationsdaten zu einem Präzisionsverlust. Obwohl die Auswirkung eines höheren Nullenüberschusses auf die Verzerrung nicht eindeutig bestimmt werden kann, steigt die mittlere quadratische Abweichung doch insgesamt an. Dem entgegen scheinen die relativen Fehler zu sinken.  
Der Schätzer  $\hat{\varphi}$  bleibt von einer Änderung des Zero-Inflation-Anteils im wesentlichen unberührt.

Eine Ursache der schlechten Schätzgüte von  $\hat{\alpha}_0$  und  $\hat{\alpha}_1$  bei Design A könnte die Verwendung der selben Designmatrix für  $\mu$ - und  $\omega$ -Regression sein. Mit Design B erreichen wir auch tatsächlich eine erheblich bessere Güte der Schätzer  $\hat{\alpha}_0$  und  $\hat{\alpha}_1$ . Die Präzisionssteigerung ist jedoch nicht dafür ausreichend, dass  $\hat{\alpha}_0$  und  $\hat{\alpha}_1$  nun das Erwartungswert-Varianz-Kriterium (8.2) erfüllen. Wählen wir Design C, so verbessern sich zwar auch die Schätzer  $\hat{\alpha}_0$  und  $\hat{\alpha}_1$ , sehr deutlich fällt aber die Verbesserung der Schätzgüte von  $\hat{\beta}_0$  und  $\hat{\beta}_1$  aus. Das Erwartungswert-Varianz-Kriterium (8.2) erfüllen  $\hat{\alpha}_0$  und  $\hat{\alpha}_1$  wiederum nicht.

Zusammenfassung:

In den meisten Fällen sind die Schätzer  $\hat{\alpha}_0$  und  $\hat{\alpha}_1$ , welche mit der  $\omega$ -Regression korrespondieren, unbrauchbar. Diese sind in hohem Maße von dem Parameterbereich  $[\omega_{\min}, \omega_{\max}]$  abhängig. Die Verwendung unterschiedlicher Kovariablen für  $\omega$ - und  $\mu$ -Regression, führen bei vergleichbarem Parameterbereich  $\Omega$  zwar auf zum Teil erheblich bessere Schätzungen, das Identifizierbarkeitsproblem der Schätzer  $\hat{\alpha}_0$  und  $\hat{\alpha}_1$  wird dadurch aber nicht gelöst. Die Parameter  $\varphi$ ,  $\beta_0$  und  $\beta_1$  werden für alle betrachteten Designs und Parameterkonstellationen im Wesentlichen zufriedenstellend geschätzt. Damit das Verfahren für alle Parameter akzeptable Schätzer liefert scheint insgesamt ein höherer Stichprobenumfang erforderlich zu sein.

$\varphi$	$\mu$	$\Omega$	Parameter	Schätzer					
				Bias		MSE		relativer Fehler	
1.25	3	[0.08, 0.18]	$\varphi = 1.25$	-0.02	(0.01)	0.02	(0.01)	-0.02	(0.01)
			$\mu = 3$	0.03	(0.02)	0.06	(0.05)	0.01	(0.01)
			$\alpha_0 = -2$	-18.84	(18.22)	33551.22	( $10^9$ )	9.42	(9.11)
			$\alpha_1 = 0.5$	19.50	(18.79)	35692.64	( $10^9$ )	38.99	(37.58)
	[0.02, 0.50]	$\varphi = 1.25$	0.00	(0.01)	0.02	(0.01)	0.00	(0.01)	
		$\mu = 3$	-0.02	(0.03)	0.09	(0.07)	-0.01	(0.01)	
		$\alpha_0 = -2$	-1.03	(0.47)	23.06	(12937.33)	0.51	(0.23)	
		$\alpha_1 = 2$	1.15	(0.55)	31.78	(24463.78)	0.57	(0.28)	
7	[0.08, 0.18]	$\varphi = 1.25$	-0.02	(0.01)	0.01	(0.00)	-0.02	(0.01)	
		$\mu = 7$	0.96	(0.04)	1.09	(0.82)	0.14	(0.01)	
		$\alpha_0 = -2$	-0.08	(0.04)	0.15	(0.10)	0.04	(0.02)	
		$\alpha_1 = 0.5$	0.02	(0.06)	0.40	(0.46)	0.04	(0.13)	
[0.02, 0.50]	$\varphi = 1.25$	-0.00	(0.01)	0.01	(0.00)	-0.00	(0.01)		
	$\mu = 7$	0.99	(0.04)	1.13	(0.79)	0.14	(0.01)		
	$\alpha_0 = -2$	-0.14	(0.05)	0.25	(0.15)	0.07	(0.02)		
	$\alpha_1 = 2$	0.22	(0.07)	0.58	(1.14)	0.11	(0.04)		
2.00	3	[0.08, 0.18]	$\varphi = 2$	-0.03	(0.03)	0.07	(0.04)	-0.02	(0.01)
			$\mu = 3$	1.83	(0.05)	3.63	(4.23)	0.61	(0.02)
			$\alpha_0 = -2$	-28.17	(27.75)	77788.00	( $10^{10}$ )	14.09	(13.87)
			$\alpha_1 = 0.5$	30.19	(29.74)	89339.80	( $10^{10}$ )	60.37	(59.47)
	[0.02, 0.50]	$\varphi = 2$	0.01	(0.02)	0.05	(0.02)	0.00	(0.01)	
		$\mu = 3$	1.85	(0.06)	3.77	(5.25)	0.62	(0.02)	
		$\alpha_0 = -2$	-0.94	(0.38)	15.40	(3621.89)	0.47	(0.19)	
		$\alpha_1 = 2$	1.14	(0.44)	20.98	(6836.82)	0.57	(0.22)	
7	[0.08, 0.18]	$\varphi = 2$	-0.05	(0.02)	0.03	(0.01)	-0.02	(0.01)	
		$\mu = 7$	5.75	(0.07)	33.47	(62.11)	0.82	(0.01)	
		$\alpha_0 = -2$	-0.10	(0.03)	0.13	(0.13)	0.05	(0.02)	
		$\alpha_1 = 2$	0.07	(0.06)	0.39	(0.60)	0.15	(0.12)	
[0.02, 0.50]	$\varphi = 2$	0.01	(0.02)	0.03	(0.02)	0.00	(0.01)		
	$\mu = 7$	5.84	(0.08)	34.75	(87.13)	0.83	(0.01)		
	$\alpha_0 = -2$	-0.06	(0.05)	0.21	(0.28)	0.03	(0.02)		
	$\alpha_1 = 2$	0.09	(0.08)	0.71	(4.32)	0.05	(0.04)		

Tabelle 8.7: Bias, MSE und relativer Fehler der Maximum-Likelihood-Schätzer im ZIGP $\omega$ -Regressionsmodell bei Stichprobenumfang n=100

## 8.5 ZIGP $\omega$ -Regression

### 8.5.1 Datensimulation

Wird gegenüber der ZIGP $\mu\omega$ -Regression ein konstanter Parameter  $\mu$  unterstellt, so erhalten wir ZIGP $\omega$ -Regression. Für  $\omega_i$  gilt:

$$\omega_i \stackrel{(6.55)}{=} \frac{\exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\}}{1 + \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\}}$$

Die Erwartungswertfunktion lautet im ZIGP $\omega$ -Modell:

$$E(Y_i | \mathbf{x}_i) \stackrel{(6.56)}{=} \frac{\mu}{1 + \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\}}$$

Mit der S-Plus Funktion `response.zigp3` (Anhang C.3.5) erzeugen wir je 100 Responsevektoren  $\mathbf{Y}$  für

$$\begin{aligned} \mu' &= 2 \\ \mu'' &= 4 \\ \mu''' &= 7 \end{aligned}$$

Dispersionsgrad und Regressionsparameter  $\boldsymbol{\alpha}$  wählen wir wie bei Design A in Abschnitt 8.4.1, als Stichprobenumfang verwenden wir  $n=100$ . Die Maximum-Likelihood-Schätzer der Parameter  $\varphi$ ,  $\mu$ ,  $\alpha_0$  und  $\alpha_1$  bestimmen wir mit der S-Plus Funktion `mle.zigp3` (Anhang C.3.6).

### 8.5.2 Güte der Maximum-Likelihood-Schätzer

Den Simulationsdaten (siehe Tabelle 8.7) entnehmen wir, dass die Schätzer  $\hat{\alpha}_0$  und  $\hat{\alpha}_1$  bei  $\mu = 3$  und geringer Zero-Inflation ( $\Omega = [0.08, 0.18]$ ) das Erwartungs-Varianz-Kriterium (8.2) verletzen.  $\hat{\alpha}_0$  und  $\hat{\alpha}_1$  scheinen bei der genannten Parameterkonstellation nicht identifizierbar zu sein. Der Schätzer  $\hat{\mu}$  scheint systematisch zu überschätzen. In den meisten Fällen liegt das 95%-Konfidenzintervall seiner Verzerrung (8.7) komplett im positiven Wertebereich. Unter allen Schätzern besitzt  $\hat{\varphi}$  die höchste Güte. Stets weist  $\hat{\varphi}$  die höchste Präzision und die niedrigste mittlere quadratische Abweichung aus, meist ist auch die Verzerrung und der relative Fehler dieses Schätzers am geringsten. Betrachten wir erneut den Einfluss der einzelnen Parameterkonstellationen auf die Güte der Schätzer:

- Dispersionsparameter  $\varphi$ :  
Die Erhöhung des Dispersionsgrades hat nur eine geringe Auswirkung auf die Güte des Schätzers  $\hat{\varphi}$ . Mit  $\varphi$  steigen Verzerrung, Standardabweichung, mittlere quadratische Abweichung und der relative Fehler nur leicht. Ganz im Gegensatz dazu erhöhen sich entsprechende Größen von  $\hat{\mu}$ , allen voran die Verzerrung, recht deutlich wenn der Dispersionsgrad steigt. Die Simulationsdaten von  $\hat{\alpha}_0$  und  $\hat{\alpha}_1$  zeigen ein geteiltes Bild. Während Verzerrung, Standardabweichung, mittlere quadratische Abweichung und der relative Fehler für steigendes  $\varphi$  bei  $\mu = 3$  stark anwachsen, sinken diese Größen, wenn den Daten der Parameter  $\mu = 7$  zu Grunde liegt.
- Parameterbereich  $\Omega$ :  
Durch Ausweitung des Nullenüberschusses werden im ZIGP $\omega$ -Modell der Regression mehr

Daten zur Verfügung gestellt. Wir erwarten daher eine Verbesserung der Schätzgüte von  $\hat{\alpha}_0$  und  $\hat{\alpha}_1$ , wenn wir  $\Omega = [0.08, 0.18]$  gegen  $\Omega = [0.02, 0.50]$  austauschen. Dies trifft auch zu, falls  $\mu = 3$  gilt. Dem entgegen steigen Verzerrung, Standardabweichung, mittlere quadratische Abweichung und der relative Fehler, falls  $\mu = 7$  der wahre Parameter ist. Für beide Werte von  $\mu$  erfährt der Schätzer  $\hat{\varphi}$  eine Verbesserung seiner Güte bei Ausweitung des  $\omega$ -Bereichs, während die Standardabweichung, die mittlere quadratische Abweichung, der relative Fehler und meist auch die Verzerrung von  $\hat{\mu}$  leicht steigen.

- Parameter  $\mu$ :

Auffallend ist die ausgeprägte Verschlechterung der Güte von  $\hat{\mu}$ , wenn der wahre Parameter  $\mu$  erhöht wird. Die Güte der übrigen Schätzer steigt in diesem Fall. Bei  $\hat{\varphi}$  ist dies deutlicher zu beobachten als bei  $\hat{\alpha}_0$  und  $\hat{\alpha}_1$ .

Zusammenfassung:

Der Schätzer  $\hat{\varphi}$  besitzt eine hohe Güte. Bei geringem Nullenüberschuss und kleinem Parameter  $\mu$  treten Schwierigkeiten bei der Identifizierbarkeit der Schätzer  $\hat{\alpha}_0$  und  $\hat{\alpha}_1$  auf. Für die übrigen Parameterkonstellationen werden akzeptable Schätzungen geliefert. Der Schätzer  $\hat{\mu}$  ist zum Teil stark verzerrt.

$\lambda$	$\alpha$	Parameter	Schätzer					
			Bias		MSE		relativer Fehler	
0.2	$\alpha'$	$\lambda = 0.2$	-0.07	(0.01)	0.02	(0.01)	-0.34	(0.66)
		$\alpha_0 = 1$	11.73	(5.69)	3378.34	( $10^8$ )	11.73	(5.69)
		$\beta_0 = -1$	1.86	(0.03)	3.56	(1.36)	-1.86	(0.03)
		$\beta_1 = 3$	-5.98	(0.03)	35.89	(16.14)	-1.99	(0.11)
	$\alpha'''$	$\lambda = 0.2$	-0.07	(0.01)	0.02	(0.01)	-0.33	(0.06)
		$\alpha_0 = 1$	8.49	(4.44)	2043.72	( $10^8$ )	8.49	(4.44)
		$\alpha_1 = 0.5$	6.36	(4.82)	2361.56	( $10^8$ )	12.73	(9.46)
		$\beta_0 = -1$	1.85	(0.03)	3.55	(1.42)	-1.85	(0.03)
		$\beta_1 = 3$	-6.00	(0.04)	26.16	(18.54)	-2.00	(0.01)

Tabelle 8.8: Bias, MSE und relativer Fehler der Maximum-Likelihood-Schätzer im ZHGP-Regressionsmodell bei Stichprobenumfang  $n=20$



$\lambda$	$\alpha$	Parameter	Schätzer					
			Bias		MSE		relativer Fehler	
0.2	$\alpha'$	$\lambda = 0.2$	-0.03	(0.01)	0.01	(0.00)	-0.13	(0.03)
		$\alpha_0 = 1$	-0.20	(0.01)	0.06	(0.01)	-0.20	(0.01)
		$\beta_0 = -1$	-0.20	(0.06)	0.37	(0.25)	0.20	(0.06)
		$\beta_1 = 3$	0.02	(0.06)	0.41	(0.44)	0.01	(0.02)
	$\alpha''$	$\lambda = 0.2$	-0.05	(0.01)	0.02	(0.00)	-0.25	(0.06)
		$\alpha_0 = -1$	-0.28	(0.03)	0.20	(0.05)	0.28	(0.03)
		$\alpha_1 = 3$	0.21	(0.06)	0.43	(0.45)	0.07	(0.02)
		$\beta_0 = -1$	-0.32	(1.12)	1.46	(19.16)	0.32	(0.12)
	$\alpha'''$	$\beta_1 = 3$	0.12	(0.12)	1.51	(20.13)	0.04	(0.04)
		$\lambda = 0.2$	-0.02	(0.01)	0.00	(0.00)	-0.09	(0.03)
		$\alpha_0 = 1$	16.70	(14.26)	20600.08	( $10^9$ )	16.70	(14.26)
		$\alpha_1 = 0.5$	15.08	(14.47)	21174.71	( $10^9$ )	30.17	(28.95)
0.8	$\alpha'$	$\beta_0 = -1$	-0.23	(0.06)	0.41	(0.52)	0.23	(0.06)
		$\beta_1 = 3$	0.02	(0.06)	0.42	(0.83)	0.01	(0.02)
		$\lambda = 0.8$	-0.03	(0.01)	0.01	(0.00)	-0.04	(0.01)
		$\alpha_0 = 1$	-1.62	(0.02)	2.64	(0.22)	-1.62	0.02
	$\alpha''$	$\beta_0 = -1$	-2.97	(0.48)	31.94	(4552.17)	2.97	(0.48)
		$\beta_1 = 3$	-1.63	(0.54)	32.17	(5058.67)	-0.54	0.18
		$\lambda = 0.8$	-0.09	(0.01)	0.03	(0.01)	-0.11	(0.02)
		$\alpha_0 = -1$	-1.81	(0.07)	3.81	(13.03)	1.81	(0.07)
	$\alpha'''$	$\alpha_1 = 3$	0.26	(0.10)	1.06	(10.03)	0.09	(0.03)
		$\beta_0 = -1$	-3.65	(0.77)	71.94	(34100.28)	3.65	(0.77)
		$\beta_1 = 3$	1.39	(0.85)	74.92	(35578.11)	0.46	(0.28)
		$\lambda = 0.8$	-0.02	(0.01)	0.00	(0.00)	-0.02	(0.01)
$\alpha''$	$\alpha_0 = 1$	-0.60	(0.01)	0.38	(0.01)	-0.60	(0.01)	
	$\alpha_1 = 0.5$	0.01	(0.02)	0.03	(0.00)	0.01	(0.04)	
	$\beta_0 = -1$	-2.63	(0.39)	22.47	(2415.47)	2.63	(0.39)	
	$\beta_1 = 3$	0.02	(0.48)	23.50	(2932.28)	0.01	(0.16)	

Tabelle 8.9: Bias, MSE und relativer Fehler der Maximum-Likelihood-Schätzer im ZHGP-Regressionsmodell bei Stichprobenumfang  $n=100$

## 8.6 ZHGP-Regression

### 8.6.1 Datensimulation

Mit der S-Plus Funktion `response.zhgp` (Anhang C.4.1) erzeugen wir entsprechend Abschnitt 8.2.1 für unterschiedlichen Stichprobenumfang  $n$  und verschiedene Parameterwerte von  $\varphi$  je 100 unabhängig identisch verteilte Responsevektoren  $\mathbf{Y}$ . Im ZHGP-Regressionsmodell wird  $Y_i|\mathbf{x}_i \sim ZHGP(\gamma_i, \theta_i, \lambda)$  angenommen. Die Regressionsparameter werden wie folgt spezifiziert:

$$\begin{aligned}\gamma_i &\stackrel{(7.26)}{=} \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\} \\ \theta_i &\stackrel{(7.27)}{=} \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}\end{aligned}$$

Die Erwartungswertfunktion lautet:

$$E(Y_i|\mathbf{x}_i) \stackrel{(7.28)}{=} \frac{1 - \exp\{-\exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\}\}}{1 - \exp\{-\exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}\}} \frac{\exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}}{1 - \lambda}$$

Für den zusätzlichen Regressionsparameter  $\boldsymbol{\alpha}$ , welcher mit dem Parameter  $\gamma$  der Hürdenverteilung korrespondiert (vgl. Abschnitt 7.4), wählen wir

$$\boldsymbol{\alpha}' = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \boldsymbol{\alpha}'' = \begin{pmatrix} -1 \\ 3 \end{pmatrix}, \quad \boldsymbol{\alpha}''' = \begin{pmatrix} 1 \\ 0.5 \end{pmatrix}$$

als Regressionsparameter für die  $\theta$ -Regression wird

$$\boldsymbol{\beta}'' = \begin{pmatrix} -1 \\ 3 \end{pmatrix}$$

verwendet.  $\mathbf{X}$  dient uns erneut als Modellmatrix.

Zunächst wollen wir beobachten, wie sich die Schätzer verhalten, wenn auf  $\gamma$  keine Regression angewendet wird. Dazu dient Vektor  $\boldsymbol{\alpha}'$ . In diesem Fall ist die Wahrscheinlichkeit des Ereignisses Null für alle Beobachtungen gleich hoch. Für Beobachtungen mit Parameter  $\theta_i > \gamma = e^1 \approx 2.72$  (17% aller Beobachtungen) liegt deshalb Nullenüberschuss vor, ist  $\theta_i$  kleiner als  $\gamma$  ist die Wahrscheinlichkeit des Ereignisses Null für diese Beobachtung geringer als im Ursprungsmodell (vgl. Abschnitt 7.4).

Verwenden wir für die  $\gamma$ -Regression den Regressionsparameter  $\boldsymbol{\alpha}''$ , so stimmen Ursprungs- und Hürdenverteilung überein. Mit dem dritten Regressionsparameter ( $\boldsymbol{\alpha}'''$ ) modellieren wir Daten, die in allen Beobachtungen Nullenüberschuss aufweisen, da für alle  $x_i$  die Ungleichung  $\gamma_i < \theta_i$  gilt.

Die Maximum-Likelihood-Schätzer bestimmen wir jeweils mit der S-Plus Funktion `mle.zhgp` (Anhang C.4.2).

### 8.6.2 Güte der Maximum-Likelihood-Schätzer

In den Tabellen 8.8 und 8.9 sind für einige Parameterkonstellationen die Kennzahlen für die Maximum-Likelihood-Schätzer im ZHGP-Modell angegeben. Für alle Parameterkonstellationen besitzt der Schätzer  $\hat{\lambda}$  die höchste Güte. Dies liegt vor allem daran, dass er stets die höchste Präzision besitzt und meist auch die geringste Verzerrung aufweist.  $\lambda$  wird allerdings stets unterschätzt. Für das Verhalten der Schätzer bei wechselnden Parameterkonstellationen und unterschiedlichem Stichprobenumfang können wir folgende Aussagen treffen:

- Stichprobenumfang  $n$ :  
Deutlich ist ein Gütezuwachs der Schätzer für ansteigenden Stichprobenumfang zu verzeichnen. Besonders deutlich sinkt die mittlere quadratische Abweichung für die Schätzer  $\hat{\alpha}_0$  und  $\hat{\alpha}_1$ .
- Dispersionsparameter  $\lambda$ :  
Bis auf einen Fall können wir einen Anstieg der mittleren quadratischen Abweichung bei Vergrößerung des Dispersionsgrades feststellen. Am stärksten ist dieser Anstieg bei  $\hat{\beta}_0$  und  $\hat{\beta}_1$  zu beobachten, wogegen die Verschlechterung bei  $\hat{\lambda}$  nur minimal ist. Während die Präzision in fast allen Fällen abnimmt, für  $\hat{\beta}_0$  und  $\hat{\beta}_1$  am deutlichsten, steigt die Verzerrung meist an. Eine Ausnahme liegt bei der Wahl von  $\alpha'''$  vor. Steigt der Dispersionsgrad, so erfahren die Schätzer von  $\alpha_0$  und  $\alpha_1$  hier eine Verbesserung.
- $\gamma$ -Regressionsparameter:  
Auf die Güte des Schätzers  $\hat{\lambda}$  hat die Wahl des Parameters  $\alpha$  kaum eine Auswirkung. Am besten wird  $\lambda$  jedoch bei  $\alpha'''$  geschätzt, am schlechtesten bei  $\alpha''$ . Einen größeren Einfluss hat  $\alpha$ , wie zu erwarten war, auf die Schätzer  $\hat{\alpha}_0$  und  $\hat{\alpha}_1$ . Dabei lassen sich zum Teil große Unterschiede in der Schätzgüte ausmachen. Meist besitzen diese Schätzer bei  $\alpha'$  die höchste und bei  $\alpha'''$  die geringste Güte. Gilt  $\alpha'''$  so erfüllen  $\hat{\alpha}_0$  und  $\hat{\alpha}_1$  in einigen das Erwartungswert-Varianz-Kriterium (8.2) nicht. Die Parameter  $\beta_0$  und  $\beta_1$  werden bei der Wahl von  $\alpha'$  meist am besten und bei  $\alpha''$  am schlechtesten geschätzt.
- $\theta$ -Regressionsparameter:  
Wie bei  $\alpha$  kann auch für verschiedene  $\beta$  kein nennenswerter Unterschied der Schätzgüte von  $\hat{\lambda}$  ausgemacht werden. Meist ist die Verzerrung bei  $\beta'$  jedoch etwas geringer als bei  $\beta''$ . Für  $\hat{\alpha}_0$  und  $\hat{\alpha}_1$  lässt sich keine eindeutige Entwicklung verzeichnen, tendenziell sinkt die mittlere quadratische Abweichung jedoch bei dem Übergang von  $\beta'$  zu  $\beta''$ . Für  $\hat{\beta}_0$  und  $\hat{\beta}_1$  können wir dabei in den meisten Fällen eine Erhöhung der Präzision beobachten, die zu einer Reduktion der mittleren quadratischen Abweichung führt. Bei hohem Dispersionsgrad sind die Schätzer  $\hat{\beta}_1$  nicht signifikant von Null verschieden.

Zusammenfassung:

Der Dispersionsparameter  $\lambda$  wird stets sehr gut geschätzt. Um für die Parameter  $\alpha_0$  und  $\alpha_1$  gute Schätzungen zu erhalten, ist ein hoher Stichprobenumfang erforderlich. Die Schätzer für  $\beta_0$  und  $\beta_1$  sind bei hohem Dispersionsgrad kaum noch zu gebrauchen.

## 8.7 Missspezifikation

In den vorangegangenen Abschnitten wurden die verschiedenen Schätzmethoden jeweils auf Datensätze angewendet, deren Responsevektoren gemäß dem jeweiligen Modell (im folgenden Originalmodell genannt) erzeugt wurden. Nun wollen wir unseren Blick auf die Anpassungsgüte von Modellen mit falscher Spezifikation richten. Da wir auch nicht eingebettete Regressionsmodelle miteinander vergleichen wollen, scheidet das Akaike-Informationskriterium zur Messung der Anpassungsgüten aus. Statt dessen wollen wir die Fehlerquadratsumme

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8.25)$$

verwenden. Die angepassten Werte sind hier mit  $\hat{y}_i$  bezeichnet. Für den Vergleich der Anpassungsqualität des misspezifizierten Regressionsmodells  $\tilde{M}$  mit der Anpassungsqualität des Originalmodells  $M$  bilden wir den Quotienten der beiden Fehlerquadratsummen  $RSS^{\tilde{M}}$  und  $RSS^M$ :

$$Q = \frac{RSS^{\tilde{M}}}{RSS^M} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i^{\tilde{M}})^2}{\sum_{i=1}^n (y_i - \hat{y}_i^M)^2} \quad (8.26)$$

$M$  bezeichnet dabei das Originalmodell,  $\tilde{M}$  das misspezifizierte Modell. Für die Statistik (8.26) gilt  $0 < Q < \infty$ . Das misspezifizierte Modell ist bezüglich seiner Anpassungsgüte mit dem Originalmodell vergleichbar, wenn wir  $Q \approx 1$  erhalten. Ist  $Q$  wesentlich größer als 1, schätzt das Originalmodell deutlich besser.

Für die Modelle GP, ZIGP $\mu$ , ZIGP $\mu\omega$ , ZIGP $\omega$  und ZHGP erzeugen wir jeweils  $k=100$  unabhängige identisch verteilte Responsevektoren  $\mathbf{Y}_j = (y_{1j}, \dots, y_{nj})^t$  mit Stichprobenumfang  $n=100$ . Der Dispersionsparameter beträgt  $\varphi = 1.25$ . Weiterhin wurden  $\mu = 7/M = [0.02, 7.39]$  und  $\omega = 0.5/\Omega = [0.02, 0.50]$  und die Modellmatrix  $\mathbf{X}$  (8.18) gewählt. Vergleichen wir die Anpassungsgüte eines misspezifizierten Modells  $\tilde{M}$  mit einem Originalmodell  $M$ , so wird für jede der Realisierungen  $\mathbf{Y}_j$  der Quotient  $Q_j = Q$  berechnet. In Tabelle 8.10 sind die arithmetischen Mittel  $\tilde{Q} = \frac{1}{k} \sum_{j=1}^k Q_j$  für jede Kombination aus misspezifiziertem Modell und Originalmodell zu finden.

Daten- erzeugung	Schätzverfahren				
	GP	ZIGP $\mu$	ZIGP $\mu\omega$	ZIGP $\omega$	ZHGP
GP	1.000	0.997	0.679	1.552	0.642
ZIGP $\mu$	1.818	1.000	1.344	1.139	3.893
ZIGP $\mu\omega$	1.359	2.051	1.000	1.840	5.639
ZIGP $\omega$	1.053	1.104	0.974	1.000	0.979
ZHGP	1.709	1.687	1.707	3.370	1.000

Tabelle 8.10: Quotienten aus der Fehlerquadratsumme misspezifizierter Modelle und der Fehlerquadratsumme der Originalmodelle

Aus Spalte 1 von Tabelle 8.10 entnehmen wir, dass die Anpassungsqualität des GP-Verfahrens stets schlechter ist, als die der Originalmodelle. Umgekehrt liefert die  $ZIGP_{\mu}$ -Regression für GP-Daten eine bessere Anpassung als das GP-Modell selbst, für  $ZIGP_{\mu\omega}$ -Daten ist die Fehlerquadratsumme (8.25) hingegen doppelt so hoch wie bei dem Originalmodell.  $ZIGP_{\mu\omega}$  führt bei GP- und  $ZIGP_{\omega}$ -Daten zu einer besseren, bei  $ZIGP_{\mu}$ - und ZHGP-Daten zu einer schlechteren Anpassung. Die Anpassungsqualität des  $ZIGP_{\omega}$ -Modells ist für alle Daten zum Teil wesentlich schlechter als das jeweilige Originalmodell. Das ZHGP-Regressionsmodell ist mit den übrigen Modellen am wenigsten vergleichbar. Während es für GP-Daten deutlich bessere, für  $ZIGP_{\omega}$ -Daten bessere Schätzungen liefert, ist die Anpassungsqualität für  $ZIGP_{\mu}$ - und  $ZIGP_{\mu\omega}$ -Daten erheblich schlechter als im Originalmodell.

Da der Mechanismus zur Erzeugung von  $ZIGP_{\mu\omega}$ - und ZHGP-Daten unter den betrachteten Modellen am komplexesten ist, ergibt sich für diese Daten der größte Unterschied zwischen misspezifiziertem Modell und Originalmodell. Aus der obersten Zeile von Tabelle 8.10 entnehmen wir, dass für GP-Daten von allen Modellen, mit Ausnahme von  $ZIGP_{\omega}$ , besser angepasste Werte berechnet werden. Für  $ZIGP_{\mu}$  und ZHGP ist dies auch nicht weiter verwunderlich, da das GP-Modell in diese eingebettet ist. Das selbe gilt für  $ZIGP_{\omega}$ -Daten und mit dem  $ZIGP_{\mu\omega}$ -Modell angepasste Werte.



# Kapitel 9

## Anwendungsbeispiele

Nachdem wir verschiedene, auf der verallgemeinerten Poissonverteilung basierende Regressionsmodelle hergeleitet und deren Maximum-Likelihood-Schätzer untersucht haben, wollen wir die Verfahren nun auf zwei Datensätze anwenden.

### 9.1 Patent-Datensatz

Quelle für den untersuchten Datensatz:

<http://coe.ubc.ca/users/marty>

#### 9.1.1 Datenbeschreibung

Der vorliegende Datensatz umfasst Daten zur Patentanmeldung von 70 US-amerikanischen High-Tech-Unternehmen aus dem Jahre 1976. Es sind folgende Spalten vorhanden:

Spalte	Bezeichnung	Beschreibung
1	<code>name</code>	Name des Unternehmens
2	<code>patents</code>	Anzahl der 1976 angemeldeten Patente
3	<code>rdsales</code>	Verhältnis zwischen Ausgaben für Forschung & Entwicklung und dem Verkaufserlös
4	<code>logrd</code>	Logarithmierte Ausgaben für Forschung & Entwicklung, wobei diese in der Einheit 1 Million US-Dollar angegeben werden
5	<code>logrd2</code>	Quadrierte Werte aus Spalte 4

Tabelle 9.1: Spalten des Patent-Datensatzes

Die Daten in Spalten 2 bis 5 liegen auf einer metrischen Skala, sie sind daher ungruppiert. Als abhängige Größe betrachten wir die Einträge in Spalte `patents`. Die Beobachtungen in dieser Spalte beziehen sich alle samt auf das Jahr 1976. Auf einen Offset kann somit verzichtet werden. Da die Anzahl der Patente nicht-negativ und ganzzahlig ist, liegen Zähldaten vor.

#### 9.1.2 Explorative Datenanalyse

Im Vorgriff auf die Modellierung, betrachten wir im folgenden den Zusammenhang der unabhängigen Größen mit der logarithmierten Anzahl an Patentanmeldungen. Dazu führen wir

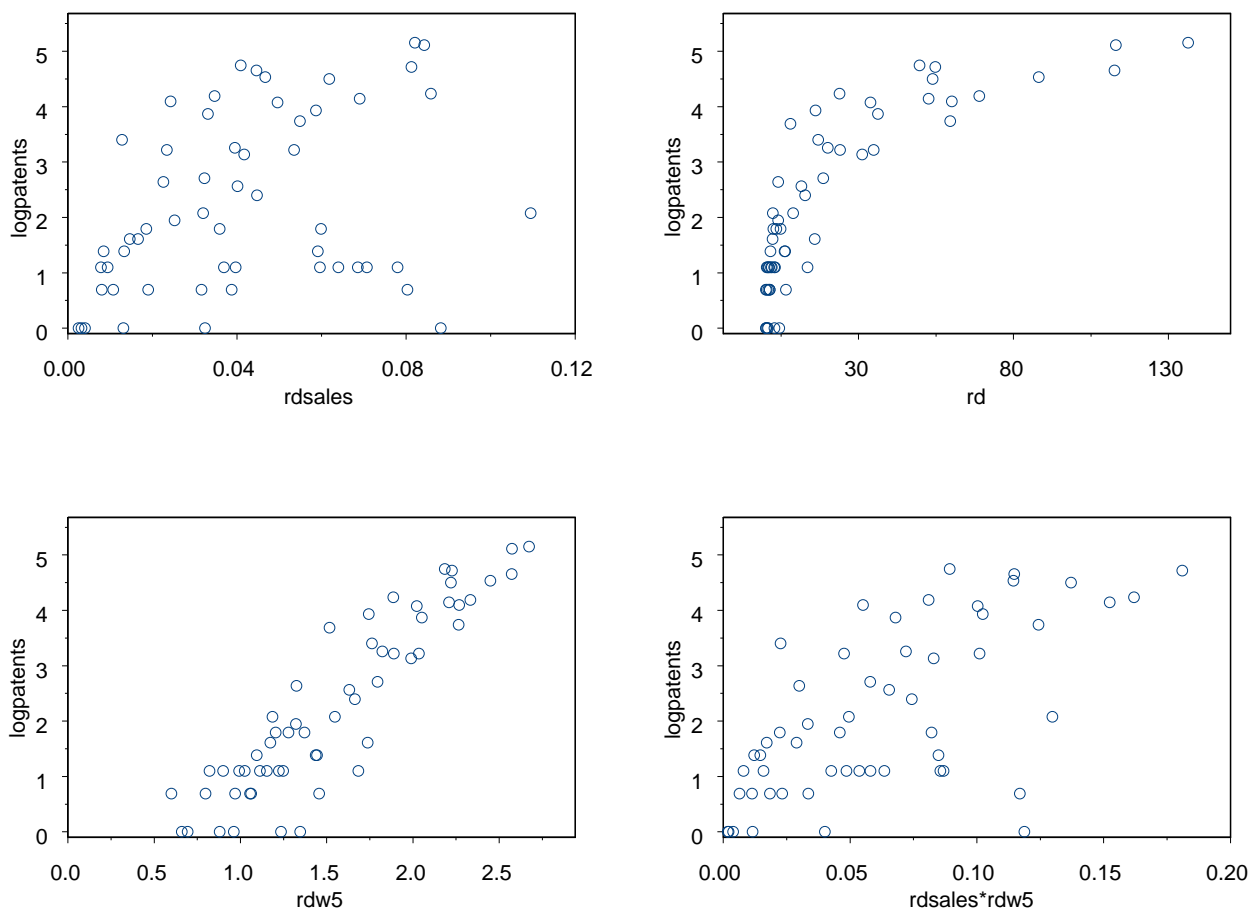


Abbildung 9.1: Explorative Datenanalyse der Patentdaten

die transformierte Zielgröße `logpatents` ein. Tragen wir `rdsales` gegen `logpatents` ab, so erkennen wir nur sehr schwach einen linearen Zusammenhang (siehe Abbildung 9.1). Recht deutlich ist jedoch der Zusammenhang zwischen den Ausgaben für Forschung & Entwicklung und `logpatents`. Aus der Abbildung `rd` gegen `logpatents` können wir einen Verlauf erkennen, der dem Graph einer Wurzel- oder Logarithmusfunktion ähnelt. Dabei stellt die Variable `rd` eine Rücktransformation der Größe `logrd` auf ihre Ursprungseinheit dar. Um einen linearen Zusammenhang mit `logpatents` herzustellen, müssen wir daher eine weitere Transformation der Variable `rd` durchführen. Dazu eignet sich sowohl die Wurzel- als auch die Logarithmusfunktion. Sikora (2002) wählt bei ihrer Analyse der Patentdaten als Transformation die 5. Wurzel. Die neue Variable `rdw5` zeigt einen annähernd linearen Zusammenhang mit `logpatents`. Um Vergleichbarkeit herzustellen, wollen auch wir diese Transformation verwenden. Zusätzlich zu den Kovariablen `rdsales` und `rdw5` scheint deren Interaktion eine Einflussgröße von `patents` zu sein (siehe Abbildung 9.1 rechts unten).

### 9.1.3 Poissonregression

Die Verteilungsannahme im gewöhnlichen Poisson-Regressionsmodell lautet:

$$\text{patents}[i] \sim \text{Poi}(\mu_i)$$



Als Verknüpfung wird der Logarithmus verwendet ( $\ln(\mu_i) = \mathbf{x}_i^t \boldsymbol{\beta}$ ). Für die Patent-Daten bedeutet dies:

$$\begin{aligned} \ln(\text{patents}[i]) &= \text{logpatents}[i] \\ &= \beta_0 + \beta_1 \text{rdsales}[i] + \beta_2 \text{rdw5}[i] + \beta_3 (\text{rdsales}[i] * \text{rdw5}[i]) \end{aligned} \quad (9.1)$$

Aus diesem Grunde haben wir bei der explorativen Datenanalyse die Größe `logpatents` anstatt `patents` verwendet, als wir einen linearen Zusammenhang mit Kovariablen nachweisen wollten. Mit S-Plus wollen wir nun das Poissonmodell (9.1) berechnen. Die Kommandozeile dafür lautet:

$$\text{glm}(\text{patents} \sim \text{rdsales} + \text{rdw5} + \text{rdsales} * \text{rdw5}, \text{family}=\text{poisson}, \text{link}=\text{log}) \quad (9.2)$$

Die Auswertung des so angepassten Modells ergibt für alle Kovariablen (incl. Intercept) einen sehr geringen p-Wert (siehe Tabelle 9.2). Dieser gibt das Signifikanzniveau an, bis zu dem die Hypothese “Regressor beeinflusst die Kovariable” verworfen werden kann. Alle Kovariablen sind daher hoch signifikant.

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-4.81634	-1.719084	-0.8486778	0.1408976	7.279557

Coefficients:

	Value	Std. Error	t value	Pr(>  t )
(Intercept)	-1.165792	0.1850514	-6.299826	2.98e-10
rdsales	-5.500628	1.9909157	-2.762863	0.00573
rdw5	2.299744	0.1035849	22.201529	< 2e-16
rdsales:rdw5	3.903443	1.3119114	2.975386	0.00293

Tabelle 9.2: summary-Tabelle des Poissonmodells der Patentdaten

Werfen wir einen Blick auf die Residuen, so erkennen wir zum Teil große Unterschiede zwischen den wahren und den angepassten Werten (siehe Abbildung 9.2 rechts oben). Mittlere Werte scheinen eher über- während kleine und große Responsewerte eher unterschätzt werden. Die Grafik unten links in Abbildung 9.2 verdeutlicht, dass die Devianzresiduen für kleine Responsewerte im Verhältnis recht hoch sind. Dies deutet darauf hin, dass gerade kleine Werte schlecht geschätzt werden. Wir wollen deshalb untersuchen, ob die Verknüpfung richtig gewählt ist. Dazu tragen wir die geschätzten Werte des linearen Prädiktors (9.1) gegen die logarithmierte Anzahl der Patente (`logpatents`) ab. Ist die Linkspezifikation korrekt, so besteht ein linearer Zusammenhang und die Prädiktorwerte streuen um die erste Winkelhalbierende. Dies trifft für die vorliegenden Daten zu, wie aus der oberen linken Grafik in Abbildung 9.2 ersichtlich ist. Die Linkspezifizierung dürfte also korrekt sein.

Eine weitere Ursache für unzureichende Anpassungsqualität kann im Vorhandensein von Überdispersion liegen. Zur Überprüfung dieser Eigenschaft tragen wir die geschätzten Erwartungswerte  $\hat{\mu}_i$  gegen die geschätzten Varianzen  $(y_i - \hat{\mu}_i)^2$  ab. Bei äquidispersionierten Daten konzentrieren sich die so gebildeten Punkte um die erste Winkelhalbierende, da für poissonverteilte Daten  $\widehat{E}(y_i) \approx \widehat{Var}(y_i)$  gelten muss. In Abbildung 9.3 ist deutlich zu erkennen, dass die Varianz einiger Beobachtungen deutlich höher geschätzt wird, als deren Erwartungswert. Die Patentdaten sind überdispersioniert. Wir können daher versuchen die Anpassungsqualität zu verbessern, indem wir ein verallgemeinertes Poissonmodell aufstellen.

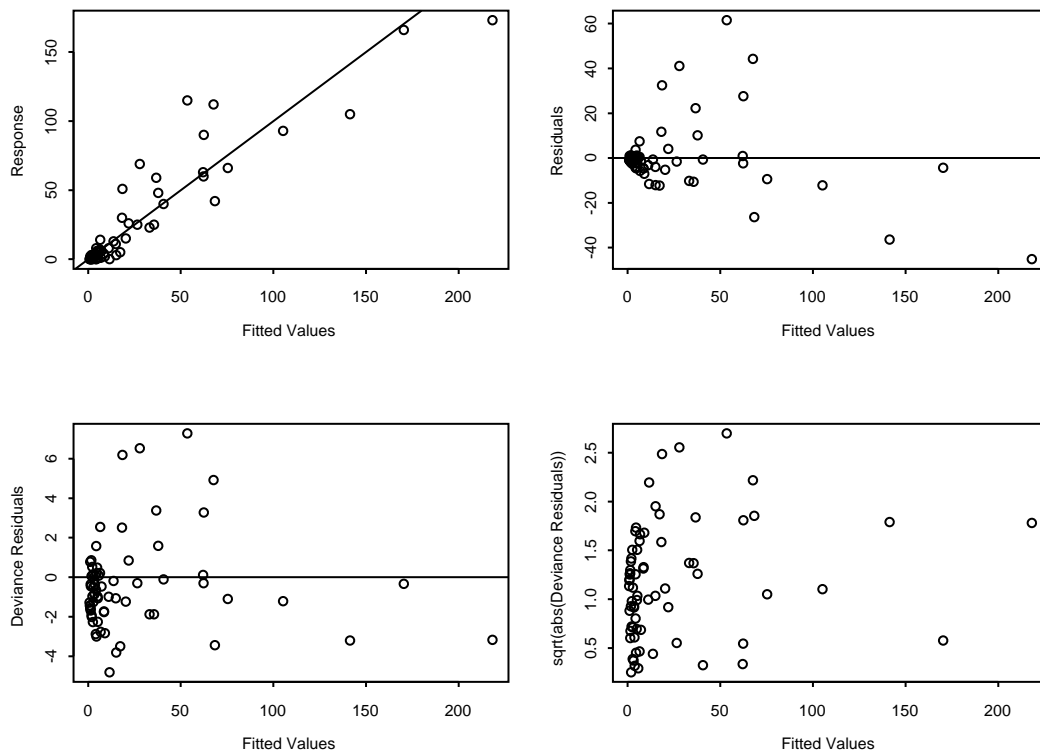


Abbildung 9.2: Angepasste Werte und Residuen des Poissonmodells der Patentdaten

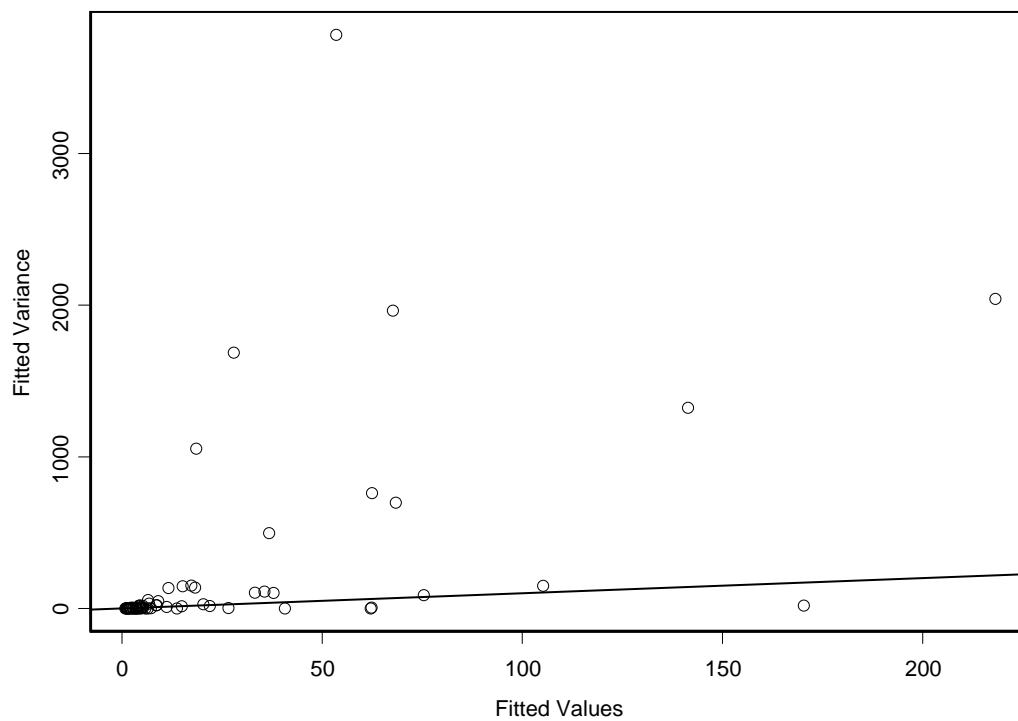


Abbildung 9.3: Überdispersion im Poissonmodell der Patentdaten

Um die Anpassungsqualität quantitativ vergleichen zu können, berechnen wir für das vorliegende Poissonmodell wie in Abschnitt 3.4 beschrieben die Fehlerquadratsumme und die Akaike-Information. Dabei ist die Fehlerquadratsumme RSS (3.33) ein reines Maß für die Qualität der Anpassung, während die Akaike-Information AIC (3.37) zusätzlich die Komplexität des Modells in die Bewertung miteinbezieht. Die ermittelten Werte sind in den Tabellen 9.10 und 9.11 zu finden<sup>1</sup>.

#### 9.1.4 Verallgemeinerte Poissonregression

Gelte nun die Annahme  $\text{patents}[i] \sim GP^*(\mu_i, \varphi)$ . Linearer Prädiktor und Verknüpfung stimmen mit (9.1) überein. Wir erhalten ein verallgemeinertes Poisson-Regressionsmodell. Das gewöhnliche Poissonmodell ist in dieses Modell eingebettet. Die Untersuchung auf Signifikanz der Regressoren muss daher nicht erneut vorgenommen werden. Wir können davon ausgehen, dass alle vier Kovariablen signifikant sind. Das entsprechende Regressionsmodell berechnen wir in S-Plus über den Programmaufruf:

```
mle.gpois(patents,
+         matrix(c(rep(1,70),rdsales,rdw5,rdsales*rdw5),70,4),rep(1,70)) (9.3)
```

Die S-Plus Funktion `summary.gpois` liefert uns für das berechnete Modell folgende Ausgabe:

	Min	1Q	Median	3Q	Max
Deviance Residuals:	-2.605943	-0.778651	-0.448552	0.107495	2.522061
Coefficients:	-0.881293	-7.557558	2.141986	5.269138	
Dispersion Parameter:	2.747298				
RSS:	14571.69				
AIC:	-10220.38				
Residual Deviance:	63.19094				
Range of mu:	0.555204	216.8011			
Message:	RELATIVE FUNCTION CONVERGENCE				

Tabelle 9.3: `summary.gpois`-Tabelle des verallgemeinerten Poissonmodells der Patentdaten

Sowohl der Schwankungsbereich der Devianzresiduen als auch die Devianz gehen recht deutlich zurück, was sich jedoch zum großen Teil mit dem zusätzlichen Dispersionsparameter erklären lässt, welcher den Wert 2.75 annimmt. An dem Bereich der Standardresiduen ändert sich nicht viel im Vergleich zum gewöhnlichen Poissonmodell. Dies können wir zum einen an den Werten in Tabelle 9.10 ablesen, zum anderen ist es aus der Abbildung 9.4 ersichtlich. Die Anpassungsqualität verbessert sich jedoch, da sich eine geringere Fehlerquadratsumme ergibt. Auch die

<sup>1</sup>Auf Grund numerischer Probleme (Speicherüberlauf) bei Berücksichtigung des Summanden  $\ln(y_i!)$  in den Likelihoodfunktionen der betrachteten Modelle, wird dieser konstante Ausdruck weggelassen. Die Vergleichbarkeit eingebetteter Modelle über AIC ist dennoch gewährleistet, da sich alle AIC-Werte im selben Umfang erhöhen.

Akaike-Information fällt um einiges günstiger aus, als bei der gewöhnlichen Poissonregression, obwohl ein zusätzlicher Parameter im Modell vorhanden ist.

### 9.1.5 ZIGP $\mu$ -Regression

Die Ursache der im vorangegangenen Modell ermittelte Überdispersion kann Nullenüberschuss sein. In Form der ZIP-Regression betrachten wir ein Modell, mit welchem Nullenüberschuss explizit modelliert werden kann. Wiederum verwenden wir dazu den linearen Prädiktor und die Verknüpfung (9.1). Für die Zielvariable gelte nun  $\text{patents}[i] \sim \text{ZIGP}(\mu_i, \varphi, \omega)$ . Wir berechnen das entsprechende ZIGP $\mu$ -Modell in S-Plus über die Kommandozeile:

```
mle.zigp1(patents,
+         matrix(c(rep(1,70),rdsales,rdw5,rdsales*rdw5),70,4),rep(1,70)) (9.4)
```

Mit `summary.zigp1` erhalten wir folgende Zusammenfassung:

```
ZI.Parameter:          0.015203
Coefficients:        -0.841596  -7.517641  2.127049  5.238711
Dispersion Parameter: 2.675360
RSS:                  14218.67
AIC:                  -10218.89
Range of mu:          0.573752   216.0171
Message:              RELATIVE FUNCTION CONVERGENCE
```

Tabelle 9.4: `summary.zigp1`-Tabelle des ZIGP $\mu$ -Modells der Patentdaten

Das Modell berechnet einen Nullenüberschuss, welcher dem Auswahlparameter 0.015 entspricht. Dieser führt zu einer Reduktion des Dispersionsparameter von 2.747298 im verallgemeinerten Poissonmodell auf nun 2.675360. Der Zero-Inflation-Anteil an der Überdispersion scheint also nicht sehr hoch zu sein. Die Verteilung der Residuen entspricht ihrer Verteilung im vorangegangenen Modell (siehe Abbildung 9.5). Der Mittelwert  $\bar{r}$  der Residuen liegt nun jedoch deutlich über Null (vgl. Tabelle 9.10). Trotzdem ist die Anpassungsqualität gemessen an der Fehlerquadratsumme höher als zuvor. Diese Verbesserung überträgt sich nicht auf die Akaike-Information. Der zusätzliche Parameter  $\omega$  führt dazu, dass die Akaike-Information nicht weiter sinkt, sondern leicht über dem AIC-Wert der verallgemeinerten Poissonregression liegt.

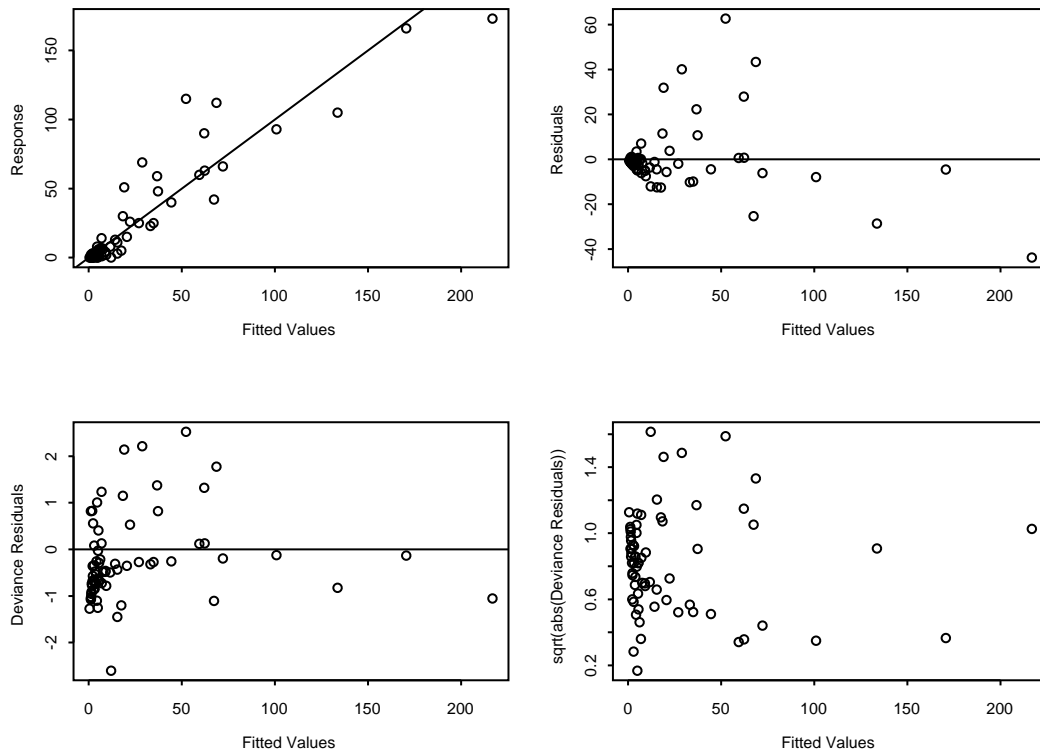


Abbildung 9.4: Angepasste Werte und Residuen des verallgemeinerten Poissonmodells der Patentdaten

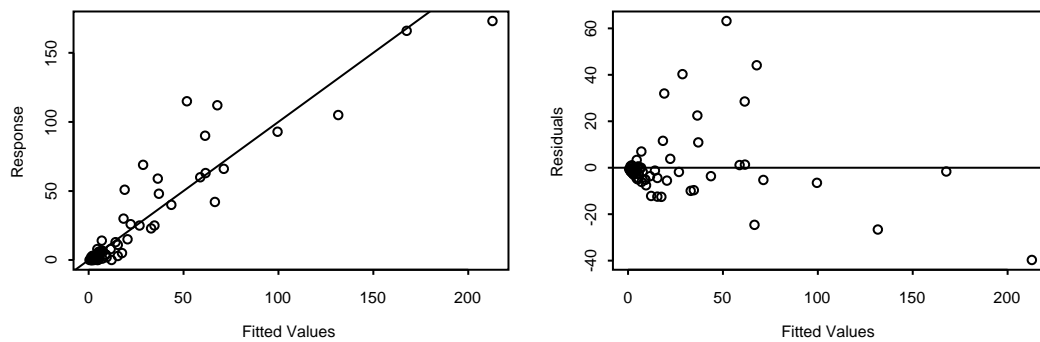


Abbildung 9.5: Angepasste Werte und Residuen des ZIP $\mu$ -Modells der Patentdaten

### 9.1.6 ZIP-Regression

Ein einfacheres Modell um Nullenüberschuss zu modellieren ist Zero-Inflated Poisson. Das ZIP-Modell führt die gesamte Überdispersion auf Nullenüberschuss zurück. Für die Zielvariable verwenden wir hier die Verteilungsannahme  $\text{patents}[i] \sim ZIP(\mu_i, \omega)$  sowie den linearen Prädiktor und die Verknüpfung (9.1). Um das ZIP-Modell in S-Plus berechnen zu können, treffen wir in der Schätzroutine `mle.zigp1` die Restriktion  $\varphi = 1$ . In die Akaike-Information fließt die Anzahl der Parameter mit ein. Das ZIP-Modell enthält einen Parameter weniger als das ZIGP $\mu$ -Modell. Die Berechnungsformel von AIC in der Routine `mle.zigp1` ändern wir daher entsprechend ab. Mit der S-Plus Kommandozeile (9.4) erhalten wir dann ein ZIP-Modell mit folgender Zusammenfassung:

```

ZI.Parameter:           0.0915072

Coefficients:          -0.9210854  -6.339783  2.185856  4.443186

Dispersion Parameter:           1

RSS:                    14126.05

AIC:                    -10038.98

Range of mu:           0.752524  215.9793

Message:                RELATIVE FUNCTION CONVERGENCE

```

Tabelle 9.5: `summary.zigp1`-Tabelle des ZIP-Modells der Patentdaten

Das ZIP-Modell schätzt den Auswahlparameter auf 0.09. Der Anteil des Nullenüberschusses wird hier erwartungsgemäß höher geschätzt als im ZIGP $\mu$ -Modell, da die gesamte Überdispersion in den Daten zusätzlichen Nullen zugeschrieben wird. Wie schon bei dem Übergang vom Poisson zum verallgemeinertem Poissonmodell, wirkt sich der Übergang vom ZIP- zum ZIGP $\mu$ -Modell vor allem auf den Schätzer des zu `rdsales` gehörenden Regressionsparameter aus. Vor allem diese Kovariable scheint daher die Überdispersion zu steuern. Die Residuen sind im Vergleich zum ZIGP $\mu$ -Modell etwas nach oben verschoben (vgl. Tabelle 9.10), besitzen aber eine ähnliche Verteilung (vgl. Abbildung 9.6). Die Fehlerquadratsumme geht zurück. Trotzdem steigt im Vergleich zur ZIGP $\mu$ -Regression die Akaike-Information an.

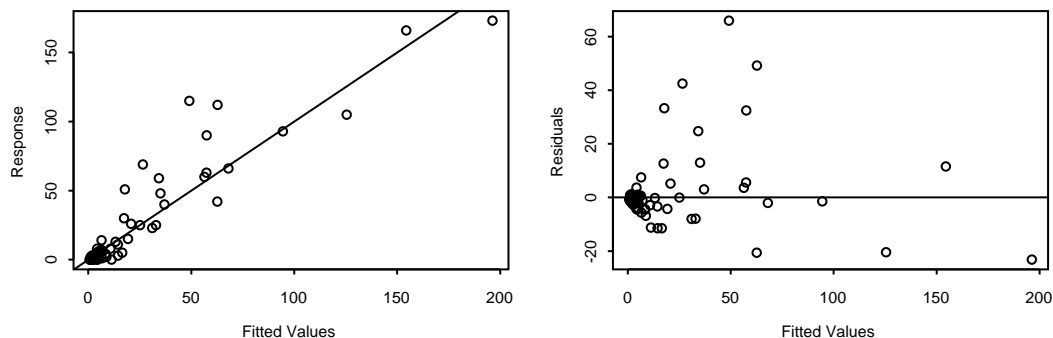


Abbildung 9.6: Angepasste Werte und Residuen des ZIP-Modells der Patentdaten

### 9.1.7 ZIGP $\mu\omega$ -Regression

Im Vergleich zum ZIGP $\mu$ -Modell wird beim ZIGP $\mu\omega$ -Modell auch für den Auswahlparameter Regression zugelassen. Es gelte  $\text{patents}[i] \sim \text{ZIGP}(\mu_i, \varphi, \omega_i)$ . Setzen wir sowohl die Parameter  $\mu_i$  wie auch  $\omega_i$  in Beziehung zu den vier Kovariablen aus den bisherigen Modellen, berechnen wir das zugehörige ZIGP $\mu\omega$ -Modell über den S-Plus Programmaufruf:

```
mle.zigp2(patents,
+         matrix(c(rep(1,70),rdsales,rdw5,rdsales*rdw5),70,4),
+         matrix(c(rep(1,70),rdsales,rdw5,rdsales*rdw5),70,4),rep(1,70)) (9.5)
```

Eine Zusammenfassung des Modells erhalten wir über die Funktion `summary.zigp2`:

```
Coefficients.ZI:      -7.997876   132.2676   2.491076  -87.75858

Coefficients.GP:     -0.922079  -5.713425   2.187157   4.048188

Dispersion Parameter:  2.659953

RSS:                  14047.85

AIC:                  -10213.92

Range of mu:         0.897313   209.0180

Range of omega:      0.000059           1

Message:              RELATIVE FUNCTION CONVERGENCE
```

Tabelle 9.6: `summary.zigp2`-Tabelle des ZIGP $\mu\omega$ -Modells der Patentdaten

Nur unmerklich nimmt der geschätzte Dispersionsanteil im Vergleich zu den vorigen Modellen weiter ab, während der Zero-Inflation-Anteil ansteigt. Auch die Verteilung der Residuen ändert sich kaum (vgl. Abbildung 9.7). Eine deutlichere Verbesserung ist jedoch bei dem Wertebereich der Residuen zu verzeichnen, welcher ein kleineres Intervall umfasst und stärker um Null zentriert ist als beim ZIGP $\mu$ -Modell. Auch die Fehlerquadratsumme ist geringer. Da nun aber 9

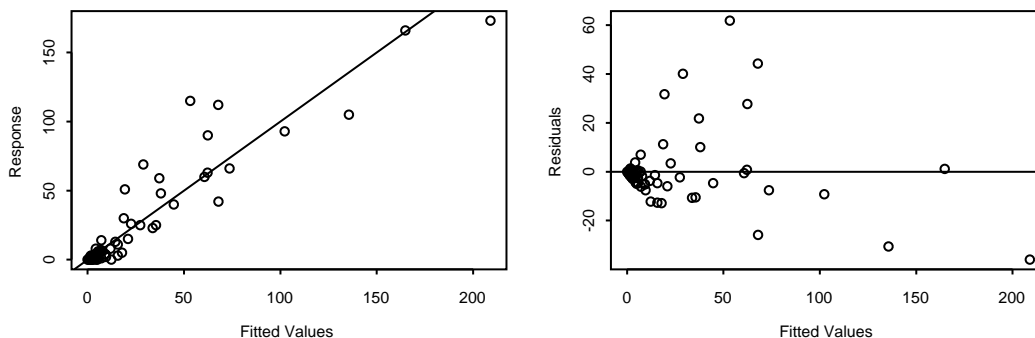


Abbildung 9.7: Angepasste Werte und Residuen des ZIGP $\mu\omega$ -Modells der Patentdaten

Parameter im Modell vorhanden sind, liegt die Akaike-Information höher als bei den beiden vorangegangenen Modellen. Anzumerken ist, dass die Größe AIC hier nicht ohne weiteres mit der entsprechenden Größe der vorangegangenen Modelle verglichen werden darf, da  $ZIGP_{\mu\omega}$  nicht in diesen Modellen eingebettet ist. Gleiches gilt für die beiden folgenden Modelle.

### 9.1.8 $ZIGP_{\omega}$ -Regression

Unbrauchbar für die vorliegenden Daten ist das  $ZIGP_{\omega}$ -Regressionsmodell. Charakteristisch für diese Regressionsart ist der übereinstimmende Parameter  $\mu$  für alle Beobachtungen, während der Zero-Inflation-Anteil über Regression von den Kovariablen abhängt. Mit der Kommandozeile

```
mle.zigp3(patents,
+         matrix(c(rep(1,70),rdsales,rdw5,rdsales*rdw5),70,4),rep(1,70)) (9.6)
```

erhalten wir ein  $ZIGP_{\omega}$ -Regressionsmodell mit unten stehender Zusammenfassung.

```
Coefficients.ZI:          806.65  9869.75  -1772.20  -6144.57

mu:                    24.87879

Dispersion Parameter:   12.54147

RSS:                   103173

AIC:                   -10109.85

Range of omega:        2.220446e-016      1

Message:               RELATIVE FUNCTION CONVERGENCE
```

Tabelle 9.7: `summary.zigp3`-Tabelle des  $ZIGP_{\omega}$ -Modells der Patentdaten

Im Vergleich zur  $ZIGP_{\mu\omega}$ -Regression fällt der geschätzte Dispersionsparameter auf. Dieser liegt nun wesentlich höher als zuvor. Die Schätzer  $\hat{\omega}_i$  bewegen sich im selben Bereich. Für die Daten wird ein gemeinsamer Parameter  $\hat{\mu} = 24.87879$  geschätzt. Von diesem Wert weichen 95% der angepassten Zielgrößen `patents[i]` nur minimal ab. In Abbildung 9.8 erkennen wir diese Konzentration. Lediglich 4 der 70 angepassten Patentanmeldungen `patents[i]` nehmen (positive) Werte sehr nahe bei Null an.

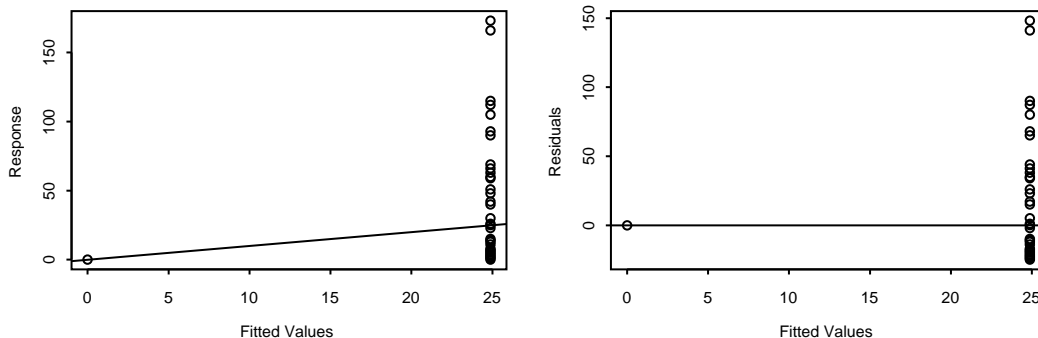


Abbildung 9.8: Angepasste Werte und Residuen des  $ZIGP_{\omega}$ -Modells der Patentdaten



### 9.1.9 ZHGP-Regression

Als letztes Modell wollen wir das ZHGP-Regressionsmodell auf die Patentdaten anwenden. Wir rufen die Schätzroutine über die Kommandozeile

```
mle.zhgp(patents,
+         matrix(c(rep(1,70),rdsales,rdw5,rdsales*rdw5),70,4),rep(1,70)) (9.7)
```

auf. Die Funktion `summary.zhgp` liefert folgende Zusammenfassung des Modells:

```
Coefficients.GP1:      -1.188492  -12.97728   1.530498   8.696002
Coefficients.GP2:      -2.060978  -5.106859   2.257431   3.655855
Dispersion Parameter:   0.630514
RSS:                    14491
AIC:                    -10213.62
Message.GP1:           relative function convergence
Message.GP2:           RELATIVE FUNCTION CONVERGENCE
```

Tabelle 9.8: `summary.zhgp`-Tabelle des ZHGP-Modells der Patentdaten

Zu beachten ist, dass der Dispersionsparameter in Tabelle 9.8 den Parameter  $\lambda$  aus (1.1) bezeichnet. Zum Vergleich mit den vorangegangenen Regressionsmodellen, müssen wir daher umparametrisieren:

$$\varphi = \frac{1}{1 - \lambda} = \frac{1}{1 - 0.630514} = 2.706462$$

Der Dispersionsparameter fällt daher etwas niedriger als im verallgemeinerten Poissonmodell und etwas höher als bei den Modellen  $ZIGP\mu$  und  $ZIGP\mu\omega$  aus. Abbildung 9.9 zeigt erneut keine wesentliche Änderung der Verteilung der Residuen. Betrachten wir die Struktur des Modells,

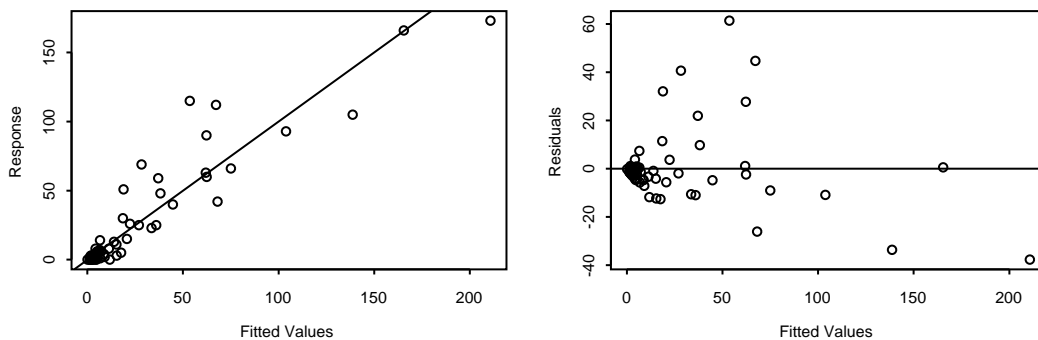


Abbildung 9.9: Angepasste Werte und Residuen des ZHGP-Modells der Patentdaten

so ist das ZHGP-Modell am besten mit  $ZIGP_{\mu\omega}$  vergleichbar. Die Anzahl der Parameter beträgt in beiden Modellen 9. In beiden Modellen liegt Regression sowohl für den Erwartungswert vor, als auch für den Mechanismus der den Nullenüberschuss steuert. Dementsprechend ist das Intervall  $[r_{\min}, r_{\max}]$  und der Mittelwert  $\bar{r}$  bei beiden Modellen ähnlich (vgl. Tabelle 9.10). Das ZHGP-Modell ist bei den Patentdaten dem  $ZIGP_{\mu\omega}$ -Modell jedoch unterlegen, da die Fehlerquadratsumme um einiges höher ist als bei  $ZIGP_{\mu\omega}$ -Regression. Für das ZHGP-Modell spricht im direkten Vergleich hingegen der Zeitaufwand für die Berechnung des Modells. Dieser ist trotz der selben Parameteranzahl auf Grund der getrennten Bestimmung der Schätzer der Hürden- und Ursprungsverteilung im ZHGP-Modell wesentlich geringer als bei  $ZIGP_{\mu\omega}$ -Regression. Deren Zeitaufwand lag bei der Berechnung der Modelle für die Patentdaten um den Faktor 7.5 über dem des ZHGP-Regressionsmodells.

### 9.1.10 ZHP-Regression

Wie zuvor bei der  $ZIGP_{\mu}$ - und der ZIP-Regression wollen wir auch das ZHGP-Modell mit dem darin eingebetteten, einfacheren ZHP-Modell vergleichen. Wiederum ändern wir die vorhandene S-Plus Funktion durch Restriktion des Dispersionsparameter dahingehend ab, dass wir mit der Kommandozeile (9.7) nun ein ZHP-Modell bestimmen. Die Berechnung der Akaike-Information wurde ebenfalls entsprechend angepasst. Mit `summary.zhgp` erhalten wir folgende Zusammenfassung:

```

Coefficients.GP1:      -1.188492  -12.97729   1.530498   8.696004

Coefficients.GP2:      -0.839969  -6.277778   2.152145   4.397197

Dispersion Parameter:           0

RSS:                          14080.22

AIC:                          -10042.34

Message.GP1:                 relative function convergence

Message.GP2:                 RELATIVE FUNCTION CONVERGENCE

```

Tabelle 9.9: `summary.zhgp`-Tabelle des ZHP-Modells der Patentdaten

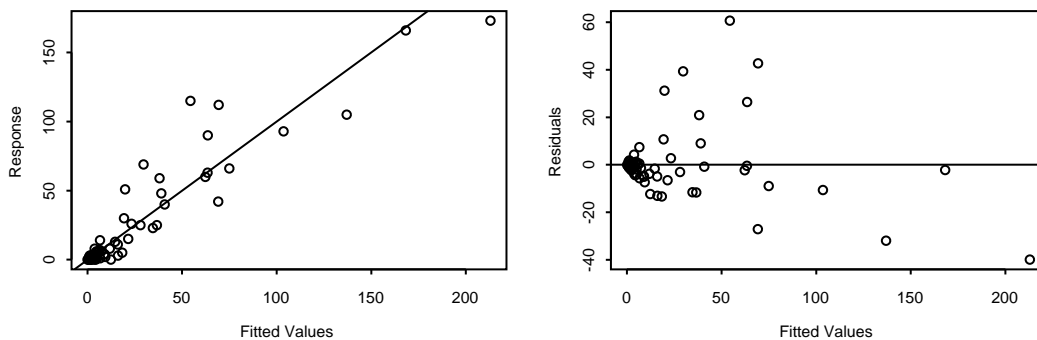


Abbildung 9.10: Angepasste Werte und Residuen des ZHP-Modells der Patentdaten

Wie wir bereits in Abschnitt 7.4.1 gezeigt hatten, wird die Hürdenverteilung nicht durch den Dispersionsparameter der kombinierten Verteilung beeinflusst. Wir erhalten für die Regressionsparameter der Hürdenverteilung dieselben Schätzer wie im ZHGP-Modell. Dagegen weichen die Schätzer für die Ursprungsverteilung zwischen beiden Modellen zum Teil deutlich voneinander ab. Sowohl Verteilung als auch die Werte der Residuen ändern sich kaum. Dennoch erfährt RSS eine Verbesserung. Die Akaike-Information des ZHP-Regressionsmodells ist größer als bei ZHGP-Regression. ZHP-Regression bietet also hier keine Verbesserung.

### 9.1.11 Zusammenfassung

In Tabelle 9.10 sind für alle betrachteten Modelle die Größen

- Anzahl der im Modell enthaltenen Parameter ( $k$ )
- Minimale Abweichung zwischen angepasstem und wahren Wert ( $r_{\min}$ )

$$r_{\min} = \min_{1 \leq i \leq 70} r_i \stackrel{(3.26)}{=} \min_{1 \leq i \leq 70} \{y_i - \hat{\mu}_i\}$$

- Arithmetisches Mittel der Residuen ( $\bar{r}$ )

$$\bar{r} = \frac{1}{70} \sum_{i=1}^{70} (y_i - \hat{\mu}_i)$$

- Maximale Abweichung zwischen angepasstem und wahren Wert ( $r_{\max}$ )

$$r_{\max} = \max_{1 \leq i \leq 70} \{y_i - \hat{\mu}_i\}$$

- Fehlerquadratsumme (RSS)

$$\text{RSS} \stackrel{(3.33)}{=} \sum_{i=1}^{70} (y_i - \hat{\mu}_i)^2$$

- Relative Fehlerquadratsumme (RSS\*)

$$\text{RSS}^* = \sum_{i=1}^{70} \left( \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} \right)^2$$

angegeben. In den Tabellen 9.11 und 9.12 sind der Wert der Log-Likelihoodfunktion, ausgewertet an der Stelle der Maximum-Likelihood-Schätzer, sowie die Akaike-Information für die ineinander eingebetteten Modelle Poisson  $\subset$  GP  $\subset$  ZIGP $_{\mu}$  bzw. Poisson  $\subset$  GP  $\subset$  ZHGP zu finden. Für diese Modelle kann das Akaike-Informationskriterium angewendet werden.

Vergleichen wir die acht betrachteten Regressionsmodelle, so weist das Poissonmodell mit 4 Parametern die niedrigste Komplexität auf. Es vernachlässigt jedoch Dispersion. Durch den Übergang zur verallgemeinerten Poissonregression erreichen wir sowohl eine Reduktion der Fehlerquadratsumme als auch der Akaike-Information. Im ZIGP $_{\mu}$ -Modell erklären wir einen Teil der Dispersion durch Nullenüberschuss, dies führt zu einer weiteren Reduktion der Fehlerquadratsumme, die Akaike-Information steigt jedoch wieder leicht an. Die Modelle ZIGP $_{\mu\omega}$  und ZHGP weisen zwar eine niedrigere Fehlerquadratsumme als das verallgemeinerte Poissonmodell

auf, sie sind mit 9 Parametern jedoch wesentlich komplexer und führen zu einer höheren relativen Fehlerquadratsumme. Das ZIGP $\omega$ -Modell stellt sich für die Poissondaten als unpassend heraus. Da die Überdispersion nur zu einem kleinen Teil von Nullenüberschuss herrühren dürfte, ergeben die eingebetteten Modelle ZIP bzw. ZHP eine schlechtere Anpassung als ZIGP $\mu$  bzw. ZHGP. Insgesamt scheint es sinnvoll, bei der Modellierung der Patentdaten der verallgemeinerten Poissonregression den Vorzug zu geben.

Modell	k	$r_{\min}$	$\bar{r}$	$r_{\max}$	RSS	RSS*
Poisson	4	-45.18	$-2.5 * 10^{-7}$	61.49	15328.96	30.80
GP	5	-43.80	$-2.7 * 10^{-6}$	62.67	14571.69	30.00
ZIGP $\mu$	6	-39.73	0.23	63.15	14218.67	30.11
ZIP	5	-23.22	1.78	65.91	14126.05	32.44
ZIGP $\mu\omega$	9	-36.01	0.04	61.75	14047.85	30.26
ZIGP $\omega$	6	-24.88	$-5.1 * 10^{-6}$	148.12	103173.00	170.69
ZHGP	9	-37.75	-0.01	61.38	14491.00	30.36
ZHP	8	-39.98	-0.01	60.64	14080.22	33.75

Tabelle 9.10: Minimum, Maximum und arithmetisches Mittel der Residuen sowie Fehlerquadratsummen RSS und RSS\* der Regressionsmodelle für die Patentdaten im Vergleich

Modell	k	$\mathcal{L}$	AIC
Poisson	4	5011.59	-10015.18
GP	5	5115.19	-10220.38
ZIGP $\mu$	6	5115.45	-10218.89

Tabelle 9.11: Log-Likelihoodfunktion und Akaike-Information der Maximum-Likelihood-Schätzer des ZIGP $\mu$ - und darin eingebetteter Regressionsmodelle für die Patentdaten

Modell	k	$\mathcal{L}$	AIC
Poisson	4	5011.59	-10015.18
GP	5	5115.19	-10220.38
ZHGP	9	5030.17	-10042.34

Tabelle 9.12: Log-Likelihoodfunktion und Akaike-Information der Maximum-Likelihood-Schätzer des ZHGP- und darin eingebetteter Regressionsmodelle für die Patentdaten

## 9.2 KFZ-Datensatz

### 9.2.1 Datenbeschreibung

Bei den vorliegenden Daten handelt es sich um einen Teildatensatz mit dem Umfang von 16307 Beobachtungen zu KFZ-Versicherungen in Bayern. Der Vektor `response` enthält die Schadenzahl der Policen. Diese nimmt den Wert 0, 1, 2, 3 oder 4 an. Die unterschiedliche Versicherungsdauer ist in Tagen (1-360 Tage) angegeben und im Vektor `exposure` festgehalten. Weiterhin liegt uns die Matrix `xdesign` vor. Sie enthält folgende Kovariablen:

Spalte	Bezeichnung	Beschreibung
1	<code>intercept</code>	Interceptspalte
2	<code>age2</code>	Quadriertes Alter des Versicherten
3	<code>age3</code>	Alter des Versicherten zur dritten Potenz
4	<code>huk</code>	Schadenfreie Jahre
5	<code>da</code>	Deckungsart
6	<code>sex</code>	Geschlecht des Versicherten
7	<code>bauj</code>	Baujahr des Fahrzeuges
8	<code>km</code>	Jahresfahrleistung
9	<code>huk*km</code>	Interaktion zwischen schadenfreien Jahren und Jahresfahrleistung
10	<code>s.age</code>	Standardisiertes Alter des Versicherten

Tabelle 9.13: Spalten der Designmatrix des KFZ-Datensatzes

Die Kovariablen `age2`, `age3`, `huk`, `da` und `km` sind metrisch skaliert und zentriert. `s.age` geht aus der Standardisierung  $\frac{a-\bar{a}}{s}$  mit  $s^2 = \frac{1}{n-1} \sum_{i=1}^n (a - \bar{a})^2$  und  $a = \text{age}$  hervor. Die Kovariablen `sex` und `bauj` sind kategorielle Variablen. `sex` nimmt die Werte 0 und 1 an, wobei 0 für männliche Versicherungsnehmer, 1 für weibliche Versicherungsnehmer steht. Die Codierung 0 der Variable `bauj` bezeichnet Fahrzeuge mit Baujahr vor 1998, 1 bezeichnet Fahrzeuge deren Baujahr 1998 oder danach ist.

#### Bemerkung:

- Aus dem gegebenen Teildatensatz soll eine Stichprobe mit Umfang 3000 untersucht werden. Dazu liegt uns die Indexmenge `index` vor. Die Stichprobe, welche durch `index` induziert wird, weist eine ähnliche Responsestruktur auf wie der Ausgangsdatensatz (vgl. Tabelle 9.14). Es treten zwar nur noch maximal 2 Schadensfälle auf, die Stichprobe enthält aber prozentual in etwa genausoviele Nullen in der Zielvariable wie der Ausgangsdatensatz.
- Auf eine explorative Datenanalyse kann an dieser Stelle verzichtet werden, da an das vorhandene Poisson-Regressionsmodell

$$\ln(\text{response}[i]) = \beta_0 + \beta_1 \text{huk}[i] + \beta_2 \text{da}[i] + \beta_3 \text{km}[i] + \beta_4 \text{age2} \quad (9.8)$$

angeknüpft werden soll, welches bereits ausschließlich signifikante Kovariablen aufweist. Wir fassen in S-Plus diese Kovariablen zur neuen Modellmatrix `xdesign2` zusammen:

$$\text{xdesign2} \leftarrow \text{matrix}(c(\text{intercept}, \text{huk}, \text{da}, \text{km}, \text{age2}), 16307, 5) \quad (9.9)$$

Schadenzahl	0	1	2	3	4
Ausgangsdatensatz	15576 (95.52%)	692 (4.24%)	36 (0.22%)	2 (0.02%)	1 (0.01%)
Stichprobe ( <code>index</code> )	2861 (95.37%)	134 (4.47%)	5 (0.17%)	0	0

Tabelle 9.14: Responsestruktur der KFZ-Daten (Anzahl und prozentualer Anteil der auftretenden Schadenszahlen pro Police im Ausgangsdatensatz ( $n=16307$ ) und der Stichprobe mit Umfang 3000)

- Bei der Modellierung der  $\omega$ -Regression in den Modellen  $ZIGP_{\mu\omega}$  und  $ZIGP_{\omega}$  wird ein logit-Link angesetzt. Da dieser nicht linear ist, scheint es sinnvoll zu sein, keine Transformation des Alters vorzunehmen. Deshalb wird in die Modellmatrix `xdesign3` der  $\omega$ -Regression anstatt `age2` das standardisierte Alter `s.age` aufgenommen:

```
xdesign3 <- matrix(c(intercept, huk, da, km, s.age),16307,5) (9.10)
```

### 9.2.2 Poissonregression

Mit der Kommandozeile

```
glm(response[index] ~ offset(log(exposure[index]))+huk[index]+da[index]
+km[index]+age2[index], family=poisson, link=log) (9.11)
```

berechnen wir das zu (9.8) gehörende Poisson-Regressionsmodell.

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.9997867	-0.365044	-0.2698152	-0.1651545	3.847807

Coefficients:

	Value	Std. Error	t value
(Intercept)	-8.5462003	0.09356088	-91.343734
huk	0.1914042	0.07592307	2.521028
da	-0.4334843	0.12319876	-3.518577
km	0.1567109	0.07112546	2.203302
age2	31.1819671	7.61655500	4.093973

Tabelle 9.15: `summary`-Tabelle des Poissonmodells (9.11)

Zunächst fällt auf, dass die angepassten Schadenszahlen alle samt sehr klein sind. Der Bereich der durch (9.11) angepassten Werte lautet  $[1.696 * 10^{-5}, 0.716]$ . Damit werden zwar alle Beobachtungen mit Schadenszahl Null im wesentlichen korrekt geschätzt, die angepassten Werte für echt positive Schadenszahlen sind aber ebenfalls annähernd Null. Dementsprechend ergibt sich ein maximaler Fehler von knapp unter 2 (vgl. Tabelle 9.23). Wie wir der Zusammenfassung in Tabelle 9.11 entnehmen können, ist der Intercept der signifikanteste Term im angepassten Poissonmodell. Für  $\hat{\beta}_0$  wird ungefähr -8.5 angenommen. Dies entspricht einem gemeinsamen Mittelwert von 0.0002.

### 9.2.3 Verallgemeinerte Poissonregression

Bestimmen wir nun ein verallgemeinertes Poisson-Regressionsmodell. Die Verteilungsannahme lautet hier  $\text{response}[i] \sim GP^*(\mu_i, \varphi)$ . Wir berechnen das Modell in S-Plus über den Programmaufruf

```
mle.gpois(response[index], xdesign2[index], exposure[index])
```

 (9.12)

Die Designmatrix `xdesign2` (9.9) enthält dabei in der selben Reihenfolge die Kovariablen welche in das Modell (9.11) eingehen. Die Funktion `summary.gpois` liefert uns für das berechnete Modell die Ausgabe in Tabelle 9.12.

	Min	1Q	Median	3Q	Max
Deviance Residuals:	-18.6927	-6.62099	-4.00622	-1.46005	21.95841
Coefficients:	-8.54490	0.186835	-0.43267	0.159751	30.89652
Dispersion Parameter:	1.011666				
RSS:	144.9914				
AIC:	2637.682				
Residual Deviance:	81906.94				
Range of mu:	0.000017	0.695795			
Message:	RELATIVE FUNCTION CONVERGENCE				

Tabelle 9.16: `summary.gpois`-Tabelle des verallgemeinerten Poissonmodells der KFZ-Daten

Das Modell schätzt den Dispersionsparameter auf rund 1.01. Das resultierende Modell unterscheidet sich daher bzgl. den geschätzten Regressionsparametern kaum vom Poissonmodell 9.11. Auch ändert sich die Abbildung der angepassten Werte gegen die tatsächlichen Werte kaum (siehe Abbildung 9.12). Im Mittel sinken die Fehler minimal, was zu einer etwas geringeren Fehlerquadratsumme führt (vgl. Tabelle 9.23). Die Verbesserung der Anpassungsgenauigkeit reicht jedoch nicht aus um den Komplexitätszuwachs in Form des zusätzlichen Parameters wettzumachen, so dass die Akaike-Information höher ist als bei dem vorangegangenen Poissonmodell.

### 9.2.4 ZIGP $\mu$ -Regression

Beziehen wir nun Zero-Inflation in das Modell mit ein. Wir treffen die Verteilungsannahme  $\text{response}[i] \sim ZIGP(\mu_i, \varphi, \omega)$ . Das entsprechende Modell berechnen wir in S-Plus über den Kommandozeile

```
mle.zigp1(response[index], xdesign2[index], exposure[index])
```

 (9.13)

und erhalten mit `summary.zigp1` die Zusammenfassung in Tabelle 9.17.

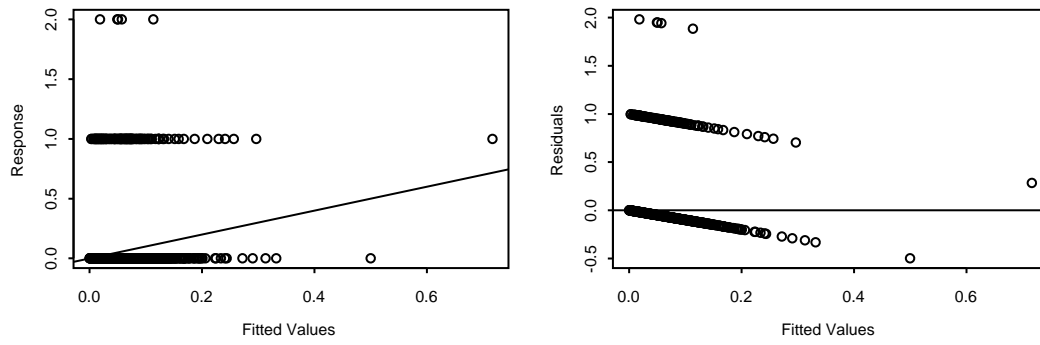


Abbildung 9.11: Angepasste Werte und Residuen des Poissonmodells der KFZ-Daten

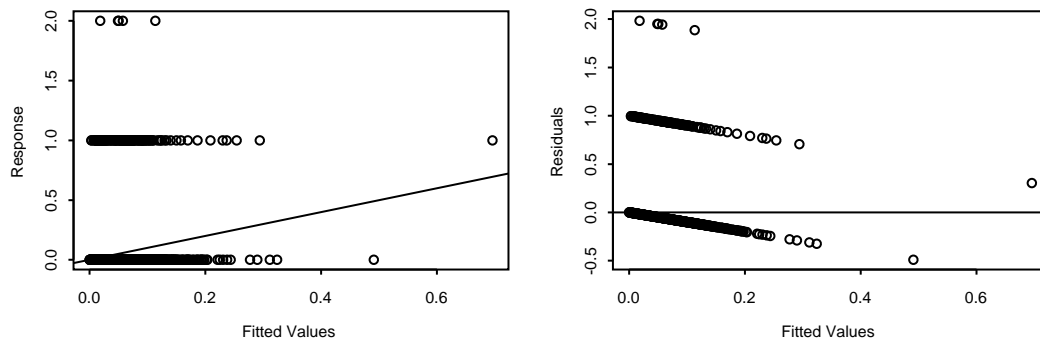


Abbildung 9.12: Angepasste Werte und Residuen des verallgemeinerten Poissonmodells der KFZ-Daten

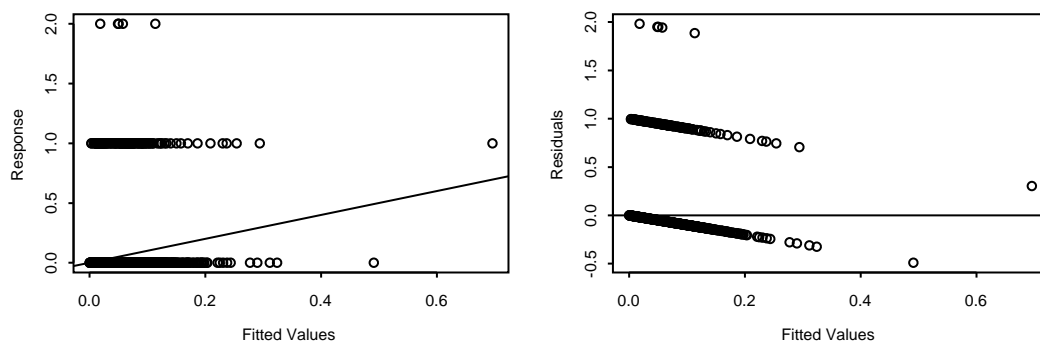
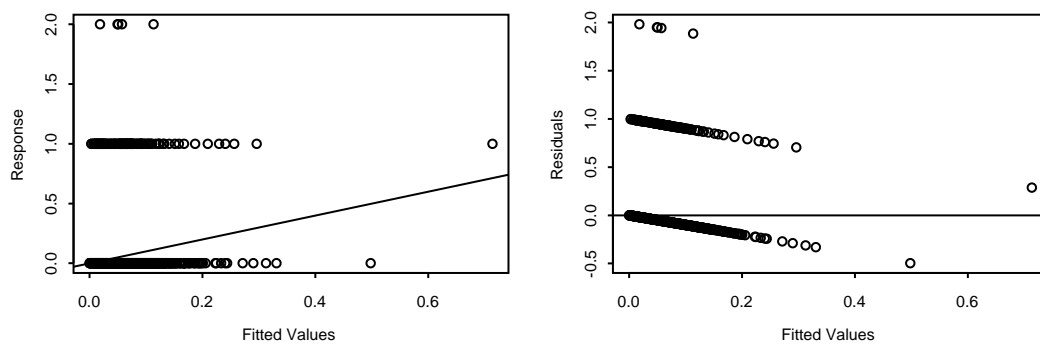
Abbildung 9.13: Angepasste Werte und Residuen des ZIGP $\mu$ -Modells der KFZ-Daten

Abbildung 9.14: Angepasste Werte und Residuen des ZIP-Modells der KFZ-Daten



```

ZI.Parameter:                0

Coefficients:               -8.54490  0.186837  -0.43267  0.159751  30.89738

Dispersion Parameter:      1.011667

RSS:                        144.9914

AIC:                        2639.682

Range of mu:                0.000017  0.695833

Message:                    RELATIVE FUNCTION CONVERGENCE

```

Tabelle 9.17: `summary.zigp1`-Tabelle des  $ZIGP_{\mu}$ -Regressionsmodells der KFZ-Daten

Das Modell erkennt in den Daten keinen Nullenüberschuss. Der Auswahlparameter  $\omega$  wird daher auf Null geschätzt. Dementsprechend stimmen die übrigen Schätzwerte im wesentlichen mit denen im verallgemeinerten Poisson-Regressionsmodell überein. Auf Grund des zusätzlichen Parameters ist die Akaike-Information dennoch höher als zuvor (vgl. Tabelle 9.23).

### 9.2.5 ZIP-Regression

Wie in Abschnitt 9.1.6 modifizieren wir die Funktion `mle.zigp1` um ein ZIP-Modell für die KFZ-Daten berechnen zu können. Mit der S-Plus Befehlszeile (9.13) erhalten wir das zur Tabelle 9.18 gehörende ZIP-Modell.

```

ZI.Parameter:                0.042303

Coefficients:               -8.50271  0.191217  -0.43387  0.15664  31.1203

Dispersion Parameter:      1

RSS:                        144.997

AIC:                        2638.57

Range of mu:                0.000018  0.745182

Message:                    RELATIVE FUNCTION CONVERGENCE

```

Tabelle 9.18: `summary.zigp1`-Tabelle des ZIP-Regressionsmodells der KFZ-Daten

Da die Überdispersion welche durch das GP- und  $ZIGP_{\mu}$ -Modell ausgewiesen wurde nicht mehr durch den Dispersionsparameter  $\varphi$  modelliert werden kann, ermittelt das ZIP-Modell einen Nullenüberschuss, der dem Auswahlparameter 0.042303 entspricht. An den Schätzern der Regressionsparameter, den Residuen und der Akaike-Information ändert sich im Vergleich zur  $ZIGP_{\mu}$ -Regression nicht sehr viel.

### 9.2.6 ZIGP $\mu\omega$ -Regression

Mit dem Programmaufruf

```
mle.zigp2(response[index], xdesign3[index],
          xdesign2[index], exposure[index])
```

 (9.14)

wird ein Regressionsmodell an die KFZ-Daten angepasst, welches für den Parameter  $\omega$  ebenfalls Regression annimmt. Die Verteilungsannahme lautet:  $\text{response}[i] \sim \text{ZIGP}(\mu_i, \varphi, \omega_i)$ . Mit (9.14) erhalten wir das ZIGP $\mu\omega$ -Modell dessen Zusammenfassung in Tabelle 9.19 zu finden ist.

```
Coefficients.ZI:      -1.21768  0.266268  0.071758  -0.64409  -0.06920
Coefficients.GP:      -8.50973  0.204762  -0.37528  0.144395  30.48879
Dispersion Parameter:  1.01101
RSS:                  144.909
AIC:                  1155.98
Range of mu:          0.000024  0.727467
Range of omega:        0.000135  0.694301
Message:              RELATIVE FUNCTION CONVERGENCE
```

Tabelle 9.19: `summary.zigp2`-Tabelle des ZIGP $\mu\omega$ -Regressionsmodells der KFZ-Daten

Auch hier weichen die ermittelten Schätzer der Regressionsparameter und des Dispersionsparameters kaum von den vorangegangenen Modellen ab. Die Fehlerquadratsumme ist zwar etwas geringer, die Komplexität liegt aber auf Grund der zusätzlichen Regression mit 11 Parametern wesentlich höher als zuvor.

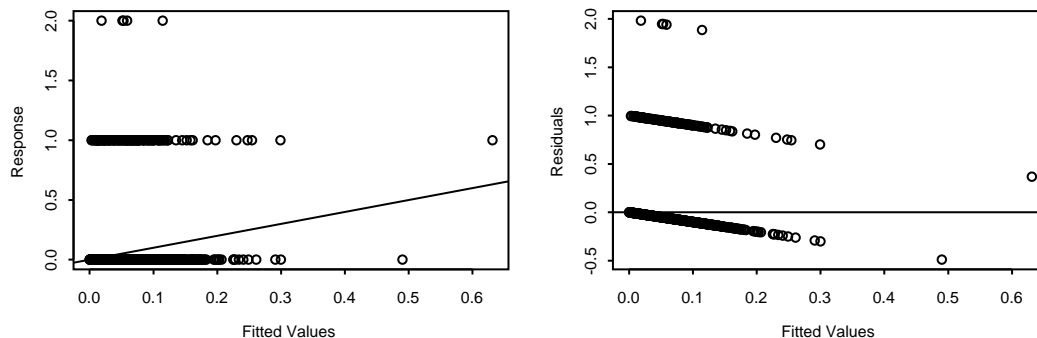


Abbildung 9.15: Angepasste Werte und Residuen des ZIGP $\mu\omega$ -Modells der KFZ-Daten

### 9.2.7 ZIGP $\omega$ -Regression

Setzen wir nun einen gemeinsamen Parameter  $\mu$  voraus und wenden Regression ausschließlich auf den Auswahlparameter  $\omega$  an. Dies entspricht der Verteilungsannahme  $\text{response}[i] \sim \text{ZIGP}(\mu, \varphi, \omega_i)$ . Das mit dem Programmaufruf

$$\text{mle.zigp3}(\text{response}[\text{index}], \text{xdesign3}[\text{index}], \text{exposure}[\text{index}]) \quad (9.15)$$

an die KFZ-Daten angepasste ZIGP $\omega$ -Modell ist in Tabelle 9.20 zusammengefasst.

```

Coefficients.ZI:      -0.02839  -0.29744  0.772195  -0.36282  0.017932

mu:                  0.000432

Dispersion Parameter: 1.003189

RSS:                 146.1533

AIC:                 1167.783

Range of omega:      0.011671  0.985439

Message:              RELATIVE FUNCTION CONVERGENCE

```

Tabelle 9.20: `summary.zigp3`-Tabelle des ZIGP $\omega$ -Regressionsmodells der KFZ-Daten

Der Parameter  $\mu$  wird mit 0.000432 geschätzt. Der geschätzte Dispersionsgrad ist ähnlich nahe Äquidispersion, wie zuvor. Auf Grund der recht unterschiedlichen Schätzungen der ZI-Regressionsparameter, ist das Maximum der geschätzten Auswahlparameter  $\omega_i$  mit 0.99 deutlich höher als bei der ZIGP $\mu\omega$ -Regression. Dennoch sind die angepassten Werte in beiden Modellen recht ähnlich - lediglich das Maximum der angepassten Werte ist im ZIGP $\omega$ -Modell kleiner (s. Abbildung 9.16). Die Fehler sind daher mit den vorigen Modellen vergleichbar (vgl. Tabelle 9.23).

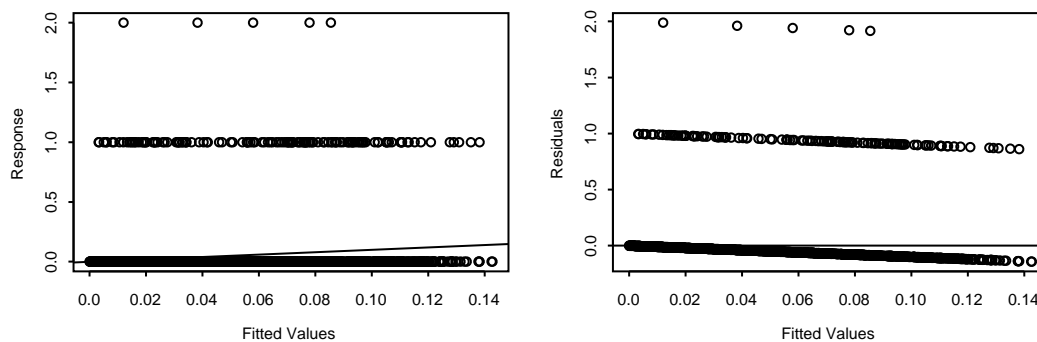


Abbildung 9.16: Angepasste Werte und Residuen des ZIGP $\omega$ -Modells der KFZ-Daten

### 9.2.8 ZHGP-Regression

Die Verteilungsannahme  $\text{response}[i] \sim \text{ZHGP}(\theta_{1i}, \theta_{2i}, \lambda)$  führt auf das ZHGP-Regressionsmodell. Mit dem Programmaufruf

$$\text{mle.zhgp}(\text{response}[\text{index}], \text{xdesign2}[\text{index}], \text{exposure}[\text{index}]) \quad (9.16)$$

passen wir solch ein Regressionsmodell an die KFZ-Daten an. Die Zusammenfassung wird in Tabelle 9.21 wiedergegeben.

```

Coefficients.GP1:      -8.55205  0.185890  -0.44322  0.169336  33.1051

Coefficients.GP2:      -21.4770  1.011638  -0.02841  -0.02760  30.6718

Dispersion Parameter:  0.034724

RSS:                   144.9855

AIC:                   1157.621

Message.GP1:          relative function convergence

Message.GP2:          FALSE CONVERGENCE

```

Tabelle 9.21: `summary.zhgp`-Tabelle des ZHGP-Regressionsmodells der KFZ-Daten

Zunächst fällt auf, dass der Schätzalgorithmus für die Ursprungsverteilung nicht konvergiert. Dies dürfte daran liegen, dass nur 139 der 3000 Beobachtungen dieser Teilregression zur Verfügung stehen. Das starke Übergewicht der Schadenzahl Null kommt durch die zur Hürdenverteilung gehörenden geschätzten Regressionsparameter zum Ausdruck. Diese weisen eine hohe Ähnlichkeit zu den geschätzten Regressionsparameter im Poisson-Regressionsmodell auf. Dort wird die Regression somit im wesentlichen auch nur von den Nullen in der Zielgröße bestimmt. Der geschätzte Dispersionsparameter stimmt ebenfalls ungefähr mit den vorigen Modellen überein:

$$\hat{\phi} = \frac{1}{1 - \hat{\lambda}} = \frac{1}{1 - 0.034724} = 1.035973$$

Dies gilt auch für die angepassten Werte (vgl. Abbildung 9.17)

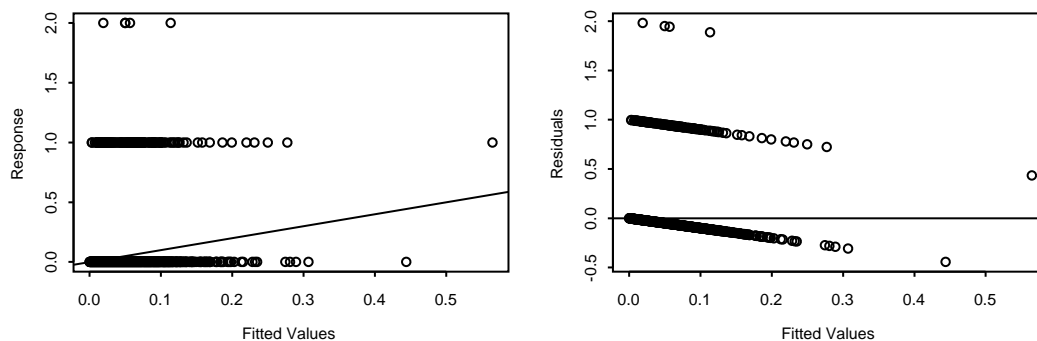


Abbildung 9.17: Angepasste Werte und Residuen des ZHGP-Modells der KFZ-Daten

### 9.2.9 ZHP-Regression

Entsprechend Abschnitt 9.1.10 modifizieren wir die Funktion `mle.zhgp` um ein ZHP-Modell an die KFZ-Daten anpassen zu können. Mit der S-Plus Befehlszeile

```
mle.zhgp(response[index], xdesign2[index], exposure[index]) (9.17)
```

erhalten wir dann ein ZHP-Modell dessen Zusammenfassung in Tabelle 9.22 zu finden ist.

```

Coefficients.GP1:      -8.55205  0.185890  -0.44322  0.169336  33.1051
Coefficients.GP2:      -8.39463  0.424104  -0.09940  -0.39429  4.43146
Dispersion Parameter:          0
RSS:                      145.0871
AIC:                      1157.124
Message.GP1:              relative function convergence
Message.GP2:              RELATIVE FUNCTION CONVERGENCE

```

Tabelle 9.22: `summary.zhgp`-Tabelle des ZHP-Regressionsmodells der KFZ-Daten

Im Gegensatz zur ZHGP-Regression konvergiert der Schätzalgorithmus hier auch für die Ursprungsverteilung. Für deren Intercept erhalten wir einen ähnlichen Parameterschätzer wie für den Interceptterm der Hürdenverteilung, während sich die übrigen Schätzer zwischen Ursprungs- und Hürdenverteilung unterscheiden. An den Residuen tritt erneut kaum eine Änderung ein. Die Akaike-Information ist minimal kleiner als bei dem vorangegangenen ZHGP-Modell.

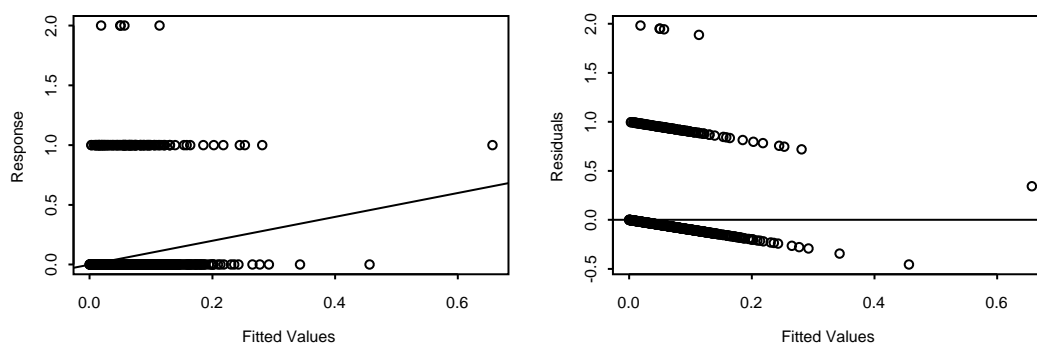


Abbildung 9.18: Angepasste Werte und Residuen des ZHP-Modells der KFZ-Daten

### 9.2.10 Zusammenfassung

Modell	k	$r_{\min}$	$\bar{r}$	$r_{\max}$	RSS	RSS*
Poisson	5	-0.4997	$1.99 * 10^{-7}$	1.9817	145.000	405035.1
GP	6	-0.4910	$3.70 * 10^{-8}$	1.9817	144.991	407027.9
ZIGP $\mu$	7	-0.4910	$2.27 * 10^{-8}$	1.9818	144.991	407028.7
ZIP	6	-0.4981	$-1.30 * 10^{-5}$	1.9817	144.997	404761.0
ZIGP $\mu\omega$	11	-0.4901	$2.15 * 10^{-5}$	1.9813	144.909	377117.2
ZIGP $\omega$	7	-0.1427	$-3.26 * 10^{-4}$	1.9880	146.153	440255.0
ZHGP	11	-0.4438	$1.99 * 10^{-4}$	1.9808	144.989	394496.6
ZHP	10	-0.4561	$-1.43 * 10^{-4}$	1.9813	145.087	416313.9

Tabelle 9.23: Minimum, Maximum und arithmetisches Mittel der Residuen sowie Fehlerquadratsummen RSS und RSS\* der Regressionsmodelle für die KFZ-Daten im Vergleich

Offensichtlich liegt in den KFZ-Daten trotz des erheblichen Anteils an Nullen in der Zielgröße kein Nullenüberschuss und nur marginale Überdispersion vor. Durch den hohen Nullenanteil wird die Modellierung jedoch erschwert. Da sich die Anpassungsgenauigkeit gemessen an der Fehlerquadratsumme zwischen den betrachteten Modellen kaum unterscheidet (vgl. Tabelle 9.23<sup>2</sup>), dürfte für die Güte der Anpassung gemäß dem Akaike-Informationskriterium vor allem die Komplexität des Modells ausschlaggebend sein. Trotz ihrer hohen Parameteranzahl weisen die Zero-Hurdle-Regressionmodelle eine wesentlich niedrigere Akaike-Information aus als die übrigen Modelle (vgl. Tabellen 9.24, 9.25 und 9.22). Dem entgegen ist die relative Fehlerquadratsumme RSS\* des ZHP-Modells recht hoch. Das ZIGP $\mu\omega$ -Modell liefert unter diesem Gesichtspunkt die beste Anpassung. Insgesamt dürfte das ZHGP-Modell auf Grund der niedrigen Werte von AIC und RSS\* am besten zur Modellierung der KFZ-Daten geeignet sein.

Modell	k	$\mathcal{L}$	AIC
Poisson	5	-1313.55	2637.105
GP	6	-1312.84	2637.682
ZIGP $\mu$	7	-1312.84	2639.682

Tabelle 9.24: Log-Likelihoodfunktion und Akaike-Information der Maximum-Likelihood-Schätzer des ZIGP $\mu$ - und darin eingebetteter Regressionsmodelle für die KFZ-Daten

Modell	k	$\mathcal{L}$	AIC
Poisson	5	-1313.55	2637.105
GP	6	-1312.84	2637.682
ZHGP	11	-567.81	1157.621

Tabelle 9.25: Log-Likelihoodfunktion und Akaike-Information der Maximum-Likelihood-Schätzer des ZHGP- und darin eingebetteter Regressionsmodelle für die KFZ-Daten

<sup>2</sup>Die Werte in den Tabellen 9.23, 9.24 und 9.25 werden äquivalent zu den Werten in den Tabellen 9.10, 9.11 und 9.12 berechnet.

# Kapitel 10

## Zusammenfassung

Wie wir in Kapitel 1 gesehen hatten, besitzt die verallgemeinerte Poissonverteilung einige wünschenswerte Eigenschaften. Insbesondere trifft dies auf ihr flexibles Verhalten bezüglich Dispersion zu, da sie sowohl Über- als auch Unterdispersion abbilden kann. Die Erzeugung verallgemeinert poissonverteilter Zufallsvariablen ist wie in Abschnitt 1.5 gesehen recht einfach zu realisieren.

Nicht unproblematisch ist jedoch die Verwendung der verallgemeinerten Poissonverteilung als angenommene Responseverteilung in Regressionsmodellen. Auf Grund ihrer fehlenden Zugehörigkeit zur Familie der exponentiellen Verteilungen, liegt kein verallgemeinertes lineares Modell vor. Die Standardtheorie dieser Regressionsklasse kann deshalb nicht auf das in Kapitel 5 vorgestellte verallgemeinerte Poisson-Regressionsmodell angewendet werden.

Da keine Aussagen über die Definitheit der Fisher-Information vorhanden sind, ist die Existenz und Eindeutigkeit von Maximum-Likelihood-Schätzern nicht sichergestellt. Für die im Rahmen dieser Arbeit betrachteten Daten ergab sich an der Stelle des ermittelten Schätzwertes allerdings stets negative Definitheit der beobachteten Fisher-Information, was zumindest auf die Existenz von Maximum-Likelihood-Schätzern hindeutet. Die Betrachtung der Likelihoodfunktion für den einfachen Fall eines einzigen Regressionsparameters in Abschnitt 5.3 lässt auf die Existenz eines eindeutigen Maximum-Likelihood-Schätzers bei solch einer Spezifikation schließen. Im Allgemeinen besitzen die Maximum-Likelihood-Schätzer wie wir den Simulationsresultaten in Abschnitt 8.2.1 entnehmen konnten eine hohe Güte. Wird deren Verteilung betrachtet (vgl. Abschnitt 8.2.3), so scheint die Vermutung asymptotisch normalverteilter Maximum-Likelihood-Schätzer gerechtfertigt zu sein.

Aufbauend auf das verallgemeinerte Poisson-Regressionsmodell aus Kapitel 5 wurden in Kapitel 6 verschiedene verallgemeinerte Zero-Inflated Poisson-Regressionsmodelle (ZIGP-Regressionsmodelle) spezifiziert. Diese ermöglichen die Modellierung von Nullenüberschuss, welcher wie in Abschnitt 4.4 gezeigt wurde, eine mögliche Ursache von Überdispersion darstellt. Die ZIGP-Modelle unterscheiden zwischen Nullenüberschuss und "Rest-Überdispersion", wobei der Anteil des Nullenüberschusses von einem Auswahlparameter gesteuert wird. Im ersten Ansatz (Abschnitt 6.3) wurde ein gemeinsamer Auswahlparameter für alle Beobachtungen vorausgesetzt. Die Güte der Maximum-Likelihood-Schätzer für das resultierende ZIGP $\mu$ -Modell ist meist hoch. Bei steigendem Dispersionsgrad und/oder Zero-Inflation-Anteil musste, wie zu erwarten war, der Stichprobenumfang erhöht werden um akzeptable Schätzer zu erhalten.

In Abschnitt 6.4 wurde der Vorschlag von Lambert (1992) aufgegriffen, für die einzelnen Beobachtungen individuelle Auswahlparameter zuzulassen. In dem resultierenden ZIGP $\mu\omega$ -Modell liegen daher zwei Regressionsarten vor: zum einen für die Parameter  $\mu_i$  der verallgemeinerten Poissonverteilung, zum anderen für die Auswahlparameter  $\omega_i$ . Die Simulationsstudie in Ab-

schnitt 8.4 ergab hier unabhängig vom gewählten Design ein Identifizierbarkeitsproblem für die mit der  $\omega$ -Regression korrespondierenden Maximum-Likelihood-Schätzer. Die übrigen Schätzer im  $ZIGP_{\mu\omega}$ -Modell waren mit den entsprechenden Schätzern im  $ZIGP_{\mu}$ -Modell vergleichbar. Insgesamt schien der verwendete Stichprobenumfang von 100 Beobachtungen zu gering zu sein. Ein weiteres verallgemeinertes Zero-Inflated Poisson Regressionsmodell wurde in Abschnitt 6.5 spezifiziert. Das  $ZIGP_{\omega}$ -Modell verwendet für alle Beobachtungen einen gemeinsamen Parameter  $\mu$ . Regression wird hier nur noch für die Auswahlparameter  $\omega_i$  zugelassen. Die Betrachtung der Maximum-Likelihood-Schätzer in Abschnitt 8.5 zeigte uns, dass die mit der  $\omega$ -Regression korrespondierenden Schätzer auch in diesem Modell in einigen Fällen nicht identifizierbar sind. Zudem wurde der gemeinsame Parameter  $\mu$ , welcher dem Erwartungswert der zu Grunde liegenden verallgemeinerten Poissonverteilung entspricht, häufig stark überschätzt.

Wie für das verallgemeinerte Poisson-Regressionsmodell liegen auch für die verallgemeinerten Zero-Inflated Poisson-Regressionsmodelle bisher keine Erkenntnisse über die Asymptotik der Maximum-Likelihood-Schätzer vor. Deshalb ist es auch hier nicht möglich die Modellspezifikation über statistische Hypothesentests zu untersuchen.

In Form von Zero-Hurdle-Modellen wurde in Kapitel 7 ein weiterer Ansatz zur Modellierung von Nullenüberschuss betrachtet. Solche Modelle nehmen eine Unterscheidung der Beobachtungen in zwei Gruppen vor: Beobachtungen mit Response Null und Beobachtungen mit echt positivem Response. Bei dem konstruierten verallgemeinerten Zero-Hurdle Poissonmodell (ZHGP-Modell) wurde für beide Klassen verallgemeinerte Poissonverteilung (mit unterschiedlichen Parametern) angenommen. Auch für dieses Modell sind noch keine Aussagen zum asymptotischen Verhalten der Maximum-Likelihood-Schätzer bekannt. Die Betrachtung dieser Schätzer in Abschnitt 8.6.1 zeigte, dass wir je nach Anteil des Dispersionsgrades bzw. des Nullenüberschusses entweder für die Regressionsparameter der Ursprungsverteilung (positive Ereignisse) oder für die Regression der Hürdenverteilung (Null-Ereignisse) schlechte Schätzungen erhalten.

In Kapitel 9 wurden die genannten Modelle auf zwei Datensätze aus der Praxis angewendet. Die Ursache der Überdispersion in den betrachteten Patentdaten scheint nicht durch Nullenüberschuss hervorgerufen zu werden. Dem entsprechend stellt sich das  $ZIGP_{\omega}$ -Modell für diese Daten als unbrauchbar heraus, während das verallgemeinerte Poisson-Regressionsmodell, welches Nullenüberschuss nicht explizit berücksichtigt, die beste Anpassungsgüte besaß. Trotz dem hohen Anteil von Nullen im zweiten betrachteten Datensatz konnte durch die Verwendung von GP-, ZIGP- und ZHGP-Regression keine nennenswerte Verbesserung der Anpassungsgüte gegenüber dem gewöhnlichen Poisson-Regressionsmodell erreicht werden.

Einen Überblick über alle verwendeten Regressionsmodelle ist in Anhang D zu finden.



## Anhang A

# Grundlegende Verteilungen und Verteilungsfamilien

### A.1 Familie der exponentiellen Verteilungen

**Definition A.1 (Exponentielle Familie)** Eine Familie  $\{P_x(\theta, \varphi), (\theta, \varphi) \in \Theta\}$  von Wahrscheinlichkeitsverteilungen heißt *exponentielle Familie*, falls es reellwertige Funktionen  $a$ ,  $b$  und  $c$  gibt, sodass sich die Dichte  $P_x(\theta, \varphi)$  in der Form

$$P_x(\theta, \varphi) = \exp \left\{ \frac{\theta x - b(\theta)}{a(\varphi)} + c(x, \varphi) \right\} \quad (\text{A.1})$$

schreiben lässt.  $\theta$  wird als *natürlicher oder kanonischer Parameter* bezeichnet,  $\varphi$  ist ein *Stör- oder Skalenparameter*.  $\Theta$  bezeichnet den *zugehörigen (natürlichen) Parameterraum* der Verteilungsfamilie.

### A.2 Poissonverteilung

Bei der Poissonverteilung handelt es sich um eine diskrete Verteilung, welche die Anzahl des Eintritts eines bestimmten Ereignisses während eines vorgegebenen Zeitintervalls zählt. Als Beispiel für einen solchen Zählvorgang kann die Anzahl der Schadensmeldungen bei einer Versicherung innerhalb eines Jahres genannt werden.

Die Wahrscheinlichkeit, dass ein Ereignis in einem sehr kleinen Zeitintervall  $\Delta t$  eintritt, ist hierbei sehr klein ( $\approx \theta \Delta t$ ). Die Poissonverteilung wird daher als Verteilung der seltenen Ereignisse bezeichnet. Der Parameter  $\theta$  heißt (Intensitäts-)Rate oder auch nur Intensität.

**Definition A.2 (Poissonverteilung)** Eine Zufallsvariable  $X$  mit der Wahrscheinlichkeitsfunktion

$$P_x(\theta) = P(X = x) = \begin{cases} \frac{\theta^x}{x!} e^{-\theta} & x = 0, 1, 2, \dots \\ 0 & \text{sonst} \end{cases} \quad (\text{A.2})$$

heißt *poissonverteilt mit Intensität  $\theta > 0$* . Kurz:  $X \sim \text{Poi}(\theta)$ .

Eine poissonverteilte Zufallsvariable kann mit Hilfe eines stochastischen Prozesses mit exponentialverteilten Zwischenankunftszeiten (Poissonprozess) konstruiert werden. Dies wird z.B. von Resnick (1992, S.303 ff.) beschrieben.

**Satz A.3 (Momenterzeugende Funktion der Poissonverteilung)** Sei  $X \sim Poi(\theta)$ . Dann wird die momenterzeugende Funktion der Zufallsvariablen  $X$  durch

$$M_X(v) = e^{\theta(e^v-1)} \quad (\text{A.3})$$

bestimmt.

*Beweis:*

$$\begin{aligned} M_X(v) &= E(e^{vX}) = \sum_{x=0}^{\infty} e^{vx} \frac{\theta^x}{x!} e^{-\theta} \\ &= e^{-\theta} \sum_{x=0}^{\infty} \frac{(e^v \theta)^x}{x!} = e^{-\theta} e^{e^v \theta} = e^{\theta(e^v-1)} \end{aligned}$$

□

Hervorzuheben ist die spezielle Beziehung, die zwischen dem Erwartungswert und der Varianz der Poissonverteilung besteht.

**Satz A.4 (Erwartungswert und Varianz der Poissonverteilung)** Sei  $X$  poissonverteilte Zufallsvariable mit Intensität  $\theta$ . Für den Erwartungswert und die Varianz von  $X$  gilt:

$$E(X) = \text{Var}(X) = \theta \quad (\text{A.4})$$

*Beweis:*

Der Erwartungswert folgt aus der ersten Ableitung der momenterzeugenden Funktion (A.3).

$$\begin{aligned} E(X) &= \mu_1 = M'(0) \\ &= e^{\theta(e^v-1)} \theta e^v \Big|_{v=0} = \theta \end{aligned}$$

Für das zweite nicht zentrierte Moment  $\mu_2$  von  $X$  gilt:

$$\begin{aligned} E(X^2) &= \mu_2 = M''(0) \\ &= e^{\theta(e^v-1)} (\theta e^v)^2 + e^{\theta(e^v-1)} \theta e^v \Big|_{v=0} \\ &= \theta^2 + \theta \end{aligned} \quad (\text{A.5})$$

Nun lässt sich die Varianz leicht bestimmen:

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = \theta$$

□

Die in Satz A.4 beschriebene Gleichheit von Erwartungswert und Varianz wird als Äquidispersion bezeichnet.

**Satz A.5 (Verteilungsklasse der Poissonverteilung)** Die Poissonverteilung gehört zur Klasse der einparametrischen exponentiellen Verteilungen.

*Beweis:*

Die Wahrscheinlichkeitsfunktion der Poissonverteilung lautet:

$$\begin{aligned} P_x(\theta) &= \frac{\theta^x}{x!} e^{-\theta} \\ &= \exp\{x \ln(\theta) - \ln(x!) - \theta\} \end{aligned} \quad (\text{A.6})$$

mit  $a(\varphi) := \varphi$ ,  $b(\theta) := e^\theta$ ,  $c(x, \varphi) := -\ln(x!)$ ,  $\tilde{\theta} := \ln(\theta)$  und  $\varphi := 1$  folgt:

$$(A.6) = \exp \left\{ \frac{\tilde{\theta}x - b(\tilde{\theta})}{a(\varphi)} + c(x, \varphi) \right\}$$

□

### A.3 Gammaverteilung

Bildet man die Summe aus  $n$  unabhängig exponentialverteilten Zufallsvariablen mit Ausfallrate  $\lambda$ , so erhält man eine Zufallsvariable mit  $P(X = x) = \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x}$  (Råde and Westergren, 1997, S.417).

**Definition A.6 (Gammaverteilung)** Eine Zufallsvariable  $X$  mit der Dichte

$$P_x(n, \lambda) = P(X = x) = \begin{cases} \frac{\lambda^n}{\Gamma(n)} x^{n-1} e^{-\lambda x} & x \in \mathbb{R}_+ \\ 0 & \text{sonst} \end{cases} \quad (\text{A.7})$$

heißt *gammaverteilt* mit Parametern  $n > 0$  und  $\lambda > 0$ . Kurz:  $X \sim \Gamma(n, \lambda)$ .

$\Gamma(n)$  bezeichnet hierbei die Gammafunktion  $\Gamma(n) = \int_0^\infty t^{n-1} e^{-t} dt$ .

**Satz A.7 (Erwartungswert und Varianz der Gammaverteilung)** .

Es gelte  $X \sim \Gamma(n, \lambda)$ . Der Erwartungswert und die Varianz der Zufallsvariablen  $X$  sind dann gegeben durch

$$E(X) = \frac{n}{\lambda} \quad (\text{A.8})$$

$$\text{Var}(X) = \frac{n}{\lambda^2} \quad (\text{A.9})$$

*Beweis:*

Kredler (1998, S.42 bzw. S.121)



## Anhang B

# Stochastische Prozesse

Zur Herleitung der verallgemeinerten Poissonverteilung wird in Abschnitt 1.5 auf stochastische Prozesse zurückgegriffen. An dieser Stelle sollen daher die wichtigsten Definitionen und Eigenschaften stochastischer Prozesse aufgeführt werden, die in Abschnitt 1.5 Verwendung finden.

**Definition B.1 (Stochastischer Prozess)** *Ein stochastischer Prozess  $N$  ist eine, durch eine Indexmenge  $T$  induzierte Ansammlung von Zufallsvariablen*

$$\{N(t), t \in T\} \tag{B.1}$$

*die auf einem gemeinsamen Wahrscheinlichkeitsraum  $\Omega$  definiert sind.*

### Bemerkung:

- Die Indexmenge  $T$  beschreibt die Entwicklung des Systems  $\{N(t), t \in T\}$ . Oft liegen die Indexmengen
    - \*  $T = [0, \infty)$  für Systeme in stetiger Zeit
    - \*  $T = \{0, 1, 2, \dots\}$  für Systeme in diskreter Zeit
    - \*  $T = \{(a, b), a < b, a, b \in \mathbb{R}^n\}$  für Systeme mit räumlicher Entwicklung
- vor.

- Der stochastische Prozess  $N$  ist eine Funktion in zwei Variablen:
  - Für einen festen Zeitpunkt  $t \in T$  ist  $N$  eine Zufallsvariable

$$N(t) = N_\omega(t), \quad \omega \in \Omega \tag{B.2}$$

- Für ein festes Ereignis  $\omega \in \Omega$  ist  $N$  eine Funktion der Zeit

$$N(t) = N_\omega(t), \quad t \in T \tag{B.3}$$

In diesem Fall wird  $N(t)$  als Realisierung, Trajektorie oder Pfad des stochastischen Prozesses bezeichnet.

- Liegt ein stochastischer Zählprozess vor, so nimmt  $N(t)$  nur ganzzahlige Werte an (Resnick, 1992, S.301).

## B.1 Der Poissonprozess

Einer der wichtigsten Zählprozesse ist der Poissonprozess. Er gibt die Anzahl an Ergebnissen an, die seit Beginn der Beobachtungsperiode eingetreten sind. Beispiele solcher Ereignisse sind z.B. Telefonanrufe in einem Call-Center oder Schadensmeldungen in einem Versicherungsportfolio. Bevor wir den Poissonprozess definieren, wollen wir das Landausymbol  $o$  einführen:

**Definition B.2 (Landausymbol  $o$ )** Die Funktion  $f$  sei in einer Umgebung von  $a$  definiert. Wir schreiben

$$f(x) = o(x) \quad \text{für } x \rightarrow a, \quad \text{falls } \lim_{x \rightarrow a} \frac{f(x)}{x} = 0 \quad (\text{B.4})$$

Für  $a=0$  schreiben wir kurz  $f(x) = o(x)$ .

$f(x) = o(x)$  bedeutet in diesem Fall, dass für  $x \rightarrow 0$  die Funktion  $f$  schneller gegen Null strebt, als  $x$  selbst.

**Definition B.3 (Poissonprozess)** Ein stochastischer Prozess  $N = \{N(t), t \geq 0\}$  heißt Poissonprozess mit Intensität  $\theta$ , falls gilt:

(i) Der Prozess startet zum Zeitpunkt Null in Zustand Null.

$$N(0) = 0, \quad \forall \omega \in \Omega \quad (\text{B.5})$$

(ii) Die Zuwächse von  $N$  sind

- *unabhängig:*  
 $N(t_1 + h_1) - N(t_1)$  und  $N(t_2 + h_2) - N(t_2)$  sind unabhängige Zufallsvariablen, falls  $(t_1, t_1 + h_1] \cap (t_2, t_2 + h_2] = \emptyset$  gilt.
- *stationär:*  
 $N(t + h) - N(t) = N(h), \quad \forall h > 0, t \geq 0$

(iii)  $N$  hat rechtseitig stetige Pfade mit Sprüngen der Höhe 1, d.h. in einem hinreichend kleinen Intervall der Länge  $h$  tritt höchstens ein Ereignis auf.

$$P[N(t+h) - N(t) = k] = o(h), \quad \forall k \geq 2 \quad (\text{B.6})$$

(iv) Die Wahrscheinlichkeit, dass in einem Intervall der Länge  $h$  mindestens ein Ereignis eintritt ist  $\theta h + o(h)$ .

$$P[N(t+h) - N(t) = 1] = \theta h + o(h) \quad (\text{B.7})$$

### Bemerkung:

Liegt dem Poissonprozess eine konstante Intensität  $\theta$  zu Grunde, so spricht man von einem homogenen Poissonprozess. Ein nicht homogener Poissonprozess liegt vor, wenn die Intensität eine Funktion der Zeit ist:  $\theta = \theta(t)$ . Für zufällige Intensität ( $\theta$  ist eine Zufallsvariable) heißt  $N$  gemischter Poissonprozess.

**Satz B.4 (Verteilung von  $N(t)$ )** Unter den Annahmen (i)-(iv) aus Definition B.3 ist  $N(t)$  poissonverteilt mit Intensität  $\theta t$ :

$$P_x(t) = P[N(t) = x] = \frac{(\theta t)^x}{x!} e^{-\theta t}, \quad t > 0, x = 0, 1, 2, \dots \quad (\text{B.8})$$

*Beweis:*

Leonenko (2003, Unit 1) oder Resnick (1992, S.303 ff.)

**Lemma B.5 (Eigenschaften eines Poissonprozesses)** Sei  $N$  ein homogener Poissonprozess mit Intensität  $\theta$ . Weiterhin sei  $t \geq 0$ ,  $s \geq 0$ ,  $0 \leq t_1 \leq \dots \leq t_n$  und  $0 \leq x_1 \leq \dots \leq x_n$  seien ganzzahlige konstante Werte. Es gilt:

$$(i) \quad E[N(t)] = \text{Var}[N(t)] = \theta t$$

$$(ii) \quad P[N(t_1) = x_1, \dots, N(t_n) = x_n] = \frac{e^{-\theta t_1} (\theta t_1)^{x_1}}{x_1!} \times \dots \times \frac{e^{-\theta(t_n - t_{n-1})} (\theta(t_n - t_{n-1}))^{x_n - x_{n-1}}}{(x_n - x_{n-1})!}$$

$$(iii) \quad \text{Cov}[N(t), N(s)] = \theta \min\{t, s\}$$

*Beweis:*

(i) Folgt unmittelbar aus Satz B.4

(ii) Die gemeinsame Dichte von  $N(t_1), \dots, N(t_n)$  lautet

$$\begin{aligned} & P[N(t_1) = x_1, N(t_2) = x_2, \dots, N(t_n) = x_n] \\ &= P[N(t_1) = x_1, N(t_2) - N(t_1) = x_2 - x_1, \dots, N(t_n) - N(t_{n-1}) = x_n - x_{n-1}] \\ &= P[N(t_1) = x_1] \times P[N(t_2) - N(t_1) = x_2 - x_1] \times \dots \times P[N(t_n) - N(t_{n-1}) = x_n - x_{n-1}] \end{aligned}$$

Die letzte Gleichung gilt auf Grund unabhängiger Zuwächse. Mit Satz B.4 folgt die Behauptung.

(iii) Wir zeigen zunächst:

$$\begin{aligned} E[N(t)N(t+s)] &= E[N(t)(N(t+s) - N(t)) + N(t)^2] \\ &= E[N(t)(N(t+s) - N(t))] + E[N(t)^2] \\ &\stackrel{\text{unabh. Zuw.}}{=} E[N(t)]E[N(s)] + E[N(t)^2] \\ &\stackrel{(B.8), (A.5)}{=} \theta t \theta s + (\theta^2 t^2 + \theta t) \end{aligned} \tag{B.9}$$

Damit gilt:

$$\begin{aligned} \text{Cov}[N(t), N(t+s)] &= E[N(t)N(t+s)] - E[N(t)]E[N(t+s)] \\ &\stackrel{(B.9)}{=} \theta t \theta s + \theta^2 t^2 + \theta t - \theta t (\theta t + \theta s) \\ &= \theta t \end{aligned}$$

□





# Anhang C

## S-Plus Funktionen

### C.1 Funktionen für die verallgemeinerte Poissonverteilung

#### C.1.1 Wahrscheinlichkeitsfunktion **dgpois**

```
" dgpois " = function(x, theta = stop( " no theta arg " ),
+   lambda = stop( " no lambda arg "))
{
  if(theta <= 0)
    return( " theta has to be positive ")
  if(lambda >= 1)
    return( " lambda has to be less than unity ")
  if(x < 0)
    p <- 0
  else {
+    p <- theta/factorial(x) * (theta + lambda * x) ^ (x - 1)
    * exp( - theta - lambda *x)
  }
  return(p)
}
```

### C.1.2 Verteilungsfunktion pgpois

```

"pgpois" = function(q, theta = stop("no theta arg"),
+   lambda = stop("no lambda arg"))
{
  if(theta <= 0)
    return("theta has to be positive")
  if(lambda >= 1)
    return("lambda has to be less than unity")
  p <- double(q + 1)
  p[1] <- exp(-theta)
  if(q != 0) {
    for(x in 1:q) {
      c <- theta - lambda + x * lambda
      p[x + 1] <- (c/x) * (1 + lambda/c) ^ (x - 1)
+     * exp(-lambda) * p[x]
    }
  }
  return(sum(p))
}

```

Die Eintrittswahrscheinlichkeiten werden hier unter Verwendung der Rekursionsgleichung (1.101) bestimmt.

### C.1.3 Quantilsfunktion `qgpois`

```
"qgpois" = function(p, theta = stop("no theta arg"),
+   lambda = stop("no lambda arg"))
{
  if(theta <= 0)
    return("theta has to be positive")
  if(lambda >= 1)
    return("lambda has to be less than unity")
  q <- 0
  pq <- pgpois(0, theta, lambda)
  while(pq < p) {
    q <- q + 1
    pq <- pgpois(q, theta, lambda)
  }
  return(q)
}
```

### C.1.4 Erzeugung von Zufallsvariablen `rgpois`

```

"rgpois" = function(n, theta = stop("no theta arg "), lambda = stop("no lambda arg "))
{
  ## UNTERPROGRAMME
  rgpois.branching = function(n, theta, lambda)
  {
    x <- double(n)
    for(i in 1:n) {
      y <- rpois(1, theta)
      x[i] <- y
      if(y > 0) {
        while(y > 0) {
          z <- rpois(1, y * lambda)
          x[i] <- x[i] + z
          y <- z
        }
      }
    }
    return(x)
  }
  rgpois.inversion = function(n, theta, lambda)
  {
    x <- double(n)
    for(i in 1:n) {
      y <- integer(1)
      s <- exp(-theta)
      p <- s
      u <- runif(1, 0, 1)
      while(u > s) {
        y <- y + 1
        c <- theta - lambda + y * lambda
        p <- (c/y) * (1 + lambda/c) ^ (y - 1) * exp(-lambda) * p
        s <- s + p
      }
      x[i] <- y
    }
    return(x)
  }
  ## METHODENWAHL
  if(theta <= 0) return("theta has to be positive ")
  if(lambda >= 1)
    return("lambda has to be less than unity ")
  x <- double(n)
  if(lambda == 0)
    x <- rpois(n, theta)
  if(lambda < 0)
    x <- rgpois.inversion(n, theta, lambda)
  if(lambda > 0)
    x <- rgpois.branching(n, theta, lambda)
  return(x)
}

```

## C.2 Funktionen für die verallgemeinerte Poissonregression

### C.2.1 Erzeugung verallgemeinert poissonverteilter Regressionsdaten

```
"response.gpois" = function(lambda, beta, X)
{
  if(is.matrix(X)) {
    n <- dim(X)[1]
    k <- dim(X)[2]
  }
  else {
    n <- length(X)
    k <- 1
  }
  theta <- double(n)
  eta <- double(n)
  Y <- double(n)
  for(i in 1:n) {
    if(k == 1) {
      eta[i] <- beta * X[i]
    }
    else {
      for(j in 1:k) {
        eta[i] <- eta[i] + beta[j] * X[i, j]
      }
    }
    theta[i] <- (1 - lambda) * exp(eta[i])
    Y[i] <- rgpois(1, theta[i], lambda)
  }
  return(Y)
}
```

## C.2.2 Maximum-Likelihood-Schätzung im verallgemeinerten Poisson-Regressionsmodell

```

"mle.gpois" = function(Yein, Xein, Offset, summary = TRUE, plot = FALSE,
+   MASS = FALSE, bound = TRUE)
{
  ## VARIABLENINITIALISIERUNG
  X <- Xein
  Y <- Yein
  if(is.matrix(X)) {
    n <- dim(X)[1]
    k <- dim(X)[2]
  }
  else {
    n <- length(X)
    k <- 1
  }
  t <- Offset
  hesse <- 0
  it <- 1
  ## STARTWERTE
  beta <- summary(glm(Y ~ offset(log(t)) + X - 1, family = poisson,
+   link = log))$coefficients[, 1]
  phi <- phi.moment(beta)
  delta <- c(phi, beta)
  ## MAXIMIERUNG DER LOG-LIKELIHOODFUNKTION
  start <- proc.time()[2]
  if(bound) {
    grenze <- c(1, rep(- Inf, k))
    if(MASS) {
      library(MASS)
      opt <- optim(delta, log.likelihood.gpois, method =
+       "L-BFGS-B", lower = grenze, hessian = TRUE)
      delta <- opt$par
      hesse <- opt$hessian
      it <- opt$counts
      log.likelihood <- - opt$value
      message <- opt$message
    }
    else {
      opt <- nlminb(objective = log.likelihood.gpois,
+       start = delta, lower = grenze)
      delta <- opt$parameter
      it <- opt$iterations
      log.likelihood <- - opt$object
      message <- opt$message
    }
  }
  else {
    if(MASS) {
      library(MASS)
      opt <- optim(delta, log.likelihood.gpois,
+       hessian = TRUE)
      delta <- opt$par
      hesse <- opt$hessian
      it <- opt$counts
      log.likelihood <- - opt$value
      message <- opt$message
    }
  }
}

```

```

    else {
      opt <- nlmin(log.likelihood.gpois, delta)
      delta <- opt$x
      it <- opt$iterations
      log.likelihood <- - log.likelihood.gpois(opt$x)
      message <- opt$message
    }
  }
  zeit <- proc.time()[2] - start
  ## AUSGABELISTE
  coef <- double(k + 1)
  for(i in 1:(k + 1)) {
    coef[i] <- delta[i]
  }
  phi <- coef[1]
  beta <- coef[2:(k + 1)]
  mu <- fit.gpois(delta)
  res <- double(n)
  RSS <- 0
  for(i in 1:n) {
    res[i] <- Y[i] - mu[i]
    RSS <- RSS + res[i] ^ 2
  }
  AIC <- -2 * log.likelihood + 2 * (k + 1)
  devianz <- devianz.gpois(Y, mu, phi)
  if(hesse == 0) {
    hesse <- hesse.gpois(Y, X, phi, mu)
  }
  eigen <- eigen(hesse)$value
  ausgabe <- list(Coefficients = beta, Dispersion.Parameter = phi,
+               Log.Lokelihood = log.likelihood, Iterations = it, Time = zeit,
+               Message = message, Hessematrix = hesse, Eigenwerte = eigen,
+               Range.mu = c(min(mu), max(mu)), RSS = RSS, Residuals = res,
+               AIC = AIC, Devianz = devianz$Devianz, Devianz.Residuals =
+               devianz$Devianz.Residuals.Quantile, Devianz.Residuals.Value =
+               devianz$Devianz.Residuals.Value, Fitted.Values = mu, Response
+               = Y, Design = X)
  if(plot) {
    plot.gpois(ausgabe)
    if(summary) summary.gpois(ausgabe)
  }
  else {
    if(summary) {
      summary.gpois(ausgabe)
    }
    else {
      rm(X, Y, n, k, t)
      return(ausgabe)
    }
  }
  rm(X, Y, n, k, t)
}

```

```
"phi.moment" = function(beta)
{
  eta <- double(n)
  mu <- double(n)
  chi <- 0
  for(i in 1:n) {
    if(k == 1) {
      eta[i] <- beta[1] * X[i]
    }
    else {
      for(j in 1:k) {
        eta[i] <- eta[i] + beta[j] * X[i, j]
      }
    }
    mu[i] <- exp(eta[i])
    chi <- chi + (Y[i] - mu[i]) ^ 2 / (mu[i] * (n - k))
  }
  phi <- sqrt(abs(chi))
  return(phi)
}
```



```
"log.likelihood.gpois" = function(delta)
{
  eta <- double(n)
  l <- double(1)
  for(i in 1:n) {
    if(k == 1) {
      eta[i] <- delta[2] * X[i]
    }
    else {
      for(j in 1:k) {
        eta[i] <- eta[i] + delta[j + 1] * X[i, j]
      }
    }
    l <- l + eta[i] + (Y[i] - 1) * log(t[i] * care.exp(eta[i]) +
+ (delta[1] - 1) * Y[i]) - Y[i] * log(delta[1]) -
+ (1/delta[1]) * (t[i] * care.exp(eta[i]) +
+ (delta[1] - 1) * Y[i])
  }
  return( - l)
}
```

```
"fit.gpois" = function(delta)
{
  eta <- double(n)
  mu <- double(n)
  for(i in 1:n) {
    if(k == 1) {
      eta[i] <- delta[2] * X[i]
    }
    else {
      for(j in 1:k) {
        eta[i] <- eta[i] + delta[j + 1] * X[i, j]
      }
    }
    mu[i] <- t[i] * exp(eta[i])
  }
  return(mu)
}
```

```
" devianz.gpois" = function(Y, mu, phi)
{
  dev.temp <- double(n)
  r <- double(n)
  for(i in 1:n) {
    if(Y[i] <= 0) {
      dev.temp[i] <- -1 * log(mu[i]) - log(t[i]) - (Y[i] - 1) *
+       log( (mu[i] + (phi - 1) * Y[i])/phi) +
+       log(Y[i] ^ Y[i]) + (mu[i] - Y[i])/phi
    }
    else {
      dev.temp[i] <- -1 * log(mu[i]) - log(t[i]) - (Y[i] - 1) *
+       log( (mu[i] + (phi - 1) * Y[i])/phi) +
+       Y[i] * log(Y[i]) + (mu[i] - Y[i])/phi
    }
    r[i] <- sign(Y[i] - mu[i]) * sqrt(2 * abs(dev.temp[i]))
  }
  dev <- sum(r ^ 2)
  devianz <- list(Devianz = dev, Devianz.Residuals.Quantile =
+   quantile(r), Devianz.Residuals.Value = r)
  return(devianz)
}
```

```

"fisher.gpois" = function(X, Y, phi, mu, n, k)
{
  I <- matrix(rep(0, (k + 1) ^ 2), k + 1, k + 1)
  if(k == 1) {
    for(i in 1:n) {
      I[1, 1] <- I[1, 1] + (2 * mu[i])/
+       (phi ^ 2 * (mu[i] - 2 + 2 * phi))
      I[1, 2] <- I[1, 2] - (X[i] * 2 * mu[i] * (phi - 1))/
+       (phi ^ 2 * (mu[i] - 2 + 2 * phi))
      I[2, 2] <- I[2, 2] + ((X[i]) ^ 2 * mu[i] * (mu[i] - 2 * phi
+       + 2 * phi ^ 2))/(phi ^ 2 * (mu[i] - 2 + 2 * phi))
    }
    I[2, 1] <- I[1, 2]
  }
  else {
    for(i in 1:n) {
      I[1, 1] <- I[1, 1] + (2 * mu[i])/
+       (phi ^ 2 * (mu[i] - 2 + 2 * phi))
    }
    for(r in 1:k) {
      for(i in 1:n) {
+       I[r + 1, 1] <- I[r + 1, 1] - (X[i, r] * 2 *
+       mu[i] * (phi - 1))/(phi ^ 2 * (mu[i] -
+       2 + 2 * phi))
      }
      I[1, r + 1] <- I[r + 1, 1]
      for(s in r:k) {
        for(i in 1:n) {
          I[r + 1, s + 1] <- I[r + 1, s + 1] +
            (X[i, r] * X[i, s] * mu[i] *
              (mu[i] - 2 * phi + 2 * phi ^ 2))/
            (phi ^ 2 * (mu[i] - 2 + 2 * phi))
        }
        I[s + 1, r + 1] <- I[r + 1, s + 1]
      }
    }
  }
  return(I)
}

```

```

"hesse.gpois" = function(Y, X, phi, mu)
{
  if(is.matrix(X)) {
    n <- dim(X)[1]
    k <- dim(X)[2]
  }
  else {
    n <- length(X)
    k <- 1
  }
  H <- matrix(double((k + 1) ^ 2), k + 1, k + 1)
  if(k == 1) {
    for(i in 1:n) {
      H[1, 1] <- H[1, 1] - (Y[i] ^ 2 * (Y[i] - 1))/(mu[i] +
+ (phi - 1) * Y[i]) ^ 2 - (2 * (mu[i] - Y[i]))/
+ (phi) ^ 3 + Y[i]/(phi) ^ 2
      H[1, 2] <- H[1, 2] + X[i] * mu[i] * ((1/phi) ^ 2 -
+ (Y[i] * (Y[i] - 1))/(mu[i] + (phi - 1) * Y[i]) ^ 2)
      H[2, 2] <- H[2, 2] - (X[i]) ^ 2 * mu[i] * (1/phi - (Y[i] *
+ (Y[i] - 1) * (phi - 1))/(mu[i] + (phi - 1) * Y[i]) ^ 2)
    }
    H[2, 1] <- H[1, 2]
  }
  else {
    for(i in 1:n) {
      H[1, 1] <- H[1, 1] - (Y[i] ^ 2 * (Y[i] - 1))/(mu[i] +
+ (phi - 1) * Y[i]) ^ 2 - (2 * (mu[i] - Y[i]))/
+ phi ^ 3 + Y[i]/phi ^ 2
    }
    for(r in 1:k) {
      for(i in 1:n) {
        H[r + 1, 1] <- H[r + 1, 1] + X[i, r] * mu[i]
+ (1/phi ^ 2 - (Y[i] * (Y[i] - 1))/
+ (mu[i] + (phi - 1) * Y[i]) ^ 2)
      }
      H[1, r + 1] <- H[r + 1, 1]
      for(s in r:k) {
        for(i in 1:n) {
          H[r + 1, s + 1] <- H[r + 1, s + 1] - X[i, r] * X[i, s] *
+ mu[i] * (1/phi - (Y[i] * (Y[i] - 1) *
+ (phi - 1))/(mu[i] + (phi - 1) * Y[i]) ^ 2)
        }
        H[s + 1, r + 1] <- H[r + 1, s + 1]
      }
    }
  }
  return(H)
}

```

```

"summary.gpois" = function(mle.daten)
{
  X = mle.daten$Design
  if(is.matrix(X)) {
    n <- dim(X)[1]
    k <- dim(X)[2]
  }
  else {
    n <- length(X)
    k <- 1
  }
  out0 <- matrix(double(5), 1, 5)
  kopf <- c(" Min" , " 1Q" , " Median" , " 3Q" , " Max" )
  for(i in 1:5) { out0[i] <- mle.daten$Devianz.Residuals[i]
  }
  dimnames(out0)[[2]] <- kopf
  dimnames(out0)[[1]] <- factor(c(" Deviance Residuals: "))
  print(out0)
  out1 <- matrix(double(k), 1, k)
  kopf <- factor(rep(" " , k))
  for(i in 1:k) {
    out1[1, i] <- mle.daten$Coefficients[i]
  }
  dimnames(out1)[[2]] <- kopf
  dimnames(out1)[[1]] <- factor(c(" Coefficients: "))
  print(out1)
  out2 <- matrix(double(1), 1, 1)
  out2[1, 1] <- mle.daten$Dispersion.Parameter
  dimnames(out2)[[2]] <- factor(c(" "))
  dimnames(out2)[[1]] <- factor(c(" Dispersion Parameter: "))
  print(out2)
  out3 <- matrix(double(1), 1, 1)
  out3[1, 1] <- mle.daten$RSS
  dimnames(out3)[[2]] <- factor(c(" "))
  dimnames(out3)[[1]] <- factor(c(" RSS: "))
  print(out3)
  out4 <- matrix(double(1), 1, 1)
  out4[1, 1] <- mle.daten$AIC
  dimnames(out4)[[2]] <- factor(c(" "))
  dimnames(out4)[[1]] <- factor(c(" AIC: "))
  print(out4)
  out5 <- matrix(double(1), 1, 1)
  out5[1, 1] <- mle.daten$Devianz
  dimnames(out5)[[2]] <- factor(c(" "))
  dimnames(out5)[[1]] <- factor(c(" Residual Deviance: "))
  print(out5)
  out6 <- matrix(double(2), 1, 2)
  out6[1, 1] <- mle.daten$Range.mu[1]
  out6[1, 2] <- mle.daten$Range.mu[2]
  dimnames(out6)[[2]] <- factor(c(" " , " "))
  dimnames(out6)[[1]] <- factor(c(" Range of mu: "))
  print(out6)
  out7 <- matrix(double(1), 1, 1)
  out7[1, 1] <- mle.daten$Message
  dimnames(out7)[[2]] <- factor(c(" "))
  dimnames(out7)[[1]] <- factor(c(" Message: "))
  print(out7)
  return()
}

```

```

"plot.gpois" = function(mle.daten)
{
  X = mle.daten$Design
  if(is.matrix(X)) {
    n <- dim(X)[1]
    k <- dim(X)[2]
  }
  else {
    n <- length(X)
    k <- 1
  }
  par(mfrow = c(2, 2))
  plot1 <- matrix(double(2 * n), n, 2)
  dimnames(plot1)[[2]] <- factor(c(" Fitted Values ", " Response " ))
  for(i in 1:n) {
    plot1[i, 1] <- mle.daten$Fitted.Values[i]
    plot1[i, 2] <- mle.daten$Response[i]
  }
  plot(plot1)
  abline(0, 1)
  plot2 <- matrix(double(2 * n), n, 2)
  dimnames(plot2)[[2]] <- factor(c(" Fitted Values ", " Residuals" ))
  for(i in 1:n) {
    plot2[i, 1] <- mle.daten$Fitted.Values[i]
    plot2[i, 2] <- mle.daten$Residuals[i]
  }
  plot(plot2)
  abline(0, 0)
  plot3 <- matrix(double(2 * n), n, 2)
  dimnames(plot3)[[2]] <- factor(c(" Fitted Values", " Deviance Residuals" ))
  for(i in 1:n) {
    plot3[i, 1] <- mle.daten$Fitted.Values[i]
    plot3[i, 2] <- mle.daten$Devianz.Residuals.Value[i]
  }
  plot(plot3)
  abline(0, 0)
  plot4 <- matrix(double(2 * n), n, 2)
  dimnames(plot4)[[2]] <- factor(c(" Fitted Values",
+   " sqrt(abs(Deviance Residuals))" ))
  for(i in 1:n) {
    plot4[i, 1] <- mle.daten$Fitted.Values[i]
    plot4[i, 2] <- sqrt(abs(mle.daten$Devianz.Residuals.Value[ i]))
  }
  plot(plot4)
}

```

## C.3 Funktionen für die ZIGP-Regression

### C.3.1 Erzeugung von ZIGP $\mu$ -Regressionsdaten

```
"response.zigp1" = function(omega, lambda, beta, X)
{
  if(is.matrix(X)) {
    n <- dim(X)[1]
    k <- dim(X)[2]
  }
  else {
    n <- length(X)
    k <- 1
  }
  theta <- double(n)
  eta <- double(n)
  Y <- double(n)
  for(i in 1:n) {
    if(k == 1) {
      eta[i] <- beta * X[i]
    }
    else {
      for(j in 1:k) {
        eta[i] <- eta[i] + beta[j] * X[i, j]
      }
    }
    theta[i] <- (1 - lambda) * exp(eta[i])
    Y[i] <- rbinom(1, 1, 1 - omega) * rgpois(1, theta[i], lambda)
  }
  return(Y)
}
```



### C.3.2 Maximum-Likelihood-Schätzung im ZIGP $\mu$ -Regressionsmodell

```

"mle.zigp1" = function(Yein, Xein, Offset, summary = TRUE, plot = FALSE, MASS = FALSE)
{
  X <- Xein
  Y <- Yein
  if(is.matrix(X)) {
    n <- dim(X)[1]
    k <- dim(X)[2]
  }
  else {
    n <- length(X)
    k <- 1
  }
  t <- Offset
  grenze.unten <- c(1, 0, rep(- Inf, k))
  grenze.oben <- c(Inf, 1 - 10^99, rep(Inf, k))
  ## STARTWERTE
  phi <- 1
  omega <- 0
  beta <- summary(glm(Y ~ offset(log(t)) + X - 1, family = poisson, link =
+   log))$coefficients[, 1]
  delta <- c(phi, omega, beta)
  ## MAXIMIERUNG DER LOG-LIKELIHOODFUNKTION
  if(MASS) {
    library(MASS)
    start <- proc.time()[2]
    opt <- optim(fn = log.likelihood.zigp1, par = delta, lower =
+   grenze.unten, upper = grenze.oben, hessian = TRUE,
+   method = "L-BFGS-B")
    delta <- opt$par
    hesse <- opt$hessian
    eigen <- eigen(hesse)$values
    it <- opt$counts
    log.likelihood <- - opt$value
    message <- opt$message
  }
  else {
    start <- proc.time()[2]
    opt <- nlminb(objective = log.likelihood.zigp1, start = delta,
+   lower = grenze.unten, upper = grenze.oben)
    delta <- opt$parameter
    hesse <- NA
    eigen <- NA
    it <- opt$iterations
    log.likelihood <- - opt$object
    message <- opt$message
  }
  zeit <- proc.time()[2] - start
  ## AUSGABELISTE
  coef <- double(k + 2)
  for(i in 1:(k + 2)) {
    coef[i] <- delta[i]
  }
  phi <- coef[1]
  omega <- coef[2]
  beta <- coef[3:(k + 2)]
  fit <- fit.zigp1(delta)
  res <- double(n)
}

```

```
RSS <- 0
for(i in 1:n) {
  res[i] <- Y[i] - fit$fit[i]
  RSS <- RSS + res[i]^2
}
AIC <- -2 * log.likelihood + 2 * (k + 2)
range.mu <- c(min(fit$mu), max(fit$mu))
ausgabe <- list(ZI.Parameter = omega, Coefficients = beta,
+   Dispersion.Parameter = phi, Range.mu = range.mu,
+   Log.Likelihood = log.likelihood, Residuals = res, RSS = RSS,
+   AIC = AIC, Iterations = it, Time = zeit, Message = message,
+   Hessematrix = hesse, Eigenwerte = eigen, Response = Y,
+   Fitted.Values = fit$fit, Design = X)
if(plot) {
  plot.zigp1(ausgabe)
  if(summary)
    summary.zigp1(ausgabe)
}
else {
  if(summary) {
    summary.zigp1(ausgabe)
  }
  else {
    rm(X, Y, n, k, t)
    return(ausgabe)
  }
}
rm(X, Y, n, k, t)
}
```

```
"log.likelihood.zigp1" = function(delta)
{
  eta <- double(n)
  s1 <- double(1)
  s2 <- double(1)
  for(i in 1:n) {
    if(k == 1) {
      eta[i] <- delta[3] * X[i]
    }
    else {
      for(j in 1:k) {
        eta[i] <- eta[i] + delta[j + 2] * X[i, j]
      }
    }
    if(Y[i] == 0) {
      s1 <- s1 + log(delta[2]/(1 - delta[2]) + care.exp(
+         - t[i]/delta[1] * care.exp(eta[i])))
    }
    if(Y[i] > 0) {
      s2 <- s2 + eta[i] + (Y[i] - 1) * log(t[i] * care.exp(
+         eta[i]) + (delta[1] - 1) * Y[i]) - Y[i] * log(
+         delta[1] - 1/delta[1] * (t[i] * care.exp(
+         eta[i]) + (delta[1] - 1) * Y[i])
      )
    }
  }
  l <- s1 + s2 + n * log(1 - delta[2])
  return( - l)
}
```

```
"fit.zigp1" = function(delta)
{
  eta <- double(n)
  mu <- double(n)
  fit <- double(n)
  for(i in 1:n) {
    if(k == 1) {
      eta[i] <- delta[3] * X[i]
    }
    else {
      for(j in 1:k) {
        eta[i] <- eta[i] + delta[j + 2] * X[i, j]
      }
    }
    mu[i] <- t[i] * care.exp(eta[i])
    fit[i] <- (1 - delta[2]) * mu[i]
  }
  return(list(fit = fit, mu = mu))
}
```

```
" summary.zigp1" = function(mle.daten)
{
  X = mle.daten$Design
  if(is.matrix(X)) {
    n <- dim(X)[1]
    k <- dim(X)[2]
  }
  else {
    n <- length(X)
    k <- 1
  }
  out0 <- matrix(double(1), 1, 1)
  out0[1, 1] <- mle.daten$ZI.Parameter
  dimnames(out0)[[2]] <- factor(c(" "))
  dimnames(out0)[[1]] <- factor(c(" ZI-Parameter: "))
  print(out0)
  out1 <- matrix(double(k), 1, k)
  kopf <- factor(rep(" ", k))
  for(i in 1:k) {
    out1[1, i] <- mle.daten$Coefficients[i]
  }
  dimnames(out1)[[2]] <- kopf
  dimnames(out1)[[1]] <- factor(c(" Coefficients: "))
  print(out1)
  out2 <- matrix(double(1), 1, 1)
  out2[1, 1] <- mle.daten$Dispersion.Parameter
  dimnames(out2)[[2]] <- factor(c(" "))
  dimnames(out2)[[1]] <- factor(c(" Dispersion Parameter:"))
  print(out2)
  out3 <- matrix(double(1), 1, 1)
  out3[1, 1] <- mle.daten$RSS
  dimnames(out3)[[2]] <- factor(c(" "))
  dimnames(out3)[[1]] <- factor(c(" RSS: "))
  print(out3)
  out4 <- matrix(double(1), 1, 1)
  out4[1, 1] <- mle.daten$AIC
  dimnames(out4)[[2]] <- factor(c(" "))
  dimnames(out4)[[1]] <- factor(c(" AIC: "))
  print(out4)
  out5 <- matrix(double(2), 1, 2)
  out5[1, 1] <- mle.daten$Range.mu[1]
  out5[1, 2] <- mle.daten$Range.mu[2]
  dimnames(out5)[[2]] <- factor(c(" ", " "))
  dimnames(out5)[[1]] <- factor(c(" Range of mu: "))
  print(out5)
  out6 <- matrix(double(1), 1, 1)
  out6[1, 1] <- mle.daten$Message
  dimnames(out6)[[2]] <- factor(c(" "))
  dimnames(out6)[[1]] <- factor(c(" Message: "))
  print(out6)
  return()
}
```

```
"plot.zigp1" = function(mle.daten)
{
  X = mle.daten$Design
  if(is.matrix(X)) {
    n <- dim(X)[1]
    k <- dim(X)[2]
  }
  else {
    n <- length(X)
    k <- 1
  }
  par(mfrow = c(1, 2))
  plot1 <- matrix(double(2 * n), n, 2)
  dimnames(plot1)[[2]] <- factor(c(" Fitted Values" , " Response" ))
  for(i in 1:n) {
    plot1[i, 1] <- mle.daten$Fitted.Values[i]
    plot1[i, 2] <- mle.daten$Response[i]
  }
  plot(plot1)
  abline(0, 1)
  plot2 <- matrix(double(2 * n), n, 2)
  dimnames(plot2)[[2]] <- factor(c(" Fitted Values" , " Residuals" ))
  for(i in 1:n) {
    plot2[i, 1] <- mle.daten$Fitted.Values[i]
    plot2[i, 2] <- mle.daten$Residuals[i]
  }
  plot(plot2)
  abline(0, 0)
}
```

### C.3.3 Erzeugung von ZIGP $_{\mu\omega}$ -Regressionsdaten

```
"response.zigp2" = function(alpha, beta, lambda, A, B)
{
  if(is.matrix(A)) {
    n <- dim(A)[1]
    m <- dim(A)[2]
  }
  else {
    n <- length(A)
    m <- 1
  }
  if(is.matrix(B)) {
    k <- dim(B)[2]
  }
  else {
    k <- 1
  }
  theta <- double(n)
  omega <- double(n)
  eta <- double(n)
  nu <- double(n)
  Y <- double(n)
  for(i in 1:n) {
    if(m == 1) {
      nu[i] <- alpha * A[i]
    }
    else {
      for(j in 1:m) {
        nu[i] <- nu[i] + alpha[j] * A[i, j]
      }
    }
    omega[i] <- exp(nu[i]) / (1 + exp(nu[i]))
    if(k == 1) {
      eta[i] <- beta * B[i]
    }
    else {
      for(j in 1:k) {
        eta[i] <- eta[i] + beta[j] * B[i, j]
      }
    }
    theta[i] <- (1 - lambda) * exp(eta[i])
    Y[i] <- rbinom(1, 1, 1 - omega[i]) * rgpois(1, theta[i], lambda)
  }
  return(Y)
}
```

### C.3.4 Maximum-Likelihood-Schätzung im ZIGP $\mu\omega$ -Regressionsmodell

Maximum-Likelihood-Schätzung: mle.zigp2 und Unterprogramme

```
"mle.zigp2" = function(Yein, Aein, Bein, Offset, summary = TRUE, plot = FALSE, MASS = FALSE)
{
  ## INITIALISIERUNG
  A <- Aein
  B <- Bein
  Y <- Yein
  if(is.matrix(A)) {
    n <- dim(A)[1]
    m <- dim(A)[2]
  }
  else {
    n <- length(A)
    m <- 1
  }
  if(is.matrix(B)) {
    if(dim(B)[1] != n)
      return("Dimensionen von A und B stimmen nicht ueberein")
    k <- dim(B)[2]
  }
  else {
    if(length(B) != n)
      return("Dimensionen von A und B stimmen nicht ueberein")
    k <- 1
  }
  t <- Offset
  grenze.unten <- c(1, rep( - Inf, m), rep( - Inf, k))
  grenze.oben <- c(Inf, rep(Inf, m), rep(Inf, k))
  ## STARTWERTE
  alpha <- rep(0, m)
  beta <- summary(glm(Y  offset(log(t)) + B - 1, family = poisson,
+   link = log))$coefficients[, 1]
  phi <- 1
  delta <- c(phi, alpha, beta)
  ## MAXIMIERUNG DER LOG-LIKELIHOODFUNKTION
  if(MASS) {
    library(MASS)
    start <- proc.time()[2]
    opt <- optim(fn = log.likelihood.zigp2, par = delta, lower =
+   grenze.unten, upper = grenze.oben, hessian = TRUE,
+   method = "L-BFGS-B" )
    delta <- opt$par
    hesse <- opt$hessian
    eigen <- eigen(hesse)$values
    it <- opt$counts
    log.likelihood <- - opt$value
    message <- opt$message
  }
  else {
    start <- proc.time()[2]
    opt <- nlmnb(objective = log.likelihood.zigp2, start = delta,
+   lower = grenze.unten, upper = grenze.oben)
    delta <- opt$parameter
    hesse <- NA
    eigen <- NA
    it <- opt$iterations
    log.likelihood <- - opt$object
    message <- opt$message
  }
}
```



```

}
zeit <- proc.time()[2] - start
## AUSGABELISTE
coef <- double(m + k + 1)
for(i in 1:(m + k + 1)) {
  coef[i] <- delta[i]
}
phi <- coef[1]
alpha <- coef[2:(m + 1)]
beta <- coef[(m + 2):(m + k + 1)]
fit <- fit.zigp2(delta)
res <- double(n)
RSS <- 0
for(i in 1:n) {
  res[i] <- Y[i] - fit$fit[i]
  RSS <- RSS + res[i]^2
}
AIC <- -2 * log.likelihood + 2 * (k + m + 1)
range.mu <- c(min(fit$mu), max(fit$mu))
range.omega <- c(min(fit$omega), max(fit$omega))
ausgabe <- list(Coefficients.ZI = alpha, Coefficients.GP = beta,
+   Dispersion.Parameter = phi, Range.mu = range.mu, Range.omega
+   = range.omega, Log.Likelihood = log.likelihood, Residuals =
+   res, RSS = RSS, AIC = AIC, Iterations = it, Time = zeit,
+   Message = message, Hessematrix = hesse, Eigenwerte = eigen,
+   Response = Y, Fitted.Values = fit$fit, DesignA = A, DesignB =
+   B, Offset = t)
if(plot) {
  plot.zigp2(ausgabe)
  if(summary)
    summary.zigp2(ausgabe)
}
else {
  if(summary) {
    summary.zigp2(ausgabe)
  }
  else {
    rm(A, B, Y, n, k, m, t)
    return(ausgabe)
  }
}
rm(A, B, Y, k, m, n, t)
}

```

```

" log.likelihood.zigp2" = function(delta)
{
  eta <- double(n)
  nu <- double(n)
  l <- double(1)
  s1 <- double(1)
  s2 <- double(1)
  s3 <- double(1)
  for(i in 1:n) {
    if(m == 1) {
      nu[i] <- delta[2] * A[i]
    }
    else {
      for(j in 1:m) {
        nu[i] <- nu[i] + delta[j + 1] * A[i, j]
      }
    }
    if(k == 1) {
      eta[i] <- delta[2 + m] * B[i]
    }
    else {
      for(j in 1:k) {
        eta[i] <- eta[i] + delta[j + 1 + m] * B[i,j]
      }
    }
    if(Y[i] == 0) {
      s1 <- s1 + log(care.exp(nu[i]) + care.exp(-1/delta[1]
+      * t[i] * care.exp(eta[i])))
+
    }
    if(Y[i] > 0) {
      s2 <- s2 + log(t[i] + eta[i] + (Y[i] - 1) * log(
+      t[i] * care.exp(eta[i]) + (delta[1] - 1) *
+      Y[i] - Y[i] * log(delta[1]) - 1/delta[1] *
+      (t[i] * care.exp(eta[i]) + (delta[1] - 1) *
+      Y[i]))
    }
    s3 <- s3 + log(1 + care.exp(nu[i]))
  }
  l <- s1 + s2 - s3
  return( - l)
}

```

```
" fit.zigp2" = function(delta)
{
  eta <- double(n)
  nu <- double(n)
  mu <- double(n)
  omega <- double(n)
  fit <- double(n)
  for(i in 1:n) {
    if(m == 1) {
      nu[i] <- delta[2] * A[i]
    }
    else {
      for(j in 1:m) {
        nu[i] <- nu[i] + delta[j + 1] * A[i, j]
      }
    }
    if(k == 1) {
      eta[i] <- delta[2 + m] * B[i]
    }
    else {
      for(j in 1:k) {
        eta[i] <- eta[i] + delta[j + 1 + m] * B[i, j]
      }
    }
    omega[i] <- care.exp(nu[i]) / (1 + care.exp(nu[i]))
    mu[i] <- t[i] * care.exp(eta[i])
    fit[i] <- (1 - omega[i]) * mu[i]
  }
  return(list(fit = fit, mu = mu, omega = omega))
}
```

```

" summary.zigp2" = function(mle.daten)
{
  A = mle.daten$DesignA
  B = mle.daten$DesignB
  if(is.matrix(A)) {
    n <- dim(A)[1]
    m <- dim(A)[2]
  }
  else {
    n <- length(A)
    m <- 1
  }
  if(is.matrix(B)) {
    if(dim(B)[1] != n)
      return(" Dimensionen von A und B stimmen nicht ueberein" )
    k <- dim(B)[2]
  }
  else {
    if(length(B) != n)
      return(" Dimensionen von A und B stimmen nicht ueberein" )
    k <- 1
  }
  out0 <- matrix(double(m), 1, m)
  for(i in 1:m) {
    out0[1, i] <- mle.daten$Coefficients.ZI[i]
  }
  dimnames(out0)[[2]] <- factor(rep(" ", m))
  dimnames(out0)[[1]] <- factor(c(" Coefficients.ZI: "))
  print(out0)
  out1 <- matrix(double(k), 1, k)
  kopf <- factor(rep(" ", k))
  for(i in 1:k) {
    out1[1, i] <- mle.daten$Coefficients.GP[i]
  }
  dimnames(out1)[[2]] <- kopf
  dimnames(out1)[[1]] <- factor(c(" Coefficients.GP: "))
  print(out1)
  out2 <- matrix(double(1), 1, 1)
  out2[1, 1] <- mle.daten$Dispersion.Parameter
  dimnames(out2)[[2]] <- factor(c(" "))
  dimnames(out2)[[1]] <- factor(c(" Dispersion Parameter:"))
  print(out2)
  out3 <- matrix(double(1), 1, 1)
  out3[1, 1] <- mle.daten$RSS
  dimnames(out3)[[2]] <- factor(c(" "))
  dimnames(out3)[[1]] <- factor(c(" RSS: "))
  print(out3)
  out4 <- matrix(double(1), 1, 1)
  out4[1, 1] <- mle.daten$AIC
  dimnames(out4)[[2]] <- factor(c(" "))
  dimnames(out4)[[1]] <- factor(c(" AIC: "))
  print(out4)
  out5 <- matrix(double(2), 1, 2)
  out5[1, 1] <- mle.daten$Range.mu[1]
  out5[1, 2] <- mle.daten$Range.mu[2]
  dimnames(out5)[[2]] <- factor(c(" ", " "))
  dimnames(out5)[[1]] <- factor(c(" Range of mu: "))
  print(out5)
  out6 <- matrix(double(2), 1, 2)

```

```
out6[1, 1] <- mle.daten$Range.omega[1]
out6[1, 2] <- mle.daten$Range.omega[2]
dimnames(out6)[[2]] <- factor(c(" ", " "))
dimnames(out6)[[1]] <- factor(c(" Range of omega: "))
print(out6)
out7 <- matrix(double(1), 1, 1)
out7[1, 1] <- mle.daten$Message
dimnames(out7)[[2]] <- factor(c(" "))
dimnames(out7)[[1]] <- factor(c(" Message: "))
print(out7)
return()
}
```

```
"plot.zigp2" = function(mle.daten)
{
  X = mle.daten$DesignA
  if(is.matrix(X)) {
    n <- dim(X)[1]
    k <- dim(X)[2]
  }
  else {
    n <- length(X)
    k <- 1
  }
  par(mfrow = c(1, 2))
  plot1 <- matrix(double(2 * n), n, 2)
  dimnames(plot1)[[2]] <- factor(c(" Fitted Values" , " Response" ))
  for(i in 1:n) {
    plot1[i, 1] <- mle.daten$Fitted.Values[i]
    plot1[i, 2] <- mle.daten$Response[i]
  }
  plot(plot1)
  abline(0, 1)
  plot2 <- matrix(double(2 * n), n, 2)
  dimnames(plot2)[[2]] <- factor(c(" Fitted Values" , " Residuals" ))
  for(i in 1:n) {
    plot2[i, 1] <- mle.daten$Fitted.Values[i]
    plot2[i, 2] <- mle.daten$Residuals[i]
  }
  plot(plot2)
  abline(0, 0)
}
```

### C.3.5 Erzeugung von ZIGP $\omega$ -Regressionsdaten

```
"response.zigp3" = function(alpha, theta, lambda, X)
{
  if(is.matrix(X)) {
    n <- dim(X)[1]
    m <- dim(X)[2]
  }
  else {
    n <- length(X)
    m <- 1
  }
  omega <- double(n)
  eta <- double(n)
  nu <- double(n)
  Y <- double(n)
  for(i in 1:n) {
    if(m == 1) {
      nu[i] <- alpha * X[i]
    }
    else {
      for(j in 1:m) {
        nu[i] <- nu[i] + alpha[j] * X[i, j]
      }
    }
    omega[i] <- exp(nu[i]) / (1 + exp(nu[i]))
    Y[i] <- rbinom(1, 1, 1 - omega[i]) * rgpois(1, theta, lambda)
  }
  return(Y)
}
```

### C.3.6 Maximum-Likelihood-Schätzung im ZIGP $\omega$ -Regressionsmodell

Maximum-Likelihood-Schätzung: mle.zigp3 und Unterprogramme

```
"mle.zigp3" = function(Yein, Xein, Offset, summary = TRUE, plot = FALSE, MASS = FALSE)
{
  ## INITIALISIERUNG
  X <- Xein
  Y <- Yein
  if(is.matrix(X)) {
    n <- dim(X)[1]
    m <- dim(X)[2]
  }
  else {
    n <- length(X)
    m <- 1
  }
  t <- Offset
  grenze.unten <- c(1, rep(-Inf, m), 10^-99)
  grenze.oben <- c(Inf, rep(Inf, m), Inf)
  ## STARTWERTE
  alpha <- rep(0, m)
  mu <- 1
  phi <- 1
  delta <- c(phi, alpha, mu)
  ## MAXIMIERUNG DER LOG-LIKELIHOODFUNKTION
  if(MASS) {
    library(MASS)
    start <- proc.time()[2]
    opt <- optim(fn = log.likelihood.zigp3, par = delta, lower =
      grenze.unten, upper = grenze.oben, hessian = TRUE,
      method = "L-BFGS-B" )
    delta <- opt$par
    hesse <- opt$hessian
    eigen <- eigen(hesse)$values
    it <- opt$counts
    log.likelihood <- - opt$value
    message <- opt$message
  }
  else {
    start <- proc.time()[2]
    opt <- nlminb(objective = log.likelihood.zigp3, start = delta,
      lower = grenze.unten, upper = grenze.oben)
    delta <- opt$parameter
    hesse <- NA
    eigen <- NA
    it <- opt$iterations
    log.likelihood <- - opt$object
    message <- opt$message
  }
  zeit <- proc.time()[2] - start
  ## AUSGABELISTE
  coef <- double(m + 2)
  for(i in 1:(m + 2)) {
    coef[i] <- delta[i]
  }
  phi <- coef[1]
  alpha <- coef[2:(m + 1)]
  mu <- coef[m + 2]
  fit <- fit.zigp3(delta)
}
```



```
res <- double(n)
RSS <- 0
for(i in 1:n) {
  res[i] <- Y[i] - fit$fit[i]
  RSS <- RSS + res[i]^ 2
}
AIC <- -2 * log.likelihood + 2 * (m + 2)
range.omega <- c(min(fit$omega), max(fit$omega))
ausgabe <-< list(Coefficients.ZI = alpha, mu = mu,
  Dispersion.Parameter = phi, Range.omega = range.omega,
  Log.Likelihood = log.likelihood, Residuals = res, RSS = RSS,
  AIC = AIC, Iterations = it, Time = zeit, Message = message,
  Hessematrix = hesse, Eigenvalues = eigen, Response = Y,
  Fitted.Values = fit$fit, DesignX = X)
if(plot) {
  plot.zigp3(ausgabe)
  if(summary)
    summary.zigp3(ausgabe)
}
else {
  if(summary) {
    summary.zigp3(ausgabe)
  }
  else {
    rm(X, Y, n, m, t)
    return(ausgabe)
  }
}
rm(X, Y, m, n, t)
}
```

```

" log.likelihood.zigp3" = function(delta)
{
  nu <- double(n)
  l <- double(1)
  s1 <- double(1)
  s2 <- double(1)
  s3 <- double(1)
  for(i in 1:n) {
    if(m == 1) {
      nu[i] <- delta[2] * X[i]
    }
    else {
      for(j in 1:m) {
        nu[i] <- nu[i] + delta[j + 1] * X[i, j]
      }
    }
    if(Y[i] == 0) {
      s1 <- s1 + log(care.exp(nu[i]) + care.exp((- t[i] *
        delta[m + 2])/delta[1]))
    }
    if(Y[i] > 0) {
      s2 <- s2 + log(t[i]) + log(delta[m + 2]) + (Y[i] -
        1) * log(t[i] * delta[m + 2] + (delta[1] -
        1) * Y[i]) - Y[i] * log(delta[1]) - 1/delta[
        1] * (t[i] * delta[m + 2] + (delta[1] - 1) *
        Y[i])
    }
    s3 <- s3 + log(1 + care.exp(nu[i]))
  }
  l <- s1 + s2 - s3
  return( - l)
}

```

```
" fit.zigp3" = function(delta)
{
  nu <- double(n)
  omega <- double(n)
  fit <- double(n)
  for(i in 1:n) {
    if(m == 1) {
      nu[i] <- delta[2] * X[i]
    }
    else {
      for(j in 1:m) {
        nu[i] <- nu[i] + delta[j + 1] * X[i, j]
      }
    }
    omega[i] <- care.exp(nu[i])/(1 + care.exp(nu[i]))
    fit[i] <- (1 - omega[i]) * t[i] * delta[m + 2]
  }
  return(list(fit = fit, omega = omega))
}
```

```

" summary.zigp3" = function(mle.daten)
{
  X = mle.daten$DesignX
  if(is.matrix(X)) {
    n <- dim(X)[1]
    m <- dim(X)[2]
  }
  else {
    n <- length(X)
    m <- 1
  }
  out0 <- matrix(double(m), 1, m)
  for(i in 1:m) {
    out0[1, i] <- mle.daten$Coefficients.ZI[i]
  }
  dimnames(out0)[[2]] <- factor(rep(" ", m))
  dimnames(out0)[[1]] <- factor(c(" Coefficients.ZI: "))
  print(out0)
  out1 <- matrix(double(1), 1, 1)
  kopf <- factor(c(" "))
  out1[1, 1] <- mle.daten$mu
  dimnames(out1)[[2]] <- kopf
  dimnames(out1)[[1]] <- factor(c(" mu: "))
  print(out1)
  out2 <- matrix(double(1), 1, 1)
  out2[1, 1] <- mle.daten$Dispersion.Parameter
  dimnames(out2)[[2]] <- factor(c(" "))
  dimnames(out2)[[1]] <- factor(c(" Dispersion Parameter: "))
  print(out2)
  out3 <- matrix(double(1), 1, 1)
  out3[1, 1] <- mle.daten$RSS
  dimnames(out3)[[2]] <- factor(c(" "))
  dimnames(out3)[[1]] <- factor(c(" RSS: "))
  print(out3)
  out4 <- matrix(double(1), 1, 1)
  out4[1, 1] <- mle.daten$AIC
  dimnames(out4)[[2]] <- factor(c(" "))
  dimnames(out4)[[1]] <- factor(c(" AIC: "))
  print(out4)
  out5 <- matrix(double(2), 1, 2)
  out5[1, 1] <- mle.daten$Range.omega[1]
  out5[1, 2] <- mle.daten$Range.omega[2]
  dimnames(out5)[[2]] <- factor(c(" ", " "))
  dimnames(out5)[[1]] <- factor(c(" Range of omega: "))
  print(out5)
  out6 <- matrix(double(1), 1, 1)
  out6[1, 1] <- mle.daten$Message
  dimnames(out6)[[2]] <- factor(c(" "))
  dimnames(out6)[[1]] <- factor(c(" Message: "))
  print(out6)
  return()
}

```

```
"plot.zigp3" = function(mle.daten)
{
  X = mle.daten$Design
  if(is.matrix(X)) {
    n <- dim(X)[1]
    k <- dim(X)[2]
  }
  else {
    n <- length(X)
    k <- 1
  }
  par(mfrow = c(, 2))
  plot1 <- matrix(double(2 * n), n, 2)
  dimnames(plot1)[[2]] <- factor(c(" Fitted Values" , " Response" ))
  for(i in 1:n) {
    plot1[i, 1] <- mle.daten$Fitted.Values[i]
    plot1[i, 2] <- mle.daten$Response[i]
  }
  plot(plot1)
  abline(0, 1)
  plot2 <- matrix(double(2 * n), n, 2)
  dimnames(plot2)[[2]] <- factor(c(" Fitted Values" , " Residuals" ))
  for(i in 1:n) {
    plot2[i, 1] <- mle.daten$Fitted.Values[i]
    plot2[i, 2] <- mle.daten$Residuals[i]
  }
  plot(plot2)
  abline(0, 0)
}
```

## C.4 Funktionen für die ZHGP-Regression

### C.4.1 Erzeugung von ZHGP-Regressionsdaten

```
" response.zhgp" = function(alpha, beta, lambda, X)
{
  if(is.matrix(X)) {
    n <- dim(X)[1]
    k <- dim(X)[2]
  }
  else {
    n <- length(X)
    k <- 1
  }
  theta <- double(n)
  gamma <- double(n)
  eta <- double(n)
  nu <- double(n)
  Y1 <- double(n)
  Y2 <- double(n)
  Y <- double(n)
  for(i in 1:n) {
    if(k == 1) {
      nu[i] <- alpha * X[i]
      eta[i] <- beta * X[i]
    }
    else {
      for(j in 1:k) {
        nu[i] <- nu[i] + alpha[j] * X[i, j]
        eta[i] <- eta[i] + beta[j] * X[i, j]
      }
    }
    theta[i] <- (1 - lambda[2]) * exp(eta[i])
    gamma[i] <- (1 - lambda[1]) * exp(nu[i])
    Y1[i] <- rgeom(1, theta = gamma[i], lambda = lambda[1])
    Y2[i] <- 0
    while(Y2[i] == 0) {
      Y2[i] <- rgeom(1, theta = theta[i], lambda = lambda[2])
    }
    if(Y1[i] == 0) {
      Y[i] <- Y1[i]
    }
    else {
      Y[i] <- Y2[i]
    }
  }
  return(Y)
}
```

## C.4.2 Maximum-Likelihood-Schätzung im ZHGP-Regreessionsmodell

```

"mle.zhgp" = function(Yein, Xein, Offset, summary = TRUE, plot = FALSE)
{
  ## INITIALISIERUNG
  X <- Xein
  Y <- Yein
  if(is.matrix(X)) {
    n <- dim(X)[1]
    k <- dim(X)[2]
  }
  else {
    n <- length(X)
    k <- 1
  }
  t <- Offset
  grenze.unten <- c(0, rep(- Inf, k))
  grenze.oben <- c(1, rep(Inf, k))
  ## STARTWERTE
  alpha <- summary(glm(Y   offset(log(t)) + X - 1, family = poisson, link =
+   log))$coefficients[, 1]
  beta <- alpha
  lambda <- 0
  ## MAXIMIERUNG DER LOG-LIKELIHOODFUNKTIONEN
  start <- proc.time()[2]
  opt1 <- nlmin(f = log.likelihood.zhgp1, x = alpha)
  opt2 <- nlminb(objective = log.likelihood.zhgp2, start = c(lambda,
+   beta), lower = grenze.unten, upper = grenze.oben)
  log.likelihood <- - log.likelihood.zhgp1(opt1$x) - opt2$object
  message1 <- opt1$conv.type
  message2 <- opt2$message
  zeit <- proc.time()[2] - start
  ## AUSGABELISTE
  coef <- double(2 * k + 1)
  for(i in 1:(2 * k + 1)) {
    coef[i] <- c(opt2$parameters, opt1$x)[i]
  }
  lambda <- coef[1]
  beta <- coef[2:(k + 1)]
  alpha <- coef[(k + 2):(2 * k + 1)]
  schaetzer <- c(lambda, alpha, beta)
  fit <- fit.zhgp(schaetzer)
  res <- double(n)
  RSS <- 0
  for(i in 1:n) {
    res[i] <- Y[i] - fit$fit[i]
    RSS <- RSS + res[i] ^ 2
  }
  AIC <- -2 * log.likelihood + 2 * (2 * k + 1)
  ausgabe <- list(Coefficients.GP1 = alpha, Coefficients.GP2 = beta,
+   Dispersion.Parameter = lambda, Log.Likelihood = log.likelihood,
+   Residuals = res, RSS = RSS, AIC = AIC, Time = zeit, Message1
+   = message1, Message2 = message2, Fitted.Values = fit$fit,
+   Response = Y, Design = X, Offset = t)
  if(plot) {
    plot.zhgp(ausgabe)
    if(summary)
      summary.zhgp(ausgabe)
  }
}

```

```
    else {
      if(summary) {
        summary.zhgp(ausgabe)
      }
      else {
        rm(X, Y, n, k, t)
        return(ausgabe)
      }
    }
  }
  rm(X, Y, k, n, t)
}
```



```
"log.likelihood.zhgp1" = function(alpha)
{
  nu <- double(n)
  l <- double(1)
  s1 <- double(1)
  s2 <- double(1)
  for(i in 1:n) {
    if(k == 1) {
      nu[i] <- alpha * X[i]
    }
    else {
      for(j in 1:k) {
        nu[i] <- nu[i] + alpha[j] * X[i, j]
      }
    }
    if(Y[i] == 0) {
      s1 <- s1 - t[i] * care.exp(nu[i])
    }
    if(Y[i] > 0) {
      s2 <- s2 + log(1 - care.exp(- t[i] * care.exp(nu[i])))
    }
  }
  l <- s1 + s2
  return( - l)
}
```

```

" log.likelihood.zhgp2" = function(delta)
{
  nu <- double(n)
  eta <- double(n)
  l <- double(1)
  for(i in 1:n) {
    if(k == 1) {
      eta[i] <- delta[2] * X[i]
    }
    else {
      for(j in 1:k) {
        eta[i] <- eta[i] + delta[j + 1] * X[i, j]
      }
    }
    if(Y[i] > 0) {
      l <- l - log(1 - care.exp(- t[i] * care.exp(eta[i])))
      + eta[i] + log(t[i]) + (Y[i] - 1) * log(
      + t[i] * care.exp(eta[i]) + delta[1] * Y[i]) -
      + log(factorial(Y[i])) - t[i] * care.exp(eta[i])
      + - delta[1] * Y[i]
    }
  }
  return( - l)
}

```

```
" fit.zhgp" = function(delta)
{
  eta <- double(n)
  nu <- double(n)
  rho <- double(n)
  fit <- double(n)
  for(i in 1:n) {
    if(k == 1) {
      nu[i] <- delta[2] * X[i]
      eta[i] <- delta[3] * X[i]
    }
    else {
      for(j in 1:k) {
        nu[i] <- nu[i] + delta[j + 1] * X[i, j]
        eta[i] <- eta[i] + delta[j + 1 + k] * X[i, j]
      }
    }
    rho[i] <- (1 - care.exp(- t[i] * care.exp(nu[i])))/(1 -
+   care.exp(- t[i] * care.exp(eta[i])))
    fit[i] <- (rho[i] * t[i] * care.exp(eta[i]))/(1 - delta[1])
  }
  return(list(fit = fit, rho = rho))
}
```

```

" summary.zhgp" = function(mle.daten)
{
  X = mle.daten$Design
  if(is.matrix(X)) {
    n <- dim(X)[1]
    k <- dim(X)[2]
  }
  else {
    n <- length(X)
    k <- 1
  }
  out0 <- matrix(double(k), 1, k)
  for(i in 1:k) {
    out0[1, i] <- mle.daten$Coefficients.GP1[i]
    dimnames(out0)[[2]] <- factor(rep(" ", k))
    dimnames(out0)[[1]] <- factor(c(" Coefficients.GP1: "))
    print(out0)
  }
  out1 <- matrix(double(k), 1, k)
  for(i in 1:k) {
    out1[1, i] <- mle.daten$Coefficients.GP2[i]
  }
  dimnames(out1)[[2]] <- factor(rep(" ", k))
  dimnames(out1)[[1]] <- factor(c(" Coefficients.GP2: "))
  print(out1)
  out2 <- matrix(double(1), 1, 1)
  out2[1, 1] <- mle.daten$Dispersion.Parameter
  dimnames(out2)[[2]] <- factor(c(" "))
  dimnames(out2)[[1]] <- factor(c(" Dispersion Parameter: "))
  print(out2)
  out3 <- matrix(double(1), 1, 1)
  out3[1, 1] <- mle.daten$RSS
  dimnames(out3)[[2]] <- factor(c(" "))
  dimnames(out3)[[1]] <- factor(c(" RSS: "))
  print(out3)
  out4 <- matrix(double(1), 1, 1)
  out4[1, 1] <- mle.daten$AIC
  dimnames(out4)[[2]] <- factor(c(" "))
  dimnames(out4)[[1]] <- factor(c(" AIC: "))
  print(out4)
  out5 <- matrix(double(1), 1, 1)
  out5[1, 1] <- mle.daten$Message1
  dimnames(out5)[[2]] <- factor(c(" "))
  dimnames(out5)[[1]] <- factor(c(" Message.GP1: "))
  print(out5)
  out6 <- matrix(double(1), 1, 1)
  out6[1, 1] <- mle.daten$Message2
  dimnames(out6)[[2]] <- factor(c(" "))
  dimnames(out6)[[1]] <- factor(c(" Message.GP2: "))
  print(out6)
  return()
}

```

```
"plot.zhgp" = function(mle.daten)
{
  X = mle.daten$Design
  if(is.matrix(X)) {
    n <- dim(X)[1]
    k <- dim(X)[2]
  }
  else {
    n <- length(X)
    k <- 1
  }
  par(mfrow = c(2, 1))
  plot1 <- matrix(double(2 * n), n, 2)
  dimnames(plot1)[[2]] <- factor(c(" Fitted Values" , " Response" ))
  for(i in 1:n) {
    plot1[i, 1] <- mle.daten$Fitted.Values[i]
    plot1[i, 2] <- mle.daten$Response[i]
  }
  plot(plot1)
  abline(0, 1)
  plot2 <- matrix(double(2 * n), n, 2)
  dimnames(plot2)[[2]] <- factor(c(" Fitted Values" , " Residuals" ))
  for(i in 1:n) {
    plot2[i, 1] <- mle.daten$Fitted.Values[i]
    plot2[i, 2] <- mle.daten$Residuals[i]
  }
  plot(plot2)
  abline(0, 0)
}
```



## Anhang D

# Überblick über die verwendeten Regressionsmodelle

Modell	Responseverteilung	linearer Prädiktor	Erwartungswertfunktion
Poi	$\text{Poi}(\theta_i)$	$\ln \theta_i = \mathbf{x}_i^t \boldsymbol{\beta}$	$\exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}$
GP	$\text{GP}^*(\mu_i, \varphi)$	$\ln \mu_i = \mathbf{x}_i^t \boldsymbol{\beta}$	$\exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}$
ZIP	$\text{ZIP}(\theta_i, \omega)$	$\ln \theta_i = \mathbf{x}_i^t \boldsymbol{\beta}$	$(1 - \omega) \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}$
ZIGP $\mu$	$\text{ZIGP}(\mu_i, \varphi, \omega)$	$\ln \mu_i = \mathbf{x}_i^t \boldsymbol{\beta}$	$(1 - \omega) \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}$
ZIGP $\omega$	$\text{ZIGP}(\mu, \varphi, \omega_i)$	$\text{logit} \omega_i = \mathbf{x}_i^t \boldsymbol{\alpha}$	$\frac{\mu}{1 + \exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\}}$
ZIGP $\mu\omega$	$\text{ZIGP}(\mu_i, \varphi, \omega_i)$	$\text{logit} \omega_i = \mathbf{a}_i^t \boldsymbol{\alpha}$ $\ln \mu_i = \mathbf{b}_i^t \boldsymbol{\beta}$	$\frac{\exp\{\mathbf{b}_i^t \boldsymbol{\beta}\}}{1 + \exp\{\mathbf{a}_i^t \boldsymbol{\alpha}\}}$
ZHP	$\text{ZHP}(\gamma_i, \theta_i)$	$\ln \gamma_i = \mathbf{x}_i^t \boldsymbol{\alpha}$ $\ln \theta_i = \mathbf{x}_i^t \boldsymbol{\beta}$	$\frac{1 - \exp\{\exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\}\}}{1 - \exp\{\exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}\}} \exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}$
ZHGP	$\text{ZHGP}(\gamma_i, \theta_i, \lambda)$	$\ln \gamma_i = \mathbf{x}_i^t \boldsymbol{\alpha}$ $\ln \theta_i = \mathbf{x}_i^t \boldsymbol{\beta}$	$\frac{1 - \exp\{\exp\{\mathbf{x}_i^t \boldsymbol{\alpha}\}\}}{1 - \exp\{\exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}\}} \frac{\exp\{\mathbf{x}_i^t \boldsymbol{\beta}\}}{1 - \lambda}$

Tabelle D.1: Responseverteilung, lineare Prädiktoren und Erwartungswertfunktion der verwendeten Regressionsmodelle





# Anhang E

## Abkürzungen

### E.1 Abkürzungen der verwendeten Verteilungsmodelle

GP	verallgemeinerte Poissonverteilung mit $\theta$ - $\lambda$ -Parametrisierung .	1	(Definition 1.1)
GP*	verallgemeinerte Poissonverteilung mit $\mu$ - $\varphi$ -Parametrisierung .	69	(Definition 5.4)
$\tilde{G}P$	eingeschränkte verallgemeinerte Poissonverteilung .....	67	(Definition 5.1)
$\Gamma$	Gammaverteilung .....	183	(Definition A.6)
N	Normalverteilung		
NB	negative Binomialverteilung mit stetigen Parametern .....	34	(Definition 2.6)
NB*	negative Binomialverteilung mit ganzzahligem Parameter r ...	31	(Definition 2.1)
Poi	Poissonverteilung .....	181	(Definition A.2)
ZHP	Zero-Hurdle Poissonverteilung .....	113	
ZHGP	verallgemeinerte Zero-Hurdle Poissonverteilung .....	116	
ZIP	Zero-Inflated Poissonverteilung .....	94	(Definition 6.6)
ZIGP	verallgemeinerte Zero-Inflated Poissonverteilung .....	98	(Definition 6.8)

### E.2 Abkürzungen der verwendeten Regressionsmodelle

GP	verallgemeinertes Poisson-Regressionsmodell .....	65	(Abschnitt 5.1)
Poi	Poisson-Regressionsmodell .....	55	(Abschnitt 4.1)
ZHP	Zero-Hurdle Poisson-Regressionsmodell .....	114	(Abschnitt 7.3.2)
ZHGP	verallgemeinertes Zero-Hurdle Poisson-Regressionsmodell...	118	(Abschnitt 7.4.3)
ZIP	Zero-Inflated Poisson-Regressionsmodell .....	95	(Abschnitt 6.2.2)
ZIGP $_{\mu}$	verallgemeinertes Zero-Inflated Poisson-Regressionsmodell mit gemeinsamem Auswahlparameter .....	101	(Abschnitt 6.3.3)
ZIGP $_{\mu\omega}$	verallgemeinertes Zero-Inflated Poisson-Regressionsmodell mit individuellem Auswahlparameter .....	103	(Abschnitt 6.4)
ZIGP $_{\omega}$	verallgemeinertes Zero-Inflated Poisson-Regressionsmodell mit individuellem Auswahlparameter und gemeinsamem GP-Parameter .....	105	(Abschnitt 6.5)



# Abbildungsverzeichnis

1.1	GP-Eintrittswahrscheinlichkeiten für $\lambda = -0.2$ . . . . .	3
1.2	GP-Eintrittswahrscheinlichkeiten für $\lambda = 0.5$ . . . . .	3
1.3	GP-Eintrittswahrscheinlichkeiten für $\theta = 6$ . . . . .	4
2.1	Vergleich von verallgemeinerter Poissonverteilung und negativer Binomialverteilung mit Erwartungswert 5 und Varianz 6 bzw. Varianz 15 . . . . .	39
2.2	Vergleich von verallgemeinerter Poissonverteilung und negativer Binomialverteilung mit Erwartungswert 10 und Varianz 15 bzw. Varianz 150 . . . . .	41
2.3	Vergleich von verallgemeinerter Poissonverteilung und negativer Binomialverteilung mit Erwartungswert 50 und Varianz 150 bzw. Varianz 15000 . . . . .	41
5.1	Log-Likelihoodfunktion des verallgemeinerten Poisson-Regressionsmodells mit wahren Parametern $\beta^* = 1$ und $\varphi^* = 1$ . . . . .	73
5.2	Log-Likelihoodfunktion des verallgemeinerten Poisson-Regressionsmodells mit wahren Parametern $\beta^* = 1$ und $\varphi^* = 2$ . . . . .	74
5.3	Log-Likelihoodfunktion des verallgemeinerten Poisson-Regressionsmodells mit wahren Parametern $\beta^* = 1$ und $\varphi^* = 10$ . . . . .	75
5.4	Schnitte durch die Log-Likelihoodfunktion des verallgemeinerten Poisson-Regressionsmodells mit den wahren Parametern $\beta^* = 1$ und $\varphi^* = 1$ . . . . .	77
5.5	Schnitte durch die Log-Likelihoodfunktion des verallgemeinerten Poisson-Regressionsmodells mit den wahren Parametern $\beta^* = 1$ und $\varphi^* = 2$ . . . . .	77
5.6	Schnitte durch die Log-Likelihoodfunktion des verallgemeinerten Poisson-Regressionsmodells mit den wahren Parametern $\beta^* = 1$ und $\varphi^* = 10$ . . . . .	77
6.1	Wahrscheinlichkeits- und Verteilungsfunktionen von ZIGP-Verteilungen mit $\mu = 4$ , $\varphi = 1$ und $\omega = 0, 0.1, 0.2$ und $0.5$ . . . . .	99
6.2	Wahrscheinlichkeits- und Verteilungsfunktionen von ZIGP-Verteilungen mit $\mu = 8$ , $\varphi = 2$ und $\omega = 0, 0.1, 0.2$ und $0.5$ . . . . .	99
7.1	Wahrscheinlichkeits- und Verteilungsfunktionen von ZHGP-Verteilungen mit $\lambda = 0$ , $\theta_1 = 4$ und $\theta_2 = 4, 3, 2$ und $1$ . . . . .	119
7.2	Wahrscheinlichkeits- und Verteilungsfunktionen von ZHGP-Verteilungen mit $\lambda = 0.5$ , $\theta_1 = 4$ und $\theta_2 = 4, 3, 2$ und $1$ . . . . .	119
8.1	Histogramm und QQ-Normal-Plot der Maximum-Likelihood-Schätzer im verallgemeinerten Poisson-Regressionsmodell bei $n=20$ / $\varphi=1.25$ / $M = [0.02, 7.39]$ . .	131
8.2	Histogramm und QQ-Normal-Plot der Maximum-Likelihood-Schätzer im verallgemeinerten Poisson-Regressionsmodell bei $n=100$ / $\varphi=1.25$ / $M = [0.02, 7.39]$ .	132

8.3	Histogramm und QQ-Normal-Plot der Maximum-Likelihood-Schätzer im verallgemeinerten Poisson-Regressionsmodell bei $n=500 / \varphi=1.25 / M = [0.02, 7.39]$ .	133
8.4	Histogramm und QQ-Normal-Plot der Maximum-Likelihood-Schätzer im verallgemeinerten Poisson-Regressionsmodell bei $n=20 / \varphi=5.00 / M = [0.02, 7.39]$ . .	134
8.5	Histogramm und QQ-Normal-Plot der Maximum-Likelihood-Schätzer im verallgemeinerten Poisson-Regressionsmodell bei $n=100 / \varphi=5.00 / M = [0.02, 7.39]$ .	135
8.6	Histogramm und QQ-Normal-Plot der Maximum-Likelihood-Schätzer im verallgemeinerten Poisson-Regressionsmodell bei $n=100 / \varphi=1.25 / M = [0.05, 2.72]$ .	136
9.1	Explorative Datenanalyse der Patentdaten . . . . .	156
9.2	Angepasste Werte und Residuen des Poissonmodells der Patentdaten . . . . .	158
9.3	Überdispersion im Poissonmodell der Patentdaten . . . . .	158
9.4	Angepasste Werte und Residuen des verallgemeinerten Poissonmodells der Patentdaten . . . . .	161
9.5	Angepasste Werte und Residuen des ZIGP $\mu$ -Modells der Patentdaten . . . . .	161
9.6	Angepasste Werte und Residuen des ZIP-Modells der Patentdaten . . . . .	162
9.7	Angepasste Werte und Residuen des ZIGP $\mu\omega$ -Modells der Patentdaten . . . . .	163
9.8	Angepasste Werte und Residuen des ZIGP $\omega$ -Modells der Patentdaten . . . . .	164
9.9	Angepasste Werte und Residuen des ZHGP-Modells der Patentdaten . . . . .	165
9.10	Angepasste Werte und Residuen des ZHP-Modells der Patentdaten . . . . .	166
9.11	Angepasste Werte und Residuen des Poissonmodells der KFZ-Daten . . . . .	172
9.12	Angepasste Werte und Residuen des verallgemeinerten Poissonmodells der KFZ-Daten . . . . .	172
9.13	Angepasste Werte und Residuen des ZIGP $\mu$ -Modells der KFZ-Daten . . . . .	172
9.14	Angepasste Werte und Residuen des ZIP-Modells der KFZ-Daten . . . . .	172
9.15	Angepasste Werte und Residuen des ZIGP $\mu\omega$ -Modells der KFZ-Daten . . . . .	174
9.16	Angepasste Werte und Residuen des ZIGP $\omega$ -Modells der KFZ-Daten . . . . .	175
9.17	Angepasste Werte und Residuen des ZHGP-Modells der KFZ-Daten . . . . .	176
9.18	Angepasste Werte und Residuen des ZHP-Modells der KFZ-Daten . . . . .	177

# Tabellenverzeichnis

2.1	Abweichung von GP- und NB-Wahrscheinlichkeitsfunktion . . . . .	38
8.1	Bias, MSE und relativer Fehler der Maximum-Likelihood-Schätzer im verallgemeinerten Poisson-Regressionsmodell bei Stichprobenumfang $n=20$ . . . . .	128
8.2	Bias, MSE und relativer Fehler der Maximum-Likelihood-Schätzer im verallgemeinerten Poisson-Regressionsmodell bei Stichprobenumfang $n=100$ . . . . .	128
8.3	Bias, MSE und relativer Fehler der Maximum-Likelihood-Schätzer im ZIGP $\mu$ -Regressionsmodell für den $\mu$ -Bereich $[0.02, 7.39]$ bei Stichprobenumfang $n=20$ . . . . .	138
8.4	Bias, MSE und relativer Fehler der Maximum-Likelihood-Schätzer im ZIGP $\mu$ -Regressionsmodell für den $\mu$ -Bereich $[0.02, 7.39]$ bei Stichprobenumfang $n=100$ . . . . .	138
8.5	Übersicht über die ZIGP $\mu\omega$ -Designs, die darin verwendeten Regressionsparameter und die resultierenden Parameterbereiche $\Omega$ und $M$ . . . . .	142
8.6	Bias, MSE und relativer Fehler der Maximum-Likelihood-Schätzer im ZIGP $\mu\omega$ -Regressionsmodell bei Stichprobenumfang $n=100$ . . . . .	144
8.7	Bias, MSE und relativer Fehler der Maximum-Likelihood-Schätzer im ZIGP $\omega$ -Regressionsmodell bei Stichprobenumfang $n=100$ . . . . .	146
8.8	Bias, MSE und relativer Fehler der Maximum-Likelihood-Schätzer im ZHGP-Regressionsmodell bei Stichprobenumfang $n=20$ . . . . .	148
8.9	Bias, MSE und relativer Fehler der Maximum-Likelihood-Schätzer im ZHGP-Regressionsmodell bei Stichprobenumfang $n=100$ . . . . .	149
8.10	Quotienten aus der Fehlerquadratsumme missspezifizierter Modelle und der Fehlerquadratsumme der Originalmodelle . . . . .	152
9.1	Spalten des Patent-Datensatzes . . . . .	155
9.2	<code>summary</code> -Tabelle des Poissonmodells der Patentdaten . . . . .	157
9.3	<code>summary.gpois</code> -Tabelle des verallgemeinerten Poissonmodells der Patentdaten . . . . .	159
9.4	<code>summary.zigp1</code> -Tabelle des ZIGP $\mu$ -Modells der Patentdaten . . . . .	160
9.5	<code>summary.zigp1</code> -Tabelle des ZIP-Modells der Patentdaten . . . . .	162
9.6	<code>summary.zigp2</code> -Tabelle des ZIGP $\mu\omega$ -Modells der Patentdaten . . . . .	163
9.7	<code>summary.zigp3</code> -Tabelle des ZIGP $\omega$ -Modells der Patentdaten . . . . .	164
9.8	<code>summary.zhgp</code> -Tabelle des ZHGP-Modells der Patentdaten . . . . .	165
9.9	<code>summary.zhgp</code> -Tabelle des ZHP-Modells der Patentdaten . . . . .	166
9.10	Minimum, Maximum und arithmetisches Mittel der Residuen sowie Fehlerquadratsummen RSS und RSS* der Regressionsmodelle für die Patentdaten im Vergleich . . . . .	168
9.11	Log-Likelihoodfunktion und Akaike-Information der Maximum-Likelihood-Schätzer des ZIGP $\mu$ - und darin eingebetteter Regressionsmodelle für die Patentdaten . . . . .	168

9.12	Log-Likelihoodfunktion und Akaike-Information der Maximum-Likelihood-Schätzer des ZHGP- und darin eingebetteter Regressionsmodelle für die Patentdaten . . .	168
9.13	Spalten der Designmatrix des KFZ-Datensatzes . . . . .	169
9.14	Responsestruktur der KFZ-Daten . . . . .	170
9.15	<code>summary</code> -Tabelle des Poissonmodells (9.11) . . . . .	170
9.16	<code>summary.gpois</code> -Tabelle des verallgemeinerten Poissonmodells der KFZ-Daten . .	171
9.17	<code>summary.zigp1</code> -Tabelle des ZIGP $\mu$ -Regressionsmodells der KFZ-Daten . . . . .	173
9.18	<code>summary.zigp1</code> -Tabelle des ZIP-Regressionsmodells der KFZ-Daten . . . . .	173
9.19	<code>summary.zigp2</code> -Tabelle des ZIGP $\mu\omega$ -Regressionsmodells der KFZ-Daten . . . . .	174
9.20	<code>summary.zigp3</code> -Tabelle des ZIGP $\omega$ -Regressionsmodells der KFZ-Daten . . . . .	175
9.21	<code>summary.zhgp</code> -Tabelle des ZHGP-Regressionsmodells der KFZ-Daten . . . . .	176
9.22	<code>summary.zhgp</code> -Tabelle des ZHP-Regressionsmodells der KFZ-Daten . . . . .	177
9.23	Minimum, Maximum und arithmetisches Mittel der Residuen sowie Fehlerquadratsummen RSS und RSS* der Regressionsmodelle für die KFZ-Daten im Vergleich	178
9.24	Log-Likelihoodfunktion und Akaike-Information der Maximum-Likelihood-Schätzer des ZIGP $\mu$ - und darin eingebetteter Regressionsmodelle für die KFZ-Daten . . .	178
9.25	Log-Likelihoodfunktion und Akaike-Information der Maximum-Likelihood-Schätzer des ZHGP- und darin eingebetteter Regressionsmodelle für die KFZ-Daten . . .	178
D.1	Responseverteilung, lineare Prädiktoren und Erwartungswertfunktion der verwendeten Regressionsmodelle im Überblick . . . . .	235

# Literaturverzeichnis

- Jean-François Angers and Atanu Biswas. A bayesian analysis of zero-inflated generalized poisson model. *Computational Statistics and Data Analysis*, 42:37–46, 2003.
- O. Beyer, H.-J. Girlich, and H.-U. Zschiesche. *Stochastische Prozesse und Modelle*. Verlag Harri Deutsch, Frankfurt, 1978.
- Peter C. Bickel and Kjell A. Doksum. *Mathematical Statistics - Basic Ideas and Selected Topics*. Prentice Hall, Englewood Cliffs, New Jersey, 1st edition, 1977.
- Ronald J. Bosch and Louise M. Ryan. Generalized poisson models arising from markov processes. *Statistics and Probability Letters*, 39:205–212, 1998.
- N.E. Breslow. Extra-poisson variation in log-linear models. *Applied Statistics*, 33(1):38–44, 1984.
- Jan van den Broeck. A score test for zero inflation in a poisson distribution. *Biometrics*, 51(2): 738–743, 1995.
- A.C. Cameron and P.K. Trivedi. *Regression Analysis of Count Data*. Cambridge University Press, 1998.
- G. Casella and R.L. Berger. *Statistical Inference*. Wadsworth and Brooks, 1990.
- P.C. Consul. On some models leading to the generalized poisson distribution. *Commun. Statist. - Theory Meth. (A)*, 17(2), 1988.
- P.C. Consul. *Generalized Poisson Distributions - Properties and Applications*. Marcel Dekker, New York and Basel, 1989.
- P.C. Consul and Felix Famoye. On the unimodality of generalized poisson distributions. *Statistical Neerlandica*, 40:117–122, 1986.
- P.C. Consul and Felix Famoye. Generalized poisson regression model. *Commun. Statistics - Theory Meth.*, 21(1):89–109, 1992.
- P.C. Consul and G.C. Jain. A generalization of the poisson distribution. *Technometrics*, 15(4): 791–799, 1973.
- P.C. Consul and L.R. Shenton. Some interesting properties of lagrangian distributions. *American Journal of Mathematical and Management Sciences*, 8(1), 1973.
- P.C. Consul and M.M. Shoukri. The generalized poisson distribution when the sample mean is larger than the sample variance. *Commun. Statist. - Simulation Comput.*, 14(3):677–681, 1985.

- P.C. Consul and M.M. Shoukri. Some change mechanisms generating the generalized poisson probability models. *American Journal of Mathematical and Management Sciences*, 8(1), 1988.
- Ludwig Fahrmeir and Alfred Hamerle. *Multivariate statistische Verfahren*. Walter de Gruyter, Berlin, New York, 1984.
- Ludwig Fahrmeir and Heinz Kaufmann. Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *The Annals of Statistics*, 13(1):342–368, 1985.
- Ludwig Fahrmeir, Rita Künstler, Iris Pigeot, and Gerhard Tutz. *Statistik - Der Weg zur Datenanalyse*. Springer-Verlag, Berlin, 3rd edition, 2001.
- Ludwig Fahrmeir and Gerhard Tutz. *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer-Verlag, Berlin, 2nd edition, 2001.
- Felix Famoye. Generalized poisson random variate generation. *American Journal of Mathematical and Management Sciences*, 17(3-4):219–237, 1997.
- Felix Famoye. Edf tests for the generalized poisson distribution. *J. Statist. Comput. Simul.*, 63:159–168, 1999.
- W.H. Greene. *Econometric analysis*. Prentice Hall, Englewood Cliffs, New Jersey, 4th edition, 2000.
- Pushpa L. Gupta, Ramesh C. Gupta, and Ram C. Tripathi. Analysis of zero-adjusted count data. *Computational Statistics and Data Analysis*, 23:207–218, 1996.
- Pushpa L. Gupta, Ramesh C. Gupta, and Ram C. Tripathi. Score test for zero inflated generalized poisson regression model. *Commun. Statist. - Theory Meth.*, 33(1):47–64, 2004.
- Norbert Henze and Bernhard Klar. Bootstrap based goodness of fit tests for the generalized poisson model. *Commun. Statist. - Theory Meth.*, 24(7):1875–1896, 1995.
- John Hinde and Clarice G.B. Demétrio. Overdispersion: Models and estimation. *Computational Statistics and Data Analysis*, 27:151–170, 1998.
- Josef Hofbauer. *Lagrange-Inversion*. Institut für Mathematik der Universität Wien.
- J.L.W. Jensen. Sur identité d'abel et pur d'autres formules analogues. *Acta Mathematica*, 26:307–318, 1902.
- N.L. Johnson, S. Kotz, and A.W. Kemp. *Univariate discrete distributions*. John Wiley & Sons, New York, 2nd edition, 1993.
- Konrad Königsberger. *Analysis 1*. Springer-Verlag, Berlin, 4th edition, 1999.
- Christian Kredler. *Einführung in die Wahrscheinlichkeitsrechnung und Statistik*. Technische Universität München, Zentrum Mathematik, 1998.
- Diane Lambert. Zero-inflated poisson regression with an application to defects in manufacturing. *Technometrics*, 34(1):1–14, 1992.
- J.F. Lawless. Negative binomial and mixed poisson regression. *The Canadian Journal of Statistics*, 15(3), 1987.



- Nikolai Leonenko. *Stochastic Models for Insurance*. Cardiff University, School of Mathematics, 2003.
- Bores Lerner, Amjad Lone, and Murali Rao. On generalized poisson distributions. *Probability and Mathematical Statistics*, 17(2):377–385, 1997.
- P. McCullagh and J.A. Nelder. *Generalized Linear Models*. Chapman and Hall, London, New York, 2nd edition, 1989.
- Charles E. McCulloch and Shayle R. Searle. *Generalized, Linear and Mixed Models*. John Wiley & Sons, New York, 2001.
- John Mullahy. Specification and testing in some modified count data models. *Journal of Econometrics*, 33:341–365, 1986.
- John Mullahy. Heterogeneity, excess zeros and the structure of count data models. *Journal of Applied Econometrics*, 12:337–350, 1997.
- R.H. Myers. *Classical and modern regression with applications*. Duxbury Press, Belmont, 2nd edition, 1990.
- Helmut Pruscha. *Angewandte Methoden der mathematischen Statistik*. Teubner, Stuttgart, 2nd edition, 1996.
- Lennart Råde and Bertil Westergren. *Springers Mathematische Formeln - Übersetzt und bearbeitet von P. Vachenaer*. Springer-Verlag, Berlin, 1997.
- Sidney I. Resnick. *Adventures in Stochastic Processes*. Birkhäuser, Boston, 1992.
- Inka-Diana Sikora. Quantifizierung von überdispersion. Master's thesis, Technische Universität München, Zentrum Mathematik, April 2002.
- David Stirzaker. *Elementary Probability*. Cambridge University Press, 1994.
- H.J.H. Tuentler. On the generalized poisson distribution. *Statistica Neerlandica*, 54(3):374–376, 2000.
- W.N. Venables and R.D. Ripley. *Modern Applied Statistics with S*. Springer-Verlag, Berlin, 4th edition, 2003.
- Rainer Winkelmann. Duration dependence and dispersion in count data models. *Journal of Business and Economic Statistics*, 13:467–474, 1995.
- Rainer Winkelmann. *Econometric Analysis of Count Data*. Springer-Verlag, Berlin, 4th edition, 2003.
- Anatoly Zhigljavsky. *Regression Analysis*. Cardiff University, School of Mathematics, 2003.



# Index

- $\chi^2$ -Statistik, 84
- abhängige Variable, 43
- Abhängigkeiten bei der Datenerzeugung, 63
- Aequidispersion, 13, 182
- AIC, 53
- Akaike-Information, 53
  - verallgemeinerte Poissonregression, 88
  - ZHGP-Regression, 121
  - ZIGP $\mu$ -Regression, 103
- Anpassungsgüte, 44
- Approximationsmethode, 29
- asymptotische Verteilung
  - Maximum-Likelihood-Schätzer, 130
- Ausgleichsfaktor, 57
- Auswahlparameter, 90
- Auswahlprozess, 89
- Auswahlvariable, 89, 109
  
- beobachtete Informationsmatrix, 48
- Beobachtungszeitraum, 56
- Bias, 124
- Borel-Tanner-Verteilung, 21
- Borelverteilung, 21
  
- Designmatrix, 43
- devianz.gpois, 199
- Devianzfunktion, 52
- dgpois, 189
- Dichte
  - Gammaverteilung, 183
- Diskrepanzmass, 44
- Dispersion, 111
- Dispersionsindex, 34, 64
- Dispersionsparameter, 13, 33, 64
- Doppel-Hürdenmodell, 110
  
- einfacher Verzweigungsprozess, 18
- eingebettete Modelle, 54, 103, 121
- eingeschränkte verallg. Poissonverteilung, 67
- Eintrittswahrscheinlichkeit
  - verallg. Poissonverteilung, 1, 69
- empirische Informationsmatrix, 48
- empirisches Konfidenzintervall, 125
- empirisches Quantil, 125
- endogene Selektivität, 63
- erklärende Variable, 43
- Erwartungstreue, 124
- Erwartungswert
  - Gammaverteilung, 183
  - komb. verallg. Poissonverteilung, 117
  - kombinierte Poissonverteilung, 114
  - negative Binomialverteilung, 32
  - Null-Hürdenverteilung, 110
  - Poissonverteilung, 182
  - verallg. Poissonverteilung, 5
  - ZHGP-Verteilung, 117
  - ZHP-Verteilung, 114
  - ZI-Verteilung, 93
  - ZIGP-Verteilung, 98
  - ZIP-Verteilung, 95
- Erwartungswert-Varianz-Kriterium, 124
- Erwartungswertfunktion, 46
- erzeugende Funktion, 6
  - verallg. Poissonverteilung, 7
- Erzeugung von Zufallsvariablen, 26
- Euler'sche Gammafunktion, 183
- excess of zeros, 61
- exponentielle Familie, 45, 181
  
- Faltung
  - negative Binomialverteilung, 33
  - verallg. Poissonverteilung, 13
- Fehlerquadratsumme, 44, 52, 152
- Fisher-Information, 49
- Fisher-Scoring, 48
- fisher.gpois, 200, 201
- fit.gpois, 198
- fit.zhgp, 231
- fit.zigp1, 208
- fit.zigp2, 215

- fit.zigp3, 223
- Galton-Watson-Verzweigungsprozess, 18
- Gamma-Poisson-Mischmodell, 36
- Gammafunktion, 183
- Gammaverteilung, 183
  - Erwartungswert, 183
  - Varianz, 183
- Gaussklammer, 29
- gedächtnisloser Prozess, 63
- gemischter Poissonprozess, 186
- gemischtes Regressionsmodell, 55
- gesättigtes Modell, 44
- goodness of fit, 44
- Grenzverteilung
  - Inverse Gaussverteilung, 25
  - Poissonverteilung, 35
- Hürdenmodelle, 109
- Hürdenverteilung, 61, 109
  - Wahrscheinlichkeitsfunktion, 109
- Hat-Matrix, 51
- heteroskedastisch, 56
- Histogramm, 130
- homogener Poissonprozess, 186
- hurdle model, 109
- idempotente Matrix, 51
- Identifizierbarkeitsproblem, 143
- Indexmenge, 185
- Informationsmatrix, 49
  - beobachtete, 48
  - empirische, 48
- Informationsnummer, 49
- Intensität, 181
- Intercept, 44
- Inverse Gaussverteilung, 25
- Invertierungsmethode, 27
- Invertierungstheorem, 6
- Jensen'sche Formel, 2
- Justierungsfaktor, 90
- kanonische Linkfunktion, 50
- kanonischer Link
  - Poissonmodell, 56
- kanonischer Parameter, 181
- Klassifizierungsvariable, 89
- Kleinste-Quadrate-Schätzer, 44, 46
- kombinierte Verteilung
  - Wahrscheinlichkeitsfunktion, 109
- kombinierte Verteilungsmodelle, 109
- kombiniertes
  - Poissonmodell
    - Erwartungswert, 114
    - Likelihoodfunktion, 115
    - Regression, 114
    - Varianz, 114
    - Verteilung, 113
    - Wahrscheinlichkeitsfunktion, 113
- verallg. Poissonmodell, 116
  - Akaike-Information, 121
  - Erwartungswert, 117
  - Likelihoodfunktion, 120
  - Regression, 118
  - Varianz, 117
  - Verteilung, 116
  - Wahrscheinlichkeitsfunktion, 116
- Konfidenzintervall, 125
- konsistenter Schätzer, 125
- Kovariable, 43
- Kumulanten, 22
  - Normalverteilung, 25
  - verallg. Poissonverteilung, 23
- Kumulantenfunktion, 22
- Lévy-Prozess, 63
- Lagrange'sches Invertierungstheorem, 6
- Lagrangeentwicklung, 21
- Landausymbol, 186
- Likelihoodfunktion, 47
  - komb. verallg. Poissonmodell, 120
  - kombiniertes Poissonmodell, 115
  - Null-Hürdenmodell, 112
  - verallg. Poissonregression, 71
  - ZHGP-Modell, 120
  - ZIGP $\mu\omega$ -Regression, 104
  - ZIGP $\mu$ -Regression, 101
  - ZIGP $\omega$ -Regression, 106
  - ZIP-Regression, 96
- linearer Prädiktor, 46
- Linkfunktion, 46
- Log-Likelihoodfunktion, 47
- log.likelihood.gpois, 197
- log.likelihood.zhgp1, 229
- log.likelihood.zhgp2, 230
- log.likelihood.zigp1, 207

- log.likelihood.zigp2, 214
- log.likelihood.zigp3, 222
- Maximum-Likelihood-Schätzer
  - asymptotische Verteilung, 130
  - Schätzgüte
    - verallg. Poissonregression, 129
    - ZHGP-Regression, 148
    - ZIGP $\mu\omega$ -Regression, 140
    - ZIGP $\mu$ -Regression, 139
    - ZIGP $\omega$ -Regression, 145
- Maximum-Likelihood-Schätzung, 46
  - verallg. Poissonregression, 71
- mean squared error, 124
- Messfehler, 62
- Mischmodell, 36
- Mischungsmodell, 61
- Missspezifikation, 59
  - Abhängigkeiten im Erzeugungsprozess, 63
    - Messfehler, 62
    - Nullenüberschuss, 61
    - Selektivität, 63
    - Simulationsstudie, 152
    - unbeobachtete Heterogenität, 60
    - underreporting, 63
    - Verknüpfungskomponente, 59
- mittlere quadratische Abweichung, 124
- ML-Schätzung, 46
- mle.gpois, 194
- mle.zhgp, 227
- mle.zigp1, 205
- mle.zigp2, 212
- mle.zigp3, 220
- Modellbildung, 43
- Modelle mit Nullen, 89
- Modellmatrix, 43
- Momente
  - nichtzentriert, 8
  - Rekursionsgleichung, 8, 12
  - verallg. Poissonverteilung, 8
  - zentriert, 12
- Momentenschätzer, 84
- momenterzeugende Funktion, 7
  - negative Binomialverteilung, 31
  - Poissonverteilung, 182
  - verallg. Poissonverteilung, 8
- natürliche Linkfunktion, 50
- natürlicher Parameter, 181
- natürlicher Parameterraum, 181
- negativ definit, 48
- negative Binomialverteilung
  - Definition, 31
  - Erwartungswert, 32
  - Faltung, 33
  - Grenzverteilung, 35
  - momenterzeugende Funktion, 31
  - stetige Parameter, 34
  - Varianz, 32
- nested models, 54
- Newton-Raphson-Verfahren, 48
- nicht-homogener Poissonprozess, 186
- Normalgleichungen, 51
- Normalisierungsfaktor, 2
- Normalverteilung, 24
- Normalverteilungsapproximation, 29
- Null-Hürdenmodell, 109
  - Erwartungswert, 110
  - Likelihoodfunktion, 112
  - Varianz, 110
  - Wahrscheinlichkeitsfunktion, 110
- Nullenüberschuss, 61, 109
- Nullmodell, 44, 50
- o-Notation, 186
- Offset, 57
  - Poissonregression, 56
  - verallg. Poissonregression, 82
- orthogonale Designmatrix, 141
- parametrisches Modell, 47
- parent distribution, 109
- Pfad, 185
- pgpois, 190
- phi.moment, 196
- plot.gpois, 203
- plot.zhgp, 233
- plot.zigp1, 210
- plot.zigp2, 218
- plot.zigp3, 225
- Poissonprozess, 15, 186
- Poissonregression, 55
  - gemischtes Modell, 55
  - kanonischer Link, 56
  - Offset, 56
  - Standardmodell, 55
- Poissonverteilung

- Erwartungswert, 182
- kombinierte, 113
- momenterzeugende Funktion, 182
- Varianz, 182
- Verteilungsklasse, 182
- Wahrscheinlichkeitsfunktion, 181
- positiv definit, 48
- probability integral transformation, 26
- Projektionsmatrix, 51
- Pseudozufallsvariable, 26
- Punktprozess, 55
- qgpois, 191
- QQ-Normal-Plot, 130
- Quantil, 125
- Rate, 181
- Realisierung eines stochastischen Prozesses, 185
- Regressionsmodell
  - GP, 65
  - komb. verallg. Poisson, 118
  - kombiniertes, 109
  - Poisson, 55
  - verallg. Poisson, 65
  - Zero-Hurdle, 109
  - ZHGP, 118
  - ZHP, 114
  - ZIGP $\mu\omega$ , 103
  - ZIGP $\mu$ , 98
  - ZIGP $\omega$ , 105
  - ZIP, 95
- Regressor, 43
- reguläres Modell, 47
- relativer Fehler, 91
- residual sum of squares, 52
- Residuen, 50
- Residuenquadratsumme, 52, 152
- Response, 43
- response.gpois, 193
- response.zhgp, 226
- response.zigp1, 204
- response.zigp2, 211
- response.zigp3, 219
- Responsefunktion, 46
- rgpois, 192
- saturated model, 44
- Satz von Steiner, 12
- Schatzgute, 124
- Scorefunktion, 48
- Scoregleichung, 48
- selection process, 89
- Selektivität, 63
- Semiinvariante, 22
- Skalenparameter, 181
- Smirnovtransformation, 26
- Störparameter, 64, 181
- stationärer Prozess, 186
- Stichprobenselektion, 63
- stochastische Dominanz, 90
- stochastische Komponente, 46
- stochastischer Prozess, 185
- stochastischer Zählprozess, 15
- Suchalgorithmus, 27
  - sequentielle Suche, 27
- summary.gpois, 202
- summary.zhgp, 232
- summary.zigp1, 209
- summary.zigp2, 216
- summary.zigp3, 224
- symmetrische Matrix, 51
- systematische Komponente, 46
- Trajektorie, 185
- two part model, 109
- Ueberdispersion, 13, 62, 63, 111
  - Definition, 59
- unabhängige Zuwächse, 186
- unbeobachtete Heterogenität, 60
- Underreporting, 63
- unimodal, 4
- unkalierte Devianz, 52
- Unterdispersion, 13, 63
- unverzerrt, 124
- Ursprungsprozess, 89
- Ursprungsverteilung, 109
- Variablenselektion, 53
- Varianz
  - Gammaverteilung, 183
  - komb. verallg. Poissonverteilung, 117
  - kombinierte Poissonverteilung, 114
  - negative Binomialverteilung, 32
  - Null-Hürdenverteilung, 110
  - Poissonverteilung, 182
  - verallg. Poissonverteilung, 12

- ZHGP-Verteilung, 117
- ZHP-Verteilung, 114
- ZI-Verteilung, 93
- ZIGP-Verteilung, 98
- ZIP-Verteilung, 95
- Varianz-Erwartungswert-Quotient, 111
- Varianzfunktion, 50, 56
- verallgemeinerte Poissonregression
  - Devianzresiduen, 87
  - kombinierte, 116
  - Likelihoodfunktion, 71
  - Log-Likelihoodfunktion, 71
  - Maximum-Likelihood-Schätzer
    - asymptotische Verteilung, 130
    - Schätzgüte, 129
  - Maximum-Likelihood-Schätzung, 71, 84
  - Offset, 82
  - Schätzprozeduren, 84
  - Zero-Inflation, 98
- verallgemeinerte Poissonverteilung, 1
  - Dispersionsparameter, 13
  - eingeschränkte, 67
  - Eintrittswahrscheinlichkeit, 1, 69
    - S-Plus Funktion, 189
  - Erwartungswert, 5
  - erzeugende Funktion, 7
  - Erzeugung von Zufallsvariablen, 26
  - Faltung, 13
  - Grenzverteilung, 22
    - Inverse Gaussverteilung, 25
    - Normalverteilung, 24
  - Kumulanten, 23
  - Momente, 8
  - momentenerzeugende Funktion, 8
  - Quantilsfunktion
    - S-Plus Funktion, 191
  - Varianz, 12
  - Verteilungsfunktion
    - S-Plus Funktion, 190
  - Wahrscheinlichkeitsfunktion, 1, 69
    - S-Plus Funktion, 189
- verallgemeinerter Poissonprozess, 15
- Vergleich: GP- und NB-Verteilung, 36
- Verknüpfung
  - natürliche, 50
- Verknüpfungskomponente, 46, 59
- Verteilung der seltenen Ereignisse, 181
- Verteilung mit Nullen, 89
- Verzerrung, 124
- Verzweigungsmethode, 28
- Verzweigungsprozess, 18, 28, 63
- volles Modell, 44, 50
- Wahrscheinlichkeitsfunktion
  - Hürdenverteilung, 109
  - komb. verallg. Poissonverteilung, 116
  - kombinierte Poissonverteilung, 113
  - kombinierte Verteilung, 109
  - negative Binomialverteilung, 31
  - negative Binomialverteilung (stetige Parameter), 34
  - Null-Hürdenverteilung, 110
  - Poissonverteilung, 181
  - verallg. Poissonverteilung, 1, 69
  - ZH-Verteilung, 110
  - ZHGP-Verteilung, 116
  - ZHP-Verteilung, 113
  - ZI-Verteilung, 90
  - ZIGP-Verteilung, 98
  - ZIP-Verteilung, 94
- Zähldaten, 55
- Zählprozess, 15, 55, 185
- Zählvorgang, 181
- ZA-Verteilung, 90
- Zeitintervall, 56
- zensierte Grösse, 63
- Zero-Adjusted Verteilung, 90
- Zero-Hurdle Verteilung, 109
- Zero-Inflated Verteilung, 89
- ZH-Modell, 109
- ZHGP-Modell, 116
  - Akaike-Information, 121
  - Erwartungswert, 117
  - Likelihoodfunktion, 120
  - Regression, 118
  - Schätzgüte, 148
  - Varianz, 117
  - Wahrscheinlichkeitsfunktion, 116
- ZHP-Modell
  - Erwartungswert, 114
  - Likelihoodfunktion, 115
  - Varianz, 114
  - Verteilung, 113
  - Wahrscheinlichkeitsfunktion, 113
- ZI-Modell, 89
- ZI-Verteilung

- Erwartungswert, 93
- Varianz, 93
- Wahrscheinlichkeitsfunktion, 90
- ZI-Zufallsvariable, 89
- Zielvariable, 43
- ZIGP-Modell
  - $\mu\omega$ -Regressionsmodell
    - Likelihoodfunktion, 104
    - Schätzgüte, 140
    - Spezifikation, 103
  - $\mu$ -Regressionsmodell
    - Akaike-Information, 103
    - Likelihoodfunktion, 101
    - Schätzgüte, 139
    - Spezifikation, 101
  - $\omega$ -Regressionsmodell
    - Likelihoodfunktion, 106
    - Schätzgüte, 145
    - Spezifikation, 105
  - Erwartungswert, 98
  - Varianz, 98
  - Wahrscheinlichkeitsfunktion, 98
- ZIP-Modell
  - Erwartungswert, 95
  - Likelihoodfunktion, 96
  - Varianz, 95
  - Wahrscheinlichkeitsfunktion, 94
- Zufallsvariable
  - Generierung, 26
    - Approximationsmethode, 29
    - Invertierungsmethode, 27
    - Verzweigungsprozess, 28
  - Transformation, 26
- Zufallsvariable mit Nullen, 89
- zweiseitiges Konfidenzintervall, 125