



Julia Tranninger, BSc

Capture-Recapture Schätzungen der Größe einer geschlossenen oder offenen Population

MASTERARBEIT

zur Erlangung des akademischen Grades einer Diplom-Ingenieurin

Masterstudium Technische Mathematik: Operations Research und Statistik

eingereicht an der

Technischen Universität Graz

Betreuer:

Ao. Univ.-Prof. Dipl.-Ing. Dr.techn. Herwig Friedl

Institut für Statistik

Graz, Mai 2018

Eidesstattliche Erklärung

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

Affidavit

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

Datum

Unterschrift

Zusammenfassung

Ist man an der Schätzung einer unbekanntem Populationsgröße interessiert, so kann diese ausgehend von Aufzeichnungen zu Fängen und Wiederfängen in einer Population mit Hilfe der Capture-Recapture Methode erreicht werden. In dieser Diplomarbeit werden nach einem allgemeinen Teil über die Capture-Recapture Methode der Chao- und der biaskorrigierte Chao- sowie der Zelterman-Schätzer inklusive deren Varianzen für eine geschlossene Population aus der Power Series Verteilung hergeleitet. Des Weiteren werden statistische Modelle zur Bestimmung der Populationsgröße betrachtet, die generalisierte lineare Modelle und die Bayessche Statistik verwenden. Als Grundlage gelten geschlossene Populationsmodelle. Werden diese um geographische Informationen erweitert, so entstehen vollständige spatiale Capture-Recapture Modelle. Den Abschluss bilden offene Populationsmodelle, die zusätzlich die Entwicklung einer Population im Laufe einzelner Jahre aufzeigen kann. Ein bekanntes offenes Modell ist das traditionelle Jolly-Seber Modell. Alle betrachteten Modelle werden anhand der Würfelnatter - einer besonders gefährdeten Schlangenart - veranschaulicht.

Abstract

If one is interested in the estimation of an unknown population size, it is possible to count it with the help of capture-recapture methods. At the beginning, this work provides an introduction to capture-recapture. With this knowledge the Chao, the biascorrected Chao and the Zelterman estimator for a closed population including variances can be derived from power series distributions. Furthermore the focus lies on statistical models to determine the population size. These models use generalized linear models and Bayesian analysis. The basis are closed population models, which can be expanded by geographical information to achieve fully spatial capture-recapture models. At the end, open population models considering additionally population dynamics are discussed. A popular open model is the traditional Jolly-Seber model. All of the models are illustrated by the endangered dice snake.

Danksagung

Ich möchte mich an dieser Stelle bei allen bedanken, die mir während meines Studiums zur Seite gestanden sind und mich immer wieder motiviert haben. Dank ihnen habe ich eine unvergessliche Zeit erlebt.

Ein besonderer Dank gilt natürlich meinem Betreuer, Herrn Professor Herwig Friedl. Er hat mich bei dieser Masterarbeit sowohl mit seinem fachlichen Wissen als auch mit seiner hilfsbereiten Art immer unterstützt. So leitete er beispielsweise eine Unklarheit direkt an Professor Dankmar Böhning weiter, um gemeinsam den Fehler zu beheben. Vielen Dank dafür!

Weiters bedanke ich mich bei Herrn Oliver Gebhardt für die Aufzeichnungen zu den Würfelnatter-Funden. Durch einen Workshop im Mai 2017 bekam ich zudem einen Einblick ins Leben eines Biologen und durfte mir ansehen, wie die Begehungen zum Auffinden von Würfelnattern entlang der Mur durchgeführt wurden.

Zum Abschluss möchte ich noch meinen Eltern danken, die mir mein Studium ermöglicht und mich mit aufbauenden Worten immer wieder bestärkt haben. Ein großer Dank gilt außerdem meinem Freund, meinen Freunden und meinen Studienkollegen, ohne die vieles nicht möglich gewesen wäre.

Inhaltsverzeichnis

1	Einleitung	1
2	Einführung	3
2.1	Die Würfelnatter	3
2.2	Fang-Wiederfang	5
2.3	Künstliche Versteckplätze	6
3	Die Capture-Recapture Methode	7
3.1	Allgemeine Problemstellung	8
3.2	Problemstellung Würfelnatter-Daten	11
4	Power Series Verteilung	13
4.1	Power Series	14
4.2	Ratios	16
4.3	Ratio-Plot	17
4.3.1	Ratios und Ratio-Plots der einzelnen Power Series Verteilungen	18
5	Herleitung von Schätzern für die Populationsgröße	23
5.1	Chao-Schätzer	23
5.1.1	Biaskorrigierter Chao-Schätzer	24
5.2	Zelterman-Schätzer	25
6	Herleitung von Varianzen für die Populationsgrößenschätzer	29
6.1	Chao-Schätzer	29
6.1.1	Biaskorrigierter Chao-Schätzer	34
6.2	Zelterman-Schätzer	36
7	Einführung in GLMs und in die Bayessche Statistik	43

7.1	GLMs und GLMMs	44
7.2	Die Bayessche Statistik	45
7.2.1	A-priori-Verteilungen	46
7.2.2	A-posteriori-Folgerung	47
7.3	A-posteriori-Verteilungen mittels MCMC-Simulation	48
7.4	Bayessche Statistik mit der BUGS-Sprache	51
7.5	Praktische Bayessche Statistik und MCMC	52
7.6	Poisson GLMs	53
7.7	Binomiale GLMs	54
8	Geschlossene Populationsmodelle	55
8.1	Ein einfaches geschlossenes Populationsmodell: Modell M_0	55
8.1.1	Daten-Vergrößerung	58
8.1.2	Analyse des Modells M_0	58
8.1.3	Modell M_0 für die Würfelnatter-Daten	60
8.2	Modelle mit individueller Heterogenität	62
8.2.1	Analyse des Modells M_h	64
8.2.2	Modell M_h für die Würfelnatter-Daten	65
8.3	Individuelle Kovariablen-Modelle	67
8.3.1	Analyse des Modells M_x	68
8.3.2	Modell M_x für die Würfelnatter-Daten	70
9	Spatiale Capture-Recapture Modelle	75
9.1	Das binomiale Beobachtungsmodell	76
9.1.1	Der Lebensraum-Mittelpunkt	77
9.1.2	Die latente Variable „Abstand“	77
9.2	Das binomiale Punkt-Prozessmodell	78
9.2.1	Der Zustandsraum zum Punkt-Prozess	79
9.3	Die Kernannahmen des SCR-Modells	80
9.4	Das SCR-Modell für unbekanntes N	80
9.4.1	Modell SCR_0 für die Würfelnatter-Daten	81
10	Offene Populationsmodelle	89
10.1	Populationsentwicklungen	89
10.2	Traditionelle Jolly-Seber Modelle	91
10.2.1	Jolly-Seber Modell für die Würfelnatter-Daten	94
10.3	Spatiale Jolly-Seber Modelle	98
11	Resümee	103
A	Allgemeines	105
A.1	Gestutzte Verteilung	105
A.2	Delta-Methode	106
A.3	Details zum Programm Code	108
A.3.1	Beschreibung der Matrix \mathbf{X} aus Abschnitt 3.2	108
A.3.2	Code zu Beispiel 4.1	108
A.3.3	Beschreibung der Matrix \mathbf{X} aus Abschnitt 8	109
A.3.4	Beschreibung der Versteckplätze aus Abschnitt 8.3.2	109

A.3.5 Beschreibung der Fundorte der Individuen aus Abschnitt 8.3.2 110
A.3.6 Beschreibung der Datenmatrix aus Abschnitt 9 110
A.3.7 Beschreibung der Versteckplätze aus Abschnitt 9 110
A.3.8 Beschreibung der Datenmatrix aus Abschnitt 10 110

Literaturverzeichnis **111**

KAPITEL 1

Einleitung

Die vorliegende Diplomarbeit beschäftigt sich mit der Schätzung von Bestandsgrößen einer Population. Dazu stellt die Capture-Recapture Methode eine Möglichkeit dar, die Schätzung einer unbekanntem Populationsgröße durchzuführen. Konkret interessierten sich Biologen der österreichischen Naturschutzjugend dafür, die Gesamtpopulation der gefährdeten Würfelmatter in Graz zu ermitteln. Die von ihnen gesammelten Aufzeichnungen zu diversen Würfelmatter-Funden entlang der Mur stellten die Grundlage für die folgenden Berechnungen dar.

Nach einem einführenden allgemeinen Teil über die Problemstellung und die Würfelmatter, folgt ein Kapitel über die Capture-Recapture Methode. Darin wird das Datenformat festgelegt und die Vorgehensweise zur Bestimmung der Populationsgröße beschrieben.

In weiterer Folge beschäftigt sich ein Abschnitt mit der theoretischen Herleitung von Schätzern einer geschlossenen Population und deren Varianzen. Dazu wird in Kapitel 4 als Grundlage die Familie der Power Series Verteilung behandelt. Diese wird für die Verteilungsannahmen Poisson, binomial und geometrisch untersucht und auf die Würfelmatter-Daten aus dem Jahr 2014 angewendet. Ausgehend von Power Series Verteilungen können dann in Kapitel 5 sehr einfach Schätzer für die Populationsgröße hergeleitet werden. Die behandelten Schätzer sind der Chao-, der biaskorrigierte Chao- und der Zelterman-Schätzer. Der Vollständigkeit halber werden in Kapitel 6 zusätzlich die Varianzen für diese Populationsgrößenschätzer hergeleitet.

Dadurch können anschließend die geschätzten Populationsgrößen für Würfelnatter-Daten inklusive Standardfehler visualisiert werden.

Der nächste Abschnitt formuliert Modelle zur Bestimmung von Bestandsgrößen. Ein einleitendes Kapitel gibt Informationen zu generalisierten linearen Modellen und zur Bayesschen Statistik, die die Grundlage für die folgenden Populationsmodelle darstellen. Es wird sowohl auf die Markov-Ketten-Monte-Carlo Methoden als auch auf die BUGS-Sprache eingegangen. Mit diesem Wissen werden in Kapitel 8 verschiedene geschlossene Populationsmodelle betrachtet, die eine Basis für weitere komplexere Modelle darstellen. Es werden das Prinzip der Daten-Vergrößerung erklärt und - im Hinblick auf spatiale Modelle - bereits individuelle Kovariablen eingeführt, die Fundpunkte einer Population berücksichtigen. Damit ist es nun sehr einfach, in Kapitel 9 ein vollständig spatiales Capture-Recapture Modell zu formulieren. Unter Einhaltung einiger Annahmen kann damit die unbekannte Populationsgröße durch Berücksichtigung des geografischen Aspekts geschätzt werden. Da häufig auch die Entwicklung einer Population in einzelnen Jahren interessant ist, werden abschließend in Kapitel 10 offene Populationsmodelle betrachtet. Dazu zählen das traditionelle Jolly-Seber Modell und das spatiale Jolly-Seber Modell. Jedes dieser Modelle wird zur Veranschaulichung auf die Würfelnatter-Daten angewendet und interpretiert.

KAPITEL 2

Einführung

Der Baubeginn des Wasserkraftwerks im Stadtgebiet von Graz, rückte zu Beginn des Jahres 2017 die europaweit streng geschützte Würfelnatter in den Mittelpunkt. Die Mur im Stadtgebiet von Graz ist eine Engstelle des Verbreitungsgebiets der Würfelnatter zwischen der südlichen und nördlichen Steiermark. Daher kann ein solcher Kraftwerksbau große Auswirkungen auf die lokale Population haben (Kammel et al., 2016, S.9).

2.1 Die Würfelnatter

Die Würfelnatter (*Natrix tessellata*) ist eine ungiftige Schlange aus der Familie der Nattern (*Colubridae*) und gehört zur Gattung der Europäischen Wassernattern (*Natrix*). Sie besiedelt fischreiche, warme und langsam fließende Gewässer mit steinigem Untergrund und Uferregionen von Seen. Ihre Nahrung besteht fast ausschließlich aus Fischen, daher ist sie auf Gewässer mit einer entsprechenden Fischdichte angewiesen. Dort erbeutet die Würfelnatter eine große Bandbreite an Fischen; vorzugsweise Arten, die sich in Flachwasserzonen oder am Gewässergrund strömungsarmer Bereiche aufhalten.



Abbildung 2.1: Würfelnatter (*Natrix tessellata*)
(links: Johannes Volkmer; Mitte/rechts: Laura Pabst).

Die Würfelnatter ist neben der Sandvipera (*Vipera ammodytes*) die seltenste Schlangengattung in Österreich und wird in der Roten Liste der gefährdeten Tiere Österreichs als „stark gefährdet“ eingestuft. Nach Fauna-Flora-Habitat-Richtlinie gehört sie zu den „streng geschützten“ Tierarten. Europaweit befindet sie sich in einem als „ungünstig/unzureichend“ und in Österreich als „ungünstig/schlecht“ eingestuften Erhaltungszustand.

Natrix tessellata ist ein ausgezeichneter Indikator für den Zustand eines Gewässers einschließlich des angrenzenden Umlandes, da sie empfindlich auf Eingriffe in das Ökosystem hinsichtlich Gewässergüte, Fischbestand, wasserbauliche Maßnahmen sowie Ausmaß, Gestaltung und Beeinträchtigung der ufernahen Lebensräume reagiert. Die Würfelnatter steht also dafür, welchen hohen naturschutzfachlichen Wert der Fluss inmitten der Stadt hat. Eine energiewirtschaftliche Nutzung der Mur in Form von großen Wasserkraftwerken stellt eine Bedrohung für die Bestände der Würfelnatter dar (Kammel et al., 2016, S.7ff).

Das Bauvorhaben Murkraftwerk Graz hinterlässt somit einen überwiegend letalen Einfluss auf die lokale Teilpopulation der Würfelnatter, der auch durch eine fachspezifische Bauaufsicht nicht verhindert werden kann. Hinzu kommt der nicht berücksichtigte Straßentod durch Bautätigkeiten und Baustellenverkehr, die während der gesamten Vegetationsperiode durchgeführt werden. Durch Abfangen und Aussiedeln einzelner Individuen ließe sich eine Reduktion der mortalen Einflüsse während der Bautätigkeiten erreichen. Für eine mögliche Absiedlung der Individuen zum bereits fertig gestellten Kraftwerk Gössendorf müssen aber die Lebensraumbedingungen für die Würfelnatter ausreichend gegeben sein (Kammel et al., 2016, S.5).

Neben dem Ab- bzw. Umsiedeln der Würfelnatter möchte man zusätzlich einen Überblick über die Population der Würfelnatter an der Mur im südlichen Stadtgebiet von Graz bekommen. Um ein detailliertes Bild über die Populationsgröße und -struktur zu erhalten, fanden in den Jahren 2013 bis 2016 Begehungen entlang der Mur statt. Das Ziel dieser Daten ist es, einen gegebenenfalls eingetretenen Umweltschaden durch den Bau von Kraftwerk und Sammelkanal an den lokalen Po-

pulationen der Würfelnatter besser erfassen zu können (Kammel et al., 2016, S.11). Anhand solcher Erhebungen kann anschließend die Populationsgröße mit Hilfe der *Capture-Recapture* Methode geschätzt werden.

2.2 Fang-Wiederfang

Die Methode des Fang-Wiederfangs (später *Capture-Recapture*) ermöglicht neben der Schätzung der Populationsgröße auch eine Analyse weiterer Faktoren, die für die Dynamik von Tierpopulationen wichtig sind, wie beispielsweise:

- Individuelles Wachstum
- Aktivitätsraum
- Reproduktion (Geburten)
- Immigration (Einwanderung)
- Emigration (Auswanderung)
- Mortalität (Todesfälle)

Eine direkte Zählung von Schlangenpopulationen ist aufgrund ihrer sehr versteckten Lebensweise und hohen Mobilität nicht durchführbar. Daher bietet die indirekte Methode des Fangens und Wiederfangens eine gute Möglichkeit, um die Populationsgröße zu schätzen. Individuen werden dabei gefangen und fotografiert. Aufgrund einer individuellen Musterung der Ventralseite (vorderes Körperdrittel; ähnlich dem menschlichen Fingerabdruck) kann jedes Individuum von Würfelnatter erkannt werden. Um im Zweifelsfall zusätzliche Unterscheidungsmerkmale durch eine Beschuppung zu erheben, werden außerdem Aufnahmen der Kopfseite vorgenommen. Bevor nach einer gewissen Zeit wieder ein Fang-Durchgang durchgeführt wird, werden die Tiere wieder freigelassen. Solche Fang-Durchgänge werden mehrmals wiederholt und bereits gefangene Individuen können eindeutig erkannt werden. Anschließend kann aus dem Verhältnis der wiedergefangenen Würfelnattern zu den Neufängen eine Schätzung der Bestandsgröße durchgeführt werden. Dabei gilt folgende Faustregel: Wenn wenige Tiere wiedergefangen werden, so handelt es sich um eine große Population. Sind es jedoch viele, so ist von einer kleinen Population auszugehen (Kammel et al., 2016, S.12ff).

2.3 Künstliche Versteckplätze

Damit die Nachweisbarkeit von Würfelnattern erhöht werden kann, wurden bei der vorliegenden Bestandserhebung je nach Jahr zwischen 50 und 70 künstliche Versteckplätze (kV) beidseitig der Mur in der Böschung ausgebracht. Würfelnattern ziehen sich gerne unter flachen und thermisch begünstigten Strukturen zurück. Die künstlichen Versteckplätze zielen auf das Bedürfnis der Tiere ab und dienen als Tagesversteckplätze, Nachtquartiere oder Plätze zum Aufwärmen. Gefundene Würfelnattern wurden gefangen, vermessen und zur individuellen Wiedererkennung fotografiert. Die Kombination der Verwendung von künstlichen Versteckplätzen mit der individuellen Wiedererkennung der gefangenen Würfelnattern bietet die Möglichkeit, weiterführende Aussagen über die Raumnutzung zu treffen.

Hergestellt wurden die künstlichen Versteckplätze aus 1 bis 1.3 m² großen Teichfolienstücken, die mit Lärchenholzleisten an zwei Seiten stabilisiert wurden (siehe Abbildung 2.2).

Zusätzlich zu den ausgelegten künstlichen Versteckplätzen wurde das Untersuchungsgebiet bei den Erhebungsterminen auch abseits abgegangen. Für Schlangenarten charakteristische Strukturen wie zum Beispiel Totholzhaufen, Böschungsbereiche mit ausreichender Sonneneinstrahlung (geringe Bedeckung durch Gehölze) oder der Randbereich ausgemähter Wiesen- und Staudenflächen an der Böschungsoberkante zum uferbegleitenden Radweg wurden gezielt aufgesucht (Kammel et al., 2016, S.15ff).



Abbildung 2.2: Künstlicher Versteckplatz zum Nachweis von Schlangen (Kammel et al., 2016, S.17).

Die Capture-Recapture Methode

Heutzutage ist man häufig daran interessiert, den Umfang N einer schwer zu erfassenden Zielpopulation anzugeben, der nicht einfach durch Abzählen der Einheiten erreicht werden kann. Dies können beispielsweise eine Wildtier-Population, die Population von obdachlosen oder drogensüchtigen Leuten, Software-Fehler oder Menschen mit einer bestimmten Krankheit sein. Oft wird in solchen Fällen ein Identifikationsgerät (eine Falle, ein Register, ein Screening-Test) wiederholt angewendet, um Einheiten aus der Population zu registrieren. Danach möchte man die globale Größe N der Zielpopulation schätzen (Böhning, van der Heijden und Bunge, 2017, S.4).

Die Capture-Recapture Methode (*CR-Methode*) bietet eine Möglichkeit, die Schätzung einer unbekanntem Populationsgröße durchzuführen. Die Grundidee ist es, Individuen zu fangen (*Capture*), zu markieren und zu identifizieren und danach wieder frei zu lassen, damit sie sich mit der Population vermischen können. In einer zweiten Stichprobe werden wieder Individuen gefangen und markiert. Zudem wird gezählt, wie viele Individuen bereits in der ersten Stichprobe gesehen und markiert wurden (*Recapture*). Diese Methode kann beliebig weit fortgeführt werden (Kaskasamkul, 2017, S.7). Möchte man zum Beispiel die Anzahl an Fischen in einem großen Teich bestimmen, so werden nach dem Ansatz der CR-Methode innerhalb eines bestimmten Zeitraums Stichproben aus der Population erfasst. Angenommen es wird an zwei Tagen je eine Stichprobe mittels Fischfang erfasst. Dann werden beim ersten Fisch-

fang n_1 Fische aus der unbekanntem Fischpopulation mit Umfang N gefangen und markiert. Am zweiten Tag können n_2 Fische gefangen werden. Dabei sind bereits m_2 der n_2 Fische markiert, dies bedeutet, dass m_2 Fische wiedergefangen wurden (Spendel, 2016, S.7). Später werden Schätzer für die Populationsgröße angegeben, die mit Hilfe der aufgezeichneten Wiederfänge (Häufigkeiten der Wiederfänge) den Umfang N der Zielpopulation schätzen.

Grundsätzlich unterscheidet man zwischen *geschlossenen* und *offenen* Populationen. Eine geschlossene Population liegt vor, wenn man davon ausgeht, dass es während der Untersuchung zu keinen Todesfällen, Geburten oder Ein- und Abwanderungen kommt. Das kann im Freiland der Fall sein, wenn die Untersuchung in einem sehr kurzen Zeitraum stattfindet, beispielsweise ist die oben erwähnte Fischpopulation eine geschlossene Population. Dabei treten die Auswirkungen der genannten Prozesse auf die Population nicht auf oder sind so gering, dass sie vernachlässigt werden können. Von einer offenen Population spricht man hingegen, wenn mindestens einer dieser Prozesse wesentlich eintritt. Während bei der Analyse von geschlossenen Populationen nur die Populationsgröße und gegebenenfalls die Dichte untersucht werden, ist man bei der offenen Population zunächst am Zuwachs und an der Mortalität und in zweiter Linie an der Populationsgröße interessiert. Um also ein vollständiges Verständnis über Populationsentwicklungen zu erhalten, benötigt man Informationen über gebietliche (*spatiale*) sowie zeitliche Veränderungen in der Populationsdichte und demografische Parameter. Erst dann können Prozesse modelliert werden, die Entwicklungen wie Überleben, Zuwachs/Verstärkung und Bewegung über längere zeitliche Perioden (Abwanderung, Verbreitung usw.) regeln (Royle, Chandler, Sollman und Gardner, 2014, S.401). Somit werden oftmals Modelle ausgehend von einer geschlossenen Population um spatiale Informationen erweitert. Man spricht dann von *spatialen Capture-Recapture Modellen (SCR-Modellen)* (Royle et al., 2014, S.14).

In dieser Diplomarbeit werden Capture-Recapture Modelle auf die Würfelnatter-Population aus Abschnitt 2 angewandt. Für den ersten Teil wird der Einfachheit halber eine geschlossene Würfelnatter-Population angenommen. Der zweite Teil lässt zudem spatiale Informationen, wie beispielsweise GPS-Koordinaten der Fundstelle einer Würfelnatter, einfließen. Dadurch können spatiale Capture-Recapture Modelle für die Würfelnatter-Population untersucht werden.

3.1 Allgemeine Problemstellung

Die CR-Methode zielt auf das Schätzen einer unbekanntem Populationsgröße N . Dabei werden Stichproben, die Informationen über die Population enthalten, t -mal beprobt. Die Anzahl an Stichproben in der Multiplen-Quellen-Situation ist beliebig

mit $t > 2$. Für den Sonderfall $t = 2$ können einfachere Methoden angewendet werden, diese werden hier jedoch nicht weiter betrachtet.

Die Daten einer Stichprobe j mit $j = 1, \dots, t$ liegen in Form einer Matrix vor:

$$\mathbf{X} = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1t} \\ X_{21} & X_{22} & \dots & X_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ X_{N1} & X_{N2} & \dots & X_{Nt} \end{pmatrix}. \quad (3.1)$$

Dabei sind die Einträge Bernoulliverteilte Zufallsvariablen (Fang - Nichtfang) mit

$$X_{ij} = \begin{cases} 1, & \text{wenn } i\text{-te Einheit in } j\text{-ter Stichprobe erfasst,} \\ 0, & \text{sonst} \end{cases} \quad (3.2)$$

für $i = 1, \dots, N$ und $j = 1, \dots, t$. Eine Zeile der Matrix \mathbf{X} steht für die Erfassungen der i -ten Einheit, eine Spalte für die erfassten Einheiten der Stichprobe j . Sei x_{ij} die Realisierung des Matrixeintrages X_{ij} aus der Matrix (3.1), dann ist eine weitere Darstellung der Datensituation durch Tabelle 3.1 gegeben. Die i -te Zeilensumme entspricht der Realisierung der Zufallsvariable Y_i . Dabei ist Y_i der i -te Eintrag des Zufallsvektors $\mathbf{Y} = (Y_1, \dots, Y_i, \dots, Y_N)$. Die Verteilung von Y entspricht der Randverteilung der Matrix \mathbf{X} , wobei die Einheiten $i = 1, \dots, N$ berücksichtigt werden (Spendel, 2016, S.8f).

Einheit i	Stichprobe j				
	1	2	...	t	
1	x_{11}	x_{12}	...	x_{1t}	y_1
2	x_{21}	x_{22}	...	x_{2t}	y_2
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
N	x_{N1}	x_{N2}	...	x_{Nt}	y_N

Tabelle 3.1: Darstellung der Datensituation mit t Stichproben.

Offensichtlich weiß man nur über Individuen mit $y_i = \sum_{j=1}^t x_{ij} > 0$ Bescheid, das heißt, nur für Individuen mit zumindest $j = 1, \dots, t$ kann man $x_{ij} > 0$ beobachten. Für die restlichen Einheiten gilt $x_{ij} = 0$, somit bleiben diese *unbeobachtet*. Die Anzahl der Stichproben t ist a priori bekannt. Man kann die Indizes der Einheiten umschreiben, sodass Y_1, \dots, Y_N die globale Population und Y_1, \dots, Y_S die beobachteten Einheiten bezeichnen. Hier gilt $Y_{S+1} = \dots = Y_N = 0$.

In der statistischen Modellierung für diskrete Responses ist es üblich, die Zielpopulation durch die Wahrscheinlichkeitsfunktion (j, p_j) zu beschreiben. Hierbei ist jedoch

$j = 0, 1, \dots$ beliebig und p_j bezeichnet die Wahrscheinlichkeit, dass eine Einheit aus der Population genau j -mal beobachtet wurde. Es gelten dafür die gewöhnlichen Bedingungen $p_j \geq 0$ und $\sum_{j=0}^{\infty} p_j = 1$. Die zugehörige empirische Verteilung lautet (j, f_j) , wobei f_j die Häufigkeit angibt, dass eine Einheit exakt j -mal gefangen wurde. Aufgrund des angenommenen Stichproben-Designs werden nur Einheiten mit $j > 0$ beobachtet. Somit ist die Größe der beobachteten Probe $S = \sum_{j>0} f_j$ und die zugehörige Verteilung ist *um die Null gestutzt* (Details siehe Anhang A.1). Damit sind Einheiten mit $j = 0$ nicht in den Daten enthalten (Böhning et al., 2017, S.23).

Im Sinne dieser Schreibweise können Einheiten hier maximal t -mal erfasst werden und sind jeweils durch f_j gegeben, wofür gilt

$$f_j = \sum_{i=1}^N I(Y_i = j) \text{ für alle } j = 1, \dots, t. \quad (3.3)$$

Dabei bezeichnet Y_i die i -te Zeilensumme von \mathbf{X} und $I(\cdot)$ steht für die Indikatorfunktion. Somit zählt f_1 jene Einheiten, die einmal gefangen wurden (*Capture*) und beinhaltet noch keine Wiederfänge (*Recapture*). Hingegen zählt f_2 jene Einheiten, die zweimal gefangen wurden, und beinhaltet daher alle einmal wiedergefangenen Einheiten. Die Anzahl an unterschiedlich oft erfassten Einheiten wird mit S definiert und entspricht der Summe der j -mal erfassten Einheiten über alle Stichproben (Spendel, 2016, S.9f):

$$S = \sum_{j=1}^t f_j. \quad (3.4)$$

Eine naheliegende Schätzung für p_j ist die relative Häufigkeit $\frac{f_j}{N}$, was allerdings nicht berechnet werden kann, nachdem N unbekannt ist. Jedoch bietet die empirische relative Häufigkeit $\frac{f_j}{S}$ eine Schätzung der Null-gestutzten Wahrscheinlichkeit $\frac{p_j}{1-p_0}$. Damit bleiben f_0 und $N = \sum_{j=0}^t f_j = \sum_{j=1}^t f_j + f_0 = S + f_0$ unbekannt und das Finden eines Schätzers für die Populationsgröße N auf Basis der beobachteten, Null-gestutzten Verteilung beschreibt eine spezielle Form des gewöhnlichen Capture-Recapture Problems (Böhning et al., 2017, S.23).

Angenommen p_0 ist die Wahrscheinlichkeit einer nicht beobachteten Einheit (*zero count*) und die Wahrscheinlichkeit der Einheit identifiziert zu werden ist gegeben durch $1-p_0$. Wenn man davon ausgeht, dass jede Einheit die selbe Identifikationswahrscheinlichkeit hat, so umfasst die Populationsgröße N die Anzahl der beobachteten und nicht beobachteten Einheiten

$$N = \underbrace{N(1-p_0)}_{\text{beobachtet}} + \underbrace{Np_0}_{\text{nicht beobachtet}} = S + Np_0. \quad (3.5)$$

Der beobachtete Term $N(1-p_0)$ kann durch die Anzahl der beobachteten Einheiten S mit $\mathbb{E}(S) = N(1-p_0)$ geschätzt werden. Die Wahrscheinlichkeit p_0 ist unbekannt und muss geschätzt werden. Dadurch kann die Populationsgröße N in (3.5) leicht umgeformt werden und führt zum bekannten Horvitz-Thompson-Schätzer (Kaskasamkul, 2017, S.9)

$$\hat{N} = \frac{S}{1 - \hat{p}_0}. \quad (3.6)$$

Dieser entspricht einer Null-gestutzten Verteilung. Details dazu können in Anhang A.1 gefunden werden.

Zusammenfassend besteht die Idee der CR-Methode darin, Schnittmengen von Stichproben zu verwenden, um Aussagen über die Populationsgröße N treffen zu können. Eine Stichprobe erfasst Einheiten und markiert diese. Nach dem Erfassen und Markieren wird die besagte Einheit wieder zurück in die Population entlassen (vgl. Ziehen aus einer Urne mit Zurücklegen). Beim Ziehen eines Elements einer Stichprobe j gibt es zwei Möglichkeiten bezüglich der Markierung einer Einheit: Wenn die Einheit bereits markiert ist, dann wurde sie schon in einer der vorigen Stichproben erfasst (*Recapture*). Wenn die Einheit noch nicht markiert ist, so wird sie markiert (*Capture*) (Spendel, 2016, S.10).

3.2 Problemstellung Würfelnatter-Daten

Im Falle der Würfelnatter-Daten sieht die Problemstellung folgendermaßen aus: Biologen, unter der Leitung von Oliver Gebhardt, legten künstliche Versteckplätze (siehe Abschnitt 2.3) entlang der Mur aus. In den Jahren 2013 bis 2016 führten sie Begehungen an diesen Versteckplätzen durch, jeweils beginnend mit April oder Mai und fortführend bis September. So ergaben sich insgesamt 85 Begehungen über alle vier Jahre verteilt. Diese entsprechen den oben genannten Stichproben - also den Spalten der Matrix \mathbf{X} . Während dieser vier Jahre konnten 195 Würfelnatter-Individuen identifiziert werden, die zumindest einmal gefangen wurden, was den Zeilen der Matrix \mathbf{X} entspricht. Somit ist 195×85 die Dimension der Datenmatrix \mathbf{X} . Als Beispiel wird in dieser Arbeit das Jahr 2014 mit 31 Begehungen behandelt. Details zu den Begehungstagen und zu den Wiederfängen werden später in Beispiel 4.1 aufgelistet. Die gesamte Matrix \mathbf{X} wird im Anhang A.3 detailliert beschrieben.

Da einige Populationen bis auf kurze Zeitintervalle geschlossen sind, dienen CR-Modelle für geschlossene Populationen als Basis für die Entwicklung von spatialen CR-Modellen. Daher werden im ersten Teil dieser Arbeit zuerst Schätzer der Popu-

lationsgröße unter einem geschlossenen Modell hergeleitet und diskutiert (Royle et al., 2014, S.87).

Spendel (2016, S.49ff) behandelt für die Multiple-Quellen-Situation den Chao-Schätzer und den Zelterman-Schätzer bei Annahme einer geschlossenen Population. Die folgenden drei Abschnitte beschäftigen sich mit der Ableitung dieser Schätzer aus einfachen Power Series Verteilungen, sowie mit der Herleitung der zugehörigen Varianzen dieser Schätzer.

Power Series Verteilung

Hier wird die Familie der Power Series (*Potenzreihen*) Verteilung vorgestellt. Daraus können in weiterer Folge die bekannten Chao- und Zelterman-Schätzer bei Vorliegen einer *geschlossenen Population* abgeleitet werden.

Um jedoch einen gültigen Schätzer der Größe einer geschlossenen Population hervorzubringen, müssen folgende Annahmen an die Daten gestellt werden (Spendel, 2016, S.15):

1. Geschlossenheit der Population
2. Unabhängigkeit der Stichproben
3. Homogene Wahrscheinlichkeiten

Die genannten Annahmen haben jedoch folgende Schwächen:

Ad 1) Die Geschlossenheit der Population kann nur im Falle von kurzen Studiendauern angenommen werden. Bei längerer Studiendauer kann eine Einheit sterben (Austreten aus der Population) bzw. geboren werden (Eintreten in die Population). Damit ist die Populationsgröße über die Studiendauer nicht konstant. Zur Schätzung der Würfelgatter-Population wurden die Daten eines kurzen Ausschnittes aus dem gesamten Untersuchungszeitraum herangezogen. Es wird angenommen, dass sich die Population in einer gewissen stabilen Situation befindet, in der die Geburtenrate und die Anzahl der einwandernden Individuen annähernd gleich der Sterberate und der

Anzahl der abwandernden Individuen ist.

Ad 2) Die Unabhängigkeit der Stichproben kann nicht ohne weiteres angenommen werden. Ein Aufscheinen einer Einheit in der einen Stichprobe hat oftmals einen direkten Einfluss auf das Aufscheinen in einer anderen Stichprobe (abhängige Stichproben auf Einheitsebene). Zudem können auch heterogene Fangwahrscheinlichkeiten eine Abhängigkeit in den Stichproben hervorrufen. Die Abhängigkeiten von Stichproben können zu einer Unter- bzw. Überschätzung der Populationsgröße N führen. Dieses Phänomen spiegelt sich in der Verzerrung der Schätzer wider. Im Allgemeinen wird die Unabhängigkeit der Stichproben relaxiert, womit die Capture-Recapture Methode auf praktischer Ebene einen nahezu unverzerrten Schätzer liefert. Im Falle der Würfelnatter-Daten hat die Fangerfahrung keine Auswirkungen auf das Verhalten der Tiere. Markierte (bereits gefangene) Exemplare werden nicht leichter oder schwerer gefangen als unmarkierte. Zusätzlich mischen sich markierte Individuen vollständig unter die unmarkierten.

Ad 3) Homogene Wahrscheinlichkeiten für die Häufigkeiten des Aufgegriffenwerdens sind häufig verletzt, da die Population oft aus verschiedenen Teilpopulationen besteht. Jede Teilpopulation hat dieselbe Verteilung aber unterschiedliche Parameter. Dies beschreibt die sogenannte *Heterogenität*. Hierbei ist es naheliegend, dass sich die Fangwahrscheinlichkeiten für jedes Individuum unterscheiden. Chao und Zelterman geben Schätzer für den Umfang der Population an, die Heterogenität berücksichtigen (Kaskasamkul, 2017, S.3). Diese werden hier genauer behandelt. Schätzer unter *Homogenität* können in Böhning et al. (2017, S.10ff) gefunden werden. Bei der Würfelnatter sollte die Fangwahrscheinlichkeit mit bisher nicht gefangenen Tieren gleich sein. Außerdem darf sich die Fangwahrscheinlichkeit zwischen verschiedenen Erfassungszeitpunkten nicht verändern.

Bevor die einzelnen Schätzer betrachtet werden können, benötigt man ein allgemeines Verständnis über Power Series und Ratios.

4.1 Power Series

Böhning et al. (2017, S.30) definieren eine Power Series als

$$\eta(\theta) = \sum_{x=0}^{\infty} \alpha_x \theta^x.$$

Laut dieser Definition gilt $\sum_{x=0}^{\infty} \frac{\alpha_x \theta^x}{\eta(\theta)} = 1$. Power Series Verteilungen sind diskrete Verteilungen auf \mathbb{N} mit Wahrscheinlichkeitsfunktion

$$p_x(\theta) = \mathbb{P}(X = x|\theta) = \frac{\alpha_x \theta^x}{\eta(\theta)}, \quad (4.1)$$

wobei $x = 0, 1, 2, \dots$, $\theta > 0$ und $\alpha_x > 0$. Der Koeffizient α_x ist bekannt und positiv, θ ist ein positiver Parameter, x erstreckt sich über die Menge der nicht-negativen ganzen Zahlen und $\eta(\theta) = \sum_{x=0}^{\infty} \alpha_x \theta^x$ ist die Normalisierungs-Konstante.

Die Familie der Power Series Verteilung ist sehr wichtig, da sie eine elegante und scharfsinnige Formulierung von verschiedenen klassischen diskreten Verteilungen bietet, die sowohl in der statistischen Forschung als auch im Capture-Recapture Gebiet angewendet werden (Kaskasamkul, 2017, S.90f). Vidal-Diez (2015, S.70) und Momeni (2011, S.692) geben tabellarische Überblicke über die verschiedenen Parametrisierungen der Power Series Verteilungen. Nachfolgend wird eine zusammengefasste Tabelle dargestellt.

Verteilung	originale Parameter	θ	α_x	$\eta(\theta)$	x
Poisson	$Poi(\lambda)$	λ	$\frac{1}{x!}$	$\exp(\theta)$	$0, 1, 2, \dots$
Binomial	$Bin(m, p)$	$\frac{p}{1-p}$	$\binom{m}{x}$	$(1 + \theta)^m$	$0, 1, \dots, m$
Geometrisch	$G(p)$	$1 - p$	1	$\frac{1}{1-\theta}$	$0, 1, 2, \dots$
Negativ Binomial	$NB(m, p)$	$1 - p$	$\binom{m+x-1}{x}$	$\frac{1}{(1-\theta)^m}$	$0, 1, 2, \dots$
Logarithmisch	$LS(p)$	$1 - p$	$\frac{1}{x}$	$-\ln(1 - \theta)$	$1, 2, 3, \dots$

Tabelle 4.1: Parametrisierung für die Power Series Verteilungen.

Edwin (2014, S.9) zeigt, dass eine Power Series Verteilung zur Exponentialfamilie gehört, da

$$\begin{aligned} p_x(\theta) &= \frac{\alpha_x \theta^x}{\eta(\theta)} = \exp\left(\ln\left[\frac{\alpha_x \theta^x}{\eta(\theta)}\right]\right) \\ &= \exp(\ln \alpha_x \theta^x - \ln \eta(\theta)) \\ &= \exp(\ln \alpha_x + x \ln \theta - \ln \eta(\theta)). \end{aligned}$$

Außerdem formuliert er die ersten und zweiten Ableitungen von $\eta(\theta)$ als

$$\eta'(\theta) = \frac{d\eta}{d\theta} = \sum_{x=0}^{\infty} x \alpha_x \theta^{x-1} = \sum_{x=1}^{\infty} x \alpha_x \theta^{x-1}$$

und

$$\eta''(\theta) = \frac{d^2\eta}{d\theta^2} = \sum_{x=1}^{\infty} x(x-1) \alpha_x \theta^{x-2} = \sum_{x=2}^{\infty} x(x-1) \alpha_x \theta^{x-2}$$

und folgert daraus schlussendlich den Erwartungswert und die Varianz von Power Series Verteilungen, nämlich

$$\mathbb{E}(X) = \sum_{x=0}^{\infty} x \frac{\alpha_x \theta^x}{\eta(\theta)} = \frac{\theta}{\eta(\theta)} \sum_{x=1}^{\infty} x \alpha_x \theta^{x-1} = \frac{\theta}{\eta(\theta)} \eta'(\theta), \quad (4.2)$$

$$\mathbb{E}(X(X-1)) = \sum_{x=0}^{\infty} x(x-1) \frac{\alpha_x \theta^x}{\eta(\theta)} = \frac{\theta^2}{\eta(\theta)} \sum_{x=2}^{\infty} x(x-1) \alpha_x \theta^{x-2} = \frac{\theta^2}{\eta(\theta)} \eta''(\theta),$$

und

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}(X(X-1)) + \mathbb{E}(X) - [\mathbb{E}(X)]^2 \\ &= \frac{\theta^2}{\eta(\theta)} \eta''(\theta) + \frac{\theta}{\eta(\theta)} \eta'(\theta) - \left[\frac{\theta}{\eta(\theta)} \eta'(\theta) \right]^2 \\ &= \theta \frac{\eta'(\theta)}{\eta(\theta)} + \theta^2 \left(\frac{\eta''(\theta)}{\eta(\theta)} - \left[\frac{\eta'(\theta)}{\eta(\theta)} \right]^2 \right). \end{aligned} \quad (4.3)$$

4.2 Ratios

Die Power Series Verteilung (4.1) hat eine interessante Eigenschaft. Betrachtet man die Verhältnisse (*ratios*) von benachbarten Wahrscheinlichkeiten der Power Series multipliziert mit einem bekannten Faktor über den Bereich von x , dann liefert dies einen konstanten Wert, der dem unbekanntem Parameter θ ähnlich ist. Kaskasamkul (2017, S.92) zeigt

$$\frac{p_x(\theta)}{p_{x+1}(\theta)} = \frac{\frac{\alpha_x \theta^x}{\eta(\theta)}}{\frac{\alpha_{x+1} \theta^{x+1}}{\eta(\theta)}} = \frac{\alpha_x}{\alpha_{x+1}} \frac{1}{\theta}$$

und

$$\frac{p_{x-1}(\theta)}{p_x(\theta)} = \frac{\frac{\alpha_{x-1} \theta^{x-1}}{\eta(\theta)}}{\frac{\alpha_x \theta^x}{\eta(\theta)}} = \frac{\alpha_{x-1}}{\alpha_x} \frac{1}{\theta}.$$

Dies liefert für θ

$$\theta = \frac{\alpha_x}{\alpha_{x+1}} \frac{p_{x+1}(\theta)}{p_x(\theta)} = \frac{\alpha_{x-1}}{\alpha_x} \frac{p_x(\theta)}{p_{x-1}(\theta)} = r_x. \quad (4.4)$$

In Capture-Recapture Studien sind die Verteilungen um die Null gestutzt (siehe Abschnitt A.1). Sei dazu $p_x^+(\theta) = \frac{p_x(\theta)}{1-p_0(\theta)}$ die Null-gestutzte Wahrscheinlichkeitsfunktion. Das Verhältnis r_x ist dasselbe wie das der Null-gestutzten Power Series Verteilung, da

$$r_x = \frac{\alpha_x}{\alpha_{x+1}} \frac{p_{x+1}^+(\theta)}{p_x^+(\theta)} = \frac{\alpha_x}{\alpha_{x+1}} \frac{\frac{p_{x+1}(\theta)}{1-p_0(\theta)}}{\frac{p_x(\theta)}{1-p_0(\theta)}} = \frac{\alpha_x}{\alpha_{x+1}} \frac{p_{x+1}(\theta)}{p_x(\theta)}.$$

In der Praxis wird das Verhältnis geschätzt durch

$$\hat{r}_x = \frac{\alpha_x}{\alpha_{x+1}} \frac{f_{x+1}}{f_x} = \hat{\theta}, \quad (4.5)$$

wobei f_x die Häufigkeit der Einheiten bezeichnet, die genau x -mal entdeckt wurden, für $x = 0, \dots, t$, und t die größtmögliche beobachtete Anzahl ist. Es wird benötigt, dass f_{x+1} und f_x beide positiv sind. Wenn einer der beiden Werte Null ist, so kann kein Verhältnis bestimmt werden.

4.3 Ratio-Plot

Böhning, Baksh, Lerdsuwansri und Gallagher (2013, S.138) nutzen den Graph von x gegen \hat{r}_x als Diagnoseelement für die Abweichung einer empirischen Verteilung von einem Mitglied der generalisierten Power Series Verteilung (*Generalized Power Series Distribution - kurz GPS*). Dieser Graph wird *Ratio-Plot* genannt. Dabei ist eine horizontale Linie konsistent mit homogenen Power Series Verteilungen. Abweichungen von einer horizontalen Linie hingegen erbringen den Hinweis nicht beobachteter Heterogenität.

Um nicht beobachtete Heterogenität mit dem Konzept des Ratio-Plots zu verknüpfen, betrachtet Kaskasamkul (2017, S.91) eine Mischung der Power Series Verteilung. Dazu wird nicht beobachtete Heterogenität mit einer latenten Zufallsvariable Θ beschrieben. Die gemeinsame Wahrscheinlichkeitsdichte der zählenden Zufallsvariable X und der latenten Variable Θ ist gegeben als $g(x, \theta) = g(x|\theta)g(\theta)$ und $g(x|\theta) = g(x, \theta)/g(\theta)$, wobei $g(\theta) = \int_x g(x, \theta)dx$ die Dichte von zufälligen Θ ist. Da die latente Variable unbekannt ist, wird die Randdichte $\int_\theta g(x|\theta)g(\theta)d\theta$ von x anstatt der gemeinsamen Dichte betrachtet. Definiert man die konditionale Dichte $g(x|\theta)$ durch die Power Series Dichte $p_x(\theta)$, so ist das gemischte Modell für eine Verteilung in der Power Series Familie nach Böhning et al. (2017, S.31) gegeben als

$$m_x = \mathbb{P}(X = x) = \int_\theta p_x(\theta)g(\theta)d\theta.$$

Hierbei bezeichnet $g(\theta)$ die mischende Dichte für θ und beschreibt nicht beobachtete Heterogenität. Das Verhältnis eines gemischten Modells ist dann definiert als

$$R_x = \frac{\alpha_x}{\alpha_{x+1}} \frac{m_{x+1}}{m_x}. \quad (4.6)$$

Die Mischung von Power Series Verteilungen scheint flexibler zu sein. Kaskasamkul (2017, S.93) formuliert einen Satz, der zeigt, dass der Ratio-Plot im Fall von gemischten Power Series Verteilungen kein horizontales Linienmuster mehr aufweist.

Im Gegenteil erhält man ein monoton steigendes Muster mit steigendem x .

Satz 4.1. Sei $m_x = \int_{\theta} p_x(\theta)g(\theta)d\theta$, wobei $p_x(\theta)$ ein Mitglied der Power Series Familie und $g(\theta)$ eine beliebige Dichte ist. Dann gilt für alle $x = 0, 1, \dots$ und für $R_x = \frac{\alpha_x}{\alpha_{x+1}} \frac{m_{x+1}}{m_x}$ folgende Monotonität:

$$R_x \leq R_{x+1}.$$

Damit kann der Ratio-Plot als Indikator für das Vorkommen von Heterogenität angesehen werden, wenn es ein monoton steigendes Muster im Plot gibt.

4.3.1 Ratios und Ratio-Plots der einzelnen Power Series Verteilungen

Im Folgenden werden die Verhältnisse der Verteilungen Poisson, Binomial und Geometrisch inklusive Ratio-Plot betrachtet. Dazu werden die Parameter α_x aus Tabelle 4.1 in Formel (4.5) eingesetzt.

- **Poisson**

$$\begin{aligned} \hat{r}_{xPoi} &= \frac{\alpha_x}{\alpha_{x+1}} \frac{f_{x+1}}{f_x} = \frac{\frac{1}{x!}}{\frac{1}{(x+1)!}} \frac{f_{x+1}}{f_x} \\ &= (x+1) \frac{f_{x+1}}{f_x}. \end{aligned} \tag{4.7}$$

- **Binomial**

Der Parameter m in der Binomialverteilung entspricht der Anzahl an Stichproben t in Capture-Recapture Studien. Daher wird hier nur mehr t als Wert dieses Parameters verwendet.

$$\begin{aligned} \hat{r}_{xbin} &= \frac{\alpha_x}{\alpha_{x+1}} \frac{f_{x+1}}{f_x} = \frac{\binom{t}{x}}{\binom{t}{x+1}} \frac{f_{x+1}}{f_x} = \frac{\frac{t!}{x!(t-x)!}}{\frac{t!}{(x+1)!(t-x-1)!}} \frac{f_{x+1}}{f_x} \\ &= \frac{x+1}{t-x} \frac{f_{x+1}}{f_x}. \end{aligned} \tag{4.8}$$

- **Geometrisch**

$$\begin{aligned} \hat{r}_{xgeom} &= \frac{\alpha_x}{\alpha_{x+1}} \frac{f_{x+1}}{f_x} = \frac{1}{1} \frac{f_{x+1}}{f_x} \\ &= \frac{f_{x+1}}{f_x}. \end{aligned} \tag{4.9}$$

Beispiel 4.1. Zur Veranschaulichung der Ratio-Plots werden die Würfelnettern aus dem Jahre 2014 betrachtet. Es fanden $t = 31$ Begehungen mit insgesamt 17 Wiederfängen statt. Details sind den Tabellen 4.2 und 4.3 zu entnehmen.

2014	Begehungen
Jänner	-
Februar	-
März	-
April	15./26./30.
Mai	04./07./10./14./15./18./25./30.
Juni	02./06./10./13./15./18./25./28.
Juli	03./05./12./13./15.
August	18.
September	03./05./09./14./17./21.
Oktober	-
November	-
Dezember	-
Summe	31

Tabelle 4.2: Datum und Anzahl der Begehungen im Jahr 2014.

Würfelnetter	1.Fang	2.Fang	3.Fang	4.Fang	5.Fang	Wiederfänge
WN20	15.04.	26.04.	-	-	-	1
WN24	04.05.	14.05.	15.05.	18.06.	-	3
WN29	14.05.	15.05.	-	-	-	1
WN33	25.05.	12.07.	-	-	-	1
WN37	30.05.	13.06.	-	-	-	1
WN41	02.06.	25.06.	-	-	-	1
WN42	06.06.	12.07.	03.09.	05.09.	17.09.	4
WN53	25.06.	05.07.	-	-	-	1
WN55	25.06.	05.07.	-	-	-	1
WN56	25.06.	28.06.	-	-	-	1
WN71	03.09.	05.09.	09.09.	-	-	2
Summe	11	11	3	2	1	17

Tabelle 4.3: Datum und Anzahl der Wiederfänge im Jahr 2014.

Es ist dann

$$\begin{aligned}
 f_1 &= \sum_{i=1}^N I(Y_i = 1) = 59, \\
 f_2 &= \sum_{i=2}^N I(Y_i = 2) = 8, \\
 f_3 &= \sum_{i=3}^N I(Y_i = 3) = 1, \\
 f_4 &= \sum_{i=4}^N I(Y_i = 4) = 1, \\
 f_5 &= \sum_{i=5}^N I(Y_i = 5) = 1, \\
 f_6 &= \sum_{i=6}^N I(Y_i = 6) = 0, \\
 &\vdots \\
 S &= \sum_{j=1}^{31} f_j = 70.
 \end{aligned}$$

Diese Häufigkeiten werden nochmals übersichtlich in Tabelle 4.4 zusammengefasst.

	f_0	f_1	f_2	f_3	f_4	f_5
2014	-	59	8	1	1	1

Tabelle 4.4: Häufigkeiten f der Würfelnatter-Fänge im Jahr 2014.

Nun werden die Verhältnisse (4.7) für die Poisson-Verteilung, (4.8) für die Binomialverteilung und (4.9) für die geometrische Verteilung gegen x geplottet.

Abbildung 4.1 zeigt, dass die Frequenz von vier Wiederfängen (fünf Fängen) im Jahr 2014 größer ist, als unter einem homogenen Poisson-Modell erwartet. Da auch bereits die Frequenz von drei Wiederfängen etwas höher ist, ergibt sich ein monoton steigendes Muster für das Poisson-Modell. Laut Böhning et al. (2013, S.139) sollte der Ratio-Plot unter einer beliebigen Vermengung von Poisson-Parametern ein monoton steigendes Muster aufweisen (vergleiche Satz 4.1). Auch für die geometrische Verteilung ist eine leichte Monotonie im Plot erkennbar. Nur bei Annahme einer

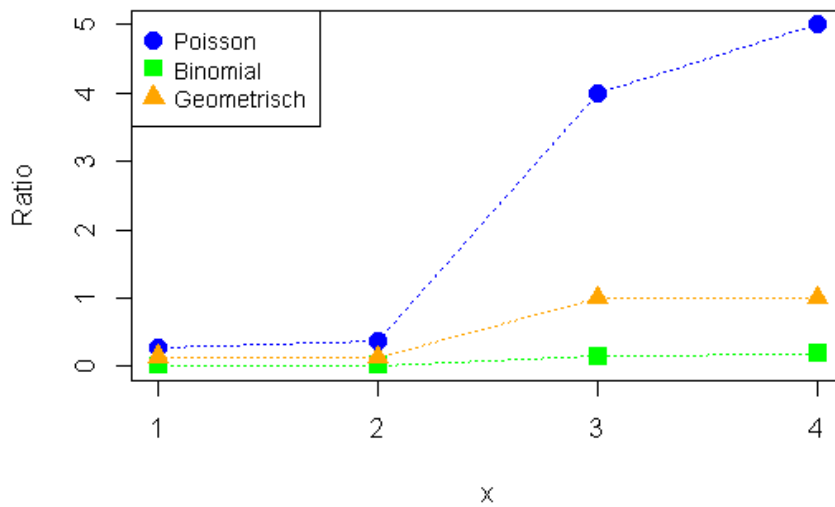


Abbildung 4.1: Ratio-Plot der Power Series Verteilungen für die Daten aus 2014.

Binomialverteilung ergeben sich keine Auffälligkeiten, sodass ein zugrunde liegendes homogenes binomiales Modell nicht verworfen werden kann.

Der Ratio-Plot der binomialen und geometrischen Verteilung zeigt also keine großen Abweichungen der Homogenitäts-Annahme. Da Erfassungswahrscheinlichkeiten in der Realität jedoch oft heterogen sind, werden im nachfolgenden Abschnitt Populationsumfangsschätzer aus den Power Series Verteilungen hergeleitet, welche die Annahme der homogenen Wahrscheinlichkeiten vernachlässigen.

Herleitung von Schätzern für die Populationsgröße

Mit Hilfe der Power Series Verteilungen können ganz einfach Populationsgrößen-schätzer für die unterschiedlichen Verteilungen hergeleitet werden. Die folgenden Abschnitte werden unterteilt in Chao-Schätzer und Zelterman-Schätzer.

5.1 Chao-Schätzer

Der Chao-Schätzer kann mit Heterogenität umgehen und wird im Allgemeinen als $\hat{N}_C = S + \hat{f}_0$ geschrieben. Kaskasamkul (2017, S.93) folgert aus dem Verhältnis (4.6) und aus dem Resultat von Satz 4.1, dass

$$\frac{\alpha_{x-1}}{\alpha_x} \frac{m_x}{m_{x-1}} \leq \frac{\alpha_x}{\alpha_{x+1}} \frac{m_{x+1}}{m_x}$$

gilt. Für $x = 1$ bedeutet das

$$\frac{\alpha_0}{\alpha_1} \frac{m_1}{m_0} \leq \frac{\alpha_1}{\alpha_2} \frac{m_2}{m_1}$$

beziehungsweise

$$m_0 \geq \frac{\alpha_0 \alpha_2}{\alpha_1^2} \frac{m_1^2}{m_2}. \quad (5.1)$$

Ersetzt man in (5.1) die theoretischen Größen m_x durch deren empirischen relativen Häufigkeiten $\frac{f_x}{N}$, so führt dies zur unteren Schranke der Chao-Schätzung für f_0 und N als

$$\hat{f}_0 = \frac{\alpha_0 \alpha_2 f_1^2}{\alpha_1^2 f_2} \quad \text{und} \quad \hat{N}_C = S + \frac{\alpha_0 \alpha_2 f_1^2}{\alpha_1^2 f_2}. \quad (5.2)$$

Hier können nun wieder je nach Verteilung die Parameter α_x aus Tabelle 4.1 eingesetzt werden und man erhält folgende Chao-Schätzer:

- **Poisson**

$$\begin{aligned} \hat{N}_{C_{Poi}} &= S + \frac{\alpha_0 \alpha_2 f_1^2}{\alpha_1^2 f_2} = S + \frac{1 \cdot \frac{1}{2} f_1^2}{1^2 f_2} \\ &= S + \frac{f_1^2}{2f_2}. \end{aligned} \quad (5.3)$$

- **Binomial**

$$\begin{aligned} \hat{N}_{C_{bino}} &= S + \frac{\alpha_0 \alpha_2 f_1^2}{\alpha_1^2 f_2} = S + \frac{1 \cdot \frac{t!}{2(t-2)!} f_1^2}{t^2 f_2} \\ &= S + \frac{(t-1) f_1^2}{t \cdot 2f_2}. \end{aligned} \quad (5.4)$$

- **Geometrisch**

$$\begin{aligned} \hat{N}_{C_{geom}} &= S + \frac{\alpha_0 \alpha_2 f_1^2}{\alpha_1^2 f_2} = S + \frac{1 \cdot 1 f_1^2}{1^2 f_2} \\ &= S + \frac{f_1^2}{f_2}. \end{aligned} \quad (5.5)$$

5.1.1 Biaskorrigierter Chao-Schätzer

Laut Kaskasamkul (2017, S.98) gilt für den biaskorrigierten Chao-Schätzer die Darstellung

$$\hat{N}_C^B = S + \frac{\alpha_0 \alpha_2 f_1 (f_1 - 1)}{\alpha_1^2 (f_2 + 1)}. \quad (5.6)$$

Für die einzelnen Verteilungen erhält man daher

- Poisson

$$\begin{aligned}\hat{N}_{C_{Poi}}^B &= S + \frac{\alpha_0 \alpha_2}{\alpha_1^2} \frac{f_1(f_1 - 1)}{(f_2 + 1)} \\ &= S + \frac{f_1(f_1 - 1)}{2(f_2 + 1)}.\end{aligned}\quad (5.7)$$

- Binomial

$$\begin{aligned}\hat{N}_{C_{bino}}^B &= S + \frac{\alpha_0 \alpha_2}{\alpha_1^2} \frac{f_1(f_1 - 1)}{(f_2 + 1)} \\ &= S + \frac{(t - 1)}{t} \frac{f_1(f_1 - 1)}{2(f_2 + 1)}.\end{aligned}\quad (5.8)$$

- Geometrisch

$$\begin{aligned}\hat{N}_{C_{geom}}^B &= S + \frac{\alpha_0 \alpha_2}{\alpha_1^2} \frac{f_1(f_1 - 1)}{(f_2 + 1)} \\ &= S + \frac{f_1(f_1 - 1)}{(f_2 + 1)}.\end{aligned}\quad (5.9)$$

5.2 Zelterman-Schätzer

Ein weiterer Schätzer, der mit Heterogenität umgehen kann, ist der Zelterman-Schätzer. Die Grundidee startet mit dem Horvitz-Thompson-Schätzer

$$\hat{N}_{HT} = \frac{S}{1 - \hat{p}_0(\theta)}, \quad (5.10)$$

wobei die Division durch $1 - p_0(\theta) = 1 - \frac{\alpha_0 \theta^0}{\eta(\theta)}$ eine Null-gestutzte Verteilung ergibt (Böhning et al., 2017, S.13). Je nach Verteilung erhält man $p_0(\theta)$ und damit auch \hat{N}_Z unterschiedlich aus Tabelle 4.1:

- Poisson

$$p_{0_{Poi}}(\theta) = \frac{\alpha_0 \theta^0}{\eta(\theta)} = \frac{1 \cdot 1}{\exp(\theta)} = \exp(-\theta) \quad \text{und} \quad (5.11)$$

$$\hat{N}_{Z_{Poi}} = \frac{S}{1 - \exp(-\hat{\theta})}. \quad (5.12)$$

- **Binomial**

$$p_{0_{bino}}(\theta) = \frac{\alpha_0 \theta^0}{\eta(\theta)} = \frac{1 \cdot 1}{(1 + \theta)^t} = (1 + \theta)^{-t} \quad \text{und} \quad (5.13)$$

$$\hat{N}_{Z_{bino}} = \frac{S}{1 - (1 + \hat{\theta})^{-t}}. \quad (5.14)$$

- **Geometrisch**

$$p_{0_{geom}}(\theta) = \frac{\alpha_0 \theta^0}{\eta(\theta)} = \frac{1 \cdot 1}{\frac{1}{1 - \theta}} = 1 - \theta \quad \text{und} \quad (5.15)$$

$$\hat{N}_{Z_{geom}} = \frac{S}{1 - (1 - \hat{\theta})} = \frac{S}{\hat{\theta}}. \quad (5.16)$$

Der fehlende Parameter θ kann mit Hilfe der Beziehung (4.5) geschätzt werden, indem jeweils $x = 1$ gesetzt wird:

- **Poisson**

$$\hat{\theta}_{Poi} = \frac{\alpha_1 f_2}{\alpha_2 f_1} = \frac{1 f_2}{\frac{1}{2} f_1} = \frac{2 f_2}{f_1} = \hat{\lambda}_{Poi}. \quad (5.17)$$

- **Binomial**

$$\hat{\theta}_{bino} = \frac{\alpha_1 f_2}{\alpha_2 f_1} = \frac{t f_2}{\frac{t(t-1)}{2} f_1} = \frac{2 f_2}{(t-1) f_1} = \frac{\hat{p}_{bino}}{1 - \hat{p}_{bino}}. \quad (5.18)$$

- **Geometrisch**

$$\hat{\theta}_{geom} = \frac{\alpha_1 f_2}{\alpha_2 f_1} = \frac{1 f_2}{1 f_1} = \frac{f_2}{f_1} = 1 - \hat{p}_{geom}. \quad (5.19)$$

Schließlich erhält man durch Einsetzen der Schätzer $\hat{\theta}$ in \hat{N}_Z die Schätzer nach Zelterman. Es resultieren

- **Poisson**

$$\hat{N}_{Z_{Poi}} = \frac{S}{1 - \exp\left(-\frac{2f_2}{f_1}\right)}. \quad (5.20)$$

- **Binomial**

$$\hat{N}_{Z_{bino}} = \frac{S}{1 - \left(1 + \frac{2f_2}{(t-1)f_1}\right)^{-t}}. \quad (5.21)$$

Da hier $\hat{\theta}_{bino} = \frac{\hat{p}_{bino}}{1 - \hat{p}_{bino}} = \frac{2f_2}{(t-1)f_1}$ gilt, kann $\hat{p}_{0_{bino}}(\theta)$ aus (5.13) noch umgeformt werden, sodass

$$\begin{aligned} \hat{p}_{0_{bino}}(\theta) &= (1 + \hat{\theta}_{bino})^{-t} = \left(1 + \frac{\hat{p}_{bino}}{1 - \hat{p}_{bino}}\right)^{-t} \\ &= \left(\frac{1}{1 - \hat{p}_{bino}}\right)^{-t} = (1 - \hat{p}_{bino})^t \end{aligned} \quad (5.22)$$

gilt. Der Schätzer (5.14) lautet somit

$$\hat{N}_{Z_{bino}} = \frac{S}{1 - (1 - \hat{p}_{bino})^t}, \quad (5.23)$$

wobei

$$\hat{p}_{bino} = \frac{\hat{\theta}_{bino}}{1 + \hat{\theta}_{bino}} = \frac{\frac{2f_2}{(t-1)f_1}}{1 + \frac{2f_2}{(t-1)f_1}} = \frac{2f_2}{(t-1)f_1 + 2f_2}. \quad (5.24)$$

Die Formeln (5.23) und (5.24) sind die gebräuchlicheren im Zusammenhang mit dem binomialen Zelterman-Schätzer. Sie tauchen so auch in diversen Literaturen auf, beispielsweise in Böhning (2010, S.223).

- **Geometrisch**

$$\hat{N}_{Z_{geom}} = \frac{S}{\frac{f_2}{f_1}} = \frac{f_1 S}{f_2}. \quad (5.25)$$

Da nun einige Schätzer für den Umfang einer Population hergeleitet wurden, gibt die nachfolgende Tabelle nochmals einen Überblick.

\hat{N}	Chao	Chao bc	Zelterman
Poisson	$S + \frac{f_1^2}{2f_2}$	$S + \frac{f_1(f_1-1)}{2(f_2+1)}$	$\frac{S}{1 - \exp\left(-\frac{2f_2}{f_1}\right)}$
Binomial	$S + \frac{(t-1)}{t} \frac{f_1^2}{2f_2}$	$S + \frac{(t-1)}{t} \frac{f_1(f_1-1)}{2(f_2+1)}$	$\frac{S}{1 - \left(1 - \frac{2f_2}{(t-1)f_1 + 2f_2}\right)^t}$
Geometrisch	$S + \frac{f_1^2}{f_2}$	$S + \frac{f_1(f_1-1)}{(f_2+1)}$	$\frac{f_1 S}{f_2}$

Tabelle 5.1: Populationsgrößenschätzer der Power Series Verteilungen.

Beispiel 5.1 (Fortsetzung von Beispiel 4.1). Es wird nun wieder ein Beispiel betrachtet, um die einzelnen Schätzer zu veranschaulichen. Die Häufigkeiten f_j und die Parameter t und S sind durch folgende Werte gegeben:

	f_0	f_1	f_2	f_3	f_4	f_5	t	S
2014	-	59	8	1	1	1	31	70

Damit ergeben sich die Schätzer aus Tabelle 5.1 zu:

\hat{N}	Chao	Chao bc	Zelterman
Poisson	287.6	260.1	294.7
Binomial	280.5	254.0	287.6
Geometrisch	505.1	450.2	516.3

Tabelle 5.2: Populationsgrößenschätzer der Power Series Verteilungen für die Daten aus 2014.

Es ist zu erkennen, dass die Schätzer unter einer Poisson- beziehungsweise Binomialverteilungsannahme sehr ähnlich sind. Die Annahme einer geometrischen Verteilung hingegen liefert Werte, die um zirka 200 Würfelgatter-Individuen größer sind. Chao-, biaskorrigierter Chao- und Zelterman-Schätzer liefern für jede Verteilungsannahme etwa gleich große Unterschiede. Aufgrund des vorher betrachteten Ratio-Plots (siehe Abbildung 4.1) scheint wohl das Ergebnis aus der Binomialverteilungsannahme das vertrauenswürdigste für einen homogenen Populationsgrößenschätzer zu sein.

Herleitung von Varianzen für die Populationsgrößenschätzer

Für die Herleitung der zugehörigen Varianz wird die in Böhning (2008, S.414) angeführte bedingte Varianzformel verwendet.

Lemma 6.1. *Seien X und Y zwei Zufallsvariablen. Dann gilt*

$$\text{Var}(X) = \mathbb{E}(\text{Var}(X|Y)) + \text{Var}(\mathbb{E}(X|Y)). \quad (6.1)$$

Dabei sind der Erwartungswert und die Varianz der Zufallsvariable X bedingt auf Y . Angewendet auf \hat{N} folgt demnach

$$\text{Var}(\hat{N}) = \mathbb{E}(\text{Var}(\hat{N}|S)) + \text{Var}(\mathbb{E}(\hat{N}|S)),$$

wobei $S \sim \text{Binomial}(N, (1 - p_0(\theta)))$.

6.1 Chao-Schätzer

Um eine Formel für die Varianz von \hat{N}_C anzugeben, setzt man in (6.1) laut Böhning (2008, S.417) für $X = \hat{N}_C = S + \hat{f}_0$ und $Y = S$. Sei $h(\hat{f}_0, S)$ die gemeinsame Dichtefunktion von \hat{f}_0 und S , die geschrieben werden kann als $h(\hat{f}_0, S) = h(\hat{f}_0|S)b(S)$, das Produkt der bedingten Wahrscheinlichkeitsfunktion von \hat{f}_0 gegeben S und der

binomialen Randwahrscheinlichkeitsfunktion $b(S)$ der Zufallsgröße S . Mit dem marginalen Erwartungswert, $\mathbb{E}(S + \hat{f}_0) = \int_1^N \int_{\mathbb{R} \setminus \{0\}} [S + \hat{f}_0] h(\hat{f}_0|S) b(S) d\hat{f}_0 dS = \sum_{S=1}^N (\int_{\mathbb{R} \setminus \{0\}} [S + \hat{f}_0] h(\hat{f}_0|S) d\hat{f}_0) b(S)$, und dem bedingten Erwartungswert, $\mathbb{E}(S + \hat{f}_0|S) = \int_{\mathbb{R} \setminus \{0\}} [S + \hat{f}_0] h(\hat{f}_0|S) d\hat{f}_0$, von \hat{N}_C gilt dann

$$\begin{aligned} \text{Var} \left(S + \hat{f}_0 \right) &= \underbrace{\sum_{S=1}^N \text{Var} \left(S + \hat{f}_0 \mid S \right) b(S)}_{=\mathbb{E}(\text{Var}(X|Y))} \\ &\quad + \underbrace{\sum_{S=1}^N \left\{ \mathbb{E} \left(S + \hat{f}_0 \mid S \right) - \mathbb{E} \left(S + \hat{f}_0 \right) \right\}^2 b(S)}_{=\text{Var}(\mathbb{E}(X|Y))} \\ &= \underbrace{\mathbb{E}_S \left(\text{Var} \left(S + \hat{f}_0 \mid S \right) \right)}_{\textcircled{1}} + \underbrace{\text{Var}_S \left(\mathbb{E} \left(S + \hat{f}_0 \mid S \right) \right)}_{\textcircled{2}}, \end{aligned} \quad (6.2)$$

wobei \mathbb{E}_S und Var_S den Erwartungswert und die Varianz der binomialen Verteilung von S kennzeichnen.

Die Terme werden nun einzeln vereinfacht.

② $\text{Var}_S \left(\mathbb{E} \left(S + \hat{f}_0 \mid S \right) \right)$:

Um die Varianz von $S + \hat{f}_0$ zu berechnen, wird zuerst der zweite Term in (6.2) betrachtet. Der Erwartungswert kann umgeformt werden zu (siehe Böhning, 2018)

$$\begin{aligned} \mathbb{E} \left(S + \hat{f}_0 \mid S \right) &= S + \mathbb{E} \left(\hat{f}_0 \mid S \right) = S + S \frac{p_0(\theta)}{1 - p_0(\theta)} \\ &= \frac{S}{1 - p_0(\theta)}. \end{aligned} \quad (6.3)$$

Mit der Varianz einer binomialverteilten Zufallsgröße, $S \sim \text{Binomial}(N, (1 - p_0(\theta)))$, ergibt sich

$$\begin{aligned} \text{Var}_S \left(\frac{S}{1 - p_0(\theta)} \right) &= \frac{1}{(1 - p_0(\theta))^2} \text{Var}_S(S) = \frac{1}{(1 - p_0(\theta))^2} N p_0(\theta) (1 - p_0(\theta)) \\ &= \frac{N p_0(\theta)}{1 - p_0(\theta)}. \end{aligned} \quad (6.4)$$

Da für den Erwartungswert von f_0 gilt, dass $\mathbb{E}(f_0) = N p_0(\theta)$ und $1 - p_0(\theta) = 1 - \frac{\mathbb{E}(f_0)}{N}$

(siehe Böhning, 2008, S.417), kann $\text{Var}_S\left(\frac{S}{1-p_0(\theta)}\right)$ geschätzt werden durch

$$\begin{aligned}\widehat{\text{Var}}_S\left(\frac{S}{1-p_0(\hat{\theta})}\right) &= \frac{\hat{N}_C p_0(\hat{\theta})}{1-p_0(\hat{\theta})} = \frac{\hat{f}_0}{1-\frac{\hat{f}_0}{S+\hat{f}_0}} = \frac{\hat{f}_0(S+\hat{f}_0)}{S} \\ &= \hat{f}_0 + \frac{\hat{f}_0^2}{S},\end{aligned}\tag{6.5}$$

mit $p_0(\hat{\theta}) = \frac{\hat{f}_0}{\hat{N}_C}$. Setzt man hier nun noch \hat{f}_0 aus (5.2) ein, so gilt für die geschätzte Varianz

$$\widehat{\text{Var}}_S\left(\frac{S}{1-p_0(\hat{\theta})}\right) = \frac{\alpha_0\alpha_2}{\alpha_1^2} \frac{f_1^2}{f_2} + \left(\frac{\alpha_0\alpha_2}{\alpha_1^2}\right)^2 \frac{f_1^4}{f_2^2 S}.\tag{6.6}$$

Anmerkung: Multipliziert man (6.5) mit $p_0^2(\hat{\theta})$, dann erhält man

$$\frac{\hat{f}_0(S+\hat{f}_0)p_0^2(\hat{\theta})}{S} = \frac{\hat{f}_0(S+\hat{f}_0)\frac{\hat{f}_0^2}{(S+\hat{f}_0)^2}}{S} = \frac{\hat{f}_0^3}{S(S+\hat{f}_0)}$$

und dies entspricht dem zweiten Teil der geschätzten Varianz von \hat{f}_0 in Böhning et al. (2017, S.390, (26.9)). Somit gilt laut Böhning et al. (2017, S.389, (26.3)) für Teil zwei der Varianz das Verhältnis

$$\widehat{\text{Var}}_S(\hat{f}_0) = p_0^2(\hat{\theta})\widehat{\text{Var}}_S(\hat{N}),$$

wobei $\hat{N} = \frac{S}{1-p_0(\hat{\theta})}$.

① $\mathbb{E}_S\left(\text{Var}\left(S+\hat{f}_0\mid S\right)\right)$:

Im ersten Teil von (6.2) wird zuerst der Erwartungswert $\mathbb{E}_S\left(\text{Var}\left(S+\hat{f}_0\mid S\right)\right)$ durch $\text{Var}\left(S+\hat{f}_0\mid S\right)$ geschätzt. Für diese Varianz gilt

$$\text{Var}(S+\hat{f}_0\mid S) = \text{Var}(\hat{f}_0\mid S).\tag{6.7}$$

Das kann mittels bivariater Δ -Methode (siehe Anhang A.2) approximiert werden

$$\widehat{\text{Var}}\left(\hat{f}_0\mid S\right) = \widehat{\text{Var}}\left(\frac{\alpha_0\alpha_2}{\alpha_1^2} \frac{f_1^2}{f_2}\mid S\right)$$

$$\approx \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \cdot \nabla g(f_1, f_2)^T \mathbf{Cov} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \nabla g(f_1, f_2), \quad (6.8)$$

wobei $g(x, y) = \frac{x^2}{y}$. Der klassische Schätzer der Kovarianz-Matrix im multinomialen Fall liefert die Kovarianz-Matrix für $\begin{pmatrix} f_1 \\ f_2 \end{pmatrix}$ laut Böhning (2008, S.416) als

$$\widehat{\mathbf{Cov}} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} = \begin{pmatrix} f_1(1 - \frac{f_1}{S}) & -\frac{f_1 f_2}{S} \\ -\frac{f_1 f_2}{S} & f_2(1 - \frac{f_2}{S}) \end{pmatrix}. \quad (6.9)$$

Eingesetzt in (6.8) folgt durch Matrixmultiplikation

$$\begin{aligned} \widehat{\mathbb{V}ar} \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \frac{f_1^2}{f_2} \middle| S \right) &\approx \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \cdot \nabla g(f_1, f_2)^T \mathbf{Cov} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \nabla g(f_1, f_2) \\ &= \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \begin{pmatrix} \frac{2f_1}{f_2} \\ -\frac{f_1^2}{f_2^2} \end{pmatrix}^T \begin{pmatrix} f_1(1 - \frac{f_1}{S}) & -\frac{f_1 f_2}{S} \\ -\frac{f_1 f_2}{S} & f_2(1 - \frac{f_2}{S}) \end{pmatrix} \begin{pmatrix} \frac{2f_1}{f_2} \\ -\frac{f_1^2}{f_2^2} \end{pmatrix} \\ &= \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \begin{pmatrix} \frac{2f_1^2}{f_2} (1 - \frac{f_1}{S}) + \frac{f_1^3}{f_2 S} \\ -\frac{2f_1^2}{S} - \frac{f_1^2}{f_2} (1 - \frac{f_2}{S}) \end{pmatrix}^T \begin{pmatrix} \frac{2f_1}{f_2} \\ -\frac{f_1^2}{f_2^2} \end{pmatrix} \\ &= \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \left[\frac{2f_1}{f_2} \left(\frac{2f_1^2}{f_2} \left(1 - \frac{f_1}{S} \right) + \frac{f_1^3}{f_2 S} \right) \right. \\ &\quad \left. - \frac{f_1^2}{f_2^2} \left(-\frac{2f_1^2}{S} - \frac{f_1^2}{f_2} \left(1 - \frac{f_2}{S} \right) \right) \right] \\ &= \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \left[\frac{f_1^3}{f_2^2} \left(4 \left(1 - \frac{f_1}{S} \right) + \frac{2f_1}{S} \right) \right. \\ &\quad \left. - \frac{f_1^3}{f_2^2} \left(-\frac{2f_1}{S} - \frac{f_1}{f_2} \left(1 - \frac{f_2}{S} \right) \right) \right] \\ &= \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \frac{f_1^3}{f_2^2} \left(4 - \frac{4f_1}{S} + \frac{2f_1}{S} + \frac{2f_1}{S} + \frac{f_1}{f_2} - \frac{f_1}{S} \right) \\ &= \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \left(\frac{4f_1^3}{f_2^2} + \frac{f_1^4}{f_2^3} - \frac{f_1^4}{f_2^2 S} \right). \quad (6.10) \end{aligned}$$

Addition von (6.10) und (6.6) liefert die Summe von ① und ② als

$$\begin{aligned} \widehat{\mathbb{V}ar}(\hat{N}_C) &= \mathbb{V}ar(\hat{f}_0 | S) + \hat{f}_0 + \frac{\hat{f}_0^2}{S} \\ &= \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \left(\frac{4f_1^3}{f_2^2} + \frac{f_1^4}{f_2^3} - \frac{f_1^4}{f_2^2 S} \right) + \frac{\alpha_0 \alpha_2}{\alpha_1^2} \frac{f_1^2}{f_2} + \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \frac{f_1^4}{f_2^2 S} \end{aligned}$$

$$= \frac{\alpha_0 \alpha_2}{\alpha_1^2} \frac{f_1^2}{f_2} + \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \left(\frac{4f_1^3}{f_2^2} + \frac{f_1^4}{f_2^3} \right). \quad (6.11)$$

Werden die Parameter α_x aus Tabelle 4.1 für die unterschiedlichen Verteilungen in (6.11) eingesetzt, so erhält man folgende Varianzen:

- **Poisson**

$$\begin{aligned} \widehat{\text{Var}}(\hat{N}_{C_{Poi}}) &= \frac{\alpha_0 \alpha_2}{\alpha_1^2} \frac{f_1^2}{f_2} + \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \left(\frac{4f_1^3}{f_2^2} + \frac{f_1^4}{f_2^3} \right) \\ &= \frac{1 \cdot \frac{1}{2}}{1^2} \frac{f_1^2}{f_2} + \left(\frac{1 \cdot \frac{1}{2}}{1^2} \right)^2 \left(\frac{4f_1^3}{f_2^2} + \frac{f_1^4}{f_2^3} \right) \\ &= \frac{1}{2} \frac{f_1^2}{f_2} + \frac{1}{4} \left(\frac{4f_1^3}{f_2^2} + \frac{f_1^4}{f_2^3} \right) \\ &= \frac{f_1^2}{2f_2} + \frac{f_1^3}{f_2^2} + \frac{f_1^4}{4f_2^3}. \end{aligned} \quad (6.12)$$

Dieses Ergebnis in (6.12) stimmt mit jenem in Chao (1987, S.786) überein.

- **Binomial**

$$\begin{aligned} \widehat{\text{Var}}(\hat{N}_{C_{bino}}) &= \frac{\alpha_0 \alpha_2}{\alpha_1^2} \frac{f_1^2}{f_2} + \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \left(\frac{4f_1^3}{f_2^2} + \frac{f_1^4}{f_2^3} \right) \\ &= \frac{1 \cdot \frac{t!}{2(t-2)!}}{t^2} \frac{f_1^2}{f_2} + \left(\frac{1 \cdot \frac{t!}{2(t-2)!}}{t^2} \right)^2 \left(\frac{4f_1^3}{f_2^2} + \frac{f_1^4}{f_2^3} \right) \\ &= \frac{(t-1)}{2t} \frac{f_1^2}{f_2} + \frac{(t-1)^2}{4t^2} \left(\frac{4f_1^3}{f_2^2} + \frac{f_1^4}{f_2^3} \right) \\ &= \frac{(t-1)}{t} \left(\frac{f_1^2}{2f_2} + \frac{(t-1)f_1^3}{tf_2^2} + \frac{(t-1)f_1^4}{4tf_2^3} \right). \end{aligned} \quad (6.13)$$

- **Geometrisch**

$$\begin{aligned} \widehat{\text{Var}}(\hat{N}_{C_{geom}}) &= \frac{\alpha_0 \alpha_2}{\alpha_1^2} \frac{f_1^2}{f_2} + \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \left(\frac{4f_1^3}{f_2^2} + \frac{f_1^4}{f_2^3} \right) \\ &= \frac{1 \cdot 1}{1^2} \frac{f_1^2}{f_2} + \left(\frac{1 \cdot 1}{1^2} \right)^2 \left(\frac{4f_1^3}{f_2^2} + \frac{f_1^4}{f_2^3} \right) \\ &= \frac{f_1^2}{f_2} + \frac{4f_1^3}{f_2^2} + \frac{f_1^4}{f_2^3}. \end{aligned} \quad (6.14)$$

Auch Kaskasamkul (2017, S.26, (2.43)) erhält dasselbe Ergebnis (6.14) für die Varianz des Chao-Schätzers unter einem geometrischen Modell.

6.1.1 Biaskorrigierter Chao-Schätzer

Der biaskorrigierte Chao-Schätzer kann analog zum Chao-Schätzer in zwei Teilen hergeleitet werden.

$$\textcircled{2} \text{ Var}_S \left(\mathbb{E} \left(S + \hat{f}_0 \mid S \right) \right):$$

Der zweite Term ist äquivalent zu (6.5) und lautet somit

$$\widehat{\text{Var}}_S \left(\frac{S}{1 - p_0(\hat{\theta})} \right) = \hat{f}_0 + \frac{\hat{f}_0^2}{S}. \quad (6.15)$$

$$\textcircled{1} \text{ E}_S \left(\text{Var} \left(S + \hat{f}_0 \mid S \right) \right):$$

Für die bedingte Varianz von $\hat{f}_0 = \frac{\alpha_0 \alpha_2}{\alpha_1^2} \frac{f_1(f_1-1)}{(f_2+1)}$ gilt

$$\begin{aligned} \widehat{\text{Var}} \left(\hat{f}_0 \mid S \right) &= \widehat{\text{Var}} \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \frac{f_1(f_1-1)}{(f_2+1)} \mid S \right) \\ &\approx \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \cdot \nabla g(f_1, f_2)^T \mathbf{Cov} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \nabla g(f_1, f_2), \end{aligned} \quad (6.16)$$

wobei $g(x, y) = \frac{x(x-1)}{(y+1)}$. Durch Matrixmultiplikation erhält man

$$\begin{aligned} \widehat{\text{Var}} \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \frac{f_1(f_1-1)}{(f_2+1)} \mid S \right) &\approx \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \cdot \nabla g(f_1, f_2)^T \mathbf{Cov} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \nabla g(f_1, f_2) \\ &= \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \begin{pmatrix} \frac{2f_1-1}{f_2+1} \\ -\frac{(f_1-1)f_1}{(f_2+1)^2} \end{pmatrix}^T \begin{pmatrix} f_1(1 - \frac{f_1}{S}) & -\frac{f_1 f_2}{S} \\ -\frac{f_1 f_2}{S} & f_2(1 - \frac{f_2}{S}) \end{pmatrix} \begin{pmatrix} \frac{2f_1-1}{f_2+1} \\ -\frac{(f_1-1)f_1}{(f_2+1)^2} \end{pmatrix} \\ &= \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \begin{pmatrix} \frac{2f_1^2 - f_1}{f_2+1} \left(1 - \frac{f_1}{S}\right) + \frac{(f_1-1)f_1^2 f_2}{(f_2+1)^2 S} \\ -\frac{2f_1^2 f_2 - f_1 f_2}{(f_2+1)S} - \frac{(f_1-1)f_1 f_2}{(f_2+1)^2} \left(1 - \frac{f_2}{S}\right) \end{pmatrix}^T \begin{pmatrix} \frac{2f_1-1}{f_2+1} \\ -\frac{(f_1-1)f_1}{(f_2+1)^2} \end{pmatrix} \\ &= \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \begin{pmatrix} \frac{(2f_1^2 - f_1)(f_2+1)S(1 - \frac{f_1}{S}) + (f_1-1)f_1^2 f_2}{(f_2+1)^2 S} \\ -\frac{(2f_1^2 f_2 - f_1 f_2)(f_2+1) - (f_1-1)f_1 f_2 S(1 - \frac{f_2}{S})}{(f_2+1)^2 S} \end{pmatrix}^T \begin{pmatrix} \frac{2f_1-1}{f_2+1} \\ -\frac{(f_1-1)f_1}{(f_2+1)^2} \end{pmatrix} \\ &= \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \left(\frac{(2f_1^2 - f_1)(f_2+1)S(1 - \frac{f_1}{S})(2f_1-1) + (f_1-1)f_1^2 f_2(2f_1-1)}{(f_2+1)^3 S} \right. \\ &\quad \left. + \frac{(2f_1^2 f_2 - f_1 f_2)(f_2+1)(f_1-1)f_1 + (f_1-1)^2 f_1^2 f_2 S(1 - \frac{f_2}{S})}{(f_2+1)^4 S} \right) \end{aligned}$$

$$\begin{aligned}
 &= \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \left(\frac{f_1(2f_1 - 1)^2(f_2 + 1)S(1 - \frac{f_1}{S}) + f_1^2 f_2(2f_1 - 1)(f_1 - 1)}{(f_2 + 1)^3 S} \right. \\
 &\quad \left. + \frac{f_1^2 f_2(2f_1 - 1)(f_2 + 1)(f_1 - 1) + f_1^2 f_2 S(f_1 - 1)^2(1 - \frac{f_2}{S})}{(f_2 + 1)^4 S} \right) \\
 &= \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \left(\frac{f_1(2f_1 - 1)^2(f_2 + 1)S - f_1^2(2f_1 - 1)^2(f_2 + 1)}{(f_2 + 1)^3 S} \right. \\
 &\quad \left. + \frac{2f_1^2 f_2(2f_1 - 1)(f_1 - 1)}{(f_2 + 1)^3 S} + \frac{f_1^2 f_2 S(f_1 - 1)^2 - f_1^2 f_2^2(f_1 - 1)^2}{(f_2 + 1)^4 S} \right) \\
 &= \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \left(\frac{f_1(2f_1 - 1)^2}{(f_2 + 1)^2} - \frac{f_1^2(2f_1 - 1)^2}{(f_2 + 1)^2 S} \right. \\
 &\quad \left. + \frac{2f_1^2 f_2(2f_1 - 1)(f_1 - 1)}{(f_2 + 1)^3 S} + \frac{f_1^2 f_2(f_1 - 1)^2}{(f_2 + 1)^4} - \frac{f_1^2 f_2^2(f_1 - 1)^2}{(f_2 + 1)^4 S} \right) \\
 &= \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \left(f_1(2f_1 - 1) \left[\frac{(2f_1 - 1)}{(f_2 + 1)^2} - \frac{f_1(2f_1 - 1)}{(f_2 + 1)^2 S} + \frac{2f_1 f_2(f_1 - 1)}{(f_2 + 1)^3 S} \right] \right. \\
 &\quad \left. + f_1^2 f_2(f_1 - 1)^2 \left[\frac{1}{(f_2 + 1)^4} - \frac{f_2}{(f_2 + 1)^4 S} \right] \right). \tag{6.17}
 \end{aligned}$$

Bildet man die Summe aus (6.15) und (6.17), so lautet die geschätzte Varianz

$$\begin{aligned}
 \widehat{\text{Var}}(\hat{N}_C^B) &= \text{Var}(\hat{f}_0 | S) + \hat{f}_0 + \frac{\hat{f}_0^2}{S} \\
 &= \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \left(f_1(2f_1 - 1) \left[\frac{(2f_1 - 1)}{(f_2 + 1)^2} - \frac{f_1(2f_1 - 1)}{(f_2 + 1)^2 S} + \frac{2f_1 f_2(f_1 - 1)}{(f_2 + 1)^3 S} \right] \right. \\
 &\quad \left. + f_1^2 f_2(f_1 - 1)^2 \left[\frac{1}{(f_2 + 1)^4} - \frac{f_2}{(f_2 + 1)^4 S} \right] \right) \\
 &\quad + \frac{\alpha_0 \alpha_2}{\alpha_1^2} \frac{f_1(f_1 - 1)}{(f_2 + 1)} + \left(\frac{\alpha_0 \alpha_2}{\alpha_1^2} \right)^2 \frac{f_1^2(f_1 - 1)^2}{(f_2 + 1)^2 S}. \tag{6.18}
 \end{aligned}$$

Auch hier können die Parameter α_x eingesetzt werden. Die Varianzen des biaskorrigierten Chao-Schätzers lauten dann:

- **Poisson**

$$\begin{aligned}
 \widehat{\text{Var}}(\hat{N}_{C_{Poi}}^B) &= \frac{1}{4} \left(f_1(2f_1 - 1) \left[\frac{(2f_1 - 1)}{(f_2 + 1)^2} - \frac{f_1(2f_1 - 1)}{(f_2 + 1)^2 S} + \frac{2f_1 f_2(f_1 - 1)}{(f_2 + 1)^3 S} \right] \right. \\
 &\quad \left. + f_1^2 f_2(f_1 - 1)^2 \left[\frac{1}{(f_2 + 1)^4} - \frac{f_2}{(f_2 + 1)^4 S} \right] \right) \\
 &\quad + \frac{1}{2} \frac{f_1(f_1 - 1)}{(f_2 + 1)} + \frac{1}{4} \frac{f_1^2(f_1 - 1)^2}{(f_2 + 1)^2 S}. \tag{6.19}
 \end{aligned}$$

• **Binomial**

$$\begin{aligned} \widehat{\text{Var}}(\hat{N}_{C_{\text{bino}}}^B) &= \frac{(t-1)^2}{4t^2} \left(f_1(2f_1-1) \left[\frac{(2f_1-1)}{(f_2+1)^2} - \frac{f_1(2f_1-1)}{(f_2+1)^2 S} + \frac{2f_1 f_2 (f_1-1)}{(f_2+1)^3 S} \right] \right. \\ &\quad \left. + f_1^2 f_2 (f_1-1)^2 \left[\frac{1}{(f_2+1)^4} - \frac{f_2}{(f_2+1)^4 S} \right] \right) \quad (6.20) \\ &\quad + \frac{(t-1)}{2t} \frac{f_1(f_1-1)}{(f_2+1)} + \frac{(t-1)^2}{4t^2} \frac{f_1^2(f_1-1)^2}{(f_2+1)^2 S}. \end{aligned}$$

• **Geometrisch**

$$\begin{aligned} \widehat{\text{Var}}(\hat{N}_{C_{\text{geom}}}^B) &= \left(f_1(2f_1-1) \left[\frac{(2f_1-1)}{(f_2+1)^2} - \frac{f_1(2f_1-1)}{(f_2+1)^2 S} + \frac{2f_1 f_2 (f_1-1)}{(f_2+1)^3 S} \right] \right. \\ &\quad \left. + f_1^2 f_2 (f_1-1)^2 \left[\frac{1}{(f_2+1)^4} - \frac{f_2}{(f_2+1)^4 S} \right] \right) \quad (6.21) \\ &\quad + \frac{f_1(f_1-1)}{(f_2+1)} + \frac{f_1^2(f_1-1)^2}{(f_2+1)^2 S}. \end{aligned}$$

6.2 Zelterman-Schätzer

Ähnlich wie beim Chao-Schätzer setzt man für die Varianz des Zelterman-Schätzers laut Böhning (2008, S.414) $X = \hat{N}_Z = \frac{S}{q_0(\hat{\theta})}$ und $Y = S$ in Formel (6.1). Dabei wurde der Einfachheit halber $q_0(\hat{\theta}) := 1 - p_0(\hat{\theta})$ gesetzt. Sei $f(\hat{\theta}, S)$ die gemeinsame Dichtefunktion von $\hat{\theta}$ und S , die geschrieben werden kann als $f(\hat{\theta}, S) = f(\hat{\theta}|S)b(S)$, das Produkt der bedingten Dichtefunktion von $\hat{\theta}$ gegeben S und der binomialen Randwahrscheinlichkeitsfunktion $b(S)$ der Zufallsgröße S . Mit dem marginalen Erwartungswert, $\mathbb{E}\left(\frac{S}{q_0(\hat{\theta})}\right) = \sum_S \int_{\hat{\theta}} \frac{S}{q_0(\hat{\theta})} f(\hat{\theta}|S) d\hat{\theta} b(S)$, und dem bedingten Erwartungswert, $\mathbb{E}\left(\frac{S}{q_0(\hat{\theta})} | S\right) = \int_{\hat{\theta}} \frac{S}{q_0(\hat{\theta})} f(\hat{\theta}|S) d\hat{\theta}$, von \hat{N}_Z gilt dann

$$\begin{aligned} \text{Var} \left(\frac{S}{q_0(\hat{\theta})} \right) &= \sum_S \left\{ \int_{\hat{\theta}} \left(\frac{S}{q_0(\hat{\theta})} - \mathbb{E} \left(\frac{S}{q_0(\hat{\theta})} \right) \right)^2 f(\hat{\theta}|S) d\hat{\theta} \right\} b(S) \\ &= \underbrace{\sum_S \text{Var} \left(\frac{S}{q_0(\hat{\theta})} \middle| S \right) b(S)}_{=\mathbb{E}(\text{Var}(X|Y))} \\ &\quad + \underbrace{\sum_S \left(\mathbb{E} \left(\frac{S}{q_0(\hat{\theta})} \middle| S \right) - \mathbb{E} \left(\frac{S}{q_0(\hat{\theta})} \right) \right)^2}_{=\text{Var}(\mathbb{E}(X|Y))} b(S). \end{aligned}$$

Dies kann folgendermaßen zusammengefasst werden.

Satz 6.1.

$$\mathbb{V}\text{ar} \left(\frac{S}{q_0(\hat{\theta})} \right) = \underbrace{\mathbb{E}_S \left(\mathbb{V}\text{ar} \left(\frac{S}{q_0(\hat{\theta})} \middle| S \right) \right)}_{\textcircled{1}} + \underbrace{\mathbb{V}\text{ar}_S \left(\mathbb{E} \left(\frac{S}{q_0(\hat{\theta})} \middle| S \right) \right)}_{\textcircled{2}}, \quad (6.22)$$

wobei \mathbb{E}_S und $\mathbb{V}\text{ar}_S$ den Erwartungswert und die Varianz der binomialen Verteilung von S (marginalen Verteilung $b(S)$) kennzeichnen.

Auch hier werden die Terme einzeln vereinfacht.

$$\textcircled{1} \quad \mathbb{E}_S \left(\mathbb{V}\text{ar} \left(\frac{S}{q_0(\hat{\theta})} \middle| S \right) \right):$$

Sei θ ein Parametervektor. Es wird der Ausdruck $S^2 \mathbb{V}\text{ar} \left(\frac{1}{q_0(\hat{\theta})} \middle| S \right)$ betrachtet und in weiterer Folge die multivariate Δ -Methode verwendet, um zu erreichen, dass

$$\mathbb{V}\text{ar} \left(\frac{1}{q_0(\hat{\theta})} \middle| S \right) \approx \left(\frac{1}{q_0(\hat{\theta})^2} \right)^2 \nabla_{q_0(\theta)^T} \mathbf{Cov}(\hat{\theta}|S) \nabla_{q_0(\theta)}.$$

Wird nun $\mathbf{Cov}(\hat{\theta}|S)$ durch eine Schätzung $\widehat{\mathbf{Cov}}(\hat{\theta}|S)$ ersetzt, sowie $q_0(\theta)$ durch $q_0(\hat{\theta})$ und $\mathbb{E}(S)$ durch S geschätzt, dann kann $\mathbb{E}_S \left(\mathbb{V}\text{ar} \left(\frac{S}{q_0(\hat{\theta})} \middle| S \right) \right)$ endgültig geschätzt werden durch

$$\left(\frac{S}{q_0(\hat{\theta})^2} \right)^2 \nabla_{q_0(\hat{\theta})^T} \widehat{\mathbf{Cov}}(\hat{\theta}|S) \nabla_{q_0(\hat{\theta})}. \quad (6.23)$$

Wenn θ ein Skalar ist - wie in unserem Fall -, dann vereinfacht sich (6.23) zu

$$S^2 \left(\frac{q'_0(\hat{\theta})}{q_0(\hat{\theta})^2} \right)^2 \widehat{\mathbb{V}\text{ar}}(\hat{\theta}|S),$$

und somit folgt schlussendlich

$$\widehat{\mathbb{E}}_S \left(\mathbb{V}\text{ar} \left(\frac{S}{q_0(\hat{\theta})} \middle| S \right) \right) \approx S^2 \left(\frac{q'_0(\hat{\theta})}{q_0(\hat{\theta})^2} \right)^2 \widehat{\mathbb{V}\text{ar}}(\hat{\theta}|S). \quad (6.24)$$

In weiterer Folge kann $\widehat{\mathbb{V}\text{ar}}(\hat{\theta}|S)$ für $\hat{\theta} = \frac{\alpha_1 f_2}{\alpha_2 f_1}$ bestimmt werden. Dazu wird die bivariate Δ -Methode verwendet und man erhält

$$\widehat{\mathbb{V}\text{ar}}(\hat{\theta}|S) = \widehat{\mathbb{V}\text{ar}} \left(\frac{\alpha_1 f_2}{\alpha_2 f_1} \middle| S \right)$$

$$\approx \left(\frac{\alpha_1}{\alpha_2}\right)^2 \cdot \nabla g(f_1, f_2)^T \mathbf{Cov} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \nabla g(f_1, f_2), \quad (6.25)$$

wobei $g(x, y) = \frac{y}{x}$. Mit obiger Kovarianz-Matrix (6.9) und Matrixmultiplikation folgt

$$\begin{aligned} \widehat{\mathbb{V}\text{ar}} \left(\frac{\alpha_1 f_2}{\alpha_2 f_1} \middle| S \right) &\approx \left(\frac{\alpha_1}{\alpha_2}\right)^2 \cdot \nabla g(f_1, f_2)^T \mathbf{Cov} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} \nabla g(f_1, f_2) \\ &= \left(\frac{\alpha_1}{\alpha_2}\right)^2 \begin{pmatrix} -\frac{f_2}{f_1^2} \\ \frac{1}{f_1} \end{pmatrix}^T \begin{pmatrix} f_1(1 - \frac{f_1}{S}) & -\frac{f_1 f_2}{S} \\ -\frac{f_1 f_2}{S} & f_2(1 - \frac{f_2}{S}) \end{pmatrix} \begin{pmatrix} -\frac{f_2}{f_1^2} \\ \frac{1}{f_1} \end{pmatrix} \\ &= \left(\frac{\alpha_1}{\alpha_2}\right)^2 \begin{pmatrix} -\frac{f_2}{f_1} \left(1 - \frac{f_1}{S}\right) - \frac{f_2}{S} \\ \frac{f_2^2}{f_1 S} + \frac{f_2}{f_1} \left(1 - \frac{f_2}{S}\right) \end{pmatrix}^T \begin{pmatrix} -\frac{f_2}{f_1^2} \\ \frac{1}{f_1} \end{pmatrix} \\ &= \left(\frac{\alpha_1}{\alpha_2}\right)^2 \left[-\frac{f_2}{f_1^2} \left(-\frac{f_2}{f_1} \left(1 - \frac{f_1}{S}\right) - \frac{f_2}{S} \right) \right. \\ &\quad \left. + \frac{1}{f_1} \left(\frac{f_2^2}{f_1 S} + \frac{f_2}{f_1} \left(1 - \frac{f_2}{S}\right) \right) \right] \\ &= \left(\frac{\alpha_1}{\alpha_2}\right)^2 \left[\frac{f_2^2}{f_1^2} \left(\frac{1}{f_1} - \frac{1}{S} + \frac{1}{S} \right) + \frac{f_2}{f_1^2} \left(\frac{f_2}{S} + \frac{f_2}{f_2} - \frac{f_2}{S} \right) \right] \\ &= \left(\frac{\alpha_1}{\alpha_2}\right)^2 \frac{f_2^2}{f_1^2} \left(\frac{1}{f_1} + \frac{1}{f_2} \right) = \hat{\theta}^2 \left(\frac{1}{f_1} + \frac{1}{f_2} \right). \quad (6.26) \end{aligned}$$

② $\mathbb{V}\text{ar}_S \left(\mathbb{E} \left(\frac{S}{q_0(\hat{\theta})} \middle| S \right) \right)$:

Angenommen der Erwartungswert $\mathbb{E} \left(\frac{S}{q_0(\hat{\theta})} \middle| S \right)$ kann durch $\frac{S}{q_0(\hat{\theta})}$ approximiert werden (siehe Böhning, 2008, S.414), dann vereinfacht sich die Varianz zu

$$\begin{aligned} \mathbb{V}\text{ar}_S \left(\mathbb{E} \left(\frac{S}{q_0(\hat{\theta})} \middle| S \right) \right) &\approx \mathbb{V}\text{ar}_S \left(\frac{S}{q_0(\hat{\theta})} \right) \\ &\stackrel{(*)}{=} N \frac{q_0(\hat{\theta})(1 - q_0(\hat{\theta}))}{q_0(\hat{\theta})^2}. \end{aligned}$$

Dabei wird in (*) der konstante Term quadratisch vor die Varianz gezogen und die Varianzformel für binomialverteilte Zufallsvariablen mit Parameter N und $q_0(\hat{\theta})$ angewandt. Eine Schätzung der Varianz von diesem Term kann erreicht werden, indem θ durch $\hat{\theta}$ und $Nq_0(\theta) = \mathbb{E}(S)$ durch S ersetzt werden. Das führt zu

$$\widehat{\mathbb{V}\text{ar}}_S \left(\mathbb{E} \left(\frac{S}{q_0(\hat{\theta})} \middle| S \right) \right) \approx S \frac{(1 - q_0(\hat{\theta}))}{q_0(\hat{\theta})^2}. \quad (6.27)$$

Anmerkung: Die Approximation $\mathbb{E}\left(\frac{S}{q_0(\hat{\theta})} \middle| S\right) \approx \frac{S}{q_0(\theta)}$ kann durch die Δ -Methode, welche hier auf das erste Moment angewendet wird, gerechtfertigt werden und behauptet, dass der Erwartungswert der transformierten Zufallsvariable approximiert werden kann durch die Transformation des Erwartungswertes. Eine Berechnung sieht wie folgt aus:

$$\frac{1}{q_0(\hat{\theta})} \approx \frac{1}{q_0(\theta)} + (\hat{\theta} - \theta) \cdot \frac{1}{q_0(\theta)^2} \cdot q_0'(\theta).$$

Wird darauf $\mathbb{E}(\cdot | S)$ angewandt, dann ergibt sich

$$\mathbb{E}\left(\frac{1}{q_0(\hat{\theta})} \middle| S\right) \approx \frac{1}{q_0(\theta)} + 0,$$

da $\mathbb{E}(\hat{\theta}|S) - \theta = 0$, falls $\hat{\theta}|S$ unverzerrt ist für θ . Multiplikation mit S liefert obige Approximation.

Auch hier für den zweiten Teil der geschätzten Varianz des Zelterman-Schätzers kann das Ergebnis aus (6.27) mit $p_0^2(\theta)$ multipliziert werden, um das Resultat aus Böhning et al. (2017, S.390, (26.8)) zu erhalten.

Die Summe von ① und ② setzt sich zusammen aus (6.24), (6.26) und (6.27) und lautet somit

$$\begin{aligned} \widehat{\text{Var}}(\hat{N}_Z) &= S^2 \left(\frac{q_0'(\hat{\theta})}{q_0(\hat{\theta})^2} \right)^2 \cdot \hat{\theta}^2 \left(\frac{1}{f_1} + \frac{1}{f_2} \right) + S \frac{(1 - q_0(\hat{\theta}))}{q_0(\hat{\theta})^2} \\ &\stackrel{(\Delta)}{=} S^2 \left(\frac{-p_0'(\hat{\theta})}{(1 - p_0(\hat{\theta}))^2} \right)^2 \cdot \hat{\theta}^2 \left(\frac{1}{f_1} + \frac{1}{f_2} \right) + S \frac{p_0(\hat{\theta})}{(1 - p_0(\hat{\theta}))^2}, \end{aligned} \quad (6.28)$$

wobei in (Δ) die Substitution $q_0(\hat{\theta}) = 1 - p_0(\hat{\theta})$ rückeringesetzt wurde und $q_0'(\theta) = -p_0'(\theta)$ gilt. Das Minus hebt sich hier allerdings mit dem Quadrat auf. Die Varianz in (6.28) stimmt mit Böhning et al. (2017, S.113, (9.8)) überein, da $\nabla_{\hat{\theta}}^T(1 - p_0(\hat{\theta})) \mathbf{Cov}(\hat{\theta}) \nabla_{\theta}(1 - p_0(\hat{\theta})) = p_0'^2(\hat{\theta}) \widehat{\text{Var}}(\hat{\theta}|S)$, wenn θ ein Skalar ist.

Sei $h(\hat{\theta}) := \frac{p_0'(\hat{\theta})}{(1 - p_0(\hat{\theta}))^2}$, dann können je nach Verteilung die in Abschnitt 5.2 berechneten Größen $\hat{\theta}$ und $p_0(\hat{\theta})$ in (6.28) eingesetzt werden:

- **Poisson**

$$\widehat{\text{Var}}(\hat{N}_{Z_{Poi}}) = Sh(\hat{\theta}_{Z_{Poi}}) \left[1 + Sh(\hat{\theta}_{Z_{Poi}}) \cdot \hat{\theta}_{Z_{Poi}}^2 \left(\frac{1}{f_1} + \frac{1}{f_2} \right) \right], \quad (6.29)$$

mit

$$h(\hat{\theta}_{Z_{Poi}}) = \frac{\exp(-\hat{\theta}_{Z_{Poi}})}{(1 - \exp(-\hat{\theta}_{Z_{Poi}}))^2} \quad \text{und} \quad \hat{\theta}_{Z_{Poi}} = \frac{2f_2}{f_1}. \quad (6.30)$$

Böhning (2008, S.416, (24)) formuliert in Korollar 1 ebenfalls Ergebnis (6.29).

• **Binomial**

$$\begin{aligned} \widehat{\text{Var}}(\hat{N}_{Z_{bino}}) &= Sh(\hat{\theta}_{Z_{bino}}) \\ &\cdot \left[\frac{(1 + \hat{\theta}_{Z_{bino}})}{-t} + Sh(\hat{\theta}_{Z_{bino}}) \cdot \hat{\theta}_{Z_{bino}}^2 \left(\frac{1}{f_1} + \frac{1}{f_2} \right) \right], \end{aligned} \quad (6.31)$$

mit

$$h(\hat{\theta}_{Z_{bino}}) = \frac{-t(1 + \hat{\theta}_{Z_{bino}})^{-t-1}}{(1 - (1 + \hat{\theta}_{Z_{bino}})^{-t})^2} \quad \text{und} \quad \hat{\theta}_{Z_{bino}} = \frac{2f_2}{(t-1)f_1}. \quad (6.32)$$

• **Geometrisch**

$$\begin{aligned} \widehat{\text{Var}}(\hat{N}_{Z_{geom}}) &= Sh(\hat{\theta}_{Z_{geom}}) \\ &\cdot \left[-(1 - \hat{\theta}_{Z_{geom}}) + Sh(\hat{\theta}_{Z_{geom}}) \cdot \hat{\theta}_{Z_{geom}}^2 \left(\frac{1}{f_1} + \frac{1}{f_2} \right) \right] \\ &= S \frac{f_1(f_1 - f_2)}{f_2^2} + S^2 \left(\frac{f_1}{f_2^2} + \frac{f_1^2}{f_2^3} \right), \end{aligned} \quad (6.33)$$

mit

$$h(\hat{\theta}_{Z_{geom}}) = \frac{-1}{\hat{\theta}_{Z_{geom}}^2} \quad \text{und} \quad \hat{\theta}_{Z_{geom}} = \frac{f_2}{f_1}. \quad (6.34)$$

Selbiges Ergebnis (6.33) formuliert Kaskasamkul (2017, S.28, (2.53)).

Um nun die einzelnen Standardfehler (geschätzten Standardabweichungen der Schätzer) zu erhalten, wird aus den eben berechneten geschätzten Varianzen die Wurzel gezogen.

Beispiel 6.1 (Fortsetzung von Beispiel 4.1). Die folgende Tabelle 6.1 zeigt eine Anwendung der Standardfehler auf die Würfelnatter-Daten aus dem Jahr 2014.

$\sqrt{\widehat{\text{Var}}(\hat{N})}$	Chao	Chao bc	Zelterman
Poisson	96.7	78.2	101.4
Binomial	96.6	75.7	98.2
Geometrisch	192.2	155.1	202.8

Tabelle 6.1: Standardfehler der Populationsgrößenschätzer für die Daten aus 2014.

Die Erkenntnisse aus den Schätzern der Populationsgröße bestätigen sich hier. Wiederum sind die Standardfehler für die Poisson- und die Binomialverteilung sehr ähnlich, während jene der geometrischen Verteilung etwa doppelt so groß sind.

Damit die hergeleiteten Populationsgrößenschätzer inklusive Standardfehler der Würfelnatter-Daten aus dem Jahr 2014 visualisiert werden können, wurde ein R-Tool entwickelt. In diesem gibt es ein Beobachtungsfenster, das am ersten Tag der Begehungen beginnt. Somit ist die linke Grenze des Fensters fixiert und bleibt immer gleich. Die rechte Grenze des Fensters wird jeweils um so viele Tage vergrößert, bis der nächste Wiederfang stattgefunden hat. Für das Jahr 2014 bedeutet dies, dass das erste Fenster vom 14.04.2014 bis zum 26.04.2014 geht. Das nächste Fenster beinhaltet dann alle Begehungen vom 14.04.2014 bis zum 14.05.2014 und so weiter. Gezeichnet werden die Schätzungen immer am Ende des jeweiligen Beobachtungsfensters, also gibt es zu jedem Begehungstag mit Wiederfang einen Schätzwert inklusive Standardfehler - sofern diese Werte berechnet werden konnten. Auf der x-Achse werden die Begehungstage in schwarz und die Wiederfang-Tage in rot durch Striche markiert. Fehlende oder sich verändernde Daten (f_1 , f_2) können zu einem scheinbaren Rückgang der Population führen. Die beiden Abbildungen 6.1 und 6.2 zeigen Chao- und Zelterman-Schätzer inklusive Standardfehler für die Annahme einer Binomialverteilung.

Die jeweils letzte Schätzung im Plot entspricht beinahe den Werten aus den Tabellen 5.2 und 6.1. Es wurde lediglich ein Begehungstag des Jahres 2014 nicht in das Beobachtungsfenster aufgenommen, da am letzten Begehungstag, 21.09.2014, kein Wiederfang stattgefunden hat.

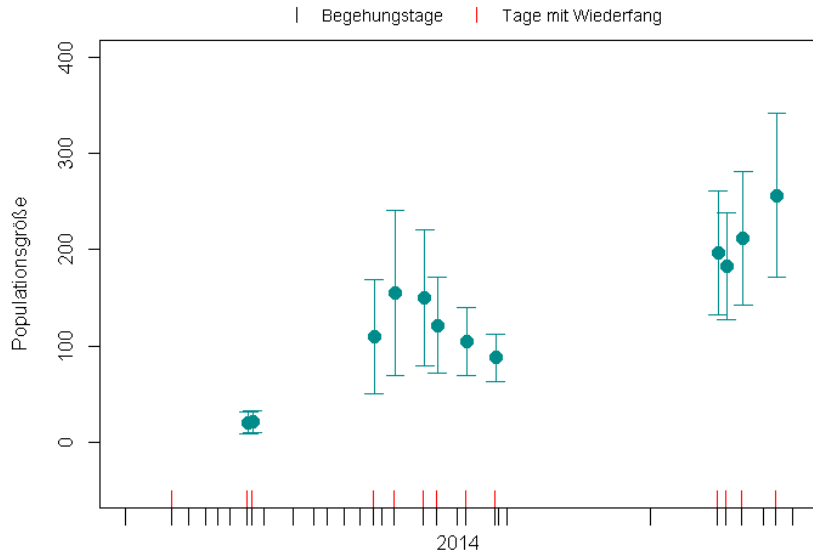


Abbildung 6.1: Visualisierung des Chao-Schätzers im Binomialfall inklusive Standardfehler für die Daten aus 2014.

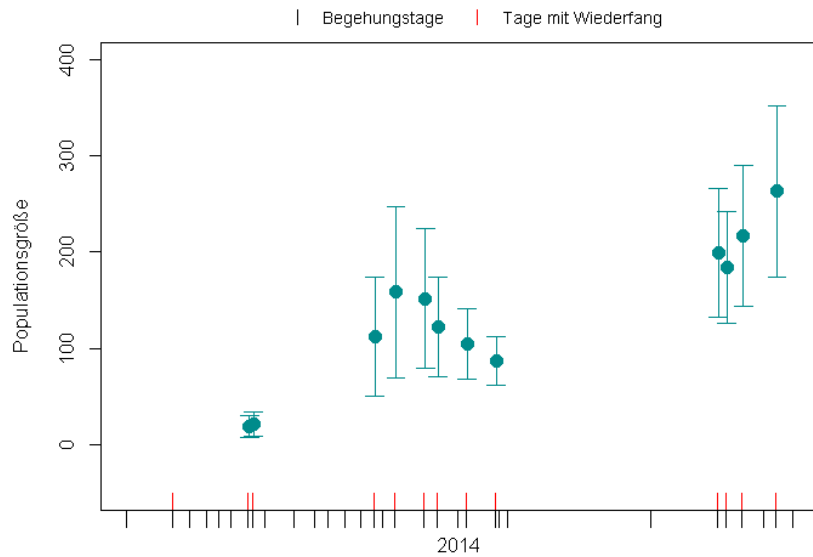


Abbildung 6.2: Visualisierung des Zelterman-Schätzers im Binomialfall inklusive Standardfehler für die Daten aus 2014.

Einführung in GLMs und in die Bayessche Statistik

Im späteren Verlauf dieser Arbeit werden *spatiale Capture-Recapture Modelle (SCR-Modelle)* behandelt, um die Würfelnatter-Daten auch unter Berücksichtigung des geografischen Aspekts zu untersuchen. Spatiale Capture-Recapture Modelle sind zum größten Teil nichts anderes als *generalisierte lineare Modelle (GLMs)*. Dabei ist die Kovariable - also der Abstand zwischen Versteckplatz oder Falle und Lebensraum-Mittelpunkt eines Individuums - nicht beobachtbar und wird daher als Zufallseffekt behandelt. Im allgemeinen Sprachgebrauch werden solche Modelle für gewöhnlich als *generalisierte lineare gemischte Modelle (GLMMs)* bezeichnet. Daher werden SCR-Modelle als eine Spezialform von GLMMs angesehen.

Mit Hilfe der *Bayesschen Statistik* können GL(M)Ms sehr gut analysiert werden, da direkt mit dem konditionalen Modell gearbeitet wird. Die Bayessche Statistik verwendet also ein Modell, das bedingt ist auf die Zufallseffekte, und verwendet Rechenmethoden, die als *Markov-Ketten-Monte-Carlo (MCMC)* bekannt sind. Als statistische Software für die Bayessche Statistik wird **WinBUGS** verwendet - eine Sprache, die auf *BUGS (Bayesian inference Using Gibbs Sampling)* basiert (Royle et al., 2014, S.47).

7.1 GLMs und GLMMs

Ein generalisiertes lineares Modell (GLM) ist eine Erweiterung von gewöhnlichen linearen Modellen, wobei hierbei die Response-Variable eine Verteilung aus der linearen Exponentialfamilie besitzt. Beispiele dazu sind die Normalverteilung sowie die Poisson-, die Binomial- oder die Gammaverteilung. Zusätzlich zu gewöhnlichen linearen Modellen steht die Response-Variable in GLMs mit den Prädiktor-Variablen (Kovariablen) in Zusammenhang, indem sie eine Link-Funktion verwenden, die meist nichtlinear ist. Ein GLM besteht somit aus den folgenden drei Komponenten:

1. Einer Wahrscheinlichkeitsverteilung aus der Exponentialfamilie für die abhängige (oder Response-) Variable y (*stochastische Komponente*).
2. Einem linearen Prädiktor $\eta = \beta_0 + \beta_1 x$, wobei x eine Prädiktor-Variable (Kovariable) ist (*systematische Komponente*).
3. Einer monotonen und zweimal stetig differenzierbaren Link-Funktion g , die den Erwartungswert von y , $\mathbb{E}(y)$, mit dem linearen Prädiktor, $\mathbb{E}(y) = \mu = g^{-1}(\eta)$, in Verbindung bringt. Daher gilt $g(\mathbb{E}(y)) = \eta = \beta_0 + \beta_1 x$.

In GLMs wird also angenommen, dass $g(\mathbb{E}(y))$ eine lineare Funktion der Prädiktor-Variablen - in diesem Fall x - mit unbekanntem und zu schätzenden Parametern β_0 und β_1 ist. Die Varianz von y ist dabei eine Funktion V des Erwartungswerts μ . Somit gilt $\text{Var}(y) = \phi V(\mu)$, wobei ϕ Dispersion genannt wird. Zwei bekannte Mitglieder aus der linearen Exponentialfamilie sind die Poisson- und die Binomialverteilung, die nun kurz betrachtet und später genauer erläutert werden. In beiden Fällen ist der Dispersionsparameter $\phi = 1$.

In einem Poisson GLM gilt beispielsweise $y_i \sim \text{Poisson}(\lambda_i)$ mit $\mathbb{E}(y_i) = \lambda_i$. Das Modell für den Erwartungswert wird beim Ziehen eines Stichprobenelements i mittels *Log-Link-Funktion*

$$\log(\lambda_i) = \beta_0 + \beta_1 x_i$$

festgelegt. Die Varianz-Funktion ist $V(y_i) = \lambda_i$.

Im binomialen GLM gilt $y_i \sim \text{Binomial}(K, p_i)$, wobei K der fixierte Parameter für die Stichprobengröße ist und $\mathbb{E}(y_i) = K \cdot p_i$. Das Modell für den Erwartungswert wird üblicherweise mittels *Logit-Link-Funktion*

$$\text{logit}(p_i) = \beta_0 + \beta_1 x_i$$

bestimmt, wobei $\text{logit}(p) = \log(p/(1-p)) = \eta$ gilt. Die inverse Logit-Funktion lautet somit $\text{logit}^{-1}(\eta) = \exp(\eta)/(1 + \exp(\eta)) = p$.

Wie in der Einleitung bereits kurz erwähnt, ist ein generalisiertes lineares gemischtes Modell (GLMM) die Erweiterung von GLMs um zufällige Effekte. Darunter versteht man das Hinzufügen eines normalen zufälligen Effekts zum linearen Prädiktor. Ein einfaches Beispiel ist die Verwendung eines zufälligen Intercepts α , womit sich das Poisson GLM ändern würde zu

$$\log(\lambda_i) = \alpha_i + \beta_1 x_i,$$

wobei

$$\alpha_i \sim \text{Normal}(\mu, \sigma^2)$$

gilt. GLMMs sind in ökologischen Modell-Anwendungen enorm nützlich. Sie können Veränderungen durch Versuchspersonen, Beobachter beziehungsweise spatialer oder zeitlicher Schichtung sowie durch Gruppierungen und Abhängigkeiten, die in jeder Gruppenstruktur auftreten, modellieren (McCullagh und Nelder, 1989; Royle et al., 2014, S.48f).

7.2 Die Bayessche Statistik

Das Hauptaugenmerk der Bayesschen Statistik ist die Verwendung einer direkten Wahrscheinlichkeit zur Charakterisierung der Ungewissheit über unbekannte Parameter der Modelle. Diese sogenannte a-priori-Wahrscheinlichkeit fasst Vorwissen und Annahmen des Beobachters in einer Wahrscheinlichkeitsverteilung zusammen. Die Bayessche Statistik nutzt die Regel von Bayes, um direkte Wahrscheinlichkeitsaussagen über Modellparameter zu machen. Bei zwei gegebenen Zufallsvariablen z und y verknüpft die Regel von Bayes die zwei konditionalen Wahrscheinlichkeitsverteilungen $[z|y]$ und $[y|z]$ durch das Verhältnis

$$[z|y] = \frac{[y|z][z]}{[y]}. \quad (7.1)$$

Dabei ist $[y|z]$ die konditionale Wahrscheinlichkeitsverteilung von y gegeben z , $[z]$ ist die marginale Verteilung von z und $[y]$ die marginale Verteilung von y . In Worten kann diese Regel formuliert werden als

$$\mathbb{P}(\text{präsent}|\text{nicht entdeckt}) = \frac{\mathbb{P}(\text{nicht entdeckt}|\text{präsent})\mathbb{P}(\text{präsent})}{\mathbb{P}(\text{nicht entdeckt})}$$

und mathematisch als

$$\mathbb{P}(z = 1|y = 0) = \frac{\mathbb{P}(y = 0|z = 1)\mathbb{P}(z = 1)}{\mathbb{P}(y = 0)}$$

$$= \frac{(1-p)^K \psi}{(1-p)^K \psi + (1-\psi)}. \quad (7.2)$$

Hier ist z eine binäre Zufallsvariable, die anzeigt, ob eine Spezies präsent ist ($z = 1$) oder nicht ($z = 0$), und es gilt $\Pr(z = 1) = \psi$, wobei ψ Eintrittswahrscheinlichkeit genannt wird. Weiters ist p die Wahrscheinlichkeit, dass eine Spezies in einer Untersuchung an einem Ort gefunden wird, wenn sie präsent ist, also $\Pr(y = 1|z = 1) = p$. Wird ein Ort K -mal beprobt, so können $y_k, k = 1, \dots, K$, als unabhängige und identisch verteilte (*iid*) Bernoulli-Versuche angesehen werden und die gesamte Anzahl an Entdeckungen, y , ist binomial verteilt mit Wahrscheinlichkeit p . Die Wahrscheinlichkeit, dass eine Spezies präsent ist, gegeben, dass diese in K Untersuchungen nicht entdeckt wurde, lautet dann $\Pr(z = 1|y_1 = 0, \dots, y_K = 0)$. Somit bezeichnet der Nenner in Formel (7.2) die Wahrscheinlichkeit, eine Spezies nicht zu entdecken, und setzt sich aus zwei Teilen zusammen: (1) die Spezies wird nicht entdeckt, gegeben sie ist präsent (Wahrscheinlichkeit $(1-p)^K \psi$), und (2) die Spezies ist nicht präsent (Wahrscheinlichkeit $1 - \psi$) (Royle et al., 2014, S.50f).

Im Kontext der statistischen Schlussfolgerung kann die Regel von Bayes folgendermaßen formuliert werden: Bezeichne y Beobachtungen und sei $[y|\theta]$ das Beobachtungsmodell (umgangssprachlich auch als *Likelihood* bezeichnet). Angenommen θ sei der gewünschte Parameter mit *a-priori*-Wahrscheinlichkeitsverteilung $[\theta]$ (oft auch nur als *Prior* bezeichnet). Die Kombination dieser beiden liefert die *a-posteriori*-Verteilung nach der Regel von Bayes als

$$[\theta|y] = \frac{[y|\theta][\theta]}{[y]}, \quad (7.3)$$

wobei $[y]$ ein Normierungsfaktor ist. Für alle statistischen Probleme sind die zwei Hauptcharakteristika der Bayesschen Statistik, dass: (1) Parameter θ als Realisierungen einer Zufallsvariable betrachtet werden, und (2) die Folgerung eine a-posteriori-Wahrscheinlichkeitsverteilung der Parameter θ bei gegebenen Daten y ist, also $[\theta|y]$ (Royle et al., 2014, S.52).

7.2.1 A-priori-Verteilungen

Die a-priori-Verteilung $[\theta]$ in (7.3) ist ein wichtiges Merkmal der Bayesschen Statistik. Sie charakterisiert vorherige Ansichten oder vorherige Informationen über einen Parameter, die sehr leicht in eine Analyse inkludiert werden können. Es gibt aber auch a-priori-Verteilungen, die zwar Informationen enthalten, aber keinen Einfluss auf die a-posteriori-Verteilung haben. Diese werden *nichtinformative* a-priori-Verteilungen genannt. Die daraus gefolgerte a-posteriori-Verteilung ist dann identisch mit der Likelihoodfunktion. Als Beispiel für eine nichtinformative Prior gilt $\text{Uniform}(0, 1)$ für

eine Wahrscheinlichkeit oder $\text{Uniform}(-\infty, \infty)$ für einen nicht beschränkten kontinuierlichen Parameter.

In SCR-Modellen gibt es meist einen Parameter, der mit der Größe des Lebensraums zusammenhängt. Hierbei kann die zusätzliche Information über die Lebensraumgröße einer Spezies als Prior-Information verwendet und damit in weiterer Folge die Parameter-Schätzung verbessert werden.

Manchmal kann es vorkommen, dass eine Prior unabsichtlich einen wesentlichen Effekt auf die Posterior eines Parameters ausübt, was nicht wünschenswert ist. Beispielsweise wird in Abschnitt 8.1.1 das Prinzip der Daten-Vergrößerung (*data augmentation*) verwendet, um die Populationsgröße N als unbekanntes Parameter zu behandeln. Dies ist äquivalent mit einer $\text{Binomial}(M, \psi)$ -Prior für N , wobei M eine ganze Zahl ist. Im Prinzip der Daten-Vergrößerung muss beachtet werden, dass M ausreichend groß ist, sodass die a-posteriori-Verteilung für N nicht beeinflusst wird. Ein weiterer wichtiger Punkt ist, dass sowohl a-priori-Verteilungen für die Transformation der Parameter als auch a-posteriori-Verteilungen *nicht* invariant sind. Somit kann es vorkommen, dass eine Prior scheinbar nichtinformativ auf der einen Skala ist, jedoch sehr informativ auf einer anderen Skala. Daher sollte immer zuerst der Einfluss der gewählten Prior beurteilt werden, denn häufig ist der Einfluss vernachlässigbar (Royle et al., 2014, S.53f). Die Wahl der a-priori-Verteilungen und Parametrisierungen ist sehr problemspezifisch und weitgehend subjektiv. Darüber hinaus beeinflusst sie das Verhalten des MCMC Algorithmus. Somit sollten idealerweise verschiedene Möglichkeiten ausprobiert werden (Royle et al., 2014, S.64).

7.2.2 A-posteriori-Folgerung

In der Bayesschen Statistik fokussiert man sich darauf, eine ganze Verteilung - die sogenannte a-posteriori-Verteilung - zu charakterisieren. Diese wird durch Vorwissen über die Verteilung des Parameters θ nach Beobachtung der Stichprobe bestimmt und ist die Grundlage zur Berechnung aller Punktschätzer und Konfidenzintervalle in der Bayesschen Statistik. Eine Punktschätzung könnte zum Beispiel der Posterior-Mittelwert, -Median oder -Mode sein (Royle et al., 2014, S.54).

7.3 A-posteriori-Verteilungen mittels MCMC-Simulation

In der Praxis ist die ständige Berechnung der marginalen Wahrscheinlichkeitsverteilung $[y]$ - der Nenner in (7.3) - nicht wirklich umsetzbar. Durch die Einführung der *Markov-Ketten-Monte-Carlo (MCMC)* Methoden wurde es einfacher, a-posteriori-Verteilungen für beinahe jedes Problem zu berechnen und zwar auf einem hinreichenden Präzisionsniveau. Vereinfacht gesagt ist MCMC eine Methode zum Ziehen einer Stichprobe aus der Ziel-Posterior-Verteilung (auch *stationäre* Verteilung genannt). Sie wird verwendet, um eine große Stichprobe aus der Posterior zu erhalten. Diese Stichprobe kann dann zur Charakterisierung von Eigenschaften der Posterior benutzt werden. Dabei muss man die a-posteriori-Verteilung nicht als eine benannte Verteilung kennen oder imstande sein, die Eigenschaften analytisch auszuwerten; das heißt, es müssen keine mathematischen Ausdrücke für Erwartungswert und Varianz entwickelt werden. Typischerweise gewinnt man aus der Stichprobe Mittelwert oder Median, kann diesen als Punktschätzung benutzen und ein Konfidenzintervall basierend auf Monte-Carlo Schätzungen der Quantile ermitteln (Royle et al., 2014, S.56f).

Bevor ein spezielles MCMC Verfahren betrachtet wird, folgt zuerst ein kurzer Überblick über Markov-Ketten:

Unter einer Markov-Kette versteht man eine Folge von Zufallsvariablen, die sich im Laufe der Zeit entwickelt. Die Kette besitzt Werte x_0, x_1, x_2, \dots aus einem Zustandsraum \mathcal{X} . Mit Hilfe einer Übergangswahrscheinlichkeit, die vom aktuellen Zustand der Kette abhängt und eine σ -Algebra \mathcal{B} verwendet, folgt schließlich die exakte Definition einer Markov-Kette:

Definition 7.1. Ein *Übergangskern* ist eine auf $\mathcal{X} \times \mathcal{B}(\mathcal{X})$ definierte Funktion K , sodass

- $\forall x \in \mathcal{X}, K(x, \cdot)$ ein Wahrscheinlichkeitsmaß ist und
- $\forall A \in \mathcal{B}(\mathcal{X}), K(\cdot, A)$ messbar ist.

Ist \mathcal{X} diskret, so spricht man von einer Übergangsmatrix K mit Elementen

$$\mathbb{P}_{xy} = \mathbb{P}(X_n = y | X_{n-1} = x), \quad x, y \in \mathcal{X}.$$

Definition 7.2. Bei gegebenem Übergangskern K nennt man eine Folge $X_0, X_1, \dots, X_n, \dots$ von Zufallsvariablen eine *Markov-Kette* (X_n) , wenn die bedingte Verteilung von X_t gegeben $x_{t-1}, x_{t-2}, \dots, x_0$ dieselbe für jedes t ist, wie die Verteilung von X_t

gegeben x_{t-1} . Somit gilt

$$\mathbb{P}(X_{k+1} \in A | x_0, x_1, x_2, \dots, x_k) = \mathbb{P}(X_{k+1} \in A | x_k) = \int_A K(x_k, dx). \quad (7.4)$$

Eine Markov-Kette lässt sich damit auch als $X_{n+1} \sim K(X_n, X_{n+1})$ schreiben. Man spricht von einer *homogenen* Markov-Kette, wenn die Verteilung von $(X_{t_1}, \dots, X_{t_k})$ gegeben x_{t_0} dieselbe ist wie die Verteilung von $(X_{t_1-t_0}, X_{t_2-t_0}, \dots, X_{t_k-t_0})$ gegeben x_0 für jedes k und jedes $t_0 \leq t_1 \leq \dots \leq t_k$. Damit sind die Übergangswahrscheinlichkeiten unabhängig vom Zeitpunkt t_k .

Zusammenfassend lässt sich also sagen, dass eine Markov-Kette ein stochastischer Prozess ist, der nur durch Kenntnis einer begrenzten Vorgeschichte des Prozesses gleich gute zukünftige Entwicklungen voraussagt wie beim Kennen der gesamten Vorgeschichte. Bei bekannter Anfangsverteilung oder bei bekanntem Anfangszustand beruht die Konstruktion einer Markov-Kette (X_n) nur auf ihrem Übergang, nämlich der bedingten Verteilung von X_n auf x_{n-1} (Robert und Casella, 1999, S.141f).

Die im MCMC Algorithmus betrachteten Ketten weisen eine *stationäre Wahrscheinlichkeitsverteilung* auf. Das heißt, es gibt eine Verteilung π , sodass für $X_n \sim \pi$ auch $X_{n+1} \sim \pi$ gilt, wenn der Kern K freie Bewegungen im gesamten Zustandsraum zulässt. Diese Freiheit wird *Irreduzibilität* genannt und man versteht darunter die Existenz von $n \in \mathbb{N}$, sodass $\mathbb{P}(X_n \in A | X_0) > 0$ für jedes A mit $\pi(A) > 0$ gilt. Die Wahrscheinlichkeit von X_n nach X_0 zu gelangen ist daher positiv. Diese Eigenschaft garantiert auch, dass die meisten Markov-Ketten in MCMC Algorithmen *rekurrent* sind, was bedeutet, dass die durchschnittliche Anzahl der Besuche einer beliebigen Menge A unendlich ist. Sie können aber auch *Harris-rekurrent* sein, das heißt, die Wahrscheinlichkeit einer unendlichen Anzahl von Rückläufen zu A ist Eins. Damit wird garantiert, dass die Kette für jeden Startwert dasselbe Grenzverhalten aufweist. Während eine Markov-Kette mit den gewünschten Eigenschaften leicht konstruiert werden kann, stellt sich das Herausfinden der benötigten Schritte, um Konvergenz zur stationären Verteilung zu erhalten, als schwieriger heraus (Robert und Casella, 1999, S.140).

Eine bekannte MCMC Methode ist das *Gibbs-Sampling*. Der Gibbs-Sampling Algorithmus erzeugt eine Folge von Stichproben der gemeinsamen Wahrscheinlichkeitsverteilung mehrerer Zufallsvariablen und beinhaltet iterative Simulation aus den vollen konditionalen Verteilungen (auch konditionale a-posteriori-Verteilungen genannt). Die volle konditionale Verteilung für eine unbekannte Anzahl ist die bedingte Verteilung dieser Anzahl, wenn jede andere Zufallsvariable - Daten und andere Parameter - im Modell gegeben ist (Royle et al., 2014, S.57).

Um den Algorithmus konkret zu formulieren, nimmt man an, dass für $p > 1$ die Zufallsvariable $\mathbf{X} \in \mathcal{X}$ als $\mathbf{X} = (X_1, \dots, X_p)$ geschrieben werden kann, wobei die

X_i 's entweder ein- oder mehrdimensional sind. Es kann nun

$$X_i | x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_p \sim f_i(x_i | x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_p)$$

für $i = 1, 2, \dots, p$ simuliert werden. Dabei sind f_1, \dots, f_p die bedingten univariaten Dichten. Der Gibbs-Sampling Algorithmus lautet demnach:

Bei gegebenem $\mathbf{x}^{(t)} = (x_1^{(t)}, \dots, x_p^{(t)})$ generiere

1. $X_1^{(t+1)} \sim f_1(x_1 | x_2^{(t)}, \dots, x_p^{(t)});$
2. $X_2^{(t+1)} \sim f_2(x_2 | x_1^{(t+1)}, x_3^{(t)}, \dots, x_p^{(t)});$
- \vdots
- p . $X_p^{(t+1)} \sim f_p(x_p | x_1^{(t+1)}, \dots, x_{p-1}^{(t+1)}).$

Die Dichten f_1, \dots, f_p bezeichnet man - wie bereits kurz erwähnt - als *volle Konditionale* (Robert und Casella, 1999, S.285f).

Konkret und unter Berücksichtigung der Schreibweise aus Formel (7.3) folgt dafür (siehe Royle et al., 2014, S.437):

Beginne mit einem Startwert θ_0 für θ . Für Iteration t

1. ziehe $\theta_1^{(t)}$ aus $[\theta_1^{(t)} | \theta_2^{(t-1)}, \dots, \theta_p^{(t-1)}];$
2. ziehe $\theta_2^{(t)}$ aus $[\theta_2^{(t)} | \theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_p^{(t-1)}];$
- \vdots
- p . ziehe $\theta_p^{(t)}$ aus $[\theta_p^{(t)} | \theta_1^{(t)}, \dots, \theta_{p-1}^{(t)}].$

Damit volle konditionale Verteilungen zum Entwickeln von MCMC Algorithmen konstruiert werden können, gibt es eine Strategie, die auf einige wenige Schritte zusammengefasst werden kann:

1. Identifiziere alle stochastischen Komponenten des Modells und sammle die Wahrscheinlichkeitsverteilungen.
2. Drücke die volle Konditionale proportional zum Produkt aller Wahrscheinlichkeitsverteilungen aus Schritt 1 aus.
3. Entferne diejenigen, die keinen zentralen Parameter beinhalten.
4. Verwende Algebra, um die resultierende Wahrscheinlichkeitsdichtefunktion (pdf) oder Wahrscheinlichkeitsfunktion (pmf) zu identifizieren.

Schritt vier ist der wichtigste und verlangt viel statistisches Geschick und Erfahrung, um die komplexen Terme zu vereinfachen. Wird statt dem Gibbs-Sampling Algorithmus der Metropolis-Hastings Algorithmus verwendet, so ist Schritt vier nicht notwendig (Royle et al., 2014, S.59).

7.4 Bayessche Statistik mit der BUGS-Sprache

Anstatt einen eigenen MCMC Algorithmus für jede Analyse zu entwickeln, können die frei zugänglichen Softwareprodukte **WinBUGS** oder **JAGS** verwendet werden. Diese werden in R mit Hilfe der Pakete **R2WinBUGS** beziehungsweise **R2jags** oder **rjags** ausgeführt. **WinBUGS** und **JAGS** verwenden eine Pseudocode-Bezeichnung von allen relevanten stochastischen und deterministischen Elementen eines Modells und erzeugen einen MCMC Algorithmus für dieses Modell. **WinBUGS/JAGS** durchläuft diesen Algorithmus und gibt den Markov-Ketten Output, also die Posterior-Stichproben der Modellparameter, aus.

Die **BUGS**-Sprache zwingt einen, jedes Element des Modells niederzuschreiben, verschiedene Annahmen ausdrücklich zuzulassen und die aktuellen Wahrscheinlichkeitsannahmen zu verstehen. Außerdem sollte man wissen, wie sich einerseits die Daten auf die latenten Variablen oder andererseits die Daten und latenten Variablen auf die Parameter beziehen. Weiters ist das Verhalten der Parameter untereinander interessant (Royle et al., 2014, S.60).

In einer BUGS-Modelldatei muss jede Variable entweder

- ein Datensatz sein, der als Input ins Modell kommt,
- oder eine Zufallsvariable, die eine Wahrscheinlichkeitsverteilung hat und mit dem Tilde-Zeichen „ \sim “ im Modell steht,
- oder ein abgeleiteter Parameter, der mit den verwendeten Variablen und Daten durch einen Zuordnungspfeil „ $<$ –“ verbunden wird.

Zur Modellanpassung werden verschiedene Datenobjekte für **WinBUGS** benötigt. Zuerst wird eine R-Liste namens **data** erzeugt, welche die in der **BUGS**-Modelldatei identifizierten Datenobjekte beinhaltet. Zudem muss eine R-Funktion **inits** erstellt werden, die eine Liste mit den Startwerten generiert und an **WinBUGS** gesendet wird. Grundsätzlich sind die Startwerte optional, dennoch wird empfohlen, sowohl für strukturelle Parameter als auch für zufällige Effekte immer sinnvolle Startwerte anzugeben. Schließlich müssen noch die Namen der Parameter bestimmt werden, die von **WinBUGS** als MCMC Output gespeichert werden sollen. **WinBUGS** wird schlussendlich mit dem R-Kommando **bugs** ausgeführt. Durch die Op-

tion `debug=TRUE` bleibt das **WinBUGS** GUI offen. Dies kann hilfreich sein, um den MCMC Output zu analysieren und einen Blick auf das Fehlerprotokoll (*error log*) zu werfen. Mit `working.dir=getwd()` werden die **WinBUGS** Output-Dateien und die Log-Datei im aktuellen R-Arbeitsverzeichnis gespeichert. Weiters sollte auch der Ort, an dem **WinBUGS** installiert wurde, durch `bugs.directory` angegeben werden. Neben diesen Objekten müssen zusätzlich die Anzahl der parallelen Markov-Ketten (`n.chains`), die Ausdünnungsrate (`n.thin`), die Anzahl der burn-in-Iterationen (`n.burnin`) und die totale Anzahl an Iterationen (`n.iter`) an die Funktion `bugs` übergeben werden. Der `print`-Befehl liefert eine grundlegende Zusammenfassung als Output. Darin helfen die beiden Spalten „Rhat“ und „n.eff“ als Konvergenz-Diagnosen und „DIC“ ist das Deviance-Informationskriterium (Royle et al., 2014, S.61ff).

7.5 Praktische Bayessche Statistik und MCMC

Markov-Ketten benötigen einige Zeit, bis sie zu ihrer stationären Verteilung - der Posterior-Verteilung $[\theta|y]$ - konvergieren. Die meisten MCMC Algorithmen garantieren nur, dass die generierten Stichproben aus der Ziel-Posterior-Verteilung stammen, die Dauer des MCMC Verfahrens ist jedoch unbekannt.

Typischerweise wird im frühen Teil des MCMC Algorithmus eine Periode der Vergänglichkeit beobachtet, welche als „burn-in“-Periode bezeichnet und verworfen wird. Die „R-hat“-Statistik überprüft die Konvergenz von mehreren Markov-Ketten. Liegt diese Statistik nahe bei 1, so haben die Markov-Ketten konvergiert und ausreichende Posterior-Stichproben wurden erlangt (Royle et al., 2014, S.65).

Manche Modelle weisen ein „schlechtes Mischverhalten“ oder „langsame Konvergenz“ der Markov-Ketten auf. Dabei stammen die Stichproben aus der a-posteriori-Verteilung, aber mischen oder bewegen sich nur langsam um die Posterior. Schlechtes Mischverhalten kann verschiedene Ursachen haben. Beispielsweise können Parameter stark korreliert sein, kaum aus den Daten identifiziert werden oder die Algorithmen sind schlecht. Langsames Mischverhalten entspricht einer hohen Autokorrelation in der Markov-Kette. Die aufeinanderfolgenden Ziehungen sind stark korreliert und daher muss der MCMC Algorithmus länger ausgeführt werden, um eine effektive Stichprobengröße zu bekommen, die hinreichend für eine Schätzung ist, oder um den Markov-Ketten Fehler auf ein tolerierbares Level zu reduzieren. Eine Möglichkeit, die Autokorrelation zu verkleinern, ist „Ausdünnung“. Dabei wird nur jeder m -te Wert des Markov-Ketten Outputs behalten. Allerdings kann Ausdünnung zwangsläufig ineffizient sein, da immer präzisere Posterior-Schätzungen möglich sind, wenn der gesamte MCMC Output unabhängig vom Level der Autokorrelation verwendet

wird.

Die nachfolgenden Stichproben, die aus der Markov-Kette generiert werden, sind aufgrund der vom Markov-Prozess eingeführten Korrelation unter Stichproben keine *unabhängigen* Proben der a-posteriori-Verteilung. Die Stichprobengröße muss angepasst werden, um die Autokorrelation in nachfolgenden Proben zu erklären. Diese angepasste Stichprobengröße wird als *effektive Stichprobengröße* bezeichnet. **WinBUGS** gibt automatisch die effektive Stichprobengröße („n.eff“) für alle beobachteten Parameter aus. Wenn die Markov-Kette nur eine sehr kleine effektive Stichprobe generiert, kann eine längere Laufzeit in Betracht gezogen werden (Royle et al., 2014, S.67).

7.6 Poisson GLMs

Das bereits kurz erwähnte *Poisson GLM* (auch bekannt als „Poisson Regression“) ist die wahrscheinlich relevanteste und wichtigste Modellklasse in der ganzen Ökologie. Das Basis-Modell betrachtet Beobachtungen y_i für $i = 1, 2, \dots, n$ einer Poisson-Verteilung mit Erwartungswert λ_i und kann als

$$y_i \sim \text{Poisson}(\lambda_i)$$

geschrieben werden. Dabei ist die Response y_i eine Anzahl an Tieren oder Pflanzen am Ort i und λ_i kann über die Orte variieren. Man ist daran interessiert, wie λ_i von Standortmerkmalen wie beispielsweise dem Lebensraum abhängt. Verfügbare Kovariablen werden typischerweise als linearer Effekt der logarithmierten Erwartung modelliert. Ist x_i eine gemessene und mit Beobachtung i in Verbindung gebrachte Kovariable, dann gilt

$$\log(\lambda_i) = \beta_0 + \beta_1 x_i.$$

Wie in allen GLMs wird hier nur die Erwartung für das Poisson-Modell direkt spezifiziert. Die Varianz ist unmittelbar mit dem Erwartungswert verbunden und lautet $\text{Var}(y_i) = \mathbb{E}(y_i) = \lambda_i$. Dadurch wächst die Varianz linear mit dem Erwartungswert (Royle et al., 2014, S.69).

7.7 Binomiale GLMs

Binomiale GLMs sind eine weitere sehr wichtige Klasse von Modellen in der Ökologie. Man verwendet binomiale Modelle für Häufigkeitsdaten. Dabei legt man die Stichprobengröße K als maximal mögliche Häufigkeit einer Stichprobe fest. Die Zufallsvariable y mit $0 \leq y \leq K$ bezeichnet dann die Häufigkeit des Eintretens aus K Versuchen. Der Parameter p in binomialen Modellen wird „Erfolgswahrscheinlichkeit“ genannt und tritt im Erwartungswert von y , $\mathbb{E}(y) = pK$, auf. Üblicherweise möchte man die Kovariablen modellieren, die den Parameter p betreffen. Solche Modelle werden binomiale GLMs, binomiale Regressionsmodelle oder logistische Regression genannt, wobei in letzterer der Logit-Link benutzt wird, um das Verhältnis zwischen p und den Kovariablen zu modellieren.

Ein bekanntes binomiales GLM ist jenes mit Stichprobengröße gleich Eins. Darin ist das Ergebnis y „Anwesenheit“ ($y = 1$) oder „Abwesenheit“ ($y = 0$) einer Spezies. In diesem Fall ist y aus einer Bernoulli-Verteilung und man spricht von einer klassischen Modellierungssituation für Spezies-Verteilungen. In Standardproblemen der binomialen Regression ist die Stichprobengröße fixiert. Ein Beispiel dafür sind Capture-Recapture Modelle einer geschlossenen Population. Dabei wird eine Population von Individuen K -mal beprobt. Jedes Auffinden eines Individuums ist ein binomiales Resultat mit Parameter (oder Begegnungswahrscheinlichkeit) p , basierend auf einer Stichprobe mit Umfang K . Des Weiteren ist die Gesamtanzahl an einzelnen beobachteten Individuen S ebenfalls eine binomiale Zufallsvariable, die auf der Populationsgröße N basiert (Royle et al., 2014, S.77f). Der folgende Abschnitt verwendet Bayessche Statistik, um Capture-Recapture Modelle einer geschlossenen Population mit Hilfe von **WinBUGS** zu analysieren, und veranschaulicht diese am Beispiel der Würfelnatter-Daten aus dem Jahr 2014.

Geschlossene Populationsmodelle

Als Einstieg in die Thematik der Capture-Recapture Modelle werden *nicht-spatiale* CR-Modelle zur Schätzung einer geschlossenen Populationsgröße betrachtet. Bei einer geschlossenen Population ändert sich deren Größe N während der gesamten Untersuchung nicht. Es gibt zwei Formen von Geschlossenheit: demografische Geschlossenheit, in der keine Geburten oder Todesfälle auftreten, und geografische Geschlossenheit, in der keine Individuen in die untersuchte Region ein- oder auswandern. Obwohl es in der Realität - bis auf kurze Zeitintervalle - nur sehr wenige Populationen gibt, die tatsächlich geschlossen sind, dienen CR-Modelle einer geschlossenen Population als Grundlage für die Entwicklung weiterer Modelle (Royle et al., 2014, S.87).

8.1 Ein einfaches geschlossenes Populationsmodell: Modell M_0

Im einfachsten geschlossenen Populationsmodell werden Individuen einer Population der Größe N in K Stichproben - beispielsweise Fang- oder Begehungstage - gefangen, markiert, freigelassen und wieder gefangen. Für jedes Individuum wird zu jeder Stichprobe eine Fanggeschichte aufgezeichnet, wobei ein Fang mit $y = 1$ sowie ein Nichtfang mit $y = 0$ vermerkt wird. Diese Beobachtungen werden zu einer Matrix

mit Einträgen y_{ik} für $i = 1, 2, \dots, N$ und $k = 1, 2, \dots, K$ zusammengefasst (vergleiche dazu die Matrix \mathbf{X} in Abschnitt 3.1; der Parameter t dort entspricht dem K hier). Die Einträge y_{ik} sind unabhängige und identisch verteilte (*iid*) Bernoulli-Zufallsvariablen und können als $y_{ik} \sim \text{Bernoulli}(p)$ geschrieben werden. In diesem einfachen Modell ist p konstant, das heißt es gibt keine individuellen oder zeitlichen Kovariablen, die p beeinflussen. Die binären Erfassungsgrößen y_{ik} können zur Gesamtanzahl der Fänge für ein Individuum, $y_i = \sum_k y_{ik}$, vereinigt werden, womit sich das Beobachtungsmodell von einer Bernoulli-Verteilung zu einer Binomialverteilung ändert, die auf der Stichprobengröße K basiert. Somit gilt für jedes Individuum $i = 1, 2, \dots, N$ in der Population

$$y_i = \sum_{k=1}^K y_{ik} \sim \text{Binomial}(K, p),$$

wobei N die Anzahl der Individuen in der Population ist (Royle et al., 2014, S.88).

Das Bernoulli-Basis-Modell, in dem ein Individuum in einer Stichprobe entweder gefangen oder nicht gefangen wird, ist der Grundstein von fast allen klassischen und auch einigen spatialen CR-Modellen und damit von zentraler Bedeutung. Es ist eine vereinfachte Version eines logistischen Regressionsmodells mit nur einem Intercept-Term ($\text{logit}(p) = \text{konstant}$). Zur leichteren Spezifizierung von Modellen wird zuerst N als bekannt angenommen, bevor der Fall für N unbekannt behandelt wird.

Wird angenommen, dass die Individuen in der Population unabhängig voneinander gefangen werden, so ist die gemeinsame Wahrscheinlichkeitsverteilung der Beobachtungen das Produkt von N Binomialverteilungen

$$\mathbb{P}(y_1, \dots, y_N | p) = \prod_{i=1}^N \text{Binomial}(y_i | K, p). \quad (8.1)$$

Dieser Ausdruck ist bedingt auf N , womit die $y_i = 0$ Beobachtungen berücksichtigt werden und die resultierenden Daten einfache binomiale iid-Zählungen sind.

Die Wahrscheinlichkeit in (8.1) kann noch weiter vereinfacht werden, wenn die Beobachtungen als Fanghäufigkeiten umgeformt werden. Sei dazu n_k die Anzahl der genau k -mal gefangenen Individuen nach K Erfassungsdurchgängen, $n_k = \sum_{i=1}^N I(y_i = k)$, wobei $I(\cdot)$ die Indikatorfunktion ist (n_k entspricht f_j in Abschnitt 3.1). Ist N bekannt, so kennt man auch die Anzahl der nicht gefangenen Individuen n_0 . Daher lautet die alternative und äquivalente Schreibweise von (8.1)

$$\mathbb{P}(y_1, \dots, y_N | p) = \prod_{k=0}^K \pi_k^{n_k}, \quad (8.2)$$

mit $\pi_k = \mathbb{P}(y = k)$ unter dem Binomialmodell mit Parameter p und Stichprobengröße K . Das wesentliche Problem in Capture-Recapture Studien ist jedoch, dass N unbekannt ist, weil die Anzahl an nicht gefangenen Individuen n_0 unbekannt ist. Folglich sind die Fanghäufigkeiten n_k nicht länger unabhängig, sondern n_0 ist eine Funktion der anderen Häufigkeiten, $n_0 = N - \sum_{k=1}^K n_k$. Deshalb ist die gemeinsame Verteilung multinomial

$$(n_0, n_1, \dots, n_K) \sim \text{Multinomial}(N, \pi_0, \pi_1, \dots, \pi_K). \quad (8.3)$$

Im Zusammenhang mit Capture-Recapture steht die multinomiale Verteilung für eine Population mit N Individuen und $K + 1$ möglichen Ausgängen, die durch die Fanghäufigkeiten - $y = 1, 2, \dots, K$ -mal gefangen oder überhaupt nicht gefangen - festgelegt werden. Die möglichen Ausgänge treten mit Wahrscheinlichkeit π_k ein und werden „Zellwahrscheinlichkeiten“ oder - in Verbindung mit Capture-Recapture - Fanggeschichten-Wahrscheinlichkeiten genannt.

Um schließlich ein Modell für *unbekanntes* N anzupassen, wird n_0 als Parameter betrachtet und die multinomiale Likelihood direkt maximiert. Eine direkte Analyse des Likelihood-Modells ist jedoch nicht immer möglich, denn viele SCR-Modelle beinhalten individuelle Kovariablen, die Unterschiede im Nachweis von Individuen mit sich bringen. Als Beispiel sei hierzu das später behandelte Modell M_x genannt (Royle et al., 2014, S.89f).

Das multinomiale und auf N bedingte Modell wird manchmal als „gemeinsame Likelihood“, „volle Likelihood“ oder „unkonditionale Likelihood“ bezeichnet. Diese Formulierung unterscheidet sich damit von der sogenannten „bedingten Likelihood“, in der die Likelihood der beobachteten Fanggeschichten bedingt ist auf den Ausgang, dass ein Individuum zumindest einmal gefangen wird. Für die Konstruktion der Likelihood muss deshalb $y > 0$ verlangt werden und damit basiert das Beobachtungsmodell auf $\mathbb{P}(y|y > 0)$. Für den einfachen Fall im Modell M_0 ergibt sich die konditionale Verteilung als „Null gestutzte“ (*zero-truncated*) Binomialverteilung, in der der Wert $y = 0$ im Datensatz nicht beobachtet werden kann. Details zu dieser Art von Verteilung werden in A.1 behandelt.

Eine weitere Formulierung von Capture-Recapture Modellen basiert auf der „Null aufgeblasenen“ (*zero-inflated*) Binomialverteilung, welche durch Anwendung von Daten-Vergrößerung (*data augmentation*) im Zuge der Bayesschen Statistik zustande kommt (Royle et al., 2014, S.91f).

8.1.1 Daten-Vergrößerung

Daten-Vergrößerung ist ein allgemeines statistisches Konzept, das in vielen verschiedenen Bereichen zur Anwendung kommt. Sie kann im Rahmen der Bayesschen Statistik sowohl für geschlossene, spatiale als auch für offene Modelle mit unbekanntem N angewendet werden.

Im Grunde bedeutet Daten-Vergrößerung eine Neuparametrisierung des gesamten, auf N konditionalen Datenmodells. Diese Neuparametrisierung wird erreicht, indem der bestehende Datensatz in einen größeren Datensatz mit $M > N$ Zeilen (entsprechen den Individuen) eingebettet und das neue Modell konditional auf M anstatt auf N ausgedrückt wird. Der Vorteil von Daten-Vergrößerung besteht darin, dass N für diese Umparametrisierung nicht bekannt sein muss. Zudem kann M für ein gegebenes Problem immer leicht gewählt werden und die Resultate sind unempfindlich gegenüber der Wahl von M . In der weiteren Datenanalyse liegt der Fokus nun auf dem vergrößerten Datensatz. Das heißt, der größere Datensatz mit M Zeilen wird durch ein geeignetes Modell beschrieben. Mit Hilfe einer Bernoulli-Stichprobe kann entschieden werden, ob ein Individuum in M auch ein Mitglied von N ist. Das Ziel ist es, das Verhältnis $\psi = \mathbb{E}(N)/M$ zu schätzen, wobei ψ der „Daten-Vergrößerungsparameter“ ist (Royle et al., 2014, S.92f). Die Wahl von M bringt eine gleichverteilte a-priori-Verteilung für N auf den Zahlen $[0, M]$ mit sich. Somit werden in Capture-Recapture Modellen $M - S$ Fanggeschichten mit allen Nullen an den Datensatz angehängt und der vergrößerte Datensatz mit insgesamt M Zeilen betrachtet. Als Faustregel sollte $M - S$ in etwa so groß gewählt werden, wie die Population erwartet wird (Royle et al., 2014, S.94).

8.1.2 Analyse des Modells M_0

Der vergrößerte Datensatz ist nun eine Null-aufgeblasene Version des Datensatzes mit bekanntem N . Dabei gehören manche Nullen der hinzugefügten Nullzeilen zur Stichprobe (die aktuellen fehlenden Individuen entsprechen) und manche sind strukturelle Nullen, die nicht den Individuen in der Population entsprechen. Für das klassische geschlossene Populationsmodell ist die resultierende Likelihood unter Daten-Vergrößerung - also für den Datensatz der Größe M - eine einfache Null-aufgeblasene binomiale Likelihood. Das Null-aufgeblasene Binomialmodell kann hierarchisch durch Einführen von binären latenten Variablen z_1, z_2, \dots, z_M beschrieben werden. Diese Variablen zeigen an, ob ein Individuum i ein Mitglied der Population mit N Individuen ($z_i = 1$) oder nicht ($z_i = 0$) ist. Es wird angenommen, dass $z_i \sim \text{Bernoulli}(\psi)$ gilt, wobei ψ der Wahrscheinlichkeit entspricht, dass ein Individuum im Datensatz der Größe M ein Mitglied der beprobten Population ist.

Das heißt, $1 - \psi$ ist die Wahrscheinlichkeit einer strukturellen Null im vergrößerten Datensatz. Das Null-aufgeblasene Binomialmodell, das unter Daten-Vergrößerung entsteht, kann formal durch die folgenden Annahmen ausgedrückt werden. Die gewählten Priors sind typisch für die Bayessche Statistik.

$$\begin{aligned} y_i | z_i = 1 &\sim \text{Binomial}(K, p) \\ y_i | z_i = 0 &\sim I(y = 0) \\ z_i &\stackrel{iid}{\sim} \text{Bernoulli}(\psi) \\ \psi &\sim \text{Uniform}(0, 1) \\ p &\sim \text{Uniform}(0, 1) \end{aligned}$$

für $i = 1, \dots, M$ und $I(y = 0)$ ist ein Massenpunkt bei $y = 0$. Manchmal ist es günstig, das auf z bedingte Beobachtungsmodell kurz und bündig in einem Schritt auszudrücken:

$$y_i | z_i \sim \text{Binomial}(K, z_i p)$$

und das bedeutet, dass y_i zwangsläufig 0 ist, wenn $z_i = 0$ gilt, weil die Erfolgswahrscheinlichkeit $z_i p = 0$ lautet.

Unter Verwendung von Daten-Vergrößerung bleibt N nicht länger ein expliziter Parameter dieses Modells. Stattdessen werden ψ und Funktionen der latenten Variablen z geschätzt. Insbesondere gilt unter den Annahmen eines Null-aufgeblasenen Modells $z_i \sim \text{Bernoulli}(\psi)$. Folglich ist N eine Funktion dieser latenten Variablen

$$N = \sum_{i=1}^M z_i.$$

Außerdem können die latenten Parameter z_i durch Integration aus dem Modell entfernt werden, womit die gemeinsame Wahrscheinlichkeitsverteilung der Daten als

$$\mathbb{P}(y_1, \dots, y_M | p, \psi) = \prod_{i=1}^M (\psi \times \text{Binomial}(y_i | K, p) + (1 - \psi) \times I(y_i = 0)) \quad (8.4)$$

gegeben ist. Als Likelihood interpretiert kann dieser Ausdruck direkt maximiert werden, um die MLEs (Maximum Likelihood Schätzer) für die strukturellen Parameter ψ und p oder jene von komplexeren Modellen zu erhalten. Für gewöhnlich wird das auf z bedingte Modell mit MCMC Methoden im **BUGS**-System analysiert (Royle et al., 2014, S.94f).

Im einfachsten Modell M_0 können die Fangdaten zu Individuen-spezifischen Fanghäufigkeiten verbunden werden. Die vergrößerten Daten sind durch den Vektor der

Häufigkeiten $(y_1, \dots, y_S, 0, 0, \dots, 0)$ gegeben, wobei die hinzugefügten Werte $y = 0$ die Fanghäufigkeiten für mögliche Individuen y_{S+1}, \dots, y_M darstellen. In Abbildung 8.1 ist die **BUGS**-Modellbeschreibung des Modells M_0 einer geschlossenen Population für die Fangbeobachtungen y_{ik} dargestellt. Die letzte Zeile in der Modellspezifizierung liefert den Ausdruck zur Berechnung von N aus den Variablen z_i der Daten-Vergrößerung (Royle et al., 2014, S.95).

```

model{
psi ~ dunif(0,1)
p ~ dunif(0,1)

for (i in 1:M){
  z[i] ~ dbern(psi)
  for (k in 1:K){
    tmp[i,k] <- p*z[i]
    y[i,k] ~ dbin(tmp[i,k],1)
  }
}

N <- sum(z[1:M])
}

```

Abbildung 8.1: **BUGS**-Beschreibung des Modells M_0 unter Daten-Vergrößerung (siehe Royle et al., 2014, S.96)

8.1.3 Modell M_0 für die Würfelnatter-Daten

Im Folgenden wird das Modell M_0 auf die Würfelnatter-Daten aus dem Jahr 2014 angewendet. Dabei wird zuerst der gesamte Datensatz `wn_data` in R eingelesen. Dieser beinhaltet die Aufzeichnungen zu den Würfelnatter-Fängen aller vier Jahre 2013 bis 2016. Anschließend können das Jahr 2014 gefiltert und nicht benötigte Nullzeilen entfernt werden, womit der Datensatz `wn_14_neu` entsteht (Details sind im Anhang A.3 zu finden). Mit dieser Matrix der Größe 70 Individuen \times 31 Begehungen kann nun weitergearbeitet werden. Für die Daten-Vergrößerung wird $M = 360$ gewählt. Dadurch erhält man für $M - S$ 290 Null-Einträge, was in etwa der Größe der Population aus Tabelle 5.2 entspricht.

```

> nind <- dim(wn_14_neu)[1]           # 70 Individuen im Jahr 2014
> K <- dim(wn_14_neu)[2]             # 31 Begehungen im Jahr 2014

```

8.1. Ein einfaches geschlossenes Populationsmodell: Modell M_0

```
> ntraps <- 58 # 58 Versteckplätze
> M <- 360
> nz <- M - nind # 290 Null-Einträge
> Yaug <- array(0, dim=c(nz,K))
> Yaug <- as.matrix(Yaug)
> colnames(Yaug) <- colnames(wn_14_neu)
> y <- rbind(wn_14_neu, Yaug)

> set.seed(2013) # jedes Mal die selben Resultate
> library(R2WinBUGS) # R2WinBUGS laden
> data0 <- list(y=y, M=M, K=K) # Daten setzen
> params0 <- c('psi','p','N') # Parameter setzen
> zst <- c(rep(1,nind),rbinom(M-nind, 1, 0.5)) # initialisieren
> inits <- function(){ list(z=zst, psi = runif(1), p = runif(1)) }

> fit0 <- bugs(data0, inits, params0, model.file="modelM0.txt",
              n.chains=3, n.iter=4000, n.burnin=2000, n.thin=1,
              debug=TRUE, working.directory=getwd(),
              bugs.directory = 'C:/Program_Win/WinBUGS14')
> print(fit0,digits=2)
```

```
Inference for Bugs model at "modelM0.txt", fit using WinBUGS,
3 chains, each with 4000 iterations (first 2000 discarded)
n.sims = 6000 iterations saved

      mean   sd  2.5%   25%   50%   75%  97.5% Rhat n.eff
psi      0.53 0.11  0.36  0.46  0.52  0.60  0.77 1.01 2200
p         0.02 0.00  0.01  0.01  0.02  0.02  0.02 1.00 3700
N        191.81 37.85 133.00 165.00 186.00 213.25 278.00 1.00 2400
deviance 905.53 33.51 844.90 882.20 903.60 927.60 974.00 1.00 2100
```

For each parameter, `n.eff` is a crude measure of effective sample size, and `Rhat` is the potential scale reduction factor (at convergence, `Rhat=1`).

DIC info (using the rule, $pD = \text{var}(\text{deviance})/2$)

`pD = 561.1` and `DIC = 1466.6`

DIC is an estimate of expected predictive error (lower deviance is better).

WinBUGS hat einen MCMC Algorithmus für dieses Modell gefunden. Beinahe für alle Parameter gilt $\hat{R} = 1$. Die effektive Stichprobengröße `n.eff` ist etwas kleiner als die Gesamtanzahl der a-posteriori-Proben `n.sims`. Der a-posteriori-Mittelwert von N unter diesem Modell beträgt 191.81 und ist damit etwas kleiner als die Werte aus Tabelle 5.2. In Abbildung 8.2 ist die a-posteriori-Verteilung für Modell M_0 zu sehen.

Die Dichte von N ist nicht abgeschnitten, somit wurde M ausreichend groß gewählt.

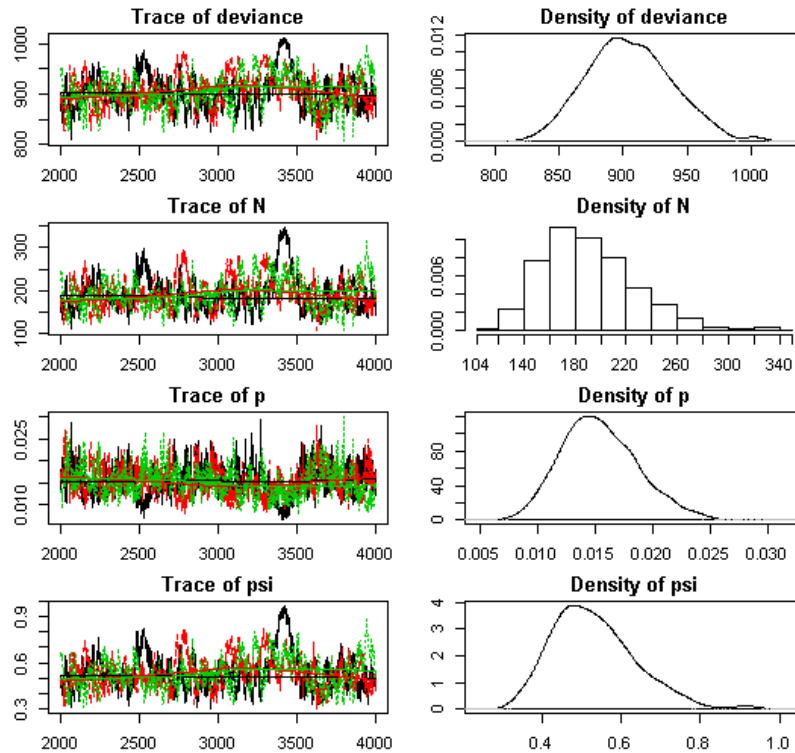


Abbildung 8.2: A-posteriori-Verteilung für Modell M_0 .

8.2 Modelle mit individueller Heterogenität

Der folgende Abschnitt beschäftigt sich mit Modellen, deren Fangwahrscheinlichkeit sich zwischen den Individuen unterscheidet. Das sogenannte Modell M_h („h“ für Heterogenität) ist eines der elementaren Capture-Recapture Modelle. Dabei wird angenommen, dass die Individuen-spezifischen Parameter p_i der Fangwahrscheinlichkeit Zufallsvariablen mit Wahrscheinlichkeitsverteilung $[p|\theta]$ sind. Diese grundlegende Annahme wird als $p_i \sim [p|\theta]$ bezeichnet. Das Verfahren in diesem Modell ist vergleichbar mit der Erweiterung eines GLMs zu einem GLMM, jedoch ist im Capture-Recapture Kontext N unbekannt. Die Basisklasse der Modelle mit Heterogenität wird oft als Modell M_h bezeichnet, aber in Wahrheit ist dies eine breite Klasse von Modellen, die durch verschiedene spezifische Verteilungen für p_i unterschieden wird. Hier wird ein konkreter Typ des Modells M_h betrachtet, der eine natürliche Erweiterung des einfachen binomialen Beobachtungsmodells M_0 ist. Es

gilt

$$\text{logit}(p_i) = \mu + \eta_i$$

mit einem fixen Parameter μ (der Erwartungswert), der geschätzt wird, und einem individuellen Zufallseffekt η_i , der normalverteilt ist

$$\eta_i \sim \text{Normal}(0, \sigma_p^2).$$

Diese beiden Schritte können auch als $\text{logit}(p_i) \sim \text{Normal}(\mu, \sigma_p^2)$ zu einer „Logit-normalen Mischung“ zusammengefasst werden (Royle et al., 2014, S.102f).

Seit man erkannte, dass ein Nebeneinander-Leben von Individuen im Bereich von Versteckplätzen oder Fallen Heterogenität in der Fangwahrscheinlichkeit verursacht, ist das Modell M_h von bedeutender historischer Wichtigkeit. Es wurde populär, verschiedene Versionen von Modell M_h anzuwenden, um N unter Berücksichtigung von spatialen Daten wie Versteckplätzen oder Fallen zu schätzen. Man erkannte, dass eine Spezies eine gewisse spatiale Treue in Form von Lebensraum oder Revier aufweist. Zudem bewegen sich Individuen in ihrem Lebensraum in zufälliger Art und Weise. Aufgrund der Überlappung des Lebensraumes mit einer beprobten Fläche wird angenommen, dass jedes Individuum in der Nähe des Untersuchungsgebiets einer unterschiedlichen Begegnungswahrscheinlichkeit ϕ unterliegt. Das Erfassen eines Individuums wird modelliert, indem $a_i = 1$ angenommen wird, wenn Individuum i während einer Stichprobe gefangen werden kann, und 0 sonst. Es gilt $\mathbb{P}(a_i = 1) = \phi$ und ϕ sollte davon abhängen, wo ein Individuum lebt, somit ist ϕ_i Individuum-spezifisch. Spatiale Capture-Recapture Modelle verwenden spatiale Informationen der Versteckplätze oder Fallen, um eine mögliche Bewegung der Individuen ins Modell aufzunehmen. Wenn Individuen mit Wahrscheinlichkeit p_0 , konditional auf $a_i = 1$, gefunden werden, dann ist die marginale Wahrscheinlichkeit, Individuum i zu entdecken gleich

$$p_i = p_0 \phi_i.$$

Daran ist erkennbar, dass individuelle Heterogenität in Fangwahrscheinlichkeiten vom Lebensraum der Individuen und der Bewegung darin verursacht wird (Royle et al., 2014, S.103f).

```

model{
p0 ~ dunif(0,1)
mup <- log(p0/(1-p0))
sigmap ~ dunif(0,2)
sigmap2 <- sigmap*sigmap
psi ~ dunif(0,1)

for (i in 1:M){
  z[i] ~ dbern(psi)
  lp[i] ~ dnorm(mup,sigmap2)
  logit(p[i]) <- lp[i]
  for (k in 1:K){
    tmp[i,k] <- p[i]*z[i]
    y[i,k] ~ dbin(tmp[i,k],1)
  }
}

N <- sum(z[1:M])
}

```

Abbildung 8.3: **BUGS**-Beschreibung des Modells M_h unter Daten-Vergrößerung (vgl. Royle et al., 2014, S.105)

8.2.1 Analyse des Modells M_h

Ähnlich wie für Modell M_0 verwendet man auch hier das Bernoulli-Modell für Null-aufgeblasene Variablen $z_i \sim \text{Bernoulli}(\psi)$ und das Beobachtungsmodell wird konditional auf diese latenten Variablen z_i ausgedrückt. Für $z_i = 1$ resultiert ein Binomialmodell mit Individuen-spezifischen p_i

$$y_i | z_i = 1 \sim \text{Binomial}(K, p_i),$$

sonst gilt $y_i | z_i = 0 \sim I(y = 0)$, das heißt $y = 0$ ist ein Massenpunkt. Weiters gilt für die Verteilung von p_i

$$\text{logit}(p_i) \sim \text{Normal}(\mu, \sigma^2).$$

Für a-priori-Verteilungen wird $p_0 = \text{logit}^{-1}(\mu) \sim \text{Uniform}(0, 1)$ angenommen und für die Standardabweichung gilt $\sigma \sim \text{Uniform}(0, B)$ für großes B . Dieses B beschreibt die Ausdehnung des Gebiets in Kilometern und lässt sich als Abstand zwischen dem

Zentrum des Gebiets aller Versteckplätze und dem entferntesten Versteckplatz berechnen. Eine weitere gebräuchliche vorgegebene Prior ist $\tau = 1/\sigma^2 \sim \text{Gamma}(.1, .1)$ (Royle et al., 2014, S.104). Abbildung 8.3 zeigt die **BUGS**-Modellbeschreibung des Modells M_h . Dabei wurde σ für B bereits an die Würfelnatter-Daten angepasst. Es entspricht $B = 2$ dem Abstand zwischen dem Zentrum aller Versteckplätze und dem entferntesten Versteckplatz in Kilometern. Im Unterschied zum Originalcode in Royle et al. (2014, S.105) gilt hier $\text{lp}[i] \sim \text{dnorm}(\text{mup}, \text{sigmap}^2)$. Es wird direkt σ^2 anstatt τ als Varianz eingesetzt.

8.2.2 Modell M_h für die Würfelnatter-Daten

In diesem Beispiel werden wiederum die Würfelnatter-Daten aus dem Jahr 2014 betrachtet. Es wird die Methode der Daten-Vergrößerung angewendet, um einen Datensatz von $M = 1400$ Individuen zu erhalten.

```
> M <- 1400
> set.seed(2013) # jedes Mal die selben Resultate
> library(R2WinBUGS) # R2WinBUGS laden
> data1 <- list(y = y, nz = nz, nind = nind, K = K) # Daten setzen
> params1 <- c('p0', 'sigmap', 'psi', 'N') # Parameter setzen
> zst <- c(rep(1, nind), rbinom(M-nind, 1, 0.5)) # initialisieren
> inits <- function(){ list(z = zst, psi=.6, p0 = runif(1),
                           sigmap = runif(1, .7, 1.2), lp = rnorm(M, -2)) }

> fit1 <- bugs(data1, inits, params1, model.file="modelMh.txt",
              n.chains=3, n.iter=3000, n.burnin=1500, n.thin=1,
              debug=TRUE, working.directory=getwd(),
              bugs.directory = 'C:/Program_Win/WinBUGS14')
> print(fit1, digits=2)
```

```
Inference for Bugs model at "modelMh.txt", fit using WinBUGS,
3 chains, each with 3000 iterations (first 1500 discarded)
n.sims = 4500 iterations saved
```

	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
p0	0.01	0.00	0.01	0.01	0.02	0.02	0.02	1.04	130
sigmap	1.76	0.16	1.43	1.64	1.78	1.89	1.99	1.09	27
psi	0.13	0.02	0.10	0.12	0.13	0.15	0.19	1.01	430
N	186.33	31.60	140.00	163.00	180.00	205.00	256.00	1.02	270
deviance	875.66	26.40	830.50	856.10	873.20	893.10	930.51	1.01	610

For each parameter, n.eff is a crude measure of effective sample size,

and R_{hat} is the potential scale reduction factor (at convergence, $R_{hat}=1$).

DIC info (using the rule, $pD = \text{var}(\text{deviance})/2$)

$pD = 347.4$ and $DIC = 1223.1$

DIC is an estimate of expected predictive error (lower deviance is better).

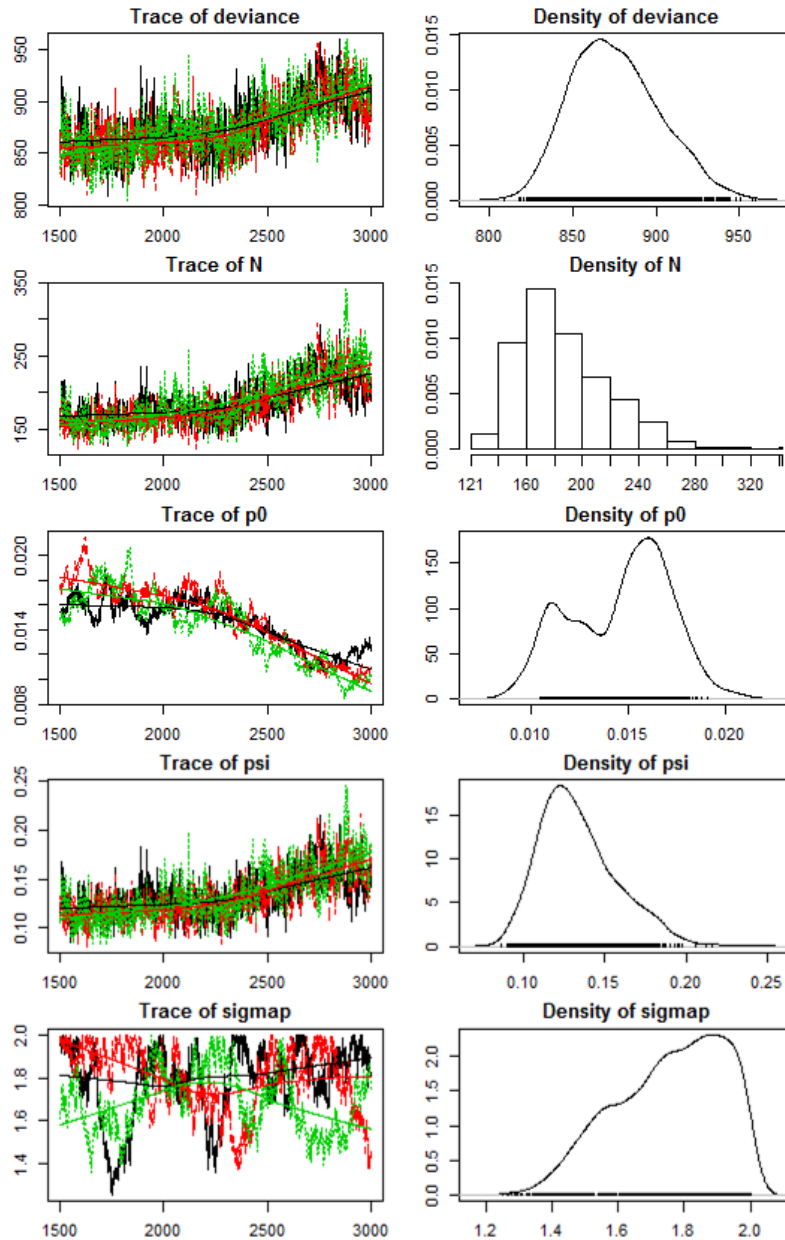


Abbildung 8.4: A-posteriori-Verteilung für Modell M_h .

Im Summary der a-posteriori-Verteilung liegt der Mittelwert für N bei 186.33, dieser ist also sehr ähnlich zum Modell M_0 von vorhin. Abbildung 8.4 liefert aus der Posterior für N keinen Hinweis darauf, dass M vergrößert werden muss.

8.3 Individuelle Kovariablen-Modelle

In Standard Capture-Recapture Situationen kann für jedes Individuum eine Kovariable gemessen werden, die möglicherweise die Fangwahrscheinlichkeit beeinflusst. Modelle, die solche Kovariablen berücksichtigen, werden „individuelle Kovariablen-Modelle“ genannt und mit M_x (x für die explizite Kovariable) bezeichnet. Individuelle Kovariablen-Modelle sind ein Wegweiser für die späteren spatialen Capture-Recapture Modelle. Wie bereits die beiden vorigen Modelle einer geschlossenen Population, startet auch dieses mit dem grundlegenden binomialen Beobachtungsmodell

$$y_i \sim \text{Binomial}(K, p_i).$$

Zur Modellierung der Kovariable wird für die Fangwahrscheinlichkeit ein Logit-Modell der Form

$$\text{logit}(p_i) = \alpha_0 + \alpha_1 x_i \tag{8.5}$$

benutzt. Dabei ist x_i der Kovariablen-Wert für Individuum i und die Parameter $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)$ sind die Regressionskoeffizienten. Klassische Beispiele von Kovariablen, die die Erfassungswahrscheinlichkeit beeinflussen, beziehen sich auf den Typ des Tiers (juvenil/adult oder männlich/weiblich), aber auch eine beständige Kovariable wie Körpermasse oder eine diskrete Kovariable wie Gruppengröße sind möglich.

Modell M_x ist dem Modell M_h in der Struktur sehr ähnlich. Jedoch werden hier die individuellen Effekte aller S Individuen, die in der Stichprobe vorkommen, *beobachtet*. Diese Modelle bilden die Grundlage von spatialen Capture-Recapture Modellen. Die Kovariable zeigt auf, *wo* sich das Individuum in Bezug auf die Fläche der Versteckplätze oder Fallen befindet. SCR-Modelle *sind* sozusagen individuelle Kovariablen-Modelle, in denen die individuelle Kovariable für jedes gefangene Individuum nur unvollständig oder teilweise beobachtet wird. Im Gegensatz zu Modell M_h gibt es in SCR-Modellen (und im Modell M_x) direkte Informationen über die latente Variable, die aus den spatialen Positionen der individuellen Wiederfänge kommen.

Üblicherweise kann die Schätzung von N im Modell M_x mit Methoden, die auf den Ideen von ungleicher Wahrscheinlichkeitsstichproben (*unequal probability sampling*)

(Horvitz-Thompson Schätzung) basieren, erreicht werden. Ein Schätzer von N ist

$$\hat{N} = \sum_{i=1}^S \frac{1}{\tilde{p}_i},$$

wobei \tilde{p}_i die Wahrscheinlichkeit bezeichnet, dass Individuum i in der Stichprobe vorkommt (*zero-truncated*). Diese Größe ist $\tilde{p}_i = \mathbb{P}(y_i > 0)$ und kann in einem Capture-Recapture Modell einer geschlossenen Population als

$$\mathbb{P}(y_i > 0) = 1 - (1 - p_i)^K$$

berechnet werden. Dabei bezeichnet p_i eine Funktion der Parameter α_0 und α_1 gemäß (8.5). In der Praxis werden die Parameter aus der konditionalen Likelihood der beobachteten Fanggeschichten geschätzt, die für Beobachtung y_i

$$\mathcal{L}_C(\boldsymbol{\alpha}|y_i) = \frac{\text{Binomial}(y_i|\boldsymbol{\alpha})}{\tilde{p}_i} \quad (8.6)$$

lautet.

Das leitet sich von der direkten Anwendung des Gesetzes der totalen Wahrscheinlichkeit ab. Es wird $\mathbb{P}(y)$ nach $\mathbb{P}(y) = \mathbb{P}(y|y > 0)\mathbb{P}(y > 0) + \mathbb{P}(y|y = 0)\mathbb{P}(y = 0)$ unterteilt. Für jeden positiven Wert von y ist der zweite Term notwendigerweise 0 und nach Umformung gilt $\mathbb{P}(y|y > 0) = \mathbb{P}(y)/\mathbb{P}(y > 0)$. Im Spezialfall, wo $\mathbb{P}(y)$ die binomiale Wahrscheinlichkeitsfunktion (pmf) ist, erhält man daraus die Formel (8.6) (Royle et al., 2014, S.108f). Auch die hergeleiteten Schätzer aus Abschnitt 5 haben den selben Aufbau wie (8.6).

8.3.1 Analyse des Modells M_x

Nun wird ein formaler Modell-basierter Ansatz der Bayesschen Statistik unter dem Einsatz von Daten-Vergrößerung, basierend auf der gemeinsamen Likelihood, für solche Modelle angewendet. Da die individuelle Kovariable für die $n_0 = N - S$ nicht gefangenen Individuen unbeobachtet ist, benötigt die Bayessche Statistik hier ein Modell, das die Veränderung in x unter den Individuen beschreibt. Im Wesentlichen muss der Stichprobe erlaubt werden, auf die Population hochgerechnet zu werden. Gibt es beispielsweise ein ständig gemessenes Merkmal an jedem Individuum, dann kann für x die Normalverteilung

$$x_i \sim \text{Normal}(\mu, \sigma^2)$$

angenommen werden. Darauf kann die Daten-Vergrößerung direkt angewendet werden. Insbesondere liefert eine Neuformulierung des Modells unter Daten-Vergrößerung für jedes $i = 1, 2, \dots, M$ ein grundlegendes Null-aufgeblasenes Binomialmodell der folgenden Form (Royle et al., 2014, S.109):

$$\begin{aligned} z_i &\sim \text{Bernoulli}(\psi) \\ y_i | z_i = 1 &\sim \text{Binomial}(K, p_i(x_i)) \\ y_i | z_i = 0 &\sim I(y = 0) \\ x_i &\sim \text{Normal}(\mu, \sigma^2). \end{aligned}$$

Im Folgenden wird noch ein für spatiale Capture-Recapture Modelle besonders relevanter Spezialtyp von Modell M_x betrachtet. Um Heterogenität in der Fangwahrscheinlichkeit zu erklären, ist eine vernünftige Kovariable für jedes Individuum das Maß des Abstandes zwischen Lebensraum-Mittelpunkt und Versteckplatz beziehungsweise Falle. Das heißt, Individuen, die vielen Versteckplätzen oder Fallen ausgesetzt sind, sollten höhere Fangwahrscheinlichkeiten haben und umgekehrt. Deshalb wird hier die skalare individuelle Kovariable x_i als der Abstand zwischen dem durchschnittlichen Fangplatz eines Individuums i , genannt \mathbf{s}_i , und dem Zentrum des Gebiets der Versteckplätze, \mathbf{x}_0 , definiert als $x_i = \|\mathbf{s}_i - \mathbf{x}_0\|$, wobei $\|u\|$ die Standardnotation für die Euklidische Norm ist. Das bedeutet, dass Individuen in der Mitte des Gebiets eine höhere Wahrscheinlichkeit haben, gefangen zu werden, umgekehrt sinkt p_i , wenn sich x_i erhöht. Ein zentraler Punkt ist, dass die \mathbf{s}_i für jedes Individuum, das nicht gefangen wird - folglich auch die x_i - fehlen. Damit ist x_i eine latente Variable und die Wahrscheinlichkeitsverteilung dafür muss spezifiziert werden. Da es sich um einen Abstand handelt, muss er zum einen positiv sein, zum anderen wird ein Individuum, das sehr weit vom Gebiet der Versteckplätze oder Fallen entfernt ist, wohl nicht gefangen werden. Daher kann x_i als gleichverteilt zwischen 0 und einer großen Zahl B

$$x_i \sim \text{Uniform}(0, B)$$

angenommen werden. Die Konstante B sollte wie im Modell M_h bereits erwähnt zumindest so groß sein wie der Abstand zwischen dem Zentrum des Gebiets aller Versteckplätze und dem entferntesten Versteckplatz (Royle et al., 2014, S.109f).

Die Implementation in der **BUGS**-Sprache ist trivial, sie ist dem Modell M_h sehr ähnlich. Die Verteilung der Kovariablen, zusammen mit den a-priori-Verteilungen dieser Parameter, muss angegeben werden. Abbildung 8.5 beinhaltet die detaillierte Beschreibung des Modells M_x .

```

model{
p0 ~ dunif(0,1)
alpha0 <- log(p0/(1-p0))
psi ~ dunif(0,1)
alpha1 ~ dnorm(0,.01)

for (i in 1:M){
  xcent[i] ~ dunif(0,B)
  z[i] ~ dbern(psi)
  lp[i] <- alpha0 + alpha1*xcent[i]
  logit(p[i]) <- lp[i]
  for (k in 1:K){
    tmp[i,k] <- p[i]*z[i]
    y[i,k] ~ dbin(tmp[i,k],1)
  }
}

N <- sum(z[1:M])
}

```

Abbildung 8.5: **BUGS**-Beschreibung des Modells M_x unter Daten-Vergrößerung
(siehe Royle et al., 2014, S.111f)

8.3.2 Modell M_x für die Würfelnatter-Daten

Analog zum Modell M_h wird M zur Modellanpassung von M_x mittels Daten-Vergrößerung auf 1400 gesetzt. Wie vorhin beträgt die Entfernung B zwischen dem Zentrum aller Versteckplätze und dem entferntesten Versteckplatz 2 Kilometer (Details zu den eingelesenen Dateien finden sich im Anhang A.3). Um einen Vergleich zu Modell M_h herzustellen, wird das folgende Modell mit $B = 2$ gerechnet, obwohl auch größere Werte möglich wären.

```

> M <- 1400
> traps <- read.csv("Versteckplaetze_2014.csv", header=TRUE, sep=";")
> gps_traps <- traps[,2:3]

> x_cent <- sum(gps_traps[,1])/58 # Mittelwert aus x-Koordinaten
> y_cent <- sum(gps_traps[,2])/58 # Mittelwert aus y-Koordinaten
> cent <- c(x_cent,y_cent)      # ermittelter Mittelpunkt

```

```

> indiv <- read.csv("WN_Koordinaten_2014_Indiv.csv", header=TRUE, sep=";")
> gps_indiv <- indiv[,2:3]

# individuelle Kovariablen
> xcent <- rep(NA,(nind+nz))
> for (i in 1:nind){
  xcent[i] <- distm(c(gps_indiv[i,1],gps_indiv[i,2]),c(cent[1],cent[2]),
    fun = distHaversine)*0.001
}

# Abstand vom Zentrum der Versteckplätze zum entferntesten Punkt (in km)
> gps_abstand <- rep(0,dim(gps_traps)[1])
> for (j in 1:dim(gps_traps)[1]){
  gps_abstand[j] <- distm(c(gps_traps[j,1], gps_traps[j,2]),
    c(cent[1], cent[2]),
    fun = distHaversine)*0.001
}
> B_entfernung <- ceiling(sort(gps_abstand)[58])

> set.seed(2013) # jedes Mal die selben Resultate
> library(R2WinBUGS) # R2WinBUGS laden
> data2 <- list(y = y, nz = nz, nind = nind, K = K, xcent = xcent,
  B = B_entfernung) # Daten setzen
> params2 <- c('p0','psi','N','alpha0') # Parameter setzen
> zst <- c(rep(1,nind),rbinom(M-nind, 1, 0.5)) # initialisieren
> inits <- function() {list(z = zst, psi = runif(1), p0 = runif(1),
  alpha1 = rnorm(1) ) }

> fit2 <- bugs(data2, inits, params2, model.file="modelMcov.txt",
  n.chains=3, n.iter=3000, n.burnin=1000, n.thin=1,
  debug=TRUE, working.directory=getwd(),
  bugs.directory = 'C:/Program_Win/WinBUGS14')
> print(fit2,digits=2)

```

```

Inference for Bugs model at "modelMcov.txt", fit using WinBUGS,
3 chains, each with 3000 iterations (first 1000 discarded)
n.sims = 6000 iterations saved

```

	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat	n.eff
p0	0.04	0.01	0.02	0.03	0.04	0.05	0.06	1.01	700
psi	0.30	0.07	0.20	0.25	0.29	0.34	0.46	1.01	2300
N	420.25	91.77	280.00	357.00	405.00	468.00	642.00	1.01	3000
alpha0	-3.21	0.28	-3.77	-3.39	-3.20	-3.03	-2.69	1.01	700

```
deviance 1023.61 35.00 961.99 999.47 1020.00 1045.00 1100.00 1.01 4100
```

For each parameter, n.eff is a crude measure of effective sample size, and Rhat is the potential scale reduction factor (at convergence, Rhat=1).

DIC info (using the rule, $pD = \text{var}(\text{deviance})/2$)

$pD = 612.4$ and $DIC = 1636.1$

DIC is an estimate of expected predictive error (lower deviance is better).

Aus dem a-posteriori-Summary ist erkennbar, dass das geschätzte N mit 420.25 um einiges größer ist als jenes aus Modell M_h . Dies kann damit zusammenhängen, dass das geschätzte N eine genau angegebene Fläche verwendet, die durch den zuvor festgelegten Wert B definiert wird. Somit ist diese Fläche eine fixe Komponente des Modells und der Parameter N steht daher in eindeutigem spatialen Zusammenhang mit der Anzahl der Individuen, deren Lebensraum-Mittelpunkt um weniger als B vom Zentrum der Versteckplätze entfernt ist. Die Fläche des Gebiets aller Versteckplätze für ein gegebenes B ist also eine genau definierte Größe, nämlich die Fläche eines Kreises mit Radius B (Royle et al., 2014, S.112). Weiters liefert die a-posteriori-Verteilung für N aus Abbildung 8.6 keinen Hinweis, dass M nochmals vergrößert werden muss. Somit wurde ein konvergentes individuelles Kovariablen-Modell gefunden.

Betrachtet man die Dichte der Schlangenpopulation $D = N/\text{area}$, so ist diese für eine gewählte gleichverteilte a-priori-Verteilung für x *nicht konstant* über das Gebiet. Modell M_x nimmt an, dass die Dichte *kleiner* wird, je weiter man sich vom Zentrum des Gebiets der Versteckplätze entfernt. Das heißt, mit $x_i \sim \text{Uniform}(B)$ ist N in jedem beliebigen Abstand vom Zentrum konstant, aber die *Fläche* jedes Abstandes ist offensichtlich anwachsend. In einer weiteren Analyse kann eine bessere Verteilung für x gefunden werden (Royle et al., 2014, S.112f). Dies wird in dieser Arbeit allerdings nicht mehr untersucht, da der Fokus auf das Erreichen eines spatialen Capture-Recapture Modells gelegt wird. Ebenso wird die Methode des *Distance Sampling* übersprungen, weil SCR-Modelle viel allgemeiner und vielseitiger sind und Daten von so gut wie allen tierischen Erhebungsdesigns anpassen können (Royle et al., 2014, S.118).

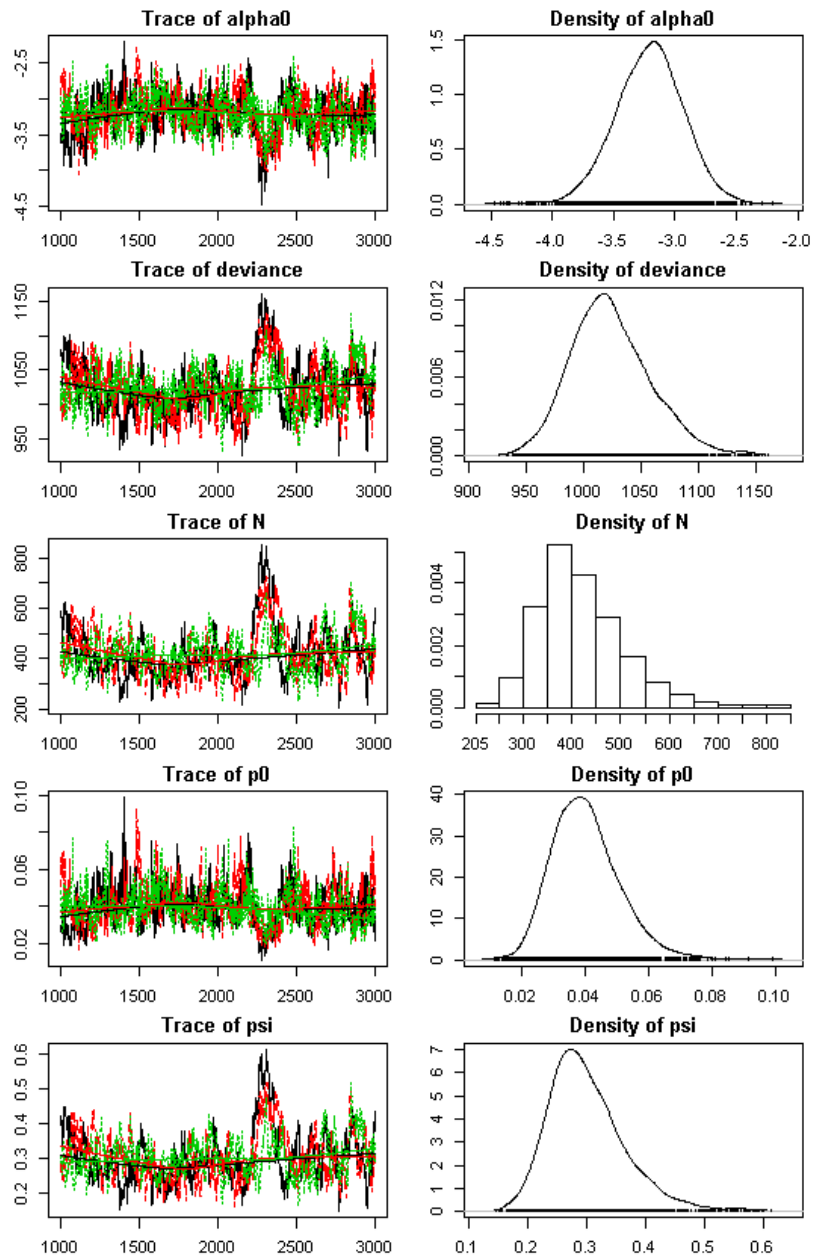


Abbildung 8.6: A-posteriori-Verteilung für Modell M_x .

Spatiale Capture-Recapture Modelle

Im vorigen Kapitel wurden Modelle diskutiert, die als primitive spatiale Capture-Recapture Modelle angesehen werden können. Das individuelle Kovariablen-Modell war in gewisser Weise „spatial“, weil es eine Charakterisierung des Lebensraumes der Individuen beinhaltet. Jedoch ist es nur ein primitives, da es keine Charakterisierung der Lage der Versteckplätze oder Fallen enthält. Es ist nur noch ein kleiner Schritt von diesem Modell zum spatialen CR-Modell, das sowohl die spatiale Information über einzelne Tiere als auch über die Lage der Fanggeräte - im Falle der Würfelnatter Versteckplätze genannt - betrachtet. Während im individuellen Kovariablen-Modell die Kovariable der Abstand zum Zentrum des Gebiets der Versteckplätze war, gibt es hier J individuelle Kovariablen, nämlich den Abstand zu jedem Versteckplatz. Und anstatt Schätzungen der einzelnen Orte \mathbf{s} zu machen, gibt es ein vollständiges hierarchisches Modell, in welchem \mathbf{s} als latente Variable mit a-priori-Verteilung angesehen wird. Um ein vollständiges SCR-Modell zu erstellen, müssen die einzelnen Fänge innerhalb und zwischen den Versteckplätzen unabhängig sein, womit Individuen- und Versteckplatz-spezifische Fänge als *unabhängige* Bernoulli-Versuche angesehen werden können (Royle et al., 2014, S.125f).

9.1 Das binomiale Beobachtungsmodell

Am Anfang betrachtet man ein einfaches Modell ohne zeitabhängige oder Fallenspezifische Kovariablen, die Begegnungen beeinflussen könnten. Zudem gibt es keine eindeutig Individuen-spezifischen Kovariablen und auch keine, die auf die Dichte der Individuen einwirken. Unter diesen Umständen können die binären Fanggeschichten aller K beprobten Situationen zusammengefasst und die Gesamtanzahl aus K Fängen aufgezeichnet werden. Diese Individuen- und Fallen-spezifischen Fanghäufigkeiten werden mit y_{ij} für $i = 1, 2, \dots, S$ gefangene Individuen und $j = 1, 2, \dots, J$ Versteckplätzen bezeichnet. Damit resultiert eine Matrix mit S Zeilen und J Spalten. Sind die y_{ij} voneinander unabhängige Ergebnisse einer binomialen Zufallsvariable, so gilt

$$y_{ij} \sim \text{Binomial}(K, p_{ij}). \quad (9.1)$$

Das entspricht dem Basismodell eines geschlossenen Populationsmodells aus dem vorigen Abschnitt. In diesem Fall sind die Fanghäufigkeiten aber individuell und Fallen-spezifisch und die Fangwahrscheinlichkeit p_{ij} hängt vom Individuum *und* von der Falle ab.

In weiterer Folge wird \mathbf{s}_i als eine zweidimensionale spatiale Koordinate definiert und steht für den Lebensraum oder den Aktivitätsmittelpunkt eines Individuums i . Spatiale Capture-Recapture Modelle fordern, dass die Fangwahrscheinlichkeit p_{ij} eine fallende Funktion des Abstandes zwischen \mathbf{s}_i und der Position der Falle j , \mathbf{x}_j (ebenfalls eine zweidimensionale spatiale Koordinate), ist. Ein Standardansatz, um binomiale Anzahlen zu modellieren, ist die logistische Regression. Darin wird die Abhängigkeit von p_{ij} auf den Abstand mittels

$$\text{logit}(p_{ij}) = \alpha_0 + \alpha_1 \|\mathbf{x}_j - \mathbf{s}_i\| \quad (9.2)$$

ausgedrückt, wobei $\|\mathbf{x}_j - \mathbf{s}_i\| = d_{ij}$ der Abstand zwischen \mathbf{x}_j und \mathbf{s}_i ist. Alternativ ist auch

$$p_{ij} = p_0 \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x}_j - \mathbf{s}_i\|^2\right) \quad (9.3)$$

ein verbreitetes Modell. Da es den Kern einer bivariaten Normal- oder Gauß-Wahrscheinlichkeits-Dichtefunktion für die Zufallsvariable „individueller Standort“ aufweist, bezeichnet man es als das „(bivariate) Normal-“ oder „Gauß-“Modell.

Modelle der Fangwahrscheinlichkeit werden manchmal parametrisiert, sodass der Abstandskoeffizient α_1 ist. Im Gauß-Modell gilt dann $\alpha_1 = 1/(2\sigma^2)$. Ein Merkmal der gemeinsamen parametrischen Form ist, dass die Modelle monoton fallend mit

dem Abstand sind, aber sich in ihrem Verhalten ändern, wenn sie sich der Entfernung = 0 nähern. Im negativen Exponential-Modell gilt $p_{ij} = p_0 \exp(-\alpha_1 \|\mathbf{x}_j - \mathbf{s}_i\|)$ und im Gauß-Hazard-Modell gilt $p_{ij} = 1 - \exp(-\lambda_0 g(\mathbf{x}_j, \mathbf{s}_i))$, wobei $g(\mathbf{x}_j, \mathbf{s}_i)$ der Gauß-Kern und $\lambda_0 > 0$ die zugrundeliegende Fangrate ist. Egal, welches Modell für die Fangwahrscheinlichkeit gewählt wird, die Aktivitätsräume \mathbf{s}_i für Individuen i , $i = 1, \dots, S$, sind immer unbeobachtete Zufallsvariablen. Damit kann das Beobachtungsmodell als

$$y_{ij} | \mathbf{s}_i \sim \text{Binomial}(K, p(\mathbf{s}_i; \boldsymbol{\alpha})) \quad (9.4)$$

formuliert werden, wobei $\boldsymbol{\alpha}$ alle Parameter im Modell beinhaltet (Royle et al., 2014, S.126ff).

9.1.1 Der Lebensraum-Mittelpunkt

Der Lebensraum eines Individuums wird als die Fläche definiert, die ein Lebewesen während einer Zeitperiode nutzt. Dazu wird der Lebensraum-Mittelpunkt (oder Aktivitätsmittelpunkt) als Zentrum der Fläche bezeichnet, die ein Individuum während der Dauer, in der die Fallen aktiv sind, bewohnt. Der Lebensraum- oder Aktivitätsmittelpunkt ist eine vergängliche Größe und gilt nur in einer bestimmten Periode der Studie (Royle et al., 2014, S.129).

9.1.2 Die latente Variable „Abstand“

Wenn man jedes \mathbf{s}_i in der Population (und die Populationsgröße N) genau kennt, so ist das durch (9.1) und (9.2) spezifizierte Modell gerade ein gewöhnliches logistisches Regressionsmodell mit Kovariable d_{ij} . Allerdings sind die Aktivitätsmittelpunkte unter keinen Umständen beobachtbar. Daher ist d_{ij} - ähnlich wie in klassischen Modellen mit zufälligen Effekten - eine nicht beobachtete Variable. Somit muss das Modell erweitert werden, um diese Zufallsvariablen als zusätzliche Modellkomponente, nämlich als Verteilung von zufälligen Effekten, aufzunehmen. Die übliche Annahme ist die sogenannte „Gleichverteilungsannahme“, die festlegt, dass die \mathbf{s}_i gleichverteilt über den Raum sind. Diese Gleichverteilungsannahme ergibt eine a-priori-Gleichverteilung für \mathbf{s}_i , das heißt, die Dichtefunktion (pdf) von \mathbf{s}_i ist konstant und damit gilt

$$\mathbb{P}(\mathbf{s}_i) \propto \text{konstant.} \quad (9.5)$$

In vielen Fällen ist diese Annahme nicht exakt genug, um SCR-Modelle zu erstellen. Eine weitere und konkretere Möglichkeit, die a-priori-Verteilung anzugeben, hängt

von der Bestimmung des „Zustandsraumes“ der Zufallsvariable \mathbf{s}_i ab (Royle et al., 2014, S.129).

9.2 Das binomiale Punkt-Prozessmodell

Wie bereits erwähnt, werden die individuellen Aktivitätsmittelpunkte nicht beobachtet und müssen als Zufallseffekte behandelt werden. Im Konkreten ist die Ansammlung von individuellen Aktivitätsmittelpunkten $\mathbf{s}_1, \dots, \mathbf{s}_N$ eine Realisierung eines *binomialen Punkt-Prozesses* (BPP). Der binomiale Punkt-Prozess ist analog zum Poisson Punkt-Prozess und stellt eine „zufällige Streuung“ von Punkten im Raum dar. Während die Gesamtanzahl der Punkte im Binomialfall fix ist, sind die Punkte im Poissonfall zufällig verteilt und haben eine Poisson-Verteilung.

Im Zusammenhang mit Capture-Recapture Modellen wird ein binomialer Punkt-Prozess betrachtet, weil er N im Modell behält und somit die direkte Verknüpfung mit geschlossenen Populationsmodellen sichert. Eigentlich sind das Modell M_0 und andere geschlossene Modelle unter einem binomialen Punkt-Prozessmodell einfache Grenzfälle von SCR-Modellen. Diese Modelle entstehen, wenn der Koeffizient für den Abstand (α_1 von oben) gegen 0 strebt.

Da das BPP-Modell ein fixes N hat, ist es nicht unmittelbar ein Modell mit „kompletter spatialer Zufälligkeit“. Werden beispielsweise Anzahlen $n(A_1), \dots, n(A_k)$ in einer Menge von disjunkten Regionen des Zustandsraums A_1, \dots, A_k gebildet, dann sind diese Anzahlen *nicht* unabhängig. Vielmehr haben sie eine multinomiale Verteilung. Somit führt das BPP-Modell eine gewisse Abhängigkeit in die Verteilung der Punkte ein. In den meisten Situationen hat dies jedoch keinen praktischen Effekt auf die Analyse. In der Praxis wird das BPP-Modell als eines mit spatialer Unabhängigkeit bezüglich individueller Aktivitätsmittelpunkte betrachtet, weil jeder Aktivitätsmittelpunkt unabhängig von einem anderen verteilt ist. Trotz dieser Unabhängigkeit weisen Realisierungen von zufällig verteilten Punkten typischerweise deutliche Ungleichförmigkeit auf. Daher erscheinen unabhängige und gleichverteilte Punkte fast nie als regelmäßig gleichverteilt oder systematisch verteilt. Da grundlegende binomiale (oder Poisson) Punkt-Prozessmodelle eine Reihe von Verteilungsmustern zulassen, ohne die Annahme von spatialer Zufälligkeit zu verletzen, sind sie enorm hilfreich in der Praxis. In SCR-Modellen gibt es einige Daten zu einer Stichprobe von Individuen, wodurch das resultierende a-posteriori-Punktmuster stark von der Gleichförmigkeit abweichen kann. Die Gleichförmigkeits-Hypothese ist nur eine a-priori-Verteilung, die direkt von der Menge und Qualität der beobachteten Daten beeinflusst wird und deutlich ungleichmäßig erscheinen kann (Royle et al., 2014, S.129ff).

9.2.1 Der Zustandsraum zum Punkt-Prozess

Um die Bayessche Statistik für ein SCR-Modell anwenden zu können, muss es möglich sein, die \mathbf{s}_i in der Population aus der a-posteriori-Verteilung zu simulieren, obwohl die individuellen Aktivitätsmittelpunkte $\mathbf{s}_1, \dots, \mathbf{s}_N$ eine unbekannt Größe sind. Dazu ist es notwendig, die Region, über die sie verteilt sind, genau zu beschreiben. Es handelt sich dabei um den oben genannten *Zustandsraum*, der im Zusammenhang mit dem Punkt-Prozess manchmal *Beobachtungsfenster* genannt wird. Der Zustandsraum wird im Folgenden mit \mathcal{S} bezeichnet und repräsentiert einen Bereich oder eine Menge von Punkten, die mögliche Werte der Zufallsvariable \mathbf{s} umfasst. Damit lautet eine äquivalente explizite Aussage der „Gleichverteilungsannahme“

$$\mathbf{s}_i \sim \text{Uniform}(\mathcal{S}),$$

wobei \mathcal{S} eine genau definierte Fläche ist. Wenn die Punkte über eine Fläche gleichverteilt sind, so wird der Punkt-Prozess üblicherweise *homogener Punkt-Prozess* genannt (Royle et al., 2014, S.131).

Wie bereits in Abschnitt 8.3 kurz angemerkt, erhöht sich N mit zunehmendem \mathcal{S} im gleichen Tempo, wie die Fläche von \mathcal{S} unter der a-priori-Annahme von konstanter Dichte steigt. Damit ist die Dichte der Population invariant zu \mathcal{S} , solange \mathcal{S} ausreichend groß ist. Um also Invarianz zu erreichen, sollte \mathcal{S} sehr groß bezüglich der typischen Lebensraumgröße der untersuchten Spezies gewählt werden. In vielen praxisnahen Situationen wird der Zustandsraum als ein Rechteck definiert, in welchem das Gebiet der Versteckplätze beziehungsweise Fallen zentriert ist. Obwohl eine Definition des Zustandsraums als regelmäßiges Vieleck rechnerische Vorteile liefert, ruft es Probleme hinsichtlich der Zulassung von Regionen hervor, die eindeutig kein Lebensraum sind, wie beispielsweise Ozeane, große Seen, Eisfelder und so weiter. Da komplizierte Gebiete schwierig in mathematischen Termen auszudrücken sind, gibt es die Möglichkeit, den Zustandsraum als diskrete Menge an Punkten mit Hilfe des R-Pakets `secr` darzustellen. Eine Definition des Zustandsraums durch eine diskrete Menge an Punkten ist nützlich, weil bestimmte Punkte gelöscht werden können, wenn sie keinen geeigneten Lebensraum repräsentieren (Royle et al., 2014, S.131f). Für das Beispiel der Würfelnatter-Daten ist ein Rechteck um das Gebiet der Versteckplätze ausreichend, da der Lebensraum der Schlangen sowohl zu Wasser wie zu Land ist.

Als eine weitere praktische Überlegung sei angemerkt, dass die Fläche des Zustandsraums \mathcal{S} die Daten-Vergrößerung beeinflusst. Nimmt die Größe von \mathcal{S} zu, so gibt es mehrere Individuen, die berücksichtigt werden müssen, und darum muss sich die Größe des erweiterten Datensatzes M erhöhen, was zu rechnerischen Konsequenzen führen kann (Royle et al., 2014, S.132f).

9.3 Die Kernannahmen des SCR-Modells

Die wichtigsten statistischen Annahmen für das *spatale* CR-Modell (*SCR-Modell*) lauten (Royle et al., 2014, S.150):

- *Demografische Geschlossenheit*: Das Modell erlaubt keine demografischen Prozesse. Es gibt keinen Zuwachs oder Eintritt in die beobachtete Population. Es gibt keine Mortalität und keinen Austritt aus der beobachteten Population.
- *Geografische Geschlossenheit*: Es wird angenommen, dass es keine dauerhafte Ein-/Auswanderung im Zustandsraum gibt. Lediglich „vorübergehende“ Bewegungen im Zustandsraum sind erlaubt. Darin liegt auch das Hauptaugenmerk von SCR-Modellen. In gewöhnlichen CR-Modellen muss geografische Geschlossenheit angenommen werden, um N in einer sinnvollen Art interpretieren zu können.
- *Die Aktivitätsmittelpunkte sind zufällig verteilt*: Das bedeutet Einheitlichkeit und Unabhängigkeit des zugrunde liegenden Punkt-Prozesses $\mathbf{s}_1, \dots, \mathbf{s}_N$.
- *Entdeckung ist eine Funktion der Entfernung*: Ein Entdeckungsmodell beschreibt, wie die Fangwahrscheinlichkeit eines Individuums abnimmt, wenn es sich von seinem Lebensraum-Mittelpunkt entfernt.
- *Unabhängigkeit der Fänge unter Individuen*: Ein Fang eines Individuums ist unabhängig von einem Fang eines anderen Individuums.
- *Unabhängigkeit der Fänge desselben Individuums*: Ein Fang eines Individuums in einer Falle ist unabhängig von einem Fang in einer anderen Falle, und unabhängig von nachfolgenden Stichproben.

9.4 Das SCR-Modell für unbekanntes N

In allen realen Anwendungen von Capture-Recapture Modellen ist N unbekannt. Um damit umzugehen wird auch hier im Falle des spatalen CR-Modells die Methode der Daten-Vergrößerung angewendet. Es wird also die beobachtete Datenmenge um Fanggeschichten mit allen Nullen erweitert, die mögliche Individuen darstellen. Bei S beobachteten Individuen gibt es $M - S$ mögliche Individuen im Datensatz. Die Daten liegen diesmal im Format einer Matrix mit S Individuen \times J Fallen vor und die Einträge y_{ij} entsprechen den Fanghäufigkeiten. An diese Matrix werden die Null-Beobachtungen drangehängt. Für Individuen daraus gilt notwendigerweise $y_{ij} = 0$ für alle j , da sie nicht beobachtet werden.

Im vergrößerten Datensatz werden binäre latente Variablen z_i eingeführt, um $\mathbb{P}(z_i = 1)$ zu beschreiben. Das ist die Wahrscheinlichkeit, dass ein Individuum aus dem vergrößerten Datensatz M ein Mitglied der Population der Größe N ist. Wenn also für eine der Null-Fanggeschichten $z_i = 1$ gilt, so handelt es sich um eine Null aus der Stichprobe. Bei $z_i = 0$ spricht man von einer strukturellen Null. Das binomiale Beobachtungsmodell wird wiederum *konditional auf* z_i als

$$y_{ij}|z_i \sim \text{Binomial}(K, z_i p_{ij})$$

formuliert. Die binomiale Wahrscheinlichkeit p_{ij} ergibt Null, wenn $z_i = 0$ ist, und Eins, wenn $z_i = 1$ ist.

Die Größe M des erweiterten Datensatzes sollte in etwa so groß sein, dass die a-posteriori-Verteilung für N nicht abgeschnitten (*truncated*) ist. Wird M andererseits zu groß gewählt, dann werden unnötige Berechnungen durchgeführt. Daher sollte M durch Trial-and-Error gewählt werden. Nach einem MCMC Durchlauf kann entschieden werden, ob M erhöht werden sollte. Der aus der Methode der Daten-Vergrößerung abgeleitete Parameter N wird mit $N = \sum_{i=1}^M z_i$ berechnet. Ebenso gilt für die Populationsdichte der Tiere $D = N/\text{area}(\mathcal{S})$ (Royle et al., 2014, S.143f). In Abbildung 9.1 ist die **BUGS**-Beschreibung des Modells *SCR0* dargestellt. Die Modellspezifikation beinhaltet M Fanggeschichten, die die vergrößerten möglichen Individuen und die Parameter z_i und ψ der Daten-Vergrößerung umfassen. Der von α_1 abgeleitete Parameter σ mit $\sigma = \sqrt{1/(2\alpha_1)}$ ist der Skalierungs-Parameter des Gauß-Kerns. Durch Umformung gilt $\alpha_1 = 1/(2\sigma^2)$ wie in (9.3) (Royle et al., 2014, S.145). In der Modellbeschreibung wurde bereits die Fläche des Zustandsraums für die Würfelnatter berücksichtigt. Zu den minimalen und maximalen x- beziehungsweise y-Koordinaten der Versteckplätze wurde ein `buffer` von 100 Metern hinzugefügt (Minus beim Minimum und Plus beim Maximum; Details sind im folgenden Modellcode zu finden). Durch Multiplikation des Abstandes zwischen minimaler und maximaler x-Koordinate inklusive `buffer` und des Abstandes zwischen minimaler und maximaler y-Koordinate inklusive `buffer` erhält man dafür als Fläche 3.45 Quadratkilometer. Somit umfasst das betrachtete Gebiet zur Bestimmung der Würfelnatter-Population eine Fläche von 3.45 Quadratkilometer.

9.4.1 Modell *SCR0* für die Würfelnatter-Daten

Um einen ersten geografischen Überblick über die Würfelnatter-Daten zu bekommen, sind in den Abbildungen 9.2 und 9.3 die Verteilung der 58 künstlichen Versteckplätze entlang der Mur sowie die Häufigkeiten der Würfelnatter-Funde an diesen Versteckplätzen dargestellt.

Im Modellcode ist die Matrix Individuen \times Versteckplätze in `wn_matrix` gespeichert. Zusätzlich sind die Koordinaten der Versteckplätze in `traplocs` zu finden (detaillierte Erklärungen dazu sind im Anhang A.3). Für das spatiale CR-Modell der Würfelnatter-Daten wird M auf 550 und der `buffer` auf 0.1 gesetzt. Damit kann der Zustandsraum definiert werden, indem zu den äußersten Versteckplätzen 100 Meter in jede geografische Richtung addiert werden.

```
model{
alpha0 ~ dnorm(0,.1)
logit(p0) <- alpha0
alpha1 ~ dnorm(0,.1)
sigma <- sqrt(1/(2*alpha1))
psi ~ dunif(0,1)

for (i in 1:M){
  z[i] ~ dbern(psi)
  s[i,1] ~ dunif(xlim[1],xlim[2])
  s[i,2] ~ dunif(ylim[1],ylim[2])
  for (j in 1:J){
    d[i,j] <- pow(pow(s[i,1]-X[j,1],2) + pow(s[i,2]-X[j,2],2),0.5)
    y[i,j] ~ dbin(p[i,j],K)
    p[i,j] <- z[i]*p0*exp(- alpha1*d[i,j]*d[i,j])
  }
}

N <- sum(z[])
D <- N/3.454275
}
```

Abbildung 9.1: **BUGS**-Beschreibung des Modells *SCR0* unter Daten-Vergrößerung (vgl. Royle et al., 2014, S.145f)

9.4. Das SCR-Modell für unbekanntes N

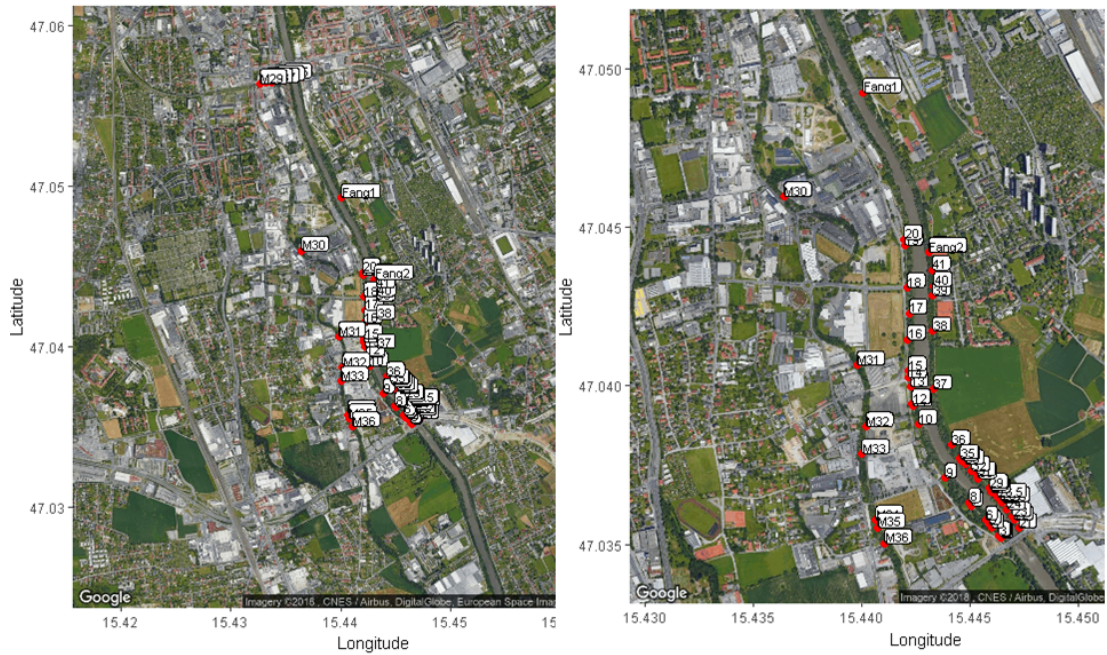


Abbildung 9.2: Lage der 58 künstlichen Versteckplätze im Jahr 2014.
Rechts: Zoom auf das Kerngebiet.

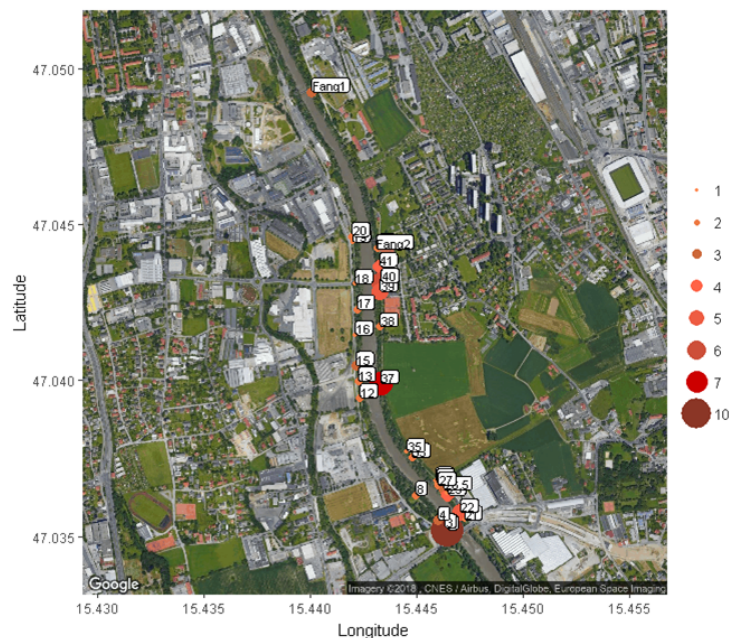


Abbildung 9.3: Häufigkeiten der Würfelnatter-Funde im Jahr 2014.

Kapitel 9. Spatiale Capture-Recapture Modelle

```
> wn_matrix <- read.csv("Matrix_Indiv_Traps_2014.csv",header=TRUE,sep=";")
> wn_matrix <- wn_matrix[,2:dim(wn_matrix)[2]]
> traplocs <- read.csv("Versteckplaetze_2014_utm.csv",header=TRUE,sep=";")
> traplocs <- traplocs[,2:3] # Matrix 58 x 2 mit Koordinaten der kVs

# Koordinaten auf Kilometer skalieren
> traplocs_scale <- cbind(traplocs[,1]*0.001,traplocs[,2]*0.001)
> buffer <- 0.1 # 100m in jede Richtung
> xlim <- c(min(traplocs_scale[,1] - buffer),
            max(traplocs_scale[,1] + buffer))
> ylim <- c(min(traplocs_scale[,2] - buffer),
            max(traplocs_scale[,2] + buffer))

> y <- as.matrix(wn_matrix) # Matrix 70 x 58 der Häufigkeiten
> nind <- dim(y)[1] # 70 Individuen im Jahr 2014
> K <- 31 # 31 Begehungen im Jahr 2014
> J <- dim(y)[2] # 58 Versteckplätze

> M <- 550
> y <- rbind(y,matrix(0,nrow=M-nind,ncol=ncol(y)))
> z <- c(rep(1,nind),rep(0,M-nind))
# Zustandsraum definieren
> sst <- cbind(runif(M,xlim[1],xlim[2]),runif(M,ylim[1],ylim[2]))
> for(i in 1:nind){
  sst[i,1] <- mean( traplocs_scale[y[i,]>0,1] )
  sst[i,2] <- mean( traplocs_scale[y[i,]>0,2] )
}

> set.seed(2013) # jedes Mal die selben Resultate
> library(R2WinBUGS) # R2WinBUGS laden
# Daten setzen
> data <- list (y=y, X=traplocs_scale, K=K, M=M, J=J, xlim=xlim, ylim=ylim)
# initialisieren
> inits <- function(){
  list (alpha0=rnorm(1,-4,.4), alpha1=runif(1,1,4), s=sst, z=z)
}
> parameters <- c("alpha0","alpha1","sigma","N","D") # Parameter setzen
> out <- bugs(data, inits, parameters, model.file = "SCROb.txt", n.chains=3,
             n.iter=6000, n.burnin=2000, n.thin=1, debug=TRUE,
             working.dir=getwd(),
             bugs.directory = 'C:/Program_Win/WinBUGS14')
> print(out,digits=2)
```

```

Inference for Bugs model at "SCR0b.txt", fit using WinBUGS,
3 chains, each with 6000 iterations (first 2000 discarded)
n.sims = 12000 iterations saved

      mean   sd  2.5%   25%   50%   75%   97.5%  Rhat  n.eff
alpha0  -7.49  0.32  -8.12  -7.71  -7.48  -7.28  -6.88  1.00  1200
alpha1   2.46  1.34   0.55   1.42   2.21   3.27   5.59  1.00   900
sigma    0.51  0.17   0.30   0.39   0.48   0.59   0.96  1.00   900
N        364.67 87.78 206.00 297.00 362.00 431.00 528.00 1.02  120
D        105.57 25.41  59.64  85.98 104.80 124.80 152.90 1.02  120
deviance 1006.87 40.72 929.40 978.70 1006.00 1035.00 1088.00 1.00  530

```

For each parameter, n.eff is a crude measure of effective sample size, and Rhat is the potential scale reduction factor (at convergence, Rhat=1).

DIC info (using the rule, $pD = \text{var}(\text{deviance})/2$)
 $pD = 825.9$ and $DIC = 1832.7$
DIC is an estimate of expected predictive error (lower deviance is better).

Für den anfangs definierten Zustandsraum liefert das SCR-Modell einen a-posteriori-Mittelwert von 364.67 für N . Damit ist dieses Ergebnis etwas geringer als jenes aus dem individuellen Kovariablen-Modell aus Abschnitt 8.3.1. Das kann damit zusammenhängen, dass im individuellen Kovariablen-Modell das Gebiet eines Kreises mit Radius 2 Kilometer betrachtet wurde. Somit ist die Kreisfläche größer als die rechteckige Fläche von 3.45 Quadratkilometern in diesem Fall. Die Dichte in Bezug auf den gewählten Zustandsraum liegt bei 105.57 Individuen pro Quadratkilometer. Wählt man den `buffer` größer, so wächst die Fläche des Zustandsraums und damit auch der Umfang der Würfelnatter-Population. Da sich die Würfelnatter jedoch hauptsächlich entlang der Mur und damit nicht außerhalb der Fallen aufhält, wurde der Bereich um die Versteckplätze nicht sehr weit vergrößert. Abbildung 9.4 beinhaltet die a-posteriori-Verteilung für das spatiale Modell. Die abgeschnittene Dichte von N bedeutet, dass M erhöht werden sollte. Dies wird den aktuellen Rechenaufwand von zirka drei Stunden nochmals deutlich erhöhen.

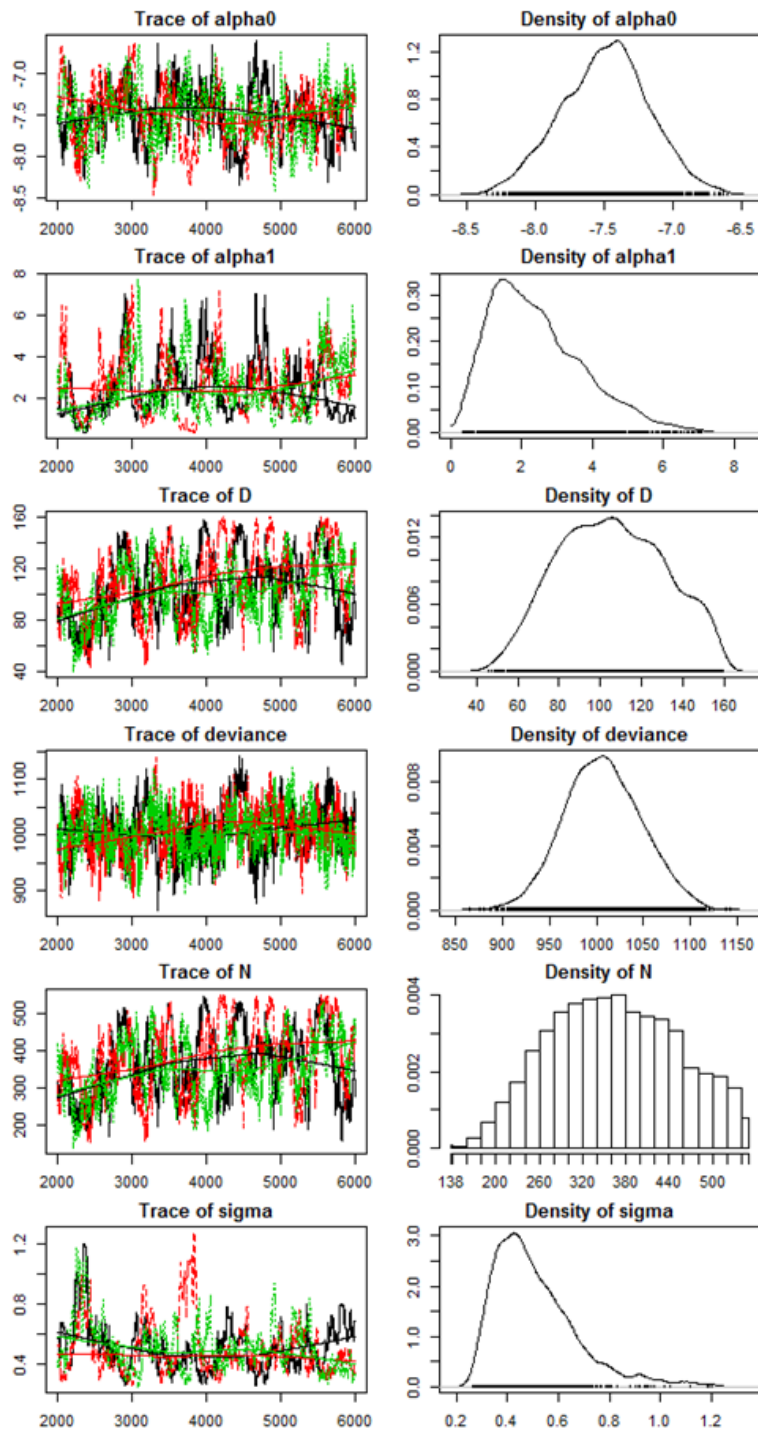


Abbildung 9.4: A-posteriori-Verteilung für Modell *SCR0*.

Der Datenvergrößerungsparameter M wird nun wie im vorigen Kapitel auf 1400 gesetzt.

```
Inference for Bugs model at "SCR0b.txt", fit using WinBUGS,
3 chains, each with 6000 iterations (first 2000 discarded)
n.sims = 12000 iterations saved

      mean      sd   2.5%   25%   50%   75%   97.5% Rhat n.eff
alpha0  -7.49   0.30  -8.11  -7.69  -7.48  -7.29  -6.94 1.01 4400
alpha1   2.55   1.60   0.61   1.26   2.15   3.52   6.46 1.02  940
sigma    0.52   0.18   0.28   0.38   0.48   0.63   0.90 1.02  940
N        376.76 131.43 194.00 279.00 354.00 449.00 689.00 1.02  200
D        109.07  38.05  56.16  80.77 102.50 130.00 199.50 1.02  200
deviance 1007.35 40.13 931.40 980.20 1006.00 1032.00 1091.00 1.00 3800
```

For each parameter, n.eff is a crude measure of effective sample size, and Rhat is the potential scale reduction factor (at convergence, Rhat=1).

```
DIC info (using the rule, pD = var(deviance)/2)
pD = 804.7 and DIC = 1812.0
DIC is an estimate of expected predictive error (lower deviance is better).
```

Auch in diesem Fall konvergieren die Markov-Ketten für 6000 Iterationen. Allerdings liegt der Rechenaufwand dafür bei etwa sechs Stunden. Das Summary der a-posteriori-Verteilung liefert als Mittelwert für N 376.76. Die Werte scheinen also tatsächlich unabhängig von der Wahl von M zu sein. Die Dichte der Würfelnat-ter-Population ist mit 109.07 Individuen pro Quadratkilometer ebenfalls mit dem vorigen Modell vergleichbar. Wird nun noch die a-posteriori-Verteilung grafisch veranschaulicht (Abbildung 9.5), so lässt sich erkennen, dass die Dichtefunktion von N nicht mehr abgeschnitten ist. Damit wurde M ausreichend groß gewählt und das Ergebnis ist zufriedenstellend.

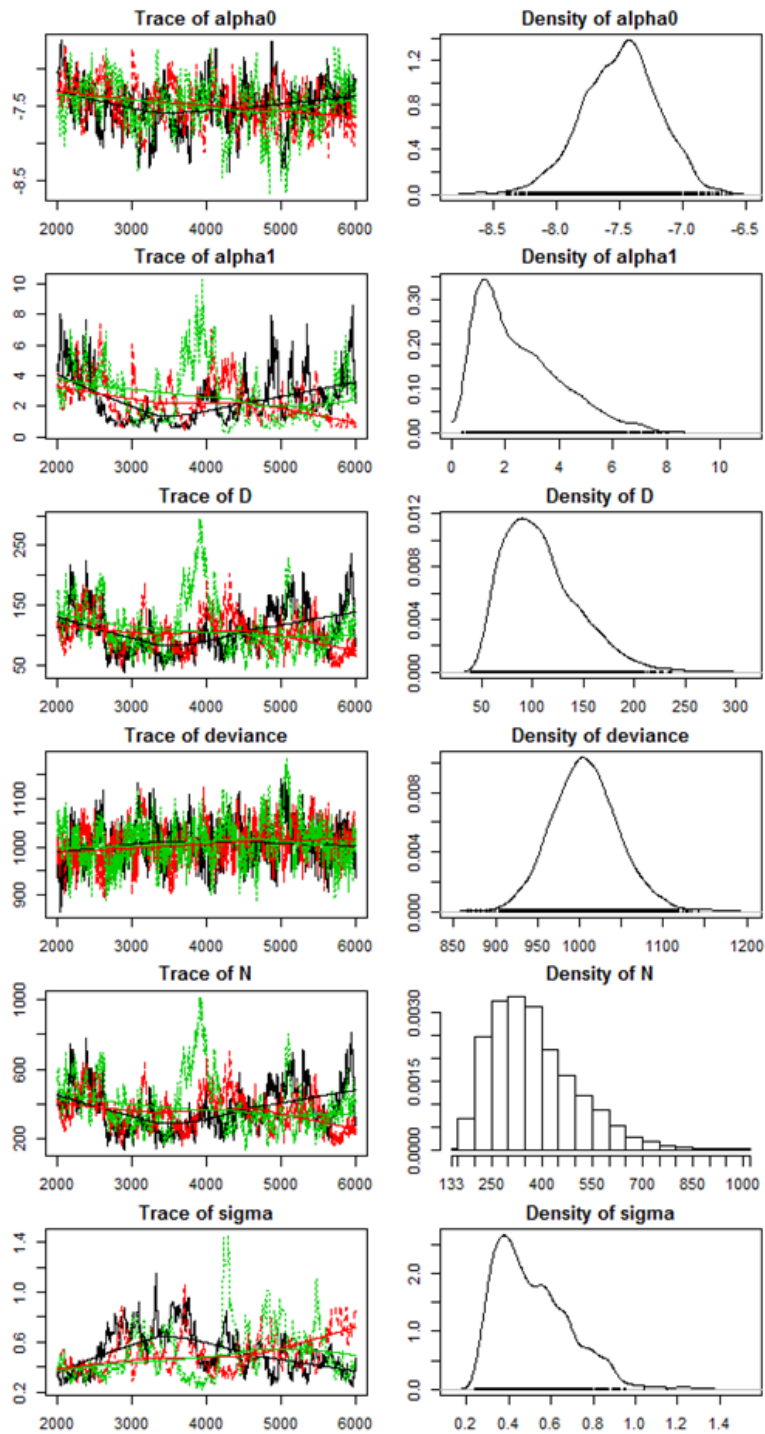


Abbildung 9.5: A-posteriori-Verteilung für Modell *SCR0* mit vergrößertem M .

Offene Populationsmodelle

Um die Entwicklung innerhalb einer Population zu verstehen, müssen sowohl spatiale als auch zeitliche Veränderungen in der Populationsdichte und in den demografischen Parametern betrachtet werden. Zur Populationsdynamik zählen unter anderem das Überleben, der Zuwachs und Bewegungen über einen größeren Zeitraum. Oft ist es notwendig, eine seltene oder schwer zu erfassende Spezies über eine längere Zeitperiode zu untersuchen, damit genug Fänge und Wiederfänge aufgezeichnet werden können. Dadurch würden die Annahmen von geschlossenen Populationsmodellen schnell verletzt werden. Daher kann in diesen Fällen ein offenes Populationsmodell verwendet werden, um mögliche Veränderungen der Population im Laufe der Zeit zu berücksichtigen. Ein offenes Modell, das mit individuellen Effekten und individuellen Kovariablen umgehen kann, ist das Jolly-Seber Modell (Royle et al., 2014, S.401f). Zuerst werden jedoch ein paar Hintergründe zur Entwicklung einer Population angegeben.

10.1 Populationsentwicklungen

Die Populationsgröße wird von vier fundamentalen Prozessen gesteuert:

- Geburten B ,

- Zuwanderungen I ,
- Todesfälle D und
- Auswanderungen E .

Somit ist die Population zum Zeitpunkt $t+1$ eine Funktion dieser vier Komponenten

$$N(t+1) = N(t) + B(t) + I(t) - D(t) - E(t),$$

wobei $N(t)$ die Populationsgröße zum Zeitpunkt t ist, $B(t)$ und $I(t)$ die Geburten und Zuwanderungen zum Zeitpunkt t bezeichnen und $D(t)$ und $E(t)$ die Todesfälle und Auswanderungen. Ein einfaches Modell zum Populationswachstum nimmt an, dass es keine Zuwanderung und keine Auswanderung gibt und lautet

$$N(t+1) = N(t) + N(t)r(t).$$

Dabei bezeichnen $b(t)$ und $d(t)$ die Geburten- und Sterberaten pro Kopf und $r(t) = b(t) - d(t)$ ist die Wachstumsrate pro Kopf.

Während man in geschlossenen Populationsmodellen an der geschätzten Populationsgröße N interessiert ist, möchte man in offenen Populationsmodellen die Entwicklung der Population zwischen verschiedenen Jahren oder Saisonen herausfinden. Somit fokussiert man sich hierbei nicht nur auf $N(t)$ sondern auch auf die Prozesse, die Veränderungen in der Population steuern. Es kann oft schwierig sein, zwischen Geburten und Zuwanderungen sowie zwischen Todesfällen und Auswanderungen zu unterscheiden, da die gesammelten Daten nichts über das Schicksal eines Individuums aussagen.

Nichtsdestotrotz definiert man die Wahrscheinlichkeit, dass ein Individuum vom Zeitpunkt t zum Zeitpunkt $t+1$ überlebt als $\phi(t)$. Die Wahrscheinlichkeit, dass ein Individuum vom Zeitpunkt t zum Zeitpunkt $t+1$ stirbt, lautet dann $1 - \phi(t)$. Der Zuwachs $\gamma(t)$ wird als die Wahrscheinlichkeit definiert, dass ein neues Individuum vom Zeitpunkt t zum Zeitpunkt $t+1$ zur Population hinzukommt, unabhängig davon, ob es dort geboren wurde oder zugewandert ist. Aufgrund dieses fehlenden Unterscheidungsmerkmals ist es nicht möglich, präzise Hypothesen über die Populationsentwicklung zu machen. Verweilt andererseits ein neues Individuum nur für einen kurzen Zeitraum in der Population und verlässt diese danach wieder, scheint es, als sei es in der Population aufgenommen worden, obwohl es vielleicht gestorben ist. Wird die Bewegung jedoch direkt mit spatialen Capture-Recapture Modellen behandelt, so können diese Einflüsse von Tieren, die in das Gebiet der Fallen kommen und wieder weggehen, besser verstanden werden. Dadurch können die Schätzungen des Überlebens verbessert werden. Außerdem haben es beispielsweise jüngere Mitglieder einer Population schwieriger, sich in das bestehende Lebensumfeld einzugliedern,

wenn dieses begrenzt ist. Daher kann das Überleben einer bestimmten Altersklasse direkt durch die Populationsstruktur beeinflusst werden. Weiters können Konkurrenzkämpfe dazu führen, dass sich Individuen auf andere Gebiete ausbreiten (Royle et al., 2014, S.402ff). Um das alles zu berücksichtigen werden im Folgenden traditionelle und spatiale Jolly-Seber Modelle betrachtet.

10.2 Traditionelle Jolly-Seber Modelle

Das *Jolly-Seber Modell* (JS Modell) kann nicht nur Erfassungen und Häufigkeiten schätzen, sondern auch Überleben und Zuwachs. Dazu gibt es primäre Ereignisse, wie Jahre oder Saisonen, zwischen denen die Population „offen“ sein darf. Das heißt, dass Individuen in den primären Ereignissen die Population durch Geburten, Todesfälle, Einwanderung oder Auswanderung betreten oder verlassen können. Innerhalb eines primären Ereignisses wird allerdings ein geschlossener Prozess angenommen. Werden in einem primären Ereignis mehrere Stichproben gezogen (z.B. verschiedene Begehungen), so spricht man von sekundären Ereignissen. Dabei können eine Veränderung in der Entdeckungswahrscheinlichkeit modelliert und die Schätzungen von N verbessert werden.

Damit ist man bereits in der Lage, ein nicht spatiales JS Modell zu formulieren: Seien y_{ikt} die Fanggeschichten für Individuen i im sekundären Ereignis k während primärem Ereignis t , mit $i = 1, \dots, S$, $k = 1, \dots, K$, $t = 1, \dots, T$. Handelt es sich um einen Bernoulli Begegnungsprozess, so lautet das Beobachtungsmodell konditional auf den „Lebenszustand“ $z(i, t)$ für Individuen i zur primären Zeit t

$$y_{ikt} | z(i, t) \sim \text{Bernoulli}(p_t z(i, t)).$$

Somit sind die Beobachtungen Bernoulli mit Entdeckungswahrscheinlichkeit p_t , wenn ein Individuum i zum Zeitpunkt t am Leben ist, also wenn $z(i, t) = 1$. Ist das i -te Individuum jedoch tot, also $z(i, t) = 0$, dann müssen die Beobachtungen mit Wahrscheinlichkeit Eins fixe Nullen sein. Die Variable z ist in diesem Modell nicht die Variable aus der Daten-Vergrößerung von zuvor.

Das Modell für die Zustandsvariablen wird konditional auf die Gesamtanzahl der Individuen, die während des Betrachtungszeitraums am Leben sind (später mit N bezeichnet), basierend auf T Perioden beschrieben. Das Zustandsmodell setzt sich aus den folgenden zwei Komponenten zusammen: Dem Anfangszustand

$$z(i, 1) \sim \text{Bernoulli}(\psi)$$

und dem Übergang eines individuellen Zustandes vom Zeitpunkt t zum Zeitpunkt $t + 1$

$$z(i, t + 1) \sim \text{Bernoulli}(\phi_t z(i, t) + \gamma_t(1 - z(i, t))).$$

Diese Parametrisierung führt zu $T - 1$ Überlebens- und Zuwachsparemetern. Die Populationsgröße N sollte für jedes t geschätzt werden. Da das eben beschriebene Zustandsmodell auf das unbekannte N bedingt ist, verwendet man wiederum das Prinzip der Daten-Vergrößerung von zuvor (Royle et al., 2014, S.404ff).

Im Unterschied zur Daten-Vergrößerung in den geschlossenen Populationsmodellen, ist bei offenen Modellen auf zwei Aspekte aufzupassen:

1. Die Daten-Vergrößerung muss groß genug sein, um alle möglichen lebenden Individuen in der Population während der gesamten Periode aufzunehmen.
2. Individuen können nicht sterben und wieder in die Population eintreten.

Die Daten-Vergrößerung reparametrisiert das Modell, indem $N(1)$ - die Anzahl der lebenden Individuen im ersten primären Ereignis - durch M ersetzt wird. Es gilt dafür $N(1) \sim \text{Binomial}(M, \psi)$. Die Zuwachs-Parameter γ_t sind relativ zur Anzahl der verfügbaren neuen Individuen in der vergrößerten Liste mit Umfang M und stehen nicht direkt mit der Populationsgröße in Beziehung. Die Populationsgröße N_t und die Anzahl neuer Individuen R_t im Jahr t können als Funktion der latenten Zustandsvariablen $z(i, t)$ abgeleitet werden. Diese lauten

$$N_t = \sum_{i=1}^M z(i, t)$$

und

$$R_t = \sum_{i=1}^M (1 - z(i, t - 1))z(i, t).$$

Die Anzahl neuer Individuen umfasst somit alle nicht lebenden Individuen zum Zeitpunkt $t - 1$ und alle lebenden Individuen zum Zeitpunkt t .

Meist gibt es mehr als zwei primäre Ereignisse. Dazu muss eine weitere latente Variable $A(i, t)$ eingeführt werden, um sicherzustellen, dass ein Individuum nur einmal in die Population eintreten kann. Wenn ein Individuum i zum Zeitpunkt t in die Population einwandern kann, gilt $A(i, t) = 1$, andernfalls ist $A(i, t) = 0$. Für $i = 1, \dots, M$ und $t > 1$ folgt

$$A(i, t) = 1 - I \left(\sum_{t=1}^{t-1} z(i, t) > 0 \right).$$

Jeder Zuwachs-Indikator ist bedingt darauf, ob ein Individuum bereits vorher gelebt hat. Dadurch kann ein Individuum kein zweites Mal in die Population einwandern. Die Zustandsvariablen $z(i, t)$ können nun durch einen Markov-Prozess beschrieben werden. Der Anfangszustand für $t = 1$ lautet

$$A(i, 1) \equiv z(i, 1)$$

und für die folgenden Zustände gilt

$$z(i, t) | z(i, t-1) \sim \text{Bernoulli}(\phi_t z(i, t-1) + \gamma_t A(i, t)).$$

Ist zum Zeitpunkt t ein Individuum in der Population, gilt also $z(i, t) = 1$, dann ist der Status des Individuums zum Zeitpunkt $t + 1$ das Ergebnis einer Bernoulli Zufallsvariable mit Überlebenswahrscheinlichkeit ϕ_t . Ist das Individuum hingegen vor dem Zeitpunkt t noch nie in der Population gewesen, das heißt $A(i, t) = 1$, dann ist die Zuwachswahrscheinlichkeit γ_t . Dieses Modell kann mit Hilfe der Funktionen `sum()` und `step()` in **JAGS** (*Just another Gibbs sampler*) implementiert werden. Die Funktion `step()` ist ein logischer Test und gibt 1 zurück, wenn `step(x ≥ 0)` wahr ist, und 0 sonst (Royle et al., 2014, S.407f). Die Modellspezifikation des nicht spatialen Jolly-Seber Modells ist in Abbildung 10.1 zu finden.

```
model{
psi ~ dunif(0,1)
phi ~ dunif(0,1)
p.mean ~ dunif(0,1)

for (t in 1:T){
  N[t] <- sum(z[1:M,t])
  gamma[t] ~ dunif(0,1)
}

for (i in 1:M){
  z[i,1] ~ dbern(psi)           # Lebenszustand für das erste Jahr
  cp[i,1] <- z[i,1]*p.mean
  Y[i,1] ~ dbinom(cp[i,1],K)   # Y sind die Anzahl der Fänge
  A[i,1] <- (1-z[i,1])
}
```

```

for (t in 2:T){
  # for-Schleife für die Jahre 2 bis T
  a1[i,t] <- sum(z[i,1:t]) # Summe aller Lebenszustände von 1 bis t
  A[i,t] <- 1-step(a1[i,t] - 1)
  # A zeigt an, ob ein Individuum in die Population
  # aufgenommen werden kann.
  mu[i,t] <- (phi*z[i,t-1]) + (gamma[t]*A[i,t-1])
  # Der Lebenszustand zum Zeitpunkt t ist abhängig
  # von phi und gamma.
  z[i,t] ~ dbern(mu[i,t])
  cp[i,t] <- z[i,t]*p.mean
  Y[i,t] ~ dbinom(cp[i,t],K)
}
}
}

```

Abbildung 10.1: **JAGS**-Beschreibung des nicht spatialen Jolly-Seber Modells unter Daten-Vergrößerung (siehe Royle et al., 2014, S.409)

10.2.1 Jolly-Seber Modell für die Würfelnatter-Daten

Damit für die Würfelnatter-Daten ein Jolly-Seber Modell gerechnet werden kann, müssen alle vier Jahre 2013 bis 2016 betrachtet werden. Dazu wird eine Matrix Y_{bin} konstruiert, die alle 195 Individuen als Zeilen und die vier Jahre als Spalten besitzt. Die Einträge sind die Fanghäufigkeiten eines Individuums pro Jahr (Details dazu stehen im Anhang A.3). Für die Daten-Vergrößerung werden die Zeilen der Matrix anschließend - analog zu den vorigen Kapiteln - auf $M = 1400$ erweitert. Die Konstante `max_tage` beinhaltet die maximalen Begehungstage in einem Jahr. In diesem Fall sind dies 31 Tage aus dem Jahr 2014. Das ist allerdings sehr groß, wenn man bedenkt, dass im Jahr 2016 nur an 10 Tagen Begehungen stattgefunden haben. Dennoch wird das Jolly-Seber Modell auf die Würfelnatter-Daten angewendet. Die Startwerte und der Modellaufruf lauten wie folgt

```

> zst <- c(rep(1, M/2), rep(0, M/2)) # Startwerte für die Lebenszustände z
> zst <- cbind(zst, zst, zst, zst)

> set.seed(2013) # jedes Mal die selben Resultate
> inits <- function() list(z=zst, gamma=runif(4,0,1)) # initialisieren
> parameters <- c("psi", "N", "phi", "p.mean", "gamma") # Parameter setzen
> data <- list(Y=Ybin, M=M, T=4, K=max_tage) # Daten setzen
> library(rjags) # rjags laden

```

```
> out1 <- jags.model("modelNSJS.txt", data, inits, n.chains=3, n.adapt=500)
> out2NSJS <- coda.samples(out1, parameters, n.iter=2000)
```

Das Summary liefert als Ausgabe:

```
Iterations = 501:2500
Thinning interval = 1
Number of chains = 3
Sample size per chain = 2000
```

1. Empirical mean and standard deviation for each variable, plus standard error of the mean:

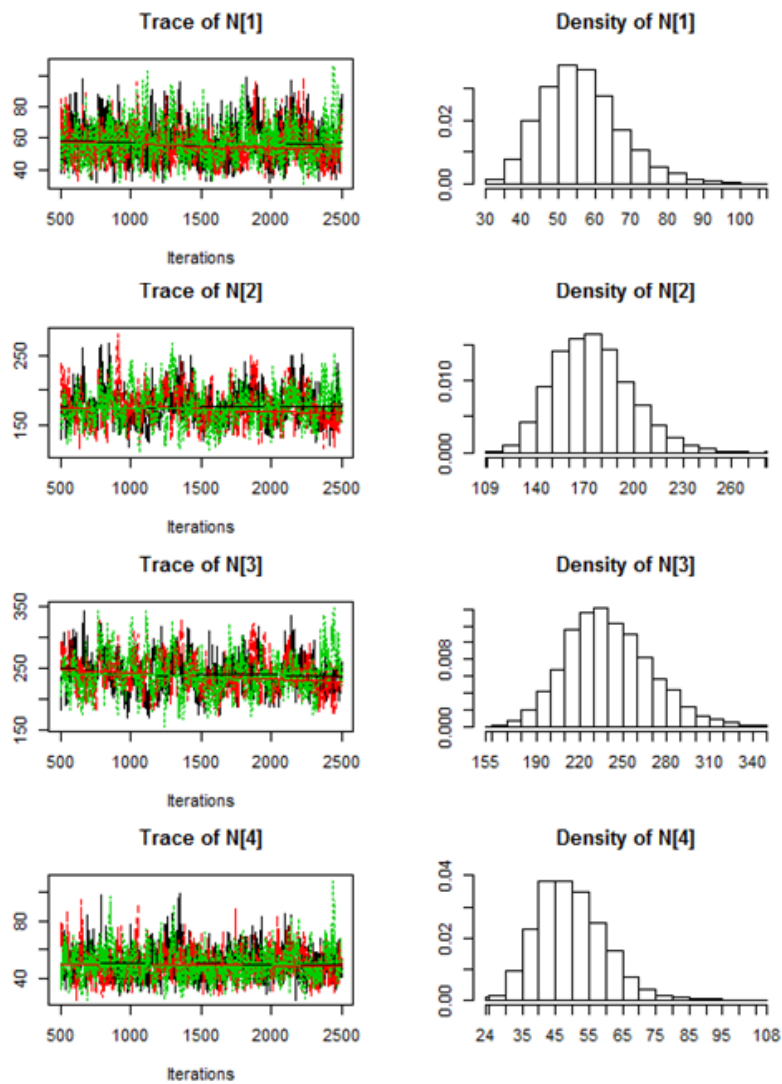
	Mean	SD	Naive SE	Time-series SE
N[1]	56.81217	10.919910	0.1409754	0.4150481
N[2]	175.21433	23.760061	0.3067411	1.4646693
N[3]	240.58150	28.879024	0.3728266	1.7442953
N[4]	50.04300	10.168106	0.1312697	0.4368336
gamma[1]	0.49573	0.289334	0.0037353	0.0036800
gamma[2]	0.12422	0.019765	0.0002552	0.0010792
gamma[3]	0.18048	0.027035	0.0003490	0.0015713
gamma[4]	0.01813	0.010487	0.0001354	0.0005484
p.mean	0.01662	0.002076	0.0000268	0.0001281
phi	0.15145	0.038790	0.0005008	0.0020089
psi	0.04126	0.009352	0.0001207	0.0003139

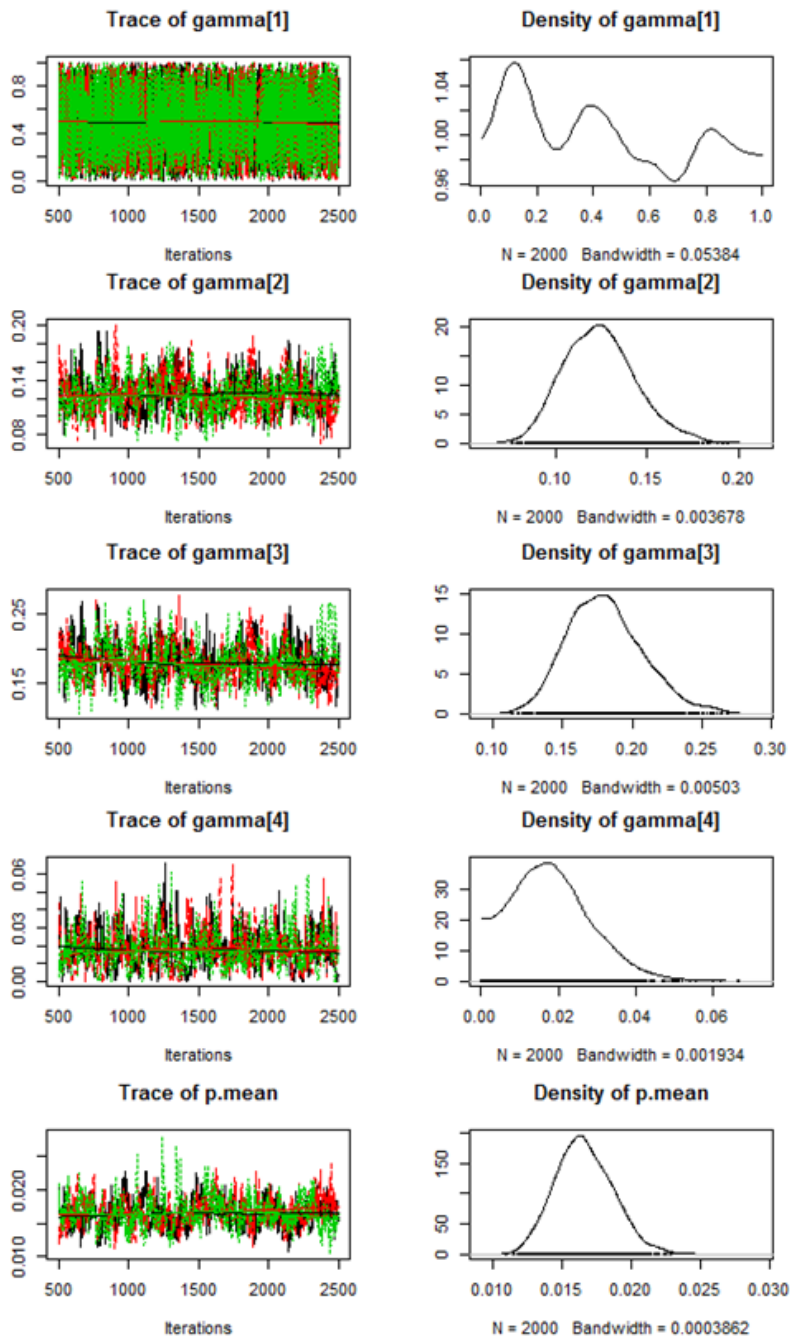
2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
N[1]	3.900e+01	49.000000	56.000000	63.000000	81.000000
N[2]	1.340e+02	158.000000	174.000000	190.000000	227.000000
N[3]	1.900e+02	220.000000	238.000000	259.000000	303.000000
N[4]	3.300e+01	43.000000	49.000000	56.000000	72.000000
gamma[1]	2.610e-02	0.24379	0.49010	0.75036	0.97445
gamma[2]	8.914e-02	0.10995	0.12333	0.13659	0.16758
gamma[3]	1.335e-01	0.16111	0.17860	0.19762	0.24018
gamma[4]	1.086e-03	0.01055	0.01732	0.02448	0.04124
p.mean	1.285e-02	0.01518	0.01649	0.01799	0.02088
phi	8.262e-02	0.12397	0.14907	0.17725	0.23369
psi	2.519e-02	0.03471	0.04051	0.04716	0.06183

Für das Jahr 2014 liefert das Jolly-Seber Modell eine geschätzte Populationsgrö-

ße von $N(2) = 175.21$. Dieser Wert ist vergleichbar mit jenen aus den geschlossenen Populationsmodellen M_0 und M_h aus Abschnitt 8. Für die restlichen Jahre sollte man mit der Interpretation der Ergebnisse vorsichtig sein, da sich die Anzahl der Begehungstage teilweise sehr von denen aus dem Jahr 2014 unterscheiden. Um die Konvergenz des Jolly-Seber Modells zu überprüfen, kann die Gelman-Rubin Diagnostik `gelman.diag()` verwendet werden. Der Test gibt die Skalierungs-Reduktionsfaktoren für jeden Parameter an. Sind diese kleiner als 1.1, so liegt ausreichende Konvergenz vor. Im Falle des Jolly-Seber Modells für die Würfelnatter-Daten scheinen die 2000 Iterationen somit ausreichend zu sein. Durch Betrachtung der a-posteriori-Verteilung (Abbildung 10.2) erkennt man zusätzlich, dass M groß genug gewählt wurde und die Histogramme für $N(t)$ nicht abgeschnitten sind.





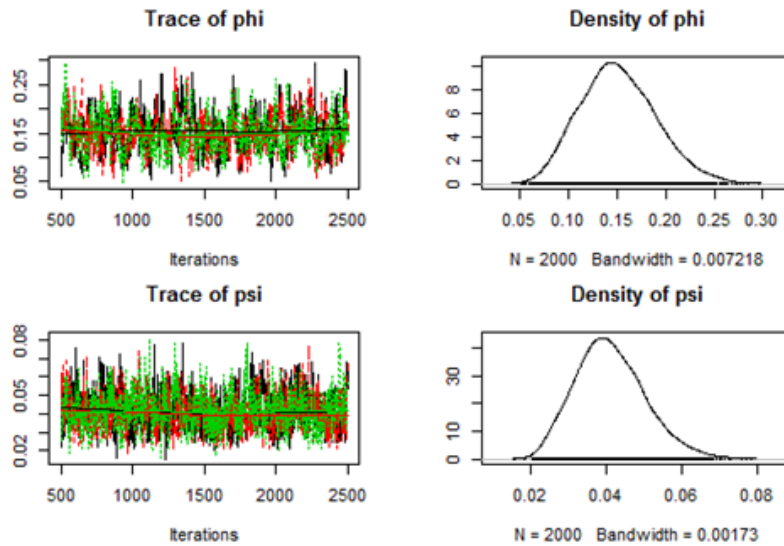


Abbildung 10.2: A-posteriori-Verteilung für Modell *NSJS*.

10.3 Spatiale Jolly-Seber Modelle

Ein großer Nachteil von traditionellen Jolly-Seber Modellen in der Schätzung von N ist der fehlende spatiale Kontext, wodurch es nicht möglich ist, die Populationsdichte zu berechnen. Damit daraus ein spatiales Modell entstehen kann, werden die Informationen über die Positionen der Versteckplätze in die Erfassungsfunktion aufgenommen. Man verwendet als Basis das SCR-Modell aus der geschlossenen Population, um die Erfassungsparameter und anfängliche Populationsgröße zu berechnen, und erweitert dieses um die offenen Komponenten (Übergang von $z(i, t)$ zu $z(i, t + 1)$) wie im nicht spatialen JS Modell. Somit lautet das auf $z(i, t)$ bedingte Bernoulli Beobachtungsmodell

$$y_{ijkt} | z(i, t) \sim \text{Bernoulli}(p_{ijk} z(i, t)),$$

wobei

$$p_{ijk} = p_0 \exp(-\alpha_1 d_{ij}^2) \tag{10.1}$$

und $d_{ij} = \|\mathbf{x}_j - \mathbf{s}_i\|$ den Abstand zwischen Aktivitätsmittelpunkt \mathbf{s}_i und Versteckplatz \mathbf{x}_j bezeichnet. Hierbei ist p_0 wieder die zugrundeliegende Fangwahrscheinlichkeit für ein Individuum, dessen Lebensraummittelpunkt direkt an einem Versteckplatz liegt. Auch der Abstandskoeffizient für das Gauß-Modell ist wiederum $\alpha_1 = 1/(2\sigma^2)$, wobei σ den Skalierungsparameter bezeichnet.

Das Modell wird für den Zeitpunkt $T = 1$ initialisiert. Danach wird der Übergang des individuellen Zustandes vom Zeitpunkt t zum Zeitpunkt $t + 1$ als

$$z(i, t + 1) \sim \text{Bernoulli}(\phi_t z(i, t) + \gamma_t(1 - z(i, t)))$$

modelliert. Analog zu vorhin kommen die latenten Indikator-Variablen $A(i, t)$ zum Einsatz. Es ist $A(i, t) = 1$, wenn ein Individuum i im Zeitintervall $(t - 1, t)$ in die Population eintreten kann, und $A(i, t) = 0$ andernfalls.

Die Anzahl der neuen Individuen kann auf Basis des Lebenszustandes der vorherigen Zeitschritte $(1, 2, \dots, t - 1)$ und des aktuellen Zeitschrittes (t) berechnet werden. Für die Neuzugänge im Zeitraum 1 auf 2 werden beispielsweise jene Individuen gezählt, die zum Zeitpunkt 1 nicht in der Population waren, das heißt $z(i, 1) = 0$, aber zum Zeitpunkt 2 am Leben sind, somit gilt $z(i, 2) = 1$. Um zu bestimmen, ob ein Individuum i die Population zum Zeitpunkt $t = 2$ betreten hat, kann die Struktur $R_{i,2} = (1 - z(i, 1))z(i, 2)$ verwendet werden. Diese wird bis M aufsummiert und gibt damit die Gesamtanzahl an möglichen Neuzugängen an. Dieser Schritt kann für alle primären Ereignisse einer Untersuchung gemacht werden. Die Aktivitätsmittelpunkte werden für jedes primäre Ereignis unabhängig modelliert, sodass $s(i, t) \sim \text{Uniform}(\mathbf{S})$ (Royle et al., 2014, S.410ff). Die detaillierte Modellspezifikation dafür ist in Abbildung 10.3 zu finden.

Da das spatiale Jolly-Seber Modell davon ausgeht, dass in den einzelnen primären Ereignissen immer die selben Versteckplätze vorliegen und diese jeweils auf Individuen untersucht werden, stößt man hierbei für die Würfelnatter-Daten an die Grenzen. In den einzelnen Jahren war die Anzahl der untersuchten Versteckplätze immer verschieden. Hinzu kommt außerdem wieder die Problematik der unterschiedlichen Anzahl an Begehungen in den betrachteten vier Jahren. Somit ist es nicht möglich, ein aussagekräftiges spatiales Jolly-Seber Modell für die Würfelnatter-Daten zu berechnen. Allerdings verwendet das JS Modell jenes Gauß-Modell, das bereits in Kapitel 9 für das spatiale CR-Modell verwendet wurde. Und nachdem die primären Ereignisse in sich geschlossen sind, wird erwartet, dass die geschätzte Populationsgröße 376.76 aus dem SCR-Modell in etwa dem Wert für $N(2)$ aus dem spatialen Jolly-Seber Modell entspricht, wenn die zugrundeliegende Fläche gleich groß gewählt wird. Im vorhin betrachteten nicht spatialen Jolly-Seber Modell ergab sich die selbe Erkenntnis: Der Wert $N(2) = 175.21$ entspricht den Ergebnissen aus den geschlossenen Populationsmodellen M_0 beziehungsweise M_h .

```

model{
psi ~ dunif(0,1)
phi ~ dunif(0,1)
alpha0 ~ dnorm(0,10)
sigma ~ dunif(0,200)
alpha1 <- 1/(2*sigma*sigma)
Area <- ((xlim[2]-xlim[1]))*((ylim[2]-ylim[1])) # Fläche des Zustandsraums

for (t in 1:T){
  N[t] <- sum(z[1:M,t]) # Populationsgröße für jedes Jahr
  D[t] <- N[t]/Area # Populationsdichte für jedes Jahr
  R[t] <- sum(R[1:M,t]) # Anzahl der Neuzugänge für jedes Jahr
  gamma[t] ~ dunif(0,1) # Prior für den Zuwachs-Parameter
}

for (i in 1:M){
  z[i,1] ~ dbern(psi) # Lebenszustand für das erste Jahr
  R[i,1] <- z[i,1] # Schätzungen für die Anzahl der Neuzugänge
  R[i,2] <- (1-z[i,1])*z[i,2]
  R[i,3] <- (1-z[i,1])*(1-z[i,2])*z[i,3]
  R[i,4] <- (1-z[i,1])*(1-z[i,2])*(1-z[i,3])*z[i,4]
  R[i,5] <- (1-z[i,1])*(1-z[i,2])*(1-z[i,3])*(1-z[i,4])*z[i,5]

  for (t in 1:T){ # for-Schleife für die Jahre 1 bis T
    S[i,1,t] ~ dunif(xlim[1],xlim[2])
    S[i,2,t] ~ dunif(ylim[1],ylim[2])

    for (j in 1:ntraps){
      d[i,j,t] <- pow(pow(S[i,1,t]-X[j,1],2)
                     + pow(S[i,2,t]-X[j,2],2),1)
    }

    for (k in 1:K){
      for (j in 1:ntraps){
        lp[i,k,j,t] <- exp(alpha0 - alpha1*d[i,j,t])*z[i,t]
        cp[i,k,j,t] <- lp[i,k,j,t]/(1+sum(lp[i,k,,t]))
      }
      cp[i,k,ntraps+1,t] <- 1-sum(cp[i,k,1:ntraps,t])
      Ycat[i,k,t] ~ dcat(cp[i,k,,t])
    }
  }
}
A[i,1] <- (1-z[i,1])

```



```
for (t in 2:T){  
  a1[i,t] <- sum(z[i,1:t])  
  A[i,t] <- 1-step(a1[i,t] - 1)  
  mu[i,t] <- (phi*z[i,t-1]) + (gamma[t]*A[i,t-1])  
  z[i,t] ~ dbern(mu[i,t])  
}  
}  
}
```

Abbildung 10.3: **JAGS**-Beschreibung des spatialen Jolly-Seber Modells unter Daten-Vergrößerung (vgl. Royle et al., 2014, S.413)

KAPITEL 11

Resümee

Das Ziel dieser Masterarbeit bestand darin, unterschiedliche Schätzmethoden für Bestandsgrößen vorzustellen. Dazu wurde das Capture-Recapture Verfahren angewendet, das aus einigen wenigen Aufzeichnungen bezüglich Fängen und Wiederfängen von Individuen einer Population die gesamte Populationsgröße schätzen kann. Als praktische Grundlage dienten die Würfelnatter-Daten, die Erfassungshäufigkeiten aus den Jahren 2013 bis 2016 dokumentieren.

Im ersten Teil wurden ausgehend von Power Series Verteilungen Populationsgrößenschätzer \hat{N} und Varianzen $\text{Var}(\hat{N})$ für den Poissonfall, den Binomialfall und den geometrischen Fall einer geschlossenen Population hergeleitet. Die betrachteten Schätzer sind der Chao-, der biaskorrigierte Chao- und der Zelterman-Schätzer. Für die Würfelnatter-Daten aus dem Jahr 2014 sind die Schätzer und Standardfehler (in Klammer) in der nachfolgenden Tabelle aufgelistet.

$\hat{N} \left(\sqrt{\widehat{\text{Var}}(\hat{N})} \right)$	Chao	Chao bc	Zelterman
Poisson	287.6 (96.7)	260.1 (78.2)	294.7 (101.4)
Binomial	280.5 (96.6)	254.0 (75.7)	287.6 (98.2)
Geometrisch	505.1 (192.2)	450.2 (155.1)	516.3 (202.8)

Tabelle 11.1: Populationsgrößenschätzer und Standardfehler der Power Series Verteilungen für die Würfelnatter-Daten aus 2014.

Es wurde festgestellt, dass das Ergebnis aus der Binomialverteilungsannahme das vertrauenswürdigste für einen homogenen Populationsgrößenschätzer ist.

Darauf aufbauend konnten im anschließenden zweiten Teil statistische Modelle zur Bestimmung der Populationsgröße betrachtet werden. Dabei dienten geschlossene Populationsmodelle als Basis für die Entwicklung weiterer Modelle. Es wurden drei Methoden vorgestellt: ein einfaches Modell M_0 , ein Modell mit individueller Heterogenität M_h und ein individuelles Kovariablen-Modell M_x . Letzteres verwendete bereits die geografischen Koordinaten der Fallen beziehungsweise Versteckplätze der Individuen. Um in weiterer Folge ein vollständig spatiales Capture-Recapture Modell zu erhalten, wurden zusätzlich die Lebensraummittelpunkte der Individuen anhand einer vorgegebenen Fläche bestimmt. Die Schätzung der Populationsgröße steht hierbei in direktem Zusammenhang mit der betrachteten Fläche. Zum Abschluss wurden noch offene Populationsmodelle untersucht und das traditionelle Jolly-Seber Modell auf die Würfelnatter-Daten angewendet. Alle Schätzungen für die Würfelnatter aus dem Jahr 2014 sind in folgender Tabelle dargestellt:

Populationsmodelle		\hat{N}	$\sqrt{\widehat{\text{Var}}(\hat{N})}$
Geschlossene Modelle	M_0	191.8	37.9
	M_h	186.3	31.6
	M_x	420.3	91.8
Spatiale CR-Modelle	$SCR0$	376.8	131.4
Offene Modelle	$NSJS$	175.2	23.8

Tabelle 11.2: Populationsgrößenschätzer und Standardfehler der Populationsmodelle für die Würfelnatter-Daten aus 2014.

Offene Populationsmodelle berechnen im Prinzip geschlossene Populationsschätzer für die einzelnen betrachteten Jahre und zusätzlich die Entwicklung der Population im Verlauf dieser Jahre. Deshalb ist es nicht verwunderlich, dass die Schätzung aus dem offenen Modell jenen aus den geschlossenen Modellen M_0 oder M_h ähnelt. Bei den spatialen Modellen muss beachtet werden, dass die zugrundeliegende Fläche des Lebensraums der Population genau definiert werden muss. Je größer die Fläche gewählt wird, desto größer werden auch die Schätzungen ausfallen.

Zusammenfassend lässt sich sagen, dass die Populationsgrößenschätzer aus den Power Series Verteilungen bereits eine erste Orientierung über die zu erwartenden Schätzungen aus den Populationsmodellen liefern. Je nach Modell und geografischer Betrachtung erhält man unterschiedlich große Schätzungen für die Populationsgröße.

A.1 Gestutzte Verteilung

Unter gestutzten Verteilungen (*truncated distributions*) versteht man Verteilungen, bei denen die Zufallsvariable nur bestimmte Werte des gesamten Trägers annehmen kann. Der Träger wird quasi um gewisse Bereiche reduziert. Damit man nach solch einer Reduktion wieder eine gültige Verteilung erhält, muss die ursprüngliche Verteilung durch die Wahrscheinlichkeit, die zum gestutzten Bereich gehört, dividiert bzw. skaliert werden. Die Fläche unter der gestutzten Dichte- oder Wahrscheinlichkeitsfunktion wird so wieder Eins. Wird nun die Verteilung einer diskreten Zufallsvariable X betrachtet, bei der man nur am Wertebereich größer k interessiert ist, dann kann deren Wahrscheinlichkeitsfunktion geschrieben werden als

$$\mathbb{P}(X = x|X > k) = \frac{\mathbb{P}(X = x)}{1 - \mathbb{P}(X \leq k)}, \quad \text{für } x = k + 1, k + 2, \dots$$

Dabei nennt man die Verteilung von X eine um den Wertebereich kleiner gleich k gestutzte Verteilung. Die Verteilungsfunktion für Zufallsvariablen, bei denen man nur an Werten kleiner k interessiert ist, kann analog angeschrieben werden. Wird die Verteilung nur um einen Wert k gestutzt, so gilt für die k -gestutzte Verteilung

$$\mathbb{P}(X = x|X \neq k) = \frac{\mathbb{P}(X = x)}{1 - \mathbb{P}(X = k)}, \quad \text{für } x = 0, \dots, k - 1, k + 1, \dots$$

In der Analyse von Capture-Recapture Modellen ist die Häufigkeit f_0 unbekannt. Das motiviert die Betrachtung einer Verteilung der Häufigkeiten, die um die Null gestutzt ist (*zero-truncated distribution*), nämlich

$$\mathbb{P}(X = x|X > 0) = \frac{\mathbb{P}(X = x)}{1 - \mathbb{P}(X = 0)}. \quad (\text{A.1})$$

Wenn X Poissonverteilt ist, d.h. $X \sim Poi(p)$, so gilt für die Null-gestutzte Verteilung

$$\mathbb{P}(X = x|X > 0) = \frac{\mathbb{P}(X = x)}{1 - \mathbb{P}(X = 0)} = \frac{\exp(-p) \frac{p^x}{x!}}{1 - \exp(-p)}$$

wobei $x = 1, 2, \dots$

Ist X hingegen binomialverteilt, d.h. $X \sim Bin(m, p)$, dann erhält man

$$\mathbb{P}(X = x|X > 0) = \frac{\mathbb{P}(X = x)}{1 - \mathbb{P}(X = 0)} = \frac{1}{1 - (1-p)^m} \binom{m}{x} p^x (1-p)^{m-x},$$

wobei $x = 1, 2, \dots, m$.

Im Falle, dass X geometrisch verteilt ist, d.h. $X \sim G(p)$, so gilt

$$\mathbb{P}(X = x|X > 0) = \frac{\mathbb{P}(X = x)}{1 - \mathbb{P}(X = 0)} = \frac{(1-p)^x p}{1-p} = (1-p)^{x-1} p,$$

wobei $x = 1, 2, \dots$

Bei der Betrachtung der Ratios von Power Series Verteilungen und der einzelnen Populationsgrößenschätzer inklusive Varianzen wird die Null-gestutzte Verteilung angewendet (Böhning et al., 2017, S.156f).

A.2 Delta-Methode

In der Statistik wird die Delta-Methode (Δ -Methode) verwendet, um die ungefähre Wahrscheinlichkeitsverteilung für eine Funktion eines asymptotisch normalen statistischen Schätzers aus der Kenntnis der Grenzvarianz dieses Schätzers zu beschreiben. Die Grundidee ist es, eine Taylorreihen-Entwicklung zu verwenden, um eine lineare Funktion zu erhalten, die eine kompliziertere Funktion approximiert.

Univariate Delta-Methode (Casella und Berger, 2002, S.243)

Eine Motivation dieser Technik kann einfach mit univariaten Termen gezeigt werden.

Satz A.1. Sei Y_n eine Folge von Zufallsvariablen, die

$$\sqrt{n}(Y_n - \theta) \xrightarrow{D} N(0, \sigma^2),$$

genügt, wobei θ und σ^2 endliche Konstanten sind und \xrightarrow{D} Konvergenz in Verteilung bezeichnet. Sei g eine reellwertige Funktion mit $g'(\theta) \neq 0$. Dann gilt

$$\sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{D} N(0, \sigma^2[g'(\theta)]^2).$$

Multivariate Delta-Methode (Casella und Berger, 2002, S.245)

Per Definition konvergiert ein konsistenter Schätzer \mathbf{Y}_n mit $\mathbf{Y}_n = (Y_{n1}, Y_{n2}, \dots, Y_{nk})^T$ in Wahrscheinlichkeit gegen seinen wahren Wert $\boldsymbol{\theta}$, wobei $\dim(\boldsymbol{\theta}) = k$. Oft kann der zentrale Grenzwertsatz angewendet werden, um die asymptotische Normalität

$$\sqrt{n}(\mathbf{Y}_n - \boldsymbol{\theta}) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Sigma})$$

zu erreichen, wobei n die Anzahl der Beobachtungen und $\boldsymbol{\Sigma}$ eine (symmetrische und positiv semi-definite) Kovarianz-Matrix ist.

Angenommen man möchte die Varianz einer Funktion $h : \mathbb{R}^k \rightarrow \mathbb{R}$ des Schätzers \mathbf{Y}_n schätzen. Indem man nur die ersten zwei Terme der Taylor-Reihe behält und die Vektor-Notation für den Gradienten verwendet, kann $h(\mathbf{Y}_n)$ approximiert werden als

$$h(\mathbf{Y}_n) \approx h(\boldsymbol{\theta}) + \nabla h(\boldsymbol{\theta})^T \cdot (\mathbf{Y}_n - \boldsymbol{\theta}).$$

Das impliziert die Varianz von $h(\mathbf{Y}_n)$ approximativ als

$$\begin{aligned} \text{Var}(h(\mathbf{Y}_n)) &\approx \text{Var}(h(\boldsymbol{\theta}) + \nabla h(\boldsymbol{\theta})^T \cdot (\mathbf{Y}_n - \boldsymbol{\theta})) \\ &= \text{Var}(h(\boldsymbol{\theta}) + \nabla h(\boldsymbol{\theta})^T \cdot \mathbf{Y}_n - \nabla h(\boldsymbol{\theta})^T \cdot \boldsymbol{\theta}) \\ &= \text{Var}(\nabla h(\boldsymbol{\theta})^T \cdot \mathbf{Y}_n) \\ &= \nabla h(\boldsymbol{\theta})^T \cdot \text{Var}(\mathbf{Y}_n) \cdot \nabla h(\boldsymbol{\theta}) \\ &= \nabla h(\boldsymbol{\theta})^T \cdot \frac{1}{n} \boldsymbol{\Sigma} \cdot \nabla h(\boldsymbol{\theta}). \end{aligned}$$

Damit lässt sich folgender Satz formulieren:

Satz A.2. Sei $\mathbf{Y}_n = (Y_{n1}, Y_{n2}, \dots, Y_{nk})^T$ eine Folge von Zufallsvektoren, sodass

$$\sqrt{n}(\mathbf{Y}_n - \boldsymbol{\theta}) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Sigma}).$$

Sei $h: \mathbb{R}^k \rightarrow \mathbb{R}$ und sei $\nabla h(\mathbf{y})^T = \left(\frac{\partial h}{\partial y_1}, \frac{\partial h}{\partial y_2}, \dots, \frac{\partial h}{\partial y_k} \right)^T$. Mit $\nabla h(\boldsymbol{\theta})^T \neq \mathbf{0}$ gilt dann

$$\sqrt{n}(h(\mathbf{Y}_n) - h(\boldsymbol{\theta})) \xrightarrow{D} N(\mathbf{0}, \nabla h(\boldsymbol{\theta})^T \cdot \boldsymbol{\Sigma} \cdot \nabla h(\boldsymbol{\theta})).$$

A.3 Details zum Programm Code

A.3.1 Beschreibung der Matrix \mathbf{X} aus Abschnitt 3.2

Wie in Abschnitt 3.2 bereits kurz erwähnt, besitzt die Matrix \mathbf{X} (*wn_data.csv*) die Dimension 195×85 . Dabei stehen in den Zeilen die 195 identifizierten Individuen und in den Spalten die 85 Begehungstage aller vier Jahre 2013 bis 2016. Die Anzahl der Begehungen für die einzelnen Jahre sind sehr unterschiedlich: 15 im Jahr 2013, 31 im Jahr 2014, 29 im Jahr 2015 und 10 im Jahr 2016. Außerdem unterscheiden sich auch die Monate, an denen Begehungen stattgefunden haben. Während in den Jahren 2014 und 2015 von Mai bis September Würfelnettern gefangen wurden, fanden in den Jahren 2013 und 2016 die Begehungen mit einigen Lücken statt. Der Eintrag in der Datenmatrix ist eine 1, wenn eine Würfelnetter entdeckt werden konnte, und eine 0 sonst.

A.3.2 Code zu Beispiel 4.1

```
> X <- read.csv("wn_data.csv", header=TRUE, sep = ";")
# z.B. X01.05.2013 ersetzen durch 01.05.2015
> colnames(X) <- lapply(colnames(X), function(x) {gsub("X", "", x)})
# in Datum umwandeln
> colnames(X) <- as.Date(colnames(X), "%d.%m.%Y")

# Jahr 2014 filtern
> j14 <- X[,grep("2014", colnames(X))]
> n_14 <- colSums(j14); y_14 <- rowSums(j14)
> f_14 <- table(y_14)[dimnames(table(y_14))$y_14!="0"]
> S_14 <- sum(f_14); t_14 <- length(n_14)
> r_poi_14 <- rep(0,4); r_bino_14 <- rep(0,4); r_geom_14 <- rep(0,4)
```



```
# Ratios berechnen
> for (i in 1:4){
  # Poisson
  r_poi_14[i] <- ((i+1)*f_14[i+1])/f_14[i]
  # Binomial
  r_bino_14[i] <- ((i+1)/(t_14-i))*(f_14[i+1]/f_14[i])
  # Geometrisch
  r_geom_14[i] <- f_14[i+1]/f_14[i]
}

# Ratio-Plot
> plot(1:4, r_poi_14, xlab = "x", ylab = "Ratio", ylim = c(0,5), pch = 16,
      col = "blue", cex = 1.5, xaxt="n")
> lines(1:4, r_poi_14, type = 'l', lty = 3, col = "blue")
> axis(1, at=1:4,labels=c("1","2","3","4"), las=0)
> points(1:4, r_bino_14, pch = 15, col = "green", cex = 1.5)
> lines(1:4, r_bino_14, type = 'l', lty = 3, col = "green")
> points(1:4, r_geom_14, pch = 17, col = "orange", cex = 1.5)
> lines(1:4, r_geom_14, type = 'l', lty = 3, col = "orange")
> legend("topleft", c("Poisson" , "Binomial", "Geometrisch"),
      pch = c(16,15,17), col = c("blue", "green", "orange"),
      text.col = "black", cex=0.9, pt.cex = 1.5, bty = "o")
```

A.3.3 Beschreibung der Matrix \mathbf{X} aus Abschnitt 8

Die Datenmatrix `wn_data` entspricht der Matrix \mathbf{X} aus Abschnitt 3.2. Werden daraus das Jahr 2014 gefiltert und nicht benötigte Nullzeilen entfernt, so entsteht der Datensatz `wn_14_neu` der Größe 70 Individuen \times 31 Begehungen. Die Einträge sind wieder 1, wenn ein Individuum gefangen wurde, und 0 andernfalls.

A.3.4 Beschreibung der Versteckplätze aus Abschnitt 8.3.2

Der Datensatz `traps` (*Versteckplaetze_2014.csv*) ist eine Matrix mit Dimension 58×3 . In der ersten Spalte stehen die Bezeichnungen der 58 Versteckplätze aus dem Jahr 2014. Die Spalten 2 und 3 beinhalten deren geografische Koordinaten *Longitude* und *Latitude* in Dezimalgrad.

A.3.5 Beschreibung der Fundorte der Individuen aus Abschnitt 8.3.2

Im Datensatz `indiv` (*WN_Koordinaten_2014_Indiv.csv*) stehen die durchschnittlichen geografischen Fangplätze der 70 Individuen. Diese Matrix besitzt die Dimension 70×3 . Zu den 70 Bezeichnungen der Individuen in Spalte 1 stehen in Spalten 2 und 3 die geografischen Koordinaten *Longitude* und *Latitude* in Dezimalgrad.

A.3.6 Beschreibung der Datenmatrix aus Abschnitt 9

Der Datensatz `wn_matrix` (*Matrix_Indiv_Traps_2014.csv*) für die spatialen Capture-Recapture Modelle sieht etwas anders aus. In den Zeilen sind die 70 Individuen aufgelistet und in den Spalten stehen die einzelnen 58 Versteckplätze. Die Matrix hat daher die Dimension 70×58 mit den Häufigkeiten der Würfelnetter-Funde als Einträge.

A.3.7 Beschreibung der Versteckplätze aus Abschnitt 9

Die Datenmatrix `traplocs` (*Versteckplaetze_2014_utm.csv*) ist analog zum Datensatz `traps` aus Abschnitt 8.3.2 aufgebaut. Allerdings wurden die GPS-Koordinaten von Dezimalgrad auf UTM-Koordinaten (*Universal Transverse Mercator*) umgerechnet. Statt *Longitude* und *Latitude* lauten die Bezeichnungen hier *Easting* (*Rechtswert*) und *Northing* (*Hochwert*). Damit ist es einfacher, Abstände zu berechnen. Die Koordinaten werden in Metern angegeben und können mit dem Faktor 0.001 auf Kilometer umgerechnet werden.

A.3.8 Beschreibung der Datenmatrix aus Abschnitt 10

Zur Konstruktion der Matrix `Ybin` mit der Dimension 195×4 verwendet man die Datenmatrix `X` aus Abschnitt 3.2 und filtert nach den einzelnen Jahren 2013 bis 2016. Anschließend können für alle 195 Individuen die Zeilensummen über die einzelnen Jahre gebildet werden, um die Fanghäufigkeiten eines Individuums pro Jahr zu erhalten.

Literaturverzeichnis

- Böhning, D. (2008). A simple variance formula for population size estimators by conditioning. *Statistical Methodology*, 5, 410-423.
- Böhning, D. (2010). Some general comparative points on Chao's and Zelterman's estimators of the population size. *Scandinavian Journal of Statistics*, 37, 221-236.
- Böhning, D. (2018). Correction of: A simple variance formula for population size estimators by conditioning. *Statistical Methodology*, 5, 410-423.
- Böhning, D., Baksh, M. F., Lerdsuwansri, R. und Gallagher, J. (2013). Use of the ratio plot in capture-recapture estimation. *Journal of Computational and Graphical Statistics*, 22, 135-155.
- Böhning, D., van der Heijden, P. G. M. und Bunge, J. (2017). *Capture-Recapture Methods for the Social and Medical Sciences*. Boca Raton: CRC Press.
- Casella, G. und Berger, R. L. (2002). *Statistical Inference* (Second Edition). Pacific Grove: Duxbury: Thomson Learning.
- Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, 43, 783-791.
- Edwin, T. K. (2014). *Power Series Distributions and Zero-Inflated Models* (Unveröffentlichte Dissertation). University of Nairobi.
- Kammel, W., Erhart, C., Fauland, A., Heimburg, H., Koglek, G., Praschk, C., ... Gebhardt, O. (2016). *Bestandserhebungen der Würfelnatter und der Schlingnatter an der Mur im südlichen Stadtgebiet von Graz 2013-2015* (Bericht). Österreichische Naturschutzjugend.

- Kaskasamkul, P. (2017). *Capture-Recapture Estimation and Modelling for One-Inflated Count Data* (Unveröffentlichte Dissertation). University of Southampton. (Version September 2017)
- McCullagh, P. und Nelder, J. A. (1989). *Generalized Linear Models* (Second Edition, Bd. 37). London: Chapman & Hall.
- Momeni, F. (2011). The generalized power series distributions and their application. *The Journal of Mathematics and Computer Science*, 2, 691-697.
- Robert, C. P. und Casella, G. (1999). *Monte Carlo Statistical Methods*. New York: Springer.
- Royle, J. A., Chandler, R. B., Sollman, R. und Gardner, B. (2014). *Spatial Capture-Recapture*. Oxford: Academic Press.
- Spendel, S. (2016). *Einführung in die Capture-Recapture Methode* (Unveröffentlichte Diplomarbeit). Technische Universität Graz.
- Vidal-Diez, A. (2015). *Development of Capture-Recapture Estimators in Closed Populations Including Individual Covariate Information* (Unveröffentlichte Dissertation). University of Southampton.