

fakultät statistik

Diplomarbeit

Zeitlich-räumliche Modellierung mit der verallgemeinerten Gamma-Poisson-Mischung

von: Tanja Katrin Rausch

vorgelegt am: 13. März 2012

Erstgutachterin: Prof. Dr. Christine Müller

Inhaltsverzeichnis

1.	Einle	eitung	1			
2.	Statistische Methoden					
	2.1. Grundlagen					
	2.2.	Verallgemeinertes Gamma-Poisson-Mischungsmodell	8			
		2.2.1. Allgemein	8			
		2.2.2. Räumlich	Ĝ			
		2.2.3. Eindimensional	13			
		2.2.4. Eindimensional mit Zeit	13			
		2.2.5. Zwei Richtungen	15			
		2.2.6. Zwei Richtungen mit Zeit	17			
3.	Simulationen					
	3.1.	Eindimensional über die Zeit	23			
	3.2.	Räumlich	26			
	3.3.	Zwei Richtungen	29			
	3.4.	Zwei Richtungen mit Zeit	32			
4.	Anw	vendung in der Biometrie	35			
	4.1.	Datenmaterial	35			
	4.2.	Statistische Auswertung	37			
5.	Zusa	ammenfassung	51			
Α.	R-C	ode	53			
В.	MA	TLAB-Code	69			
C.	Tab	ellen und Grafiken	72			
	C.1.	Simulationen	72			
	C.2.	Anwendung	78			
Lit	eratı	urverzeichnis	85			

1. Einleitung

Die Ausbreitung von Rissen in eine bestimmte Richtung in metallischen Flächen ist für die Industrie eine interessante Fragestellung. So könnten Aussagen über die Lebensdauer von stark belasteten Metallen, wie Schienen von Achterbahnen, metallischen Bestandteilen von Hüftprothesen oder anderen Materialien getroffen werden. Am Anfang dieser Diplomarbeit stand ein Anwendungsversuch aus der Technometrie. Dabei wurde eine quadratische Stahlplatte, die an allen vier Seiten befestigt worden ist, immer wieder nach oben und unten gebogen. In regelmäßigen Abständen wurde die Oberfläche dieser Platte fotografiert und die Risse, die in dieser Platte entstanden waren, festgehalten. Wenn nun ein Gitternetz über diese Platte gelegt wird und die Risse gezählt werden, die in den entstandenen Zellen aufzufinden sind, dann sind Voraussetzungen gegeben, um das verallgemeinerte Gamma-Poisson-Mischungsmodell anzuwenden, das von Hwang und He (2011) beschrieben wurde. Hwang und He (2011) und auch Müller et al. (2011) verwenden das verallgemeinerte Gamma-Poisson-Mischungsmodell, um mithilfe einer sogenannten Presence-Absence Map die Gesamtanzahl von Individuen auf einem Gebiet zu schätzen. Bei einer Presence-Absence Map handelt es sich um eine Gitternetzkarte, in der nur vermerkt ist, ob in einer Zelle betrachtete Individuen aufzufinden sind oder nicht. Die Gesamtzahl der Individuen soll in diesem Fall mit dem verallgemeinerten Gamma-Poisson-Mischungsmodell geschätzt werden. In dieser Arbeit ist die Individuenanzahl in so einem Gitternetz allerdings schon bekannt und es soll mithilfe des Modells eine Aussage über die Ausbreitungsrichtung bzw. den Zusammenhang von benachbarten Zellen getroffen werden. In der Literatur wurde das verallgemeinerte Gamma-Poisson-Mischungsmodell in ähnlicher Weise auch in anderen Bereichen angewendet. Burrell (1990) verwendet das Gamma-Poisson-Modell, um eine Vorhersage über Buchumläufe in Bibliotheken zu treffen. Toft et al. (2006) setzen das Gamma-Poisson Modell ein, um zu ermitteln, ob Mikroorganismen zufällig in einer Lebensmittelmatrix verteilt sind. In dieser Diplomarbeit wird später als Anwendungsbeispiel jedoch nicht die am Anfang beschriebenen Risse in einer Stahlplatte aufgegriffen, sondern es werden Daten analysiert, die für die Biometrie interessant sind. Es handelt sich dabei um Daten, die im Krüger-Nationalpark in Südafrika registriert wurden. In den Jahren 1977 - 1997 fanden Zählungen statt, bei denen die Tiere, die in dem Park lebten, gezählt wurden. Neben der Anzahl der Tiere wurde auch die genaue Position im Park festgehalten, an der die Tiere gesichtet wurden. In diesem Fall kann auf den Park ebenfalls ein Gitternetz gelegt werden und bestimmt werden, wie viele Tiere jeweils in den einzelnen Zellen leben. Tosh et al. (2004) haben Presence-Absence Daten aus dem Krüger-Nationalpark verwendet, um die Anzahl an großen Pflanzenfressern im Park zu schätzen. Dazu haben sie unter anderem die negative Binomialverteilung verwendet. Diese beruht auf dem Gamma-Poisson-Mischungsmodell.

Das zentrale Ergebnis aus dieser Diplomarbeit ist, dass sich das verallgemeinerte Gamma-Poisson-Mischungsmodell dazu eignet, eine Aussage über einen geringen Zusammenhang bei benachbarten Zellen zu machen. Außerdem lässt sich das Modell auf das biometrische Anwendungsbeispiel anwenden.

In dieser Diplomarbeit soll überprüft werden, ob sich das verallgemeinerte Gamma-Poisson-Mischungsmodell dazu eignet, Zusammenhänge zwischen benachbarten Zellen von Gitternetzen zu beschreiben. Dazu soll erst überprüft werden, ob dies für eine Zelle über mehrere Zeitpunkte und in einer Matrix in jeweils der x- und y-Richtung der Fall ist. Außerdem wird diese Eigenschaft im räumlichen Kontext überprüft und für eine Datenmatrix, bei der drei Richtungen, die x-, die y- und die Zeitrichtung, betrachtet werden können. Dazu werden Daten für die verschiedenen Situationen auf Grundlage des verallgemeinerten Gamma-Poisson-Mischungmodells mit vorgegebenen Parametern simuliert. Mithilfe der Momentenschätzung sollen aus den simulierten Daten die Parameter wieder herausgeschätzt und mit dem Einstichproben-t-Test die Ergebnisse überprüft werden.

Falls sich bestätigt, dass Zusammenhänge zwischen benachbarten Zellen mit dem verallgemeinerten Gamma-Poisson-Mischungsmodell beschrieben werden können, sollen die verschiedenen Zusammenhänge in einem biometrischen Datensatz mit

dem Modell bestimmt und die Ergebnisse für diesen Datensatz interpretiert werden.

In Kapitel 2 werden die statistischen Methoden vorgestellt, die in dieser Diplomarbeit zum Einsatz kommen. Dabei wird in Kapitel 2.1 auf einige für diese Diplomarbeit wesentlichen Grundlagen eingegangen und in Kapitel 2.2 das verallgemeinerte Gamma-Poisson-Mischungsmodell auf verschiedenen Hintergründen definiert.

In Kapitel 3 werden die durchgeführten Simulationen mit den Modelldefinitionen aus Kapitel 2.2 beschrieben und deren Ergebnisse ausgewertet.

Daraufhin wird in Kapitel 4 die Anwendung des Modells auf den biometrischen Datensatz durchgeführt. Dazu wird als erstes in Kapitel 4.1 der Datensatz vorgestellt und in Kapitel 4.2 mit den in Kapitel 2 vorgestellten und in Kapitel 3 simulierten Modellen ausgewertet.

Die wichtigsten Ergebnisse werden in Kapitel 5 zusammengefasst. Dabei werden auch weitere Überlegungen über zukünftige Anwendungen des in dieser Arbeit vorgestellten Modells und des Datensatzes erläutert.

2. Statistische Methoden

Das verallgemeinerte Gamma-Poisson-Mischungsmodell beruht auf der Poissonund der Gamma-Verteilung. Die Momentenmethode wird zur Schätzung der Parameter angewendet. Die Ergebnisse der Schätzungen werden mit dem Einstichproben-t-Test ausgewertet. In den folgenden Kapiteln wird auf die Verteilungen und die Methoden nochmal kurz eingegangen, bevor die Theorie des verallgemeinerten Gamma-Poisson-Mischungsmodells vorgestellt und erweitert wird.

2.1. Grundlagen

Poisson-Verteilung

Nach Fahrmeir et al. (2007, S. 260ff) eignet sich die Poisson-Verteilung zur Modellierung von Zählvorgängen.

Definition 2.1. Eine Zufallsvariable X heißt poissonverteilt mit Parameter $\lambda > 0$, falls für ihre Wahrscheinlichkeitsfunktion

$$f(x) = P(X = x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda}, & x \in \{0, 1, \dots\} \\ 0, & sonst \end{cases}$$

gilt.

Folgerung 2.2. Sei X eine poissonverteilte Zufallsvariable. Dann gilt:

$$E(X) = \lambda \ und \ Var(X) = \lambda.$$

In dem Softwarepaket R lassen sich poissonverteilte Zufallszahlen mit dem Befehl rpois(n, lambda) erzeugen.

Gammaverteilung

Nach Sachs und Hedderich (2009, S.247ff) lässt sich mit dem Modell der Gamma-Verteilung die Wartezeit bis zum Eintreten eines k-ten Ereignisses beschreiben (Poisson-Prozess).

Definition 2.3. Eine Zufallsvariable X heißt gammaverteilt mit Parameter k > 0 und $\lambda > 0$, falls die Dichtefunktion

$$f(x) = \frac{1}{\lambda^k \cdot \Gamma(k)} x^{k-1} \cdot e^{-x/\lambda}$$

 $f\ddot{u}r \ x > 0 \ ist.$

Der Parameter λ bezeichnet die Intensität des Poisson-Prozesses (Skalenparameter oder scale) und die Form der Verteilung (shape) wird durch den Parameter k bestimmt. Die Gamma-Funktion Γ wird definiert durch $\Gamma(x) := \int_0^\infty x^{x-1} \cdot e^{-t} dt$ für x > 0 und es gilt $\Gamma(n+1) = n!$ für $n \in \mathbb{N}$.

Folgerung 2.4. Sei X eine gammaverteilte Zufallsvariable. Dann gilt:

$$E(X) = \mu = k \cdot \lambda \ und \ Var(X) = \sigma^2 = k \cdot \lambda^2.$$

In dem Softwarepaket R lassen sich gammaverteilte Zufallszahlen mit dem Befehl rgamma(n, shape, rate = 1, scale = 1/rate)) ausgeben.

Momentenschätzer

In Gentschel und Becker (2005, S. 108f) wird die Momentenschätzer folgendermaßen beschrieben. Mit der Momentenmethode können für die Parameter v_1, \ldots, v_k der Verteilungsfunktion $F^X(x; v_1, \ldots, v_k)$ einer Zufallsvariablen X Punktschätzer bestimmt werden. Dabei werden die empirischen und theoretischen Momente gleichgesetzt und durch das Lösen aus dem dabei entstehenden Gleichungssystems Schätzfunktionen, sogenannte Momentenschätzer, für die Parameter v_1, \ldots, v_k zu erhalten.

Definition 2.5. Sei X eine Zufallsvariable. Mit

$$\mu_{(r)} = E(X^r) = \int_{-\infty}^{-\infty} x^r \cdot f^X(x; v_1, \dots, v_k) dx$$

ist das r-te (theoretische) Moment der Zufallsvariable X definiert. Das r-te empirische Moment ist mit

$$m_{(r)} = \frac{1}{n} \sum_{i=1}^{n} X_i^r$$

definiert.

Die Dichtefunktion von X wird mit $f^X(x; v_1, \ldots, v_k)$ bezeichnet.

Bemerkung 2.6. Für r=1 entspricht der Erwartungswert von X E(X) dem ersten Moment $\mu_{(1)}$ und das arithmetische Mittel der Stichprobenvariablen $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ entspricht dem ersten empirischen Moment $m_{(1)}$. Somit gilt insgesamt für r=1

$$\mu_{(1)} = E(X) = \frac{1}{n} \sum_{i=1}^{n} x_i = m_{(1)}.$$

Definition 2.7. X_1, \ldots, X_n seien unabhängig und identisch wie X verteilte Stichprobenvariablen mit Verteilungsfunktion $F^X(x; \vartheta_1, \ldots, \vartheta_k)$. Diese wird durch den
Parametervektor $\vartheta = (\vartheta_1, \ldots, \vartheta_k)$ charakterisiert. Des Weiteren seien die ersten k Momente von X mit $\mu_{(r)}$ bezeichnet und entsprechend die ersten k empirischen
Momente mit $m_{(r)}$ für $r = 1, \ldots, k$. Die Lösungen $\omega_1, \ldots, \omega_k$ des k-elementigen
Gleichungssystems

$$\mu_{(1)} = m_{(1)}$$

$$\mu_{(2)} = m_{(2)}$$

:

$$\mu_{(k)} = m_{(k)}$$

sind die Schätzfunktionen für die k Parameter $\vartheta_1, \ldots, \vartheta_k$ und werden als Momentenschätzer bezeichnet.

Definition 2.8. Das r-te zentrale Moment für r > 1 ist definiert als

$$\mu_{(r)}^z = E[X - E[X]]^r.$$

Das r-te empirische zentrale Moment für r > 1 ist gegeben durch

$$m_{(r)}^z = S_*^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^r.$$

Bemerkung 2.9. Für r=2 entspricht das zweite zentrale Moment $\mu_{(2)}^z$ der Varianz $Var(X)=E(X-E(X))^2$ und das zweite empirische zentrale Moment $m_{(2)}^z$ entspricht der empirischen Varianz $S_*^2=\frac{1}{n}\sum_{i=1}^n(X_i-\bar{X})^2$. Dabei wird jedoch meistens $m_{(2)}^z=S^2=\frac{1}{n-1}\sum_{i=1}^n(X_i-\bar{X})^2$ verwendet, wegen der Erwartungstreue von S^2 . Somit gilt insgesamt für r=2

$$\mu_{(2)}^z = Var(X) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = m_{(2)}^z.$$

Einstichproben-t-Test

In dieser Arbeit wird der zweiseitige Einstichproben-t-Test angewendet. Nach Genschel und Becker (2005, S. 236ff) wird der zweiseitige Einstichproben-t-Test wie folgt definiert:

Definition 2.10. Sei X eine normalverteilte Zufallsvariable mit Erwartungswert $E(X) = \mu$ und Varianz $Var(X) = \sigma^2$ oder beliebig verteilt mit Erwartungswert $E(X) = \mu$ und unbekannter Varianz $Var(X) = \sigma^2$, wobei dann der Stichprobenumfang $n \geq 30$ sein muss. Die zu untersuchende Hypothese lautet $H_0: \mu = \mu_0$ vs. $H_1: \mu \neq \mu_0$, wobei μ der Erwartungswert der Zufallsvariable X ist und μ_0 der der Stichprobe. Die Nullhypothese H_0 wird zum Niveau α abgelehnt, falls der p-Wert kleiner als α ist. Dabei lässt sich der p-Wert mit $2 \cdot P(T \geq |t_{beob}|)$ berechnen, wobei $T = \sqrt{n} \cdot \frac{\overline{X} - \mu_0}{S}$ die Teststatistik ist.

In dieser Definition ist t_{beob} der berechnete Wert der Teststatistik aus den Beobachtungen x_1, \ldots, x_n , S die Standardabweichung und \overline{X} das artihmetische Mittel der Stichprobenvariablen X_1, \ldots, X_n . Auf die Voraussetzungen zur Anwendung des Tests muss in dieser Arbeit besonders geachtet werden, da sie hier nicht unbedingt erfüllt sind. Die Zufallsvariable X ist hier nicht normalverteilt, also muss der Stichprobenumfang jeweils größer als 30 sein.

2.2. Verallgemeinertes Gamma-Poisson-Mischungsmodell

2.2.1. Allgemein

Toft et al. (2006) sagen über das Gamma-Poisson Modell, dass das Modell annimmt, dass Zählungen von Individuen einer Poisson-Verteilung $Poi(\lambda)$ folgen, es aber dem Parameter λ erlaubt, gammaverteilt $\Gamma(a,b)$ zu sein. Das verallgemeinerte Gamma-Poisson-Mischungsmodell wird von Hwang und He (2011) und Müller et al. (2011) wie folgt beschrieben.

Definition 2.11. Gegeben sei eine $I \times J$ -Matrix $(y_{ij})_{i=1,\dots,I,j=1,\dots,J}$ mit $I \cdot J$ Zellen, die ein Gitternetz darstellt. $(X_{ij})_{i=1,\dots,I,j=1,\dots,J}$ sei die Anzahl der Individuen in jeder Zelle. Weiter sei V_{st} mit $s,t=0,\pm 1,\dots$ eine u.i.v. gammaverteilte $\Gamma(a,b)$ Zufallsvariable, so dass $E(V_{st})=ab$ und $Var(V_{st})=ab^2$. Dabei ist a der Formund b der Skalenparameter. Die Zufallsvariable $X_{ij}|V_{st}=v_{st},s,t\in\mathbb{Z}$ ist bedingt poissonverteilt mit Parameter λ_{ij} , wobei

$$\lambda_{ij} = \sum_{s=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} \rho^{|i-s|+|j-t|} v_{st}$$

mit $0 \le \rho < 1$ ist. Für $(i,j) \ne (k,l)$ sind $X_{ij}|V_{st} = v_{st}, s, t \in \mathbb{Z}$ und $X_{kl}|V_{st} = v_{st}, s, t \in \mathbb{Z}$ bedingt stochastisch unabhängig.

Die Zufallsvariable V_{st} ist über dem Gitternetz verteilt, so dass in jeder Zelle eine Variable liegt. Der Parameter ρ beschreibt den Zusammenhang von räumlicher Korrelation über die Zellen, wobei i die Ausrichtung auf der y-Achse/in der Senkrechten und j die Richtung auf der x-Achse/in der Waagerechten ist. Ein großes ρ spricht für eine hohe Abhängigkeit und $\rho = 0$ für keinen Zusammenhang von benachbarten Zellen. λ_{ij} ist eine gemischte Gammaverteilung und ist bei benachbarten Zellen ähnlicher als bei Zellen, die weiter entfernt liegen.

Für die Wahrscheinlichkeitsfunktion von X_{ij} gilt

$$f(k) = P(X_{ij} = x) = \int \frac{e^{-\lambda_{ij}} \lambda_{ij}^x}{x!} g_{\lambda}(\lambda_{ij}) d\lambda_{ij}$$
$$= \int \frac{e^{-\lambda_{ij}} \lambda_{ij}^x}{x!} \prod_{l,m} \{g_v(v_{lm}) dv_{lm}\},$$

wobei g_{λ} und g_{v} die Dichtefunktionen von λ_{ij} und v_{lm} bezeichnen.

Satz 2.12. Für den Erwartungswert von X_{ij} gilt

$$E(X_{ij}) = \left(\frac{1+\rho}{1-\rho}\right)^2 ab.$$

2.2.2. Räumlich

Das verallgemeinerte Gamma-Poisson-Mischungsmodell aus Definition 2.11 beschreibt einen räumlichen Zusammenhang, das heißt keine spezielle Richtung von Zellen in einer Matrix. Zunächst werden die statistischen Kennzahlen zu diesem Modell bestimmt, die für die Momentenmethode benötigt werden. Dazu wird als erstes der Erwartungswert aus Satz 2.12 hergeleitet.

Beweis zu Satz 2.12.

$$E(X_{ij}) = E(E(X_{ij}|V_{st})) = E(\lambda_{ij})$$

$$= E\left(\sum_{s=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} \rho^{|i-s|+|j-t|} V_{st}\right)$$

$$E(aX) = aE(X) \sum_{s=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} \rho^{|i-s|+|j-t|} \underbrace{E(V_{st})}_{=ab}$$

$$= \sum_{s=-\infty}^{\infty} \rho^{|i-s|} \cdot \sum_{t=-\infty}^{\infty} \rho^{|j-t|} ab$$

$$= \sum_{n=-\infty}^{\infty} \rho^{n} + \sum_{n=-\infty}^{\infty} \rho^{n} - \sum_{n=0}^{\infty} \rho^{n} - \rho^{0}$$

$$= \left(\frac{1}{1-\rho} + \frac{1}{1-\rho} - 1\right) \cdot \sum_{t=-\infty}^{\infty} \rho^{|j-t|} ab$$

$$= \left(\frac{1+\rho}{1-\rho}\right) \sum_{t=-\infty}^{\infty} \rho^{|j-t|} ab$$

$$= \left(\frac{1+\rho}{1-\rho}\right)^{2} ab$$

wegen der geometrischen Reihe $\sum_{k=0}^{\infty} z^k = \frac{1}{1-z}$ für |z| < 1 und des bedingten Erwartungswertes E(X) = E(E(X|Y)).

Satz 2.13. Für die Varianz von X_{ij} gilt

$$Var(X_{ij}) = \left(\frac{1+\rho}{1-\rho}\right)^2 ab + \left(\frac{1+\rho^2}{1-\rho^2}\right)^2 ab^2$$

Beweis.

$$Var(X_{ij}) = E(Var(X_{ij}|V_{st})) + Var(E(X_{ij}|V_{st}))$$

$$= E(\lambda_{ij}) + Var(\lambda_{ij})$$

$$= \left(\frac{1+\rho}{1-\rho}\right)^{2} ab + Var\left(\sum_{s=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} \rho^{|i-s|+|j-t|}V_{st}\right)$$

$$Var(aX) = a^{2}Var(X) \left(\frac{1+\rho}{1-\rho}\right)^{2} ab + \left(\sum_{s=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} \rho^{|i-s|+|j-t|}\right)^{2} \underbrace{Var(V_{st})}_{=ab^{2}}$$

$$= \left(\frac{1+\rho}{1-\rho}\right)^{2} ab + \sum_{s=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} \rho^{2(|i-s|+|j-t|)} ab^{2}$$

$$= \left(\frac{1+\rho}{1-\rho}\right)^{2} ab + \sum_{s=-\infty}^{\infty} \rho^{2|i-s|} \sum_{t=-\infty}^{\infty} \rho^{2|j-t|} ab^{2}$$

$$= \left(\frac{1+\rho}{1-\rho}\right)^{2} ab + \left(\frac{1+\rho^{2}}{1-\rho^{2}}\right)^{2} ab^{2}$$

mit der bedingten Varianz Var(X) = E(Var(X|Y)) + Var(E(X|Y)) und analogen Schritten wie im Beweis zu Satz 2.12.

Satz 2.14. Für die Kovarianz $Cov(X_{ij}, X_{(i+1)(j+1)})$ gilt

$$Cov(X_{ij}, X_{(i+1)(j+1)}) = \left(\frac{2\rho}{1-\rho^2}\right)^2 ab^2.$$

Beweis.

$$Cov(X_{ij}, X_{(i+1)(j+1)}) = \underbrace{E(Cov(X_{ij}, X_{(i+1)(j+1)}|V_{st}))}_{=0} + Cov(E(X_{ij}|V_{st}), E(X_{(i+1)(j+1)}|V_{st}))$$

$$= Cov(\lambda_{ij}, \lambda_{(i+1)(j+1)})$$

$$= Cov\left(\sum_{s=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} \rho^{|i-s|+|j-t|}V_{st}, \sum_{s=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} \rho^{|i+1-s|+|j+1-t|}V_{st}\right)$$

$$= \sum_{s=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} \rho^{|i-s|+|j-t|+|i+1-s|+|j+1-t|} \underbrace{Cov(V_{st}, V_{st})}_{=Var(V_{st})=ab^2}$$

$$= \sum_{s=-\infty}^{\infty} \rho^{|i-s|+|i+1-s|} \sum_{t=-\infty}^{\infty} \rho^{|j-t|+|j+1-t|} ab^2$$

$$s=i-s=m, t=j-t=r \left(\sum_{m=0}^{\infty} \rho^{2m+1} + \sum_{n=1}^{\infty} \rho^{2n-1}\right) \left(\sum_{r=0}^{\infty} \rho^{2r+1} + \sum_{u=1}^{\infty} \rho^{2u-1}\right) ab^2$$

$$= \left(\frac{1}{1-\rho^2}\rho + \frac{1}{\rho(1-\rho^2)} - \frac{1}{\rho}\right)^2 ab^2$$

$$= \left(\frac{2\rho}{1-\rho^2}\right)^2 ab^2$$

mit $Cov(a \cdot X, b \cdot Y) = a \cdot b \cdot Cov(X, Y)$, Cov(X, Y) = 0, falls X und Y bedingt stochastisch unabhängig sind, der bedingten Kovarianz

Cov(X,Y) = E(Cov(X,Y|Z)) + Cov(E(X|Z), E(Y|Z)) und mit analogen Schritten wie im Beweis zu Satz 2.12.

Für die Momentenschätzungen werden noch die empirischen Kennwerte benötigt. Dabei kann der Erwartungswert $E(X_{ij})$ durch die Mittelwerte aller x_{ij}

$$\overline{x} = \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} x_{ij}$$

geschätzt werden. Durch die empirische Varianz

$$s^{2} = \frac{1}{IJ - 1} \sum_{i=1}^{I} \sum_{j=1}^{J} (x_{ij} - \overline{x})^{2}$$

kann die Varianz $Var(X_{ij})$ geschätzt werden und die Kovarianz $Cov(X_{ij}, X_{(i+1)(j+1)})$ kann durch die empirische Kovarianz der Paare $(X_{ij}, X_{(i+1)(j+1)}), i = 1, \ldots, I-1, j = 1, \ldots, J$

$$s_{(X_{ij},X_{(i+1)(j+1)})} = \frac{1}{(I-1)(J-1)-1} \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} (x_{ij} - \bar{x}_1)(x_{(i+1)(j+1)} - \bar{x}_2)$$

mit

$$\bar{x}_1 = \frac{1}{(I-1)(J-1)} \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} x_{ij}, \quad \bar{x}_2 = \frac{1}{(I-1)(J-1)} \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} x_{(i+1)(j+1)}$$

geschätzt werden.

Folgerung 2.15. Die Momentenmethode liefert folgende Schätzgleichungen für $a, b \ und \ \rho$:

$$\bar{x} = \left(\frac{1+\rho}{1-\rho}\right)^2 ab \qquad := E(a,b,\rho)$$

$$s^2 = \left(\frac{1+\rho}{1-\rho}\right)^2 ab + \left(\frac{1+\rho^2}{1-\rho^2}\right)^2 ab^2 \qquad := V(a,b,\rho)$$

$$s_{(X_{ij},X_{(i+1)(j+1)})} = \left(\frac{2\rho}{1-\rho^2}\right)^2 ab^2 \qquad := C(a,b,\rho)$$

 \Leftrightarrow

$$\bar{x} - \left(\frac{1+\rho}{1-\rho}\right)^2 ab = 0 \qquad := E(a,b,\rho)$$

$$s^2 - \left(\left(\frac{1+\rho}{1-\rho}\right)^2 ab + \left(\frac{1+\rho^2}{1-\rho^2}\right)^2 ab^2\right) = 0 \qquad := V(a,b,\rho)$$

$$s_{(X_{ij},X_{(i+1)(j+1)})} - \left(\frac{2\rho}{1-\rho^2}\right)^2 ab^2 = 0 \qquad := C(a,b,\rho)$$

Die Lösung der Gleichungssysteme ist äquivalent zur Lösung des Minimierungsproblems

$$\min_{\sigma=(a,b,\rho)} (\bar{x} - E(X_{ij}))^2 + (s^2 - Var(X_{ij}))^2 + (s_{(X_{ij},X_{(i+1)(j+1)})} - Cov(X_{ij},X_{(i+1)(j+1)}))^2.$$

Im folgenden Kapitel 3 wird noch genauer erklärt, aus welchem Grund die Kovarianz ebenfalls bei der Momentenmethode verwendet wird.

2.2.3. Eindimensional

Als nächstes wird, ausgehend von Definition 2.11, die Theorie für das verallgemeinerte Gamma-Poisson-Mischungsmodell für den Fall aufgestellt, dass es nur eine einzige Zelle gibt.

Definition 2.16. Sei (y_{11}) eine 1×1 -Matrix mit $M = 1 \cdot 1 = 1$ Zelle. Dabei sei $X_{11} = X$ die bekannte Anzahl von Individuen in dieser Zelle.

Die Zufallsvariable V sei u.i.v. gammaverteilt mit den Parametern a und b, so dass E(V) = ab und $Var(V) = ab^2$ ist. Die Zufallsvariable X|V = v ist bedingt poissonverteilt mit dem Parameter $\lambda = \rho \cdot v$ mit $0 \le \rho < 1$.

Korollar 2.17. Für den Erwartungswert und die Varianz von X folgt mit Satz 2.12 und Satz 2.13

$$E(X) = \rho \cdot E(V) = \rho \cdot ab$$

$$Var(X) = \rho^2 \cdot ab^2 + \rho \cdot ab.$$

2.2.4. Eindimensional mit Zeit

Die Theorie für eine Zelle aus Definition 2.16 wird im nächsten Schritt auf eine Folge von t Zellen, die eine Zeitreihe darstellen, erweitert.

Definition 2.18. Sei (y_t) eine 1×1 -Matrix mit $M = 1 \cdot 1 = 1$ Zelle zum Zeitpunkt $t, t = 1, \ldots, i$. Dabei sei X_t die bekannte Anzahl von Individuen in der Zelle zum Zeitpunkt t. Die Zufallsvariable $V_n, n = 0, -1, 1, -2, 2, \ldots$ sei u.i.v. gammaverteilt mit den Parametern a und b, so dass $E(V_n) = ab$ und $Var(V_n) = ab^2$ ist. Die Zufallsvariable $X_t|V_n = v_n, n \in \mathbb{Z}$ ist bedingt poissonverteilt mit dem Parameter $\lambda_t = \sum_{n=-\infty}^{\infty} \rho^{|t-n|} \cdot v_n$ mit $0 \le \rho < 1$. Für $t \ne r$ sind $X_t|V_n = v_n, n \in \mathbb{Z}$ und $X_r|V_n = v_n, n \in \mathbb{Z}$ bedingt stochastisch unabhängig.

Der Parameter ρ reflektiert den Zusammenhang der zeitlichen Korrelation zwischen den Zellen.

Korollar 2.19. Für den Erwartungswert von X_t gilt analog zu Satz 2.12

$$E(X_t) = E(\lambda_t) = \sum_{n = -\infty}^{\infty} \rho^{|t-n|} E(V_n) = \sum_{n = -\infty}^{\infty} \rho^{|t-n|} ab = \left(\frac{1+\rho}{1-\rho}\right) ab.$$

Korollar 2.20. Für die Varianz von X_t gilt analog zu Satz 2.13

$$Var(X_t) = E(\lambda_t) + Var(\lambda_t) = \left(\frac{1+\rho}{1-\rho}\right)ab + \sum_{n=-\infty}^{\infty} \rho^{2|t-n|} Var(V_n)$$
$$= \left(\frac{1+\rho}{1-\rho}\right)ab + \left(\frac{1+\rho^2}{1-\rho^2}\right)ab^2.$$

Korollar 2.21. Für die Kovarianz des Paares (X_t, X_{t+1}) gilt analog zu Satz 2.14

$$Cov(X_t, X_{t+1}) = Cov(\lambda_t, \lambda_{t+1}) = \sum_{n=-\infty}^{\infty} \rho^{|t-n|+|t+1-n|} Var(V_n) = \frac{2\rho}{(1-\rho^2)} ab^2.$$

Der Erwartungswert $E(X_t)$ kann geschätzt werden durch die Mittelwerte aller x_t

$$\overline{x} = \frac{1}{i} \sum_{t=1}^{i} x_t$$

und die Varianz $Var(X_t)$ durch die empirische Varianz

$$s^{2} = \frac{1}{i-1} \sum_{t=1}^{i} (x_{t} - \overline{x})^{2}.$$

Durch die empirische Kovarianz der Paare $(X_t, X_{(t+1)}), t = 1, \dots, i-1$

$$s_{(X_t, X_{(t+1)})} = \frac{1}{(i-1)-1} \sum_{t=1}^{i-1} (x_t - \bar{x}_1)(x_{(t+1)} - \bar{x}_2)$$

mit

$$\bar{x}_1 = \frac{1}{(i-1)} \sum_{t=1}^{i-1} x_t, \quad \bar{x}_2 = \frac{1}{(i-1)} \sum_{t=1}^{i-1} x_{(i+1)}$$

kann die Kovarianz $Cov(X_t, X_{(t+1)})$ geschätzt werden.

Folgerung 2.22. Die Momentenmethode liefert folgende Schätzgleichungen für $a, b \ und \ \rho$:

$$\bar{x} = \left(\frac{1+\rho}{1-\rho}\right)ab \qquad := E(a,b,\rho)$$

$$s^2 = \left(\frac{1+\rho}{1-\rho}\right)ab + \left(\frac{1+\rho^2}{1-\rho^2}\right)ab^2 \qquad := V(a,b,\rho)$$

$$s_{(X_t,X_{(t+1)})} = \frac{2\rho}{(1-\rho^2)}ab^2 \qquad := C(a,b,\rho)$$

Die Lösung der Gleichungssysteme ist äquivalent zur Lösung des Minimierungsproblems:

$$\min_{\sigma=(a,b,\rho)} (\bar{x} - E(X_t))^2 + (s^2 - Var(X_t))^2 + (s_{(X_t,X_{(t+1)})} - Cov(X_t,X_{(t+1)}))^2$$

2.2.5. Zwei Richtungen

Nun wird die Theorie für das zuerst vorgestellte räumliche Modell zur Betrachtung von zwei Richtungen übertragen. Dabei soll das Modell die Zusammenhänge der Zellen in die waagerechte und in die senkrechte Richtung betrachten können.

Definition 2.23. Sei $(y_{ij})_{i=1,...,I,j=1,...,J}$ eine IxJ-Matrix mit $M = I \cdot J$ Zellen. Die bekannte Anzahl von Individuen in jeder Zelle sei X_{ij} .

Weiter sei V_{st} , $s,t=0,-1,1,-2,2,\ldots$ eine u.i.v. gamma-verteilte Zufallsvariable mit den Parametern a und b, so dass $E(V_{st})=ab$ und $Var(V_{st})=ab^2$ ist. Die Zufallsvariable $X_{ij}|V_{st}=v_{st},s,t\in\mathbb{Z}$ ist bedingt poissonverteilt mit dem Parameter $\lambda_{ij}=\sum_{s=-\infty}^{\infty}\sum_{t=-\infty}^{\infty}\rho_1^{|i-s|}\rho_2^{|j-t|}v_{st}$ mit $0\leq\rho_1<1$ und $0\leq\rho_2<1$. Für $(i,j)\neq(k,l)$ sind $X_{ij}|V_n=v_n,n\in\mathbb{Z}$ und $X_{kl}|V_n=v_n,n\in\mathbb{Z}$ bedingt stochastisch unabhängig.

Die Parameter ρ_1 und ρ_2 reflektieren die Zusammenhänge der vertikalen und horizontalen Korrelation zwischen den Zellen.

Korollar 2.24. Für den Erwartungswert von X_{ij} gilt analog zu Satz 2.12

$$E(X_{ij}) = E(\lambda_{ij}) = \sum_{s=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} \rho_1^{|i-s|} \rho_2^{|j-t|} E(V_{st}) = \sum_{s=-\infty}^{\infty} \rho_1^{|i-s|} \sum_{t=-\infty}^{\infty} \rho_2^{|j-t|} ab$$
$$= \left(\frac{1+\rho_1}{1-\rho_1}\right) \left(\frac{1+\rho_2}{1-\rho_2}\right) ab.$$

Korollar 2.25. Für die Varianz von X_{ij} gilt analog zu Satz 2.13

$$Var(X_{ij}) = E(\lambda_{ij}) + Var(\lambda_{ij})$$

$$= \left(\frac{1+\rho_1}{1-\rho_1}\right) \left(\frac{1+\rho_2}{1-\rho_2}\right) ab + \sum_{s=-\infty}^{\infty} \rho_1^{2|i-s|} \sum_{t=-\infty}^{\infty} \rho_2^{2|j-t|} Var(V_{st})$$

$$= \left(\frac{1+\rho_1}{1-\rho_1}\right) \left(\frac{1+\rho_2}{1-\rho_2}\right) ab + \left(\frac{1+\rho_1^2}{1-\rho_1^2}\right) \left(\frac{1+\rho_2^2}{1-\rho_2^2}\right) ab^2$$

Korollar 2.26. Für die beiden Kovarianzen der Paare $(X_{ij}, X_{(i+1)j})$ und $(X_{ij}, X_{i(j+1)})$ gilt analog zu Satz 2.14

$$Cov(X_{ij}, X_{(i+1)j}) = Cov(\lambda_{ij}, \lambda_{(i+1)j}) = \sum_{s=-\infty}^{\infty} \rho_1^{|i-s|+|i+1-s|} \sum_{t=-\infty}^{\infty} \rho_2^{2|i-t|} Var(V_{st})$$
$$= \left(\frac{1+\rho_2^2}{1-\rho_2^2}\right) \left(\frac{2\rho_1}{1-\rho_1^2}\right) ab^2$$

und

$$Cov(X_{ij}, X_{i(j+1)}) = Cov(\lambda_{ij}, \lambda_{i(j+1)}) = \sum_{s=-\infty}^{\infty} \rho_1^{2|i-s|} \sum_{t=-\infty}^{\infty} \rho_2^{|i-t|+|i+1-t|} Var(V_{st})$$
$$= \left(\frac{1+\rho_1^2}{1-\rho_1^2}\right) \left(\frac{2\rho_2}{1-\rho_2^2}\right) ab^2.$$

Durch den Mittelwert aller x_{ij}

$$\bar{x} = \frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} x_{ij}$$

kann der Erwartungswert $E(X_{ij})$ geschätzt werden und die Varianz $Var(X_{ij})$ durch die empirische Varianz

$$s^{2} = \frac{1}{IJ - 1} \sum_{i=1}^{I} \sum_{j=1}^{J} (x_{ij} - \bar{x})^{2}.$$

Die Kovarianz $Cov(X_{ij}, X_{(i+1)j})$ kann durch die empirische Kovarianz der Paare $(X_{ij}, X_{(i+1)j}), \quad i = 1, \dots, I-1, \quad j = 1, \dots, J-1$

$$s_{(X_{ij},X_{(i+1)j})} = \frac{1}{(I-1)J-1} \sum_{i=1}^{I-1} \sum_{j=1}^{J} (x_{ij} - \bar{x}_1)(x_{(i+1)j} - \bar{x}_2)$$

mit

$$\bar{x}_1 = \frac{1}{(I-1)J} \sum_{i=1}^{I-1} \sum_{j=1}^{J} x_{ij}, \quad \bar{x}_2 = \frac{1}{(I-1)J} \sum_{i=1}^{I-1} \sum_{j=1}^{J} x_{(i+1)j}$$

geschätzt werden und die Kovarianz $Cov(X_{ij}, X_{i(j+1)})$ durch

$$s_{(X_{ij},X_{i(j+1)})} = \frac{1}{I(J-1)-1} \sum_{i=1}^{I} \sum_{j=1}^{J-1} (x_{ij} - \bar{x}_3)(x_{i(j+1)} - \bar{x}_4)$$

mit

$$\bar{x}_3 = \frac{1}{I(J-1)} \sum_{i=1}^{I} \sum_{j=1}^{J-1} x_{ij}, \quad \bar{x}_4 = \frac{1}{I(J-1)} \sum_{i=1}^{I} \sum_{j=1}^{J-1} x_{i(j+1)}.$$

Folgerung 2.27. Die Momentenmethode liefert folgende Schätzgleichungen für a, b, ρ_1 und ρ_2 :

$$\bar{x} = \left(\frac{1+\rho_1}{1-\rho_1}\right) \left(\frac{1+\rho_2}{1-\rho_2}\right) ab \qquad := E(a,b,\rho_1,\rho_2)$$

$$s^2 = \left(\frac{1+\rho_1}{1-\rho_1}\right) \left(\frac{1+\rho_2}{1-\rho_2}\right) ab + \left(\frac{1+\rho_1^2}{1-\rho_1^2}\right) \left(\frac{1+\rho_2^2}{1-\rho_2^2}\right) ab^2 \qquad := V(a,b,\rho_1,\rho_2)$$

$$s_{(X_{ij},X_{(i+1)j})} = \left(\frac{1+\rho_2^2}{1-\rho_2^2}\right) \frac{2\rho_1}{(1-\rho_1^2)} ab^2 \qquad := C_1(a,b,\rho_1,\rho_2)$$

$$s_{(X_{ij},X_{i(j+1)})} = \left(\frac{1+\rho_1^2}{1-\rho_1^2}\right) \frac{2\rho_2}{(1-\rho_2^2)} ab^2 \qquad := C_2(a,b,\rho_1,\rho_2)$$

Die Lösung der Gleichungssysteme ist äquivalent zur Lösung des Minimierungsproblems:

$$\min_{\sigma=(a,b,\rho_1,\rho_2)} (\bar{x} - E(X_{ij}))^2 + (s^2 - Var(X_{ij}))^2 + (s_{(X_{ij},X_{(i+1)j})} - Cov(X_{ij},X_{(i+1)j}))^2 + (s_{(X_{ij},X_{i(j+1)})} - Cov(X_{ij},X_{i(j+1)}))^2$$

2.2.6. Zwei Richtungen mit Zeit

Als letztes wird noch eine Erweiterung des zuvor vorgestellten Modells aus Kapitel 2.2.6 eingeführt. In diesem Modell werden ebenfalls die vertikale und horizontale Richtung betrachtet und zusätzlich noch eine dritte Richtung, die die Zeit beschreibt.

Definition 2.28. Sei $(y_{ijk})_{i=1,...,I,j=1,...,J,k=1,...,K}$ ein IxJxK-Datenwürfel mit $M=I\cdot J\cdot K$ Zellen. Dabei beschreibt $i=1,\ldots,I$ die senkrechte Richtung, $j=1,\ldots,J$ die waagerechte Richtung und $k=1,\ldots,K$ die Zeitrichtung. Dabei sei X_{ijk} die bekannte Anzahl von Individuen in Zelle (i,j) zum Zeitpunkt k. Weiter sei V_{stu} , $s,t,u=0,-1,1,-2,2,\ldots$ eine u.i.v. gamma-verteilte Zufallsvariable mit den Parametern a und b, so dass $E(V_{stu})=ab$ und $Var(V_{stu})=ab^2$ ist. Die Zufallsvariable $X_{ijk}|V_{stu}=v_{stu},s,t,u\in\mathbb{Z}$ ist bedingt poissonverteilt mit dem Parameter $\lambda_{ijk}=\sum_{s=-\infty}^{\infty}\sum_{t=-\infty}^{\infty}\sum_{u=-\infty}^{\infty}\rho_1^{|i-s|}\rho_2^{|j-t|}\rho_3^{|k-u|}v_{stu}$ mit $0\leq \rho_1<1$, $0\leq \rho_2<1$ und $0\leq \rho_3<1$. Für $(i,j,k)\neq (l,c,d)$ sind $X_{ijk}|V_n=v_n,n\in\mathbb{Z}$ und $X_{lcd}|V_n=v_n,n\in\mathbb{Z}$ bedingt stochastisch unabhängig.

Die Parameter ρ_1 , ρ_2 reflektieren die Zusammenhänge in die vertikale und horizontale Richtung von Nachbarzellen. Der Parameter ρ_3 beschreibt den Zusammenhang der Zellen in die Zeitrichtung.

Korollar 2.29. Für den Erwartungswert $E(X_{ijk})$ gilt analog zu Satz 2.12

$$E(X_{ijk}) = E(\lambda_{ijk}) = \sum_{s=-\infty}^{\infty} \sum_{t=-\infty}^{\infty} \sum_{u=-\infty}^{\infty} \rho_1^{|i-s|} \rho_2^{|j-t|} \rho_3^{|k-u|} E(V_{stu})$$

$$= \sum_{s=-\infty}^{\infty} \rho_1^{|i-s|} \sum_{t=-\infty}^{\infty} \rho_2^{|j-t|} \sum_{u=-\infty}^{\infty} \rho_3^{|k-u|} ab$$

$$= \left(\frac{1+\rho_1}{1-\rho_1}\right) \left(\frac{1+\rho_2}{1-\rho_2}\right) \left(\frac{1+\rho_3}{1-\rho_3}\right) ab.$$

Korollar 2.30. Für die Varianz $Var(X_{ijk})$ gilt analog zu Satz 2.13

$$Var(X_{ijk}) = E(\lambda_{ijk}) + Var(\lambda_{ijk})$$

$$= \left(\frac{1+\rho_1}{1-\rho_1}\right) \left(\frac{1+\rho_2}{1-\rho_2}\right) \left(\frac{1+\rho_3}{1-\rho_3}\right) ab$$

$$+ \sum_{s=-\infty}^{\infty} \rho_1^{2|i-s|} \sum_{t=-\infty}^{\infty} \rho_1^{2|j-t|} \sum_{u=-\infty}^{\infty} \rho_3^{2|k-u|} Var(V_{stu})$$

$$= \left(\frac{1+\rho_1}{1-\rho_1}\right) \left(\frac{1+\rho_2}{1-\rho_2}\right) \left(\frac{1+\rho_3}{1-\rho_3}\right) ab$$

$$+ \left(\frac{1+\rho_1^2}{1-\rho_1^2}\right) \left(\frac{1+\rho_2^2}{1-\rho_2^2}\right) \left(\frac{1+\rho_3^2}{1-\rho_3^2}\right) ab^2.$$

Korollar 2.31. Für die drei Kovarianzen $Cov(X_{ijk}, X_{(i+1)jk})$, $Cov(X_{ijk}, X_{i(j+1)k})$ und $Cov(X_{ijk}, X_{ij(k+1)})$ gilt ebenfalls analog zu Satz 2.14

$$Cov(X_{ijk}, X_{(i+1)jk}) = Cov(\lambda_{ijk}, \lambda_{(i+1)jk})$$

$$= \sum_{s=-\infty}^{\infty} \rho_1^{|i-s|+|i+1-s|} \sum_{t=-\infty}^{\infty} \rho_2^{2|i-t|} \sum_{u=-\infty}^{\infty} \rho_3^{2|k-u|} Var(V_{stu})$$

$$= \left(\frac{2\rho_1}{1-\rho_1^2}\right) \left(\frac{1+\rho_2^2}{1-\rho_2^2}\right) \left(\frac{1+\rho_3^2}{1-\rho_3^2}\right) ab^2,$$

$$Cov(X_{ijk}, X_{i(j+1)k}) = Cov(\lambda_{ijk}, \lambda_{i(j+1)k})$$

$$= \sum_{s=-\infty}^{\infty} \rho_1^{2|i-s|} \sum_{t=-\infty}^{\infty} \rho_2^{|i-t|+|i+1-t|} \sum_{u=-\infty}^{\infty} \rho_3^{2|k-u|} Var(V_{stu})$$

$$= \left(\frac{1+\rho_1^2}{1-\rho_1^2}\right) \left(\frac{2\rho_2}{1-\rho_2^2}\right) \left(\frac{1+\rho_3^2}{1-\rho_3^2}\right) ab^2,$$

und

$$Cov(X_{ijk}, X_{ij(k+1)}) = Cov(\lambda_{ijk}, \lambda_{ij(k+1)})$$

$$= \sum_{s=-\infty}^{\infty} \rho_1^{2|i-s|} \sum_{t=-\infty}^{\infty} \rho_2^{2|j-t|} \sum_{u=-\infty}^{\infty} \rho_3^{|k-u|+|k+1-u|} Var(V_{stu})$$

$$= \left(\frac{1+\rho_1^2}{1-\rho_1^2}\right) \left(\frac{1+\rho_2^2}{1-\rho_2^2}\right) \left(\frac{2\rho_3}{1-\rho_3^2}\right) ab^2.$$

Der Erwartungswert $E(X_{ijk})$ kann durch die Mittelwerte aller x_{ijk}

$$\bar{x} = \frac{1}{IJK} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} x_{ijk}$$

geschätzt werden und die Varianz $Var(X_{ijk})$ durch die empirische Varianz

$$s^{2} = \frac{1}{IJK - 1} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (x_{ijk} - \bar{x})^{2}.$$

Die Kovarianz $Cov(X_{ijk}, X_{(i+1)jk})$ kann durch empirische Kovarianz der Paare $(X_{ijk}, X_{(i+1)jk}), \quad i=1,\ldots,I-1, \quad j=1,\ldots,J-1, \quad k=1,\ldots,K-1$

$$s_{(X_{ijk},X_{(i+1)jk})} = \frac{1}{(I-1)JK-1} \sum_{i=1}^{I-1} \sum_{j=1}^{J} \sum_{k=1}^{K} (x_{ijk} - \bar{x}_1)(x_{(i+1)jk} - \bar{x}_2)$$

mit

$$\bar{x}_1 = \frac{1}{(I-1)JK} \sum_{i=1}^{I-1} \sum_{j=1}^{J} \sum_{k=1}^{K} x_{ijk}, \quad \bar{x}_2 = \frac{1}{(I-1)JK} \sum_{i=1}^{I-1} \sum_{j=1}^{J} \sum_{k=1}^{K} x_{(i+1)jk}$$

geschätzt werden. Analog kann die Kovarianz $Cov(X_{ijk}, X_{i(j+1)k})$ durch empirische Kovarianz der Paare $(X_{ijk}, X_{i(j+1)k}), i = 1, \ldots, I-1, \quad j = 1, \ldots, J-1,$ $k = 1, \ldots, K-1$

$$s_{(X_{ijk},X_{i(j+1)k})} = \frac{1}{I(J-1)K-1} \sum_{i=1}^{I} \sum_{j=1}^{J-1} \sum_{k=1}^{K} (x_{ijk} - \bar{x}_3)(x_{i(j+1)k} - \bar{x}_4)$$

mit

$$\bar{x}_3 = \frac{1}{I(J-1)K} \sum_{i=1}^{I} \sum_{j=1}^{J-1} \sum_{k=1}^{K} x_{ijk}, \quad \bar{x}_4 = \frac{1}{I(J-1)K} \sum_{i=1}^{I} \sum_{j=1}^{J-1} \sum_{k=1}^{K} x_{i(j+1)k}$$

und die Kovarianz $Cov(X_{ijk}, X_{ij(k+1)})$ durch empirische Kovarianz der Paare $(X_{ijk}, X_{ij(k+1)}), \quad i = 1, \dots, I-1, \quad j = 1, \dots, J-1, \quad k = 1, \dots, K-1$

$$s_{(X_{ijk},X_{ij(k+1)})} = \frac{1}{IJ(K-1)-1} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K-1} (x_{ijk} - \bar{x}_5)(x_{ij(k+1)} - \bar{x}_6)$$

mit

$$\bar{x}_5 = \frac{1}{IJ(K-1)} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K-1} x_{ijk}, \quad \bar{x}_6 = \frac{1}{IJ(K-1)} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K-1} x_{ij(k+1)}$$

geschätzt werden.

Folgerung 2.32. Die Momentenmethode liefert folgende Schätzgleichungen für a, b, ρ_1, ρ_2 und ρ_3 :

$$\begin{split} E(a,b,\rho_1,\rho_2,\rho_3) =&: \bar{x} = \left(\frac{1+\rho_1}{1-\rho_1}\right) \left(\frac{1+\rho_2}{1-\rho_2}\right) \left(\frac{1+\rho_3}{1-\rho_3}\right) ab \\ V(a,b,\rho_1,\rho_2,\rho_3) =&: s^2 = \left(\frac{1+\rho_1}{1-\rho_1}\right) \left(\frac{1+\rho_2}{1-\rho_2}\right) \left(\frac{1+\rho_3}{1-\rho_3}\right) ab \\ &+ \left(\frac{1+\rho_1^2}{1-\rho_1^2}\right) \left(\frac{1+\rho_2^2}{1-\rho_2^2}\right) \left(\frac{1+\rho_3^2}{1-\rho_3^2}\right) ab^2 \\ C_1(a,b,\rho_1,\rho_2,\rho_3) =&: s_{(X_{ijk},X_{(i+1)jk})} = \frac{2\rho_1}{(1-\rho_1^2)} \left(\frac{1+\rho_2^2}{1-\rho_2^2}\right) \left(\frac{1+\rho_3^2}{1-\rho_3^2}\right) ab^2 \\ C_2(a,b,\rho_1,\rho_2,\rho_3) =&: s_{(X_{ijk},X_{i(j+1)k})} = \left(\frac{1+\rho_1^2}{1-\rho_1^2}\right) \frac{2\rho_2}{(1-\rho_2^2)} \left(\frac{1+\rho_3^2}{1-\rho_3^2}\right) ab^2 \\ C_3(a,b,\rho_1,\rho_2,\rho_3) =&: s_{(X_{ijk},X_{ij(k+1)})} = \left(\frac{1+\rho_1^2}{1-\rho_1^2}\right) \left(\frac{1+\rho_2^2}{1-\rho_2^2}\right) \frac{2\rho_3}{(1-\rho_3^2)} ab^2 \end{split}$$

Die Lösung der Gleichungssysteme ist äquivalent zur Lösung des Minimierungsproblems

$$\min_{\sigma=(a,b,\rho_1,\rho_2,\rho_3)} (\bar{x} - E(X_{ijk}))^2 + (s^2 - Var(X_{ijk}))^2 + (s_{(X_{ijk},X_{(i+1)jk})} - Cov(X_{ijk},X_{(i+1)jk}))^2 + (s_{(X_{ijk},X_{i(j+1)k})} - Cov(X_{ijk},X_{i(j+1)k}))^2 + (s_{(X_{ijk},X_{ij(k+1)})} - Cov(X_{ijk},X_{ij(k+1)}))^2.$$

3. Simulationen

In diesem Kapitel soll überprüft werden, ob aus Daten, die mit dem verallgemeinerten Gamma-Poisson-Mischungsmodell simuliert worden sind, die Parameter geschätzt werden können, die zur Simulation verwendet wurden. Dazu wurden auf Grundlage der Modelle, die in Kapitel 2.2 definiert wurden, Daten simuliert und die zugehörigen empirischen Kennwerte berechnet. Anschließend wurde die Momentenmethode angewendet, um die Parameter aus den Daten zu schätzen. Auf die geschätzten Parameter wurde der Einstichproben-t-Test mit folgender Hypothese angewendet:

 H_0 : Der geschätzte Parameter stimmt mit dem zuvor zur Simulation der Daten verwendeten Parameter **überein**

VS.

 H_1 : Der geschätzte Parameter stimmt mit dem zuvor zur Simulation der Daten verwendeten Parameter **nicht überein**.

Falls die Nullhypothese nicht verworfen werden kann, kann nicht bestätigt werden, dass ein signifikanter Unterschied zwischen dem zur Simulation gewählten und geschätzten Parameter besteht. Es wird in diesem Fall davon ausgegangen, dass kein signifikanter Unterschied zwischen den beiden Parametern besteht und sich das verallgemeinerte Gamma-Poisson-Mischungsmodell dazu eignet die Zusammenhänge von benachbarten Zellen unter dem jeweiligen Modell zu beschreiben.

Die Modelle sind so definiert, dass eine unendlich große Menge an Daten betrachtet werden müsste. Dies ist natürlich nicht möglich. In den Abbildungen 1 und 2 auf Seite 22 ist veranschaulicht, was passiert, wenn 500 Daten mithilfe des verallgemeinerten Gamma-Poisson-Mischungsmodells aus dem eindimensionalen Kontext simuliert werden. Am Anfang der Zeitreihe steigt die Anzahl der Individuen in der

Zelle stark an, bevor sie sich stabilisiert. Zum Ende der Zeitreihe fällt die Anzahl dann wiederum ab. Durch die Randwerte werden die empirischen Kennziffern, die zur Schätzung der Parameter benötigt werden, verfälscht. Deshalb wurde in den Simulationen in diesem Kapitel jeweils eine relativ große Menge an Daten erzeugt, um dann 33 % der mittleren Daten zu betrachten. Für die Modelle, bei denen eine Datenmatrix zur Auswertung benötigt wird, werden nur 33 % der Daten aus der Mitte der Matrix betrachtet. Ebenso bei dem Modell, in dem ein Datenwürfel benötigt wird, wird auch nur der 33 %-ige, mittlere Kern betrachtet.

Für die Wahl von a und b stellte sich im Laufe der Simulationen heraus, dass

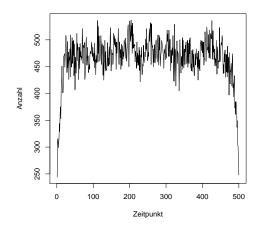


Abb. 1: Simulation: 500 Daten, $a = 50, b = 0.5, \rho = 0.9$

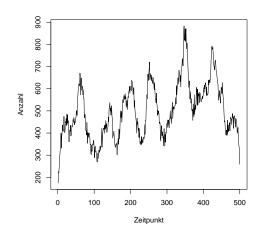


Abb. 2: Simulation: 500 Daten, $a = 0.5, b = 50, \rho = 0.9$

gelten muss: a < b und 0 < a < 1 < b. Ansonsten gab es oft nicht lösbare Gleichungen, das heißt die Lösung der Gleichungen ist komplex. In dieser Arbeit wurden nur reelle Lösungen betrachtet, da bei der Poissonverteilung nur reelle Parameter zulässig sind.

Bei dieser Arbeit soll herausgefunden werden, ob mithilfe des Parameters ρ Aussagen über den Zusammenhang von benachbarten Zellen möglich ist. Die Parameter a und b sind dabei nicht weiter von Interesse. Bei den folgenden Simulationen wurden die Ergebnisse der beiden Parameter jeweils angegeben, aber sie werden zur Modellbeurteilung nicht weiter betrachtet.

Zur Erzeugung der Daten und Auswertung mit dem Einstichproben-t-Test wurde das Software-Paket R 2.11.1 (R Development Core Team, 2010) und R 2.14.2

(R Development Core Team, 2012) verwendet. Die Schätzungen der Parameter wurden mithilfe des Softwareprogramms MATLAB 7.9.0 (R2009b) durchgeführt.

3.1. Eindimensional über die Zeit

Als Erstes wurden Daten simuliert, die die Anzahl von Individuen in einer Zelle über mehrere Zeitpunkte darstellen. Dazu wurde das Modell aus Definition 2.18 von Seite 13 verwendet. Die Daten wurden simuliert, indem die Parameter a, b und ρ vorgegeben wurden. Anschließend wurden die Parameter a, b und ρ mit der Momentenmethode aus den simulierten Daten geschätzt. Bei der Momentenmethode wurden das erste Moment und das zweite zentrale Moment verwendet und zusätzlich die Kovarianz und empirische Kovarianz des Paares (X_t, X_{t+1}) , da hier eine Aussage über den Zusammenhang von benachbarten Zellen gemacht werden soll.

In Abbildung 3 auf Seite 24 sind die mittleren 33 % der Ergebnisse von einer Simulation mit 2000 Daten abgebildet. In der linken Abbildung ist die Zeitreihe eingezeichnet, die in diesem Fall simuliert wurde. In der rechten Abbildung ist diese Zeitreihe mit einer sogenannten Hitzekarte (engl. heatmap) abgebildet. Dabei sind die Zeitpunkte von unten nach oben zu lesen. Je heller ein Streifen ist, desto weniger Individuen befinden sich zu diesem Zeitpunkt in der Zelle. Die Parameter, die bei dieser Simulation verwendet wurden, waren a = 0.5, b = 50und $\rho = 0.6$. In der rechten Abbildung ist außerdem $\hat{\rho}$ eingetragen. Zur Veranschaulichung wurden in diesem Fall die Parameter nicht mit dem Softwareprogramm MATLAB geschätzt, sondern mit dem Softwarepaket R mit dem Befehl optim(). Die geschätzten Parameter lauten: $\hat{a}=0.4762719,\,\hat{b}=49.5585385$ und $\hat{\rho} = 0.6040940$ (zum Vergleich: MATLAB gibt hier eine Lösung von $\hat{a} = 0.4763$, $\hat{b}=49.5470$ und $\hat{\rho}=0.6042$ bei den errechneten Kennwerten $\overline{x}=98.99119,$ $s^2=2617.673$ und $s_{(X_t,X_{(t+1)})}=2189.686$ aus). Im Laufe der durchgeführten Simulationen und Parameterschätzungen stellte sich heraus, dass das Softwareprogramm MATLAB für die Momentenschätzungen geeigneter ist. Gleichungen der

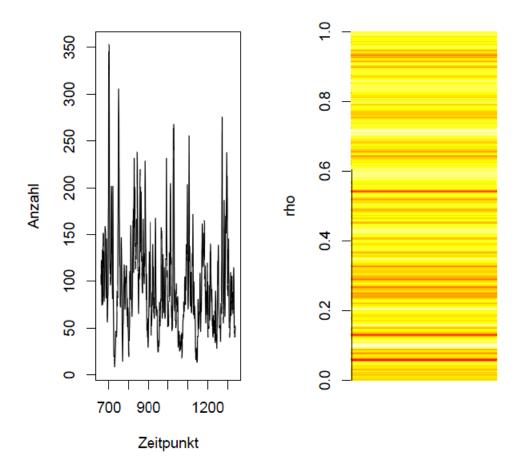


Abb. 3: Simulation 2000 Daten, nur mittleren 33 % verwendet, $\rho=0.6$, $\hat{\rho}=0.6040940$

Art

$$\bar{x} - \left(\frac{1+\rho}{1-\rho}\right)ab = 0 := E(a,b,\rho)$$

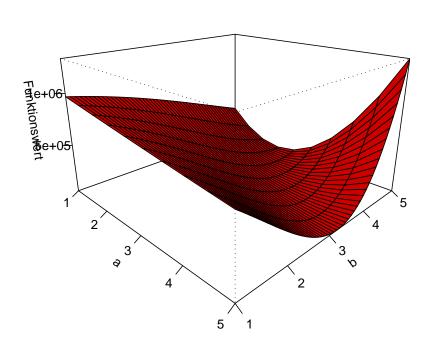
$$s^2 - \left(\left(\frac{1+\rho^2}{1-\rho^2}\right)ab^2 + \left(\frac{1+\rho}{1-\rho}\right)ab\right) = 0 := V(a,b,\rho)$$

$$s_{(X_t,X_{(t+1)})} - \frac{2\rho}{(1-\rho^2)}ab^2 = 0 := C(a,b,\rho)$$

können mit MATLAB exakt berechnet werden. Falls keine reelle Lösung existiert, wird das komplexe Ergebnis ausgegeben. In den anfangs durchgeführten Simulationen wurden die Momentenschätzungen mit dem Softwarepaket R durchgeführt. Dabei wurde nach einer minimalen Lösung von $\min_{\sigma=(a,b,\rho)} (\bar{x}-E(X_t))^2 + (s^2-Var(X_t))^2 + (s_{(X_t,X_{(t+1)})}-Cov(X_t,X_{(t+1)}))^2 \text{ gesucht. Eine minimale Lösung konnte hierbei etliche Male nicht berechnet werden, da die zu minimierende Funktion zu flach verlief und das Programm kein eindeutiges Minimum$

finden konnte. In Abbildung 4 ist dieses Problem mit einer Funktion dargestellt, bei der folgende Kennwerte verwendet wurden: $\overline{x} = 473.0164$, $s^2 = 770.8164$ und $s_{(X_t,X_{(t+1)})} = 469.9944$. Für die Abbildung wurde der Parameter ρ auf 0.9 festgelegt und die Parameter a und b konnten alle Werte aus dem Intervall [1, 5] annehmen. Auf der Höhe von b = 3 verläuft die Funktion sehr flach, so dass eine exakte Angabe des Minimums problematisch ist. Aus diesem Grund finden alle weiteren Schätzungen mit dem Softwareprogramm MATLAB statt.

Die Simulation der Daten für dieselben Parameter wurde mehrmals durchgeführt



a, b variabel, festes rho=0.9

Abb. 4: Darstellung einer zu minimierenden Funktion, dabei a,b fest, $\rho=0.9,$ $\overline{x}=473.0164,\,s^2=770.8164,\,s_{(X_t,X_{(t+1)})}=469.9944$

und auf den daraus geschätzten Parametern $\hat{a},\,\hat{b}$ und $\hat{\rho}$ wurde der Einstichprobent-Test angewendet.

Die Tabellen 2 und 3 auf den Seiten 72 und 73 zeigen für 2000 simulierte Zeitpunkte und 50 Wiederholungen die jeweiligen p-Werte für die geschätzten Parameter an, die mit dem Einstichproben-t-Test berechnet wurden. Folgende Parameter wurden zur Simulation der Daten angenommen: (a, b) = (0.5, 50), (a, b) = (0.2, 150),

 $(a,b)=(0.2,50),\ (a,b)=(0.3,93)$ und (a,b)=(2,20) mit jeweils allen ρ -Werte in 0.1-Schritten von 0.1 bis 0.9. Wie viele Ergebnisse der 50 Wiederholungen in den Einstichproben-t-Test eingegangen sind, ist in der zweiten Spalte unter n zu finden. Ausgeschlossen wurden dabei alle komplexen Ergebnisse. In der Tabelle sind alle p-Werte, bei denen die Nullhypothese zum Signifikanzniveau 5 % nicht abgelehnt werden kann, in fett dargestellt und für ein Signifikanzniveau zwischen 1 % und unter 5 % in kursiv. Ebenfalls in kursiv sind die p-Werte, die gerundet bei einem Signifikanzniveau von 1 % die Nullhypothese nicht ablehnen können. Aus den Tabellen geht hervor, dass für die Parameter a und b die Nullhypothesen bis auf wenige Ausnahmen zum 5 %-Niveau nicht abgelehnt werden kann. Bei dem Parameter ρ sieht dies anders aus. Für die 45 Parameterkombinationen kann für den Parameter ρ die Nullhypothese 22 mal zum 5 %-Niveau und 22+8 also 30 mal zum 1 % Niveau nicht abgelehnt werden. In 12 der Fälle wird die Nullhypothese abgelehnt und in weiteren drei Fällen könnte sie gerundet zum 1 %-Niveau nicht abgelehnt werden.

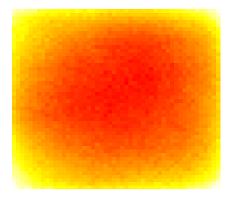
Das verallgemeinerte Gamma-Poisson-Mischungsmodell eignet sich bei der Betrachtung von Zellen über mehrere Zeitpunkte dazu, die Korrelation der nachfolgenden Zellen bei relativ hohem ρ -Wert zwischen 0.5 bis 0.8 wiederzugeben.

3.2. Räumlich

Im nächsten Schritt wurden Daten auf Grundlage des räumlichen Modells aus Definition 2.11 von Seite 8 simuliert. In diesem Modell stellt der Parameter ρ die räumliche Korrelation von benachbarten Zellen dar.

Um dieses Modell zu überprüfen, wurden 2500 Daten simuliert, die eine 50×50 -Matrix darstellen. Für die Schätzungen wurden, wie schon erwähnt, nur 33 % der mittleren Matrix verwendet. In den Abbildungen 5 bis 7 auf der Seite 27 sind Hitzekarten von simulierten 50×50 Matrizen mit den Parametern (a, b) = (0.2, 50) für verschiedene ρ -Werte abgebildet. Vor allem für die Parameter $\rho = 0.9$ und $\rho = 0.5$ ist der Zusammenhang von benachbarten Zellen sehr schön zu erkennen.

Für dieses Modell wurden ebenfalls Daten mit verschiedenen Parameterausprägungen simuliert und die Parameter aus den Daten mit der Momentenmethode



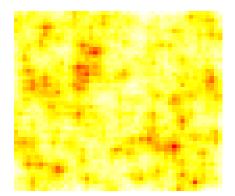


Abb. 5: Simulierte Daten mit $\rho = 0.9$

Abb. 6: Simulierte Daten mit $\rho = 0.5$

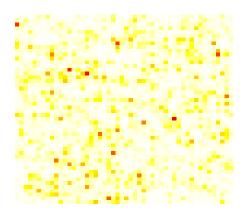


Abb. 7: Simulierte Daten mit $\rho = 0.1$

geschätzt. Dabei wurden wiederum das erste Moment und das zweite zentrale Moment verwendet. Da hier der räumliche Zusammenhang betrachtet wird, wurde außerdem die Kovarianz und die empirische Kovarianz des Paares $(X_{ij}, X_{(i+1)(j+1)})$ verwendet. Die Tabelle 4 auf Seite 74 zeigt die Ergebnisse des Einstichproben-t-Tests für die Parameter (a,b)=(0.2,50) und (a,b)=(0.1,25), jeweils für verschiedene ρ -Werte. Die fett und kursiv geschriebenen p-Werte haben wiederum die gleiche Bedeutung für die verschiedenen Signifikanzniveaus wie bereits im vorherigen Kapitel 3.1 bei der Simulation des eindimensionalen Modells beschrieben. Bei Schätzungen mit ρ -Werten von 0.9 und 0.8 wurden für die Parameter bei 50 Durchläufen nur komplexe Ergebnisse geschätzt und für ρ = 0.7 bei den Parametern (a,b)=(0.5,50) ebenfalls. Dargestellt ist dies in der Tabelle mit "NA"

Einträgen. Für die Schätzungen für $\rho=0.6$ und $\rho=0.7$ besteht ein signifikanter Unterschied zwischen dem geschätzten und dem wahren Parameter ρ . Bei geschätzten ρ -Werten zwischen 0.6 und 0.9 kann davon ausgegangen werden, dass das verallgemeinerte Gamma-Poisson-Mischungsmodell nicht zuverlässig schätzen kann.

Bei den Schätzungen für $\rho = 0.1$ gab es bei den 50 Wiederholungen 12 Schätzungen, bei denen es keine reelle Lösung gab. In all diesen Fällen war die berechnete empirische Kovarianz aus den simulierten Daten negativ.

Für die ρ -Werte 0.1, 0.3 und 0.4 kann die Nullhypothese zum 5 % Signifikanzniveau nicht abgelehnt werden. Hierbei konnte nicht nachgewiesen werden, dass es einen signifikanten Unterschied zwischen den berechneten und geschätzten Parametern gibt.

Bei den Parametern $(a,b,\rho)=(0.2,50,0.5)$ und $(a,b,\rho)=(0.1,25,0.2)$ lehnt der Einstichproben-t-Test die Nullhypothese jeweils zum 5 %-Niveau ab. Für die Parameter $(a,b,\rho)=(0.1,25,0.2)$ könnte der p-Wert auf 0.01 gerundet werden und somit kann der Test die Nullhypothese zum 1 %-Niveau nicht ablehnen. Da es sich bei den Simulationen um zufällige Werte handelt, kann in diesem Fall davon ausgegangen werden, dass der p-Wert nur zufällig diesen Wert annimmt. In Abbildung 8 auf Seite 29 sind geschätzte Parameterwerte bei 50 Wiederholungen in Histogramme abgebildet, da die geschätzten ρ -Werte möglicherweise nicht normalverteilt sind. Dies geht aus den Abbildungen nicht hervor. Hier liegt ebenfalls die Vermutung nahe, dass es Ausreißer gibt, da es sich hier um zufällig geschätzte Daten handelt.

Insgesamt eignet sich das verallgemeinerte Gamma-Poisson-Mischungsmodell dazu, im räumlichen Kontext um ρ -Parameterwerte aus dem Intervall [0,0.5] wiederzugeben.

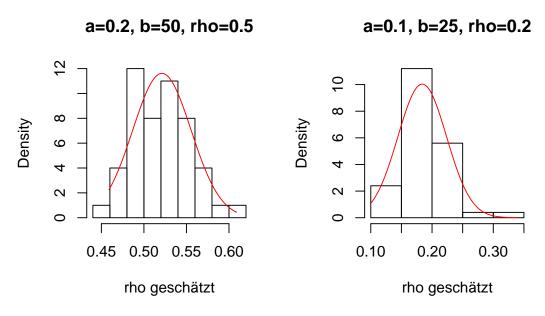
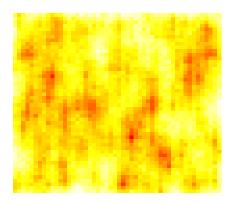


Abb. 8: Histogramme der geschätzten ρ -Werte

3.3. Zwei Richtungen

Als Vorletztes wurden Daten auf Grundlage des Modells aus Definition 2.23 von Seite 15 simuliert. Dieses Modell beinhaltet jetzt zwei Richtungsparameter ρ_1 und ρ_2 . Zu der Modellierung wurden wiederum 2500 Werte erzeugt, die eine 50 × 50-Matrix darstellen. Wie bereits weiter oben erwähnt, wurden zur Schätzung der Parameter nur die mittleren 33 % der Werte ausgewählt. Die Parameter a und b wurden diesmal nur mit den Werten (a,b)=(0.2,50) betrachtet. Für die Parameter ρ_1 und ρ_2 wurden alle möglichen Kombinationen von 0.1 bis 0.9, also 81 Kombinationen, simuliert. Jede Kombination wurde 50 mal wiederholt. In den Abbildungen 9 bis 11 auf der Seite 30 sind Beispiele für Daten, die mit diesem Modell simuliert wurden, in Hitzekarten abgebildet. Dabei wurden alle Daten, nicht nur 33 %, visualisiert. In Abbildung 9 wurden dafür die Parameterwerte $(\rho_1, \rho_2) = (0.7, 0.5)$ verwendet, in Abbildung 10 die Werte $(\rho_1, \rho_2) = (0.2, 0.3)$ und in Abbildung 11 die Werte $(\rho_1, \rho_2) = (0.1, 0.9)$. Hier ist sehr schön zu erkennen, wie sowohl bei einem hohen als auch bei einem geringen ρ -Wert benachbarte Zellen einen Einfluss aufeinander haben.

Nachdem die Daten simuliert wurden, wurden die Parameter wiederum mit der Momentenmethode geschätzt. Diesmal wurden ebenfalls das erste Moment und



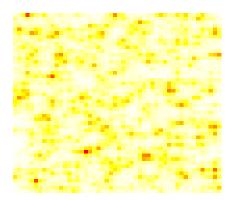


Abb. 9: Simulierte Daten mit $\rho_1=0.7,$ $\rho_2=0.5$

Abb. 10: Simulierte Daten mit $\rho_1 = 0.2,$ $\rho_2 = 0.3$

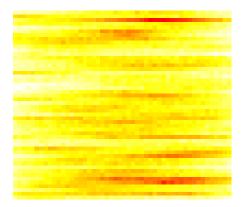


Abb. 11: Simulierte Daten mit $\rho_1 = 0.1$, $\rho_2 = 0.9$

das zweite zentrale Moment zur Schätzung gewählt und auch die Kovarianz und empirische Kovarianz der Paare $(X_{ij}, X_{(i+1)j})$ und $(X_{ij}, X_{i(j+1)})$, da diesmal die Zellennachbarn in der senkrechten und waagerechten Richtung betrachtet werden sollen. Die Ergebnisse des Einstichproben-t-Test für die vier geschätzten Parameter sind in die Tabellen 5 bis 7 auf den Seiten 75 bis 77 eingetragen. Darin sind die p-Werte, die über dem 5 % Signifikanzniveau liegen, in fett und die p-Werte, die zwischen dem 1 %- und 5 %-Niveau liegen, in kursiv dargestellt. In kursiv wurden außerdem auch die p-Werte dargestellt, die gerundet den Wert 0.01 ergeben. Bei den 81 Kombinationen fielen die Schätzungen bei den Wiederholungen nie unter die Mindestanzahl von 30 Betrachtungen, so dass davon ausgegangen werden kann,

dass die Ergebnisse normalverteilt sind und der Einstichproben-t-Test angewendet werden kann. Bei den Simulationen fiel auf, dass, falls eine berechnete empirische Kovarianz, also $s_{(X_{ij},X_{(i+1)j})}$ oder $s_{(X_{ij},X_{i(j+1)})}$, negativ war, der geschätzte ρ_1 - bzw. ρ_2 -Wert ebenfalls als negativ geschätzt wurde. Laut Definition 2.23 nehmen die Parameter ρ_1 und ρ_2 nur Werte zwischen 0 und 1 an. Da die beiden Parameter ρ_1 und ρ_2 Korrelationen zwischen Zellen wiedergeben, kann davon ausgegangen werden, dass gelten kann: $-1 < \rho_1 < 1$ und $-1 < \rho_2 < 1$. Falls die Parameterwerte negativ sind, gibt es einen negativen Zusammenhang, andernfalls einen positiven. Aus den Tabellen wird erkennbar, dass der Einstichproben-t-Test die Nullhypothese H_0 grundsätzlich zum 5 %-Niveau ablehnt, falls beide ρ -Werte groß gewählt wurden, das heißt $0.6 \le \rho_1 \le 0.9$ und $0.6 \le \rho_2 \le 0.9$. Zwischen den wahren und geschätzten Parametern besteht ein signifikanter Unterschied. In zwei Fällen würde der Test zum 1 %-Niveau die Nullhypothese nicht ablehnen. Auch hier war die Überlegung, dass die geschätzten Parameter evtl. nicht normalverteilt sind und die Nullhypothese deswegen abgelehnt werden muss. Dazu wurden Histogramme erstellt, um die Parameterwerte auf Normalverteilung zu überprüfen. Die Histogramme bestätigten, dass die geschätzten ρ -Werte normalverteilt sind. In Abbildung 12 sind zwei der resultierenden Histogramme abgebildet.

In den Fällen, in denen der Parameter kleiner gewählt wurden, das heißt

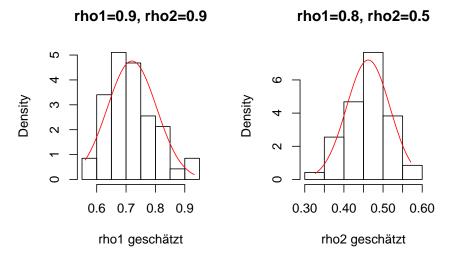


Abb. 12: Histogramme der geschätzten Parameterwerte

 $0.1 \leq \rho_1 \leq 0.6$ und $0.1 \leq \rho_1 \leq 0.4$, dann konnte in den meisten Fällen

der Test die Nullhypothese zum 5 %-Niveau nicht ablehnen. In einigen wenigen Fällen konnte der Test H_0 zwischen dem 1 %- und 4 %-Niveau nicht ablehnen.

Es lässt sich keine einheitliche Aussage machen, falls der Parameter ρ_2 den Wert 0.5 und ρ_1 alle möglichen Ausprägungen annimmt. Hierbei lehnt der Test entweder die Nullhypothese H_0 zum 5 %-Niveau oder zum Niveau zwischen 1 % und 4 % nicht ab, oder die Nullhypothese wird abgelehnt.

Aus diesen Ergebnissen lässt sich schließen, dass sich das verallgemeinerte Gamma-Poisson-Mischungsmodell dazu eignet, Korrelationen zwischen benachbarten Zellen für die beiden Parameter ρ_1 und ρ_2 wiederzugeben, falls diese einen Wert aus dem Intervall [0, 0.6] für ρ_1 und [0, 0.4] für ρ_2 annehmen.

3.4. Zwei Richtungen mit Zeit

Zuletzt wurde das dreidimensionale Modell aus Definition 2.28 von Seite 17 betrachtet. Neben der senkrechten und der waagerechten Richtung mit den Parametern ρ_1 und ρ_2 , wird hier noch eine dritte Richtung betrachtet, die Zeit, mit Parameter ρ_3 . Hierbei wurde ein $14 \times 14 \times 14$ Datenwürfel erzeugt, der insgesamt 2744 Daten enthält. Zur Auswertung wurden wiederum nur 33~% der Daten aus dem Kern des Datenwürfels ausgewertet. Die Parameter a und b wurden bei der Simulation mit (0.2,50) festgelegt. Für die drei Richtungsparameter wurden alle möglichen Kombinationen zwischen 0.1 und 0.5 angenommen. Insgesamt sind dies 125 Kombinationen. Es wurden nur Daten für geringe ρ -Werte simuliert, da sich in den vorherigen Modellen herausgestellt hat, dass bei höheren ρ -Werten der Einstichproben-t-Test die Nullhypothese oftmals ablehnte. Zur Schätzung der Parameter wurden bei der Momentenmethode das erste Moment und das zweite zentrale Moment und zusätzlich noch die Kovarianz und die empirische Kovarianz der Paare $(X_{ijk}, X_{(i+1)jk}), (X_{ijk}, X_{i(j+1)k})$ und $(X_{ijk}, X_{ij(k+1)})$ verwendet, da hier eine Aussage über die Zusammenhänge dieser drei Richtungen gemacht werden soll. Jede Kombination wurde 45 bzw. 50 mal wiederholt. Da die Datensimulationen anfänglich sehr lange dauerten, wurden die Wiederholungen heruntergesetzt. Inzwischen wurde der Programmiercode überarbeitet und für die Simulation der Daten wird weniger Zeit benötigt.

Die Schätzungen der Parameter fanden bei diesem Modell mit dem Softwarepaket R statt. Das Softwareprogramm MATLAB eignete sich bei fünf Parametern für die Schätzungen nicht. Bereits bei der Parameterschätzung von einem Datensatz musste die Prozedur manuell abgebrochen werden, da auch nach längerer Wartezeit (ca. 45 min) keine Lösung ausgegeben wurde. Bei Schätzungen von vier Parametern stand das Ergebnis bereits nach ca. sechs Sekunden fest. Da die Schätzungen mit dem Paket R zwar nur approximativ genau sind, aber wesentlich schneller (ca. eine Sekunde für einen Datensatz) durchgeführt wurden, werden die geringen Abweichungen in Kauf genommen, um die Schätzungen binnen kurzer Zeit zu erhalten.

In Tabelle 1 sind die p-Werte, die sich nach Anwendung des Einstichproben-t-Test ergaben, eingetragen. Es konnten nur 12 der 125 Simulationen ausgewertet werden, da für den Rest die Voraussetzung, dass der Stichprobenumgang mindestens 30 sein muss, zur Anwendung des Einstichproben-t-Test nicht gegeben war. Für die anderen Kombinationen ist es notwendig, weitere Datensimulationen durchzuführen.

Bei den verfügbaren Ergebnissen in Tabelle 1 wird wiederum deutlich, dass die

Anzahl Durchläufe: 45, Anzahl Daten: 2744 (14 $ imes$ 14 $ imes$ 14-Datenwürfel)										
$(\mathbf{a}, \mathbf{b}, \rho_1, \rho_2)$	n	p-Wert \hat{a}	p-Wert \hat{b}	p-Wert $\hat{\rho}_1$	p-Wert $\hat{\rho}_2$	p-Wert $\hat{\rho}_3$				
(0.2, 50, 0.3, 0.2, 0.2)	30	1.149e-06	9.129 e-07	0.0002825	0.004955	0.3548				
$(0.2,\!50,\!0.3,\!0.1,\!0.1)$	33	5.674 e - 05	0.01126	0.06805	0.3052	0.05344				
(0.2, 50, 0.2, 0.3, 0.3)	30	0.0006728	0.2319	0.03323	0.08577	0.03529				
(0.2, 50, 0.2, 0.2, 0.1)	33	0.01756	0.4075	0.1585	0.4588	0.07398				
(0.2, 50, 0.2, 0.1, 0.3)	32	0.0009563	0.09115	0.1745	0.0003282	0.1632				
(0.2, 50, 0.2, 0.1, 0.2)	32	0.2232	0.403	0.1126	0.1338	0.07006				
(0.2, 50, 0.2, 0.1, 0.1)	30	0.000141	0.0002829	0.1358	0.0485	0.08541				
(0.2, 50, 0.1, 0.3, 0.1)	35	0.002248	0.02883	0.2819	0.7129	0.5728				
(0.2, 50, 0.1, 0.2, 0.2)	31	0.02129	0.317	0.446	0.22	0.4048				
(0.2, 50, 0.1, 0.2, 0.1)	34	0.01347	0.1921	0.3254	0.7852	0.409				
(0.2, 50, 0.1, 0.1, 0.2)	32	3.154 e-05	1.712e-05	0.6706	0.02893	0.1205				
$(0.2,\!50,\!0.1,\!0.1,\!0.1)$	41	0.0001194	0.01776	0.00471	0.3226	0.02379				

Tab. 1: Schätzungen Ergebnistabelle Simulation Datenwürfel (3 Richtungen)

Nullhypothese zum 5 % Niveau oftmals nicht abgeleh
hnt werden kann. Dabei gab es Parameterkombinationen für ρ_1 , ρ_2 und ρ_3 aus dem Intervall [0.1, 0.3]. In vier der Parameterkombinationen konnte bei einem oder zwei Parametern die Nullhypothese erst zum 2 %-Niveau nicht abgelehnt werden. Lediglich in drei der Parameterkombinationen konnte bei einem oder zwei Parametern die Nullhypothese abgelehnt werden. Da es sich bei Daten wie auch schon bei den anderen Modellen um simulierte Daten handelt, könnten Ausreißer für die Ablehnung verantwortlich sein. In Abbildung 13 sind die Histogramme für die geschätzten ρ -Werte, bei denen die Nullhypothese abgelehnt wird. Hierbei kann von einer Normalverteilung bei geschätzten Parameterwerte ausgegangen werden.

Bei dem jetztigen Stand der Auswertung kann davon ausgegangen werden, dass das verallgemeinerte Gamma-Poisson-Mischungsmodell bei Betrachtung von drei Richtungsparametern die Zusammenhänge für kleine Parameterwerte wiedergeben kann.

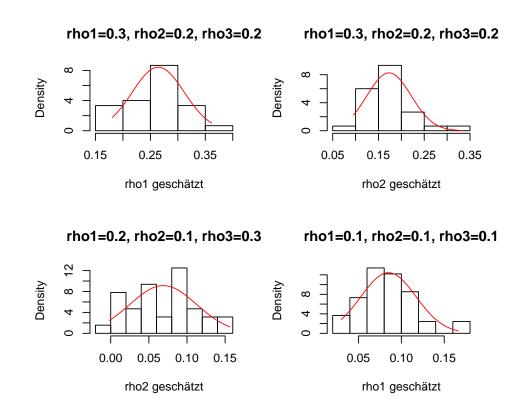


Abb. 13: Histogramme der geschätzten Parameterwerte

4. Anwendung in der Biometrie

4.1. Datenmaterial

Als Datengrundlage für die biometrische Anwendung liegen Datensätze zur Verfügung, die auf der Internetseite http://dataknp.sanparks.org/sanparks/metacat? action=read&qformat=sanparks&sessionid=0&docid=judithk.739.17 abgerufen werden können.

Auf dieser Seite sind vier Datensätze zu finden. Drei der vier Datensätze ("Ungulates85-95.txt", "ungulates81-84.txt", "ungulates77-80.txt") enthalten Daten über die Anzahl verschiedener Tiere, die im Krüger-Nationalpark, der in Südafrika liegt, leben. Die Anzahl wurde bei Zählungen, die in den Jahren 1977 bis 1997 durchgeführt wurden, bestimmt. In den Datensätzen ist die Anzahl der Tiere unter der dritten Spaltenüberschrift "NO" festgehalten. Außerdem enthalten die drei Datensätze noch das Jahr der jeweiligen Zählung, erste Spaltenüberschrift "YE-AR", und die jeweilige Position bei der Sichtung im Park, das heißt die Breitenund Längengradangabe. Diese sind unter den letzten beiden Spaltenüberschriften "LAT" und "LONG" zu finden. Die zweite Spaltenüberschrift lautet "SP". Diese enthält einen Code für die jeweilige Tierspezies, die an dieser Stelle gezählt wurde. Die Ubersetzung des Codes kann aus dem vierten Datensatz ("speciecodes.txt") der oben genannten Seite entnommen werden. Dabei steht zum Beispiel "k" für das Steppenzebra (Equus burchelli) und "kp" für die Steppengiraffe (Giraffa camelopardalis). In Tabelle 8 auf Seite 78 sind alle Spezien, die in den Datensätzen festgehalten sind, aufgeführt.

Laut der Beschreibung auf der oben genannten Seite wurden die Zählungen wie folgt durchgeführt: Der Park wurde am Anfang in kleine Quadrate, sogenannte Zensus-Blöcke, eingeteilt. Die Zählungen fanden in den frühen Morgenstunden (8 bis 12 Uhr) bei klarem Wetter statt, wobei die Areale überflogen wurden. In den

Jahren 1985 bis 1993 wurden die Positionen der Tiere in eine Karte eingezeichnet und danach digitalisiert. Ab 1994 wurden die Aufzeichnungen mit einem sich an Bord befindenen GPS-Gerätes durchgeführt. Bei der Anzahl handelt es sich immer um die Tiere in der gesamten Herde.

Nach Braack (2011, S. 5ff) wurde der Krüger-Nationalpark im Jahr 1926 gegründet und umfasst heute eine Größe von fast $20.000 \ km^2$. Dabei hat er eine durchschnittliche Breite von ca. $54 \ km$ und eine Länge von ca. $345 \ km$. In Abbildung 14 ist eine Karte des Parks in grün abgebildet, wobei mithilfe der anderen beiden Karten auch die genaue Lokalisation in Afrika ersichtlich wird. Der Park liegt im Nordosten von Südafrika. Dabei grenzt er südlich an Simbabwe und westlich an Mosambik. Im Park ist eine große Anzahl an verschiedenen Tier- und Pflanzenarten beheimatet: $336 \ \text{Bäume}$, $53 \ \text{Fische}$, $35 \ \text{Amphibien}$, $118 \ \text{Reptilien}$, $507 \ \text{Vögel}$ und $148 \ \text{Säugetiere}$. Bei der Auswertung werden in dieser Diplomarbeit nur die Steppengiraffen be-



Abb. 14: Lokalisation und Karte des Krüger-Nationalparks (grün)

Quelle: http://www.gotravel24.com/files/media/Places/Limpopo/KrugerPark_map_goafrikasafari.jpg

trachtet, die unter dem Kürzel "kp" in den Dateien zu finden sind. Ob es in den einzelnen Datensätzen fehlende Daten gibt, kann soweit nicht festgestellt werden. Natürlich kann es bei der Zählung vorgekommen sein, dass Giraffen übersehen oder doppelt gezählt worden sind. Die Jahre 1977 bis 1981 enthalten noch keine Informationen über im Park lebende Giraffen. In der Auswertung werden deswegen

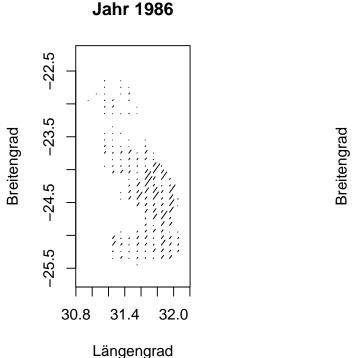
nur die Jahre 1982 bis 1997 betrachtet.

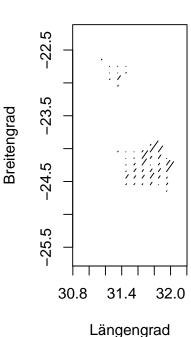
Die Giraffe lebt laut Stuart und Stuart (2001, S.180) in trockenen Savannen und Steppen. Dabei ernährt sie sich von Blättern von Bäumen und Büschen, weniger aber von Gras. Sie lebt in Herden von vier bis 30 Tieren, wobei es sich um lockere Verbände handelt. Die Herden haben keine bestimmten Territorien, in denen sie sich aufhalten, sondern wandern viel umher. Laut Braack (2011, S. 9) werden die Giraffen oft in Herden von zwei bis sechs Tieren gesichtet, wie sie Blätter von Akazienbäumen fressen.

4.2. Statistische Auswertung

Um einen ersten Eindruck der Daten zu erhalten, sind in den Abbildungen 15 und 16 auf Seite 38 die Giraffenanzahlen aus den Jahren 1986 und 1996 abgebildet. Alle weiteren Jahre sind in den Abbildungen 26 bis 30 auf den Seiten 80 und 81 zu finden. Zu erkennen ist das Gitternetz mit 0.1° × 0.1° Zellen, mit dem der Park eingeteilt wurde. Je länger die Linien in den Abbildungen sind, desto mehr Giraffen sind in der jeweiligen Zelle gezählt worden. Für die weiteren Berechnungen und Darstellungen wurde die gleiche Einteilung des Gitternetzes verwendet und es wurden zum Beispiel keine Zellen zusammengefasst. Aus den Abbildungen ist ersichtlich, dass in den Jahren 1994 bis 1997 die Zählungen unvollständig sind. Eine mögliche Ursache dafür ist, dass die Zählungen aus der Luft stattfanden und somit so teuer waren, dass nicht mehr das ganze Areal überflogen werden konnte. In der folgenden Auswertung werden diese vier Jahre ebenfalls nicht weiter betrachtet. In der oberen Spitze des Parks scheinen sich in all den Jahren nicht so viele Giraffen aufgehalten zu haben. Im Verbindungsstück des oberen und unteren Parkteils halten sich hingegen die Giraffen vermehrt auf. In Tabelle 9 auf Seite 79 sind einige Informationen der gegebenen Daten aufgeführt. Neben der Anzahl der Giraffen und der gesichteten Gruppen sind auch die Längen- und Breitengradabmessungen für das jeweilige Jahr zu finden.

Die Auswertung hatte sich als schwierig herausgestellt, da der Park eine unvorteilhafte Form hat. Zur Abhilfe wurde ein größtmögliches Rechteck in dem Park gesucht, welches sich zur Auswertung besser eignete. Das gefundene Rechteck hat





Jahr 1996

Abb. 15: Giraffenanzahlen im Jahr 1986

Abb. 16: Giraffenanzahlen im Jahr 1996

eine Größe von 13×4 Zellen und liegt in dem Mittelstück des Parks, welches den oberen und unteren Teil des Parks verbindet. In Abbildung 17 auf Seite 39 ist dieses Rechteck in rot in die Karte für ein Jahr eingezeichnet. Wie schon oben bemerkt, halten sich die Giraffen in diesem Teil des Parks über die Jahre vermehrt auf. Dieses Rechteck eignet sich deswegen sehr gut für die weiteren Auswertungen. Das grüne Rechteck in der Abbildung markiert auch einen möglichen Parkausschnitt, der betrachtet werden könnte, ist aber nur 4×9 Zellen groß. Außerdem gibt dieser Ausschnitt nur den untersten Teil des Parks wieder. In Tabelle 10 auf Seite 79 ist die neue Anzahl der Giraffen in dem 13×4 -Zellen großen Teilabschnitt für die Jahre 1982 bis 1993 eingetragen. Außerdem ist noch die durchschnittliche Anzahl der Giraffen in den Zellen für jedes Jahr mit aufgeführt.

Da die Giraffen viel umher wandern, stellt sich nun die Frage, ob sie dabei eine bestimmte Richtung bevorzugen. Mit den in Kapitel 2.2 aufgestellten Modellen soll diese Frage mit den vorliegenden Daten beantwortet werden.



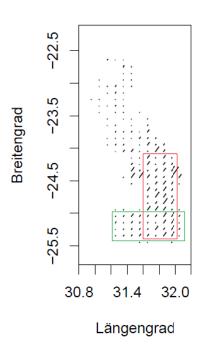


Abb. 17: Park mit möglichen Ausschnitten für die Auswertung

Eindimensional über die Zeit

Zunächst werden die 13×4 also 52 Zellen einzeln über die Jahre von 1982 bis 1993 betrachtet. Dazu wurde für jede Zelle des Rechtecks eine Zeitreihe erstellt. Das bedeutet, dass die Giraffenanzahl der 12 Jahre für jede Zelle zusammengeführt wurde und es nun 52 Zeitreihen mit jeweils 12 Werten gibt. Die Kennzahlen für jede Zelle wurden berechnet und damit anschließend die Parameter des verallgemeinerten Gamma-Poisson-Mischungsmodells mithilfe der Momentenmethode geschätzt. Die Ergebnisse sind in Tabelle 11 auf Seite 82 eingetragen. Jede Zelle dieser Tabelle entspricht jeweils der Zelle des Parkrechtecks an der entsprechenden Stelle. Dabei ist die Zelle (1,1) die Zelle, die im Parkausschnitt oben links lokalisiert ist. In den Abbildungen 19 und 20 auf den Seiten 42 und 43 sind die Zeitreihen jeweils in einer Hitzekarte abgebildet, wobei die Zeitpunkte wieder von unten nach oben gelesen werden müssen. Außerdem wurde der geschätzte Parameter ρ jeweils mit einge-

tragen. Falls der Wert negativ ist, beginnt die Linie oben in der Karte, andernfalls unten.

Aus der Tabelle ist ersichtlich, dass, falls für eine Zeitreihe eine negative Kovarianz berechnet wurde, der Wert für $\hat{\rho}$ ebenfalls negativ geschätzt wurde. Das bedeutet, dass die Giraffen diese Zelle nicht bevorzugen und diese eher vermehrt verlassen, als sich in ihr aufzuhalten.

Die größten und kleinsten positiv und negativ geschätzten Werte für den Parameter ρ und der geschätzte Wert Null wurden in der Tabelle hervorgehoben. Die Zeitreihen aus den dazugehörigen Zellen sind in der Tabelle 12 auf Seite 83 genauer aufgeführt und in Abbildung 18 auf Seite 41 eingezeichnet. Die Zelle (10,1) hat den größten $\hat{\rho}$ -Wert mit 0.37. Dieser Wert besagt, dass hier innerhalb der Zellen im Vergleich zu allen anderen 51 Zellen der größte Zusammenhang innerhalb der Zelle über den betrachteten Zeitraum besteht. Die Anzahl der Giraffen ist über die Jahre ähnlich geblieben. Dies bestätigt auch die dazugehörige Hitzekarte und die in grün abgetragene Zeitreihe. Ebenfalls eine sehr hohe, aber negative Korrelation mit -0.268 wurde für die Zelle (12,3) geschätzt. Die zugehörige Zeitreihe ist in hellblau eingetragen. Diese Zeitreihe schwankt, im Gegensatz zur grünen Zeitreihe, ein wenig mehr. Der negative Zusammenhang wird ebenfalls deutlich.

Die Zelle (7,3) ist von besonderem Interesse, da hier der $\hat{\rho}$ -Wert auf Null geschätzt wurde. Dieser Wert spricht nicht für einen Zusammenhang von benachbarten Zellen über die Jahre. In Abbildung 18 ist diese Zeitreihe in rot eingezeichnet. Im Gegensatz zu den anderen Zeitreihen, sind hier sehr hohe Schwankungen in den Giraffenanzahlen sichtbar. Auch für die Zeitreihen der Zellen (11,4) und (3,1) wurde der ρ -Wert nahe Null geschätzt, nämlich mit -0.08 und 0.005. Die jeweiligen Zeitreihen sind in dunkelblau und schwarz eingezeichnet. Bei beiden sind auch Schwankungen zu erkennen, die aber nicht so stark sind wie bei der Zelle (7,3). Eine negative Korrelation in der dunkelblauen und eine positive in der schwarzen Zeitreihe wird beim Vergleichen der beiden Reihen deutlich.

Bei den für dieses Modell durchgeführten Simulationen konnten vorgegebene Werte für den Parameter ρ mit 0.5 bis 0.8 gut geschätzt werden. Bei der Auswertung dieses Datensatzes wurden jeweils geringere Werte geschätzt. Da sich bei der Betrachtung der visualisierten Daten diese Werte bestätigt haben, kann davon aus-

gegangen werden, dass diese Schätzungen für den Datensatz richtig sind. Bei den Simulationen besteht das Problem, dass die Werte randomisiert sind, es Ausreißer geben kann und sich bei weiteren und größeren Simulationen bessere Ergebnisse zeigen könnten.

Somit gibt es Zellen, bei denen über die Jahre viele Giraffen die Zelle verlassen, sich längere Zeit in einer Zelle aufhalten oder in eine Zelle wandern. Als weitere Möglichkeit könnten sie sich in einem Teil des Parks vermehrt aufgehalten haben, dann zum Beispiel diesen Bereich wegen Bauarbeiten gemieden und sich nach Fertigstellung der Baumaßnahmen wieder vermehrt darin aufgehalten haben.

Giraffenanzahlen in den Zellen über die Jahre

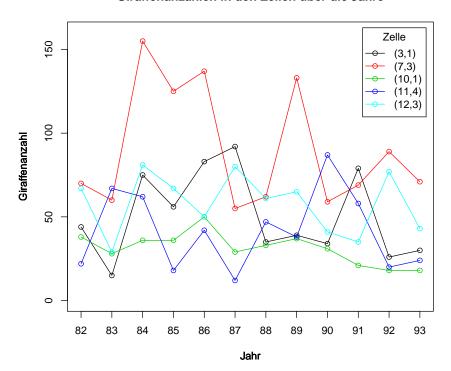


Abb. 18: Anzahlen der Giraffen in den "Extrem"-Zellen über die Jahre

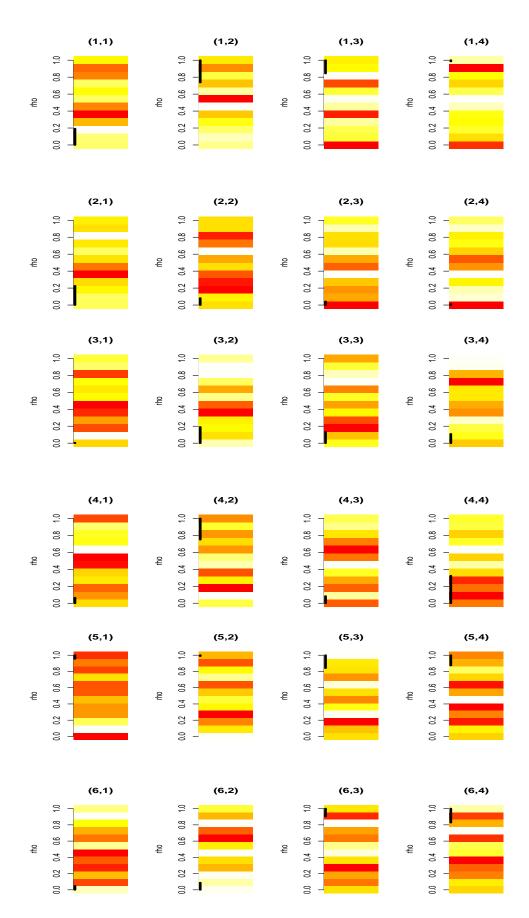


Abb. 19: Zellen des Parkausschnitts über die Jahre mit $\hat{\rho}$

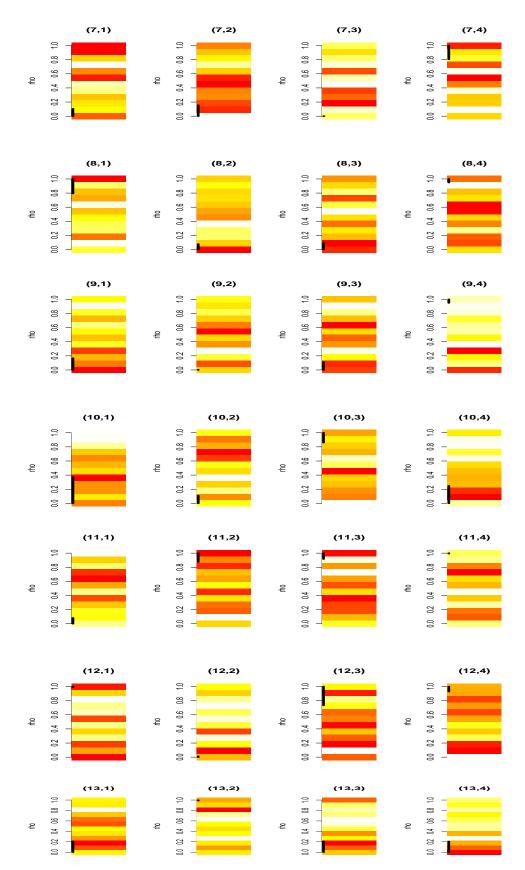


Abb. 20: Zellen des Parkausschnitts über die Jahre mit $\hat{\rho}$

Räumlich

Als Nächstes wird der oben definierte Parkausschnitt im räumlichen Kontext ausgewertet. Dazu wurden die Kennzahlen \bar{x} , s^2 und $s_{(X_{ij},X_{(i+1)(j+1)})}$ des Ausschnitts berechnet und die Parameter des verallgemeinerten Gamma-Poisson-Mischungsmodells mithilfe der Momentenmethode geschätzt. In Tabelle 13 auf Seite 83 sind die berechneten Kennzahlen und die Ergebnisse der Schätzungen eingetragen. Dabei reichen die geschätzten ρ -Werte von 0.238 bis 0.525. In Abbildung 21 sind die geschätzten ρ -Werte in einer Zeitreihe abgetragen. Mit den Simulationen für das räumliche Modell aus Definition 2.11 kann daraus geschlossen werden, dass diese geschätzten Werte die Korrelation wiedergeben, da nur für ρ -Werte größer gleich 0.6 ein signifikanter Unterschied zwischen den geschätzten und wahren Parametern bestand. In Abbildung 22 auf Seite 46 sind die Parkausschnitte in Hitzekarten abgebildet. Die geschätzten ρ -Werte sind jeweils auch eingezeichnet. Dabei muss beachtet werden, dass der Ausschnitt spiegelverkehrt dargestellt ist und die rechte obere Zelle in der Hitzekarte die linke obere Zelle in dem Parkausschnitt darstellt. Der größte ρ -Wert ist im Jahr 1984 und der niedrigste im Jahr 1993 geschätzt

Geschätzte rho-Werte

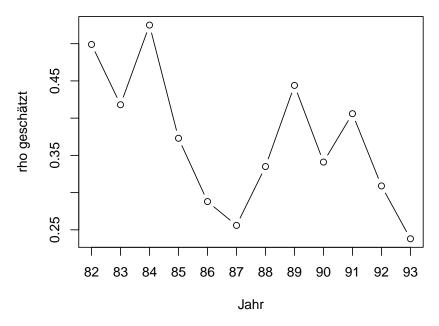
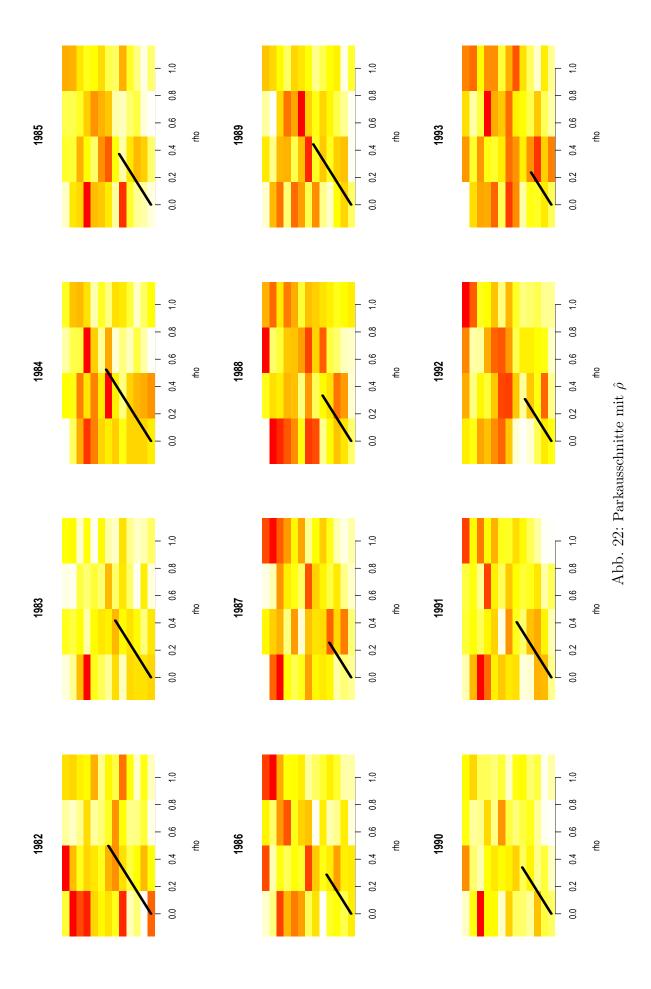


Abb. 21: Ergebnisse der Schätzungen für ρ über alle Jahre

worden. In diesen Jahren besteht ein sehr großer bzw. ein sehr niedriger räumlicher Zusammenhang für benachbarte Zellen, im Gegensatz zu den anderen Jahren. Insgesamt besteht aber in allen Jahren mit ρ -Werten zwischen 0.23 bis 0.53 ein relativ großer räumlicher Zusammenhang zwischen benachbarten Zellen. Dies ist auch in den Hitzekarten ersichtlich. Vor allem in den Jahren 1983 und 1990 sind kaum Abstufungen in den Daten sichtbar.

Aus der Zeitreihe geht hervor, dass ab dem Jahr 1984 die $\hat{\rho}$ -Werte über drei Jahre fallen. Danach steigen sie wieder und fallen in den letzten beiden Jahren nochmal stark ab.

Da für alle Jahre die räumliche Korrelation sehr groß ist, kann daraus geschlossen werden, dass die Giraffen gleichmäßig durch den Park wandern.



Zwei Richtungen

Als Vorletztes wird der Parkausschnitt wiederum in jedem der 12 Jahre einzeln betrachtet und die Zusammenhänge der Zellen in die vertikale und horizontale Richtung ausgewertet. Es wurden die benötigten Kennwerte von der Giraffenanzahl von jedem Jahr berechnet und dann mithilfe der Momentenschätzung die vier Parameter des verallgemeinerten Gamma-Poisson-Mischungsmodells geschätzt. Die Kennwerte und die Schätzergebnisse sind in der Tabelle 14 auf der Seite 84 eingetragen. In Abbildung 23 sind die Ergebnisse der Schätzungen für ρ_1 und ρ_2 in einer Zeitreihe dargestellt. Außerdem sind die Parkausschnitte in Hitzekarten in der Abbildung 24 auf Seite 49 dargestellt. In den Karten sind jeweils die geschätzten ρ_1 -und ρ_2 -Werte eingetragen. Auch hier sind die Parkausschnitte spiegelverkehrt abgebildet.

Als Erstes ist aus der Tabelle wiederum zu entnehmen, dass, falls die empiri-

Geschätzte rho-Werte

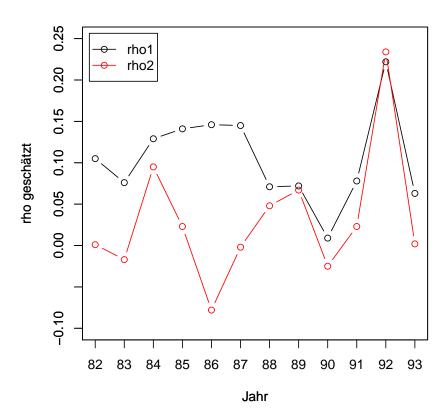


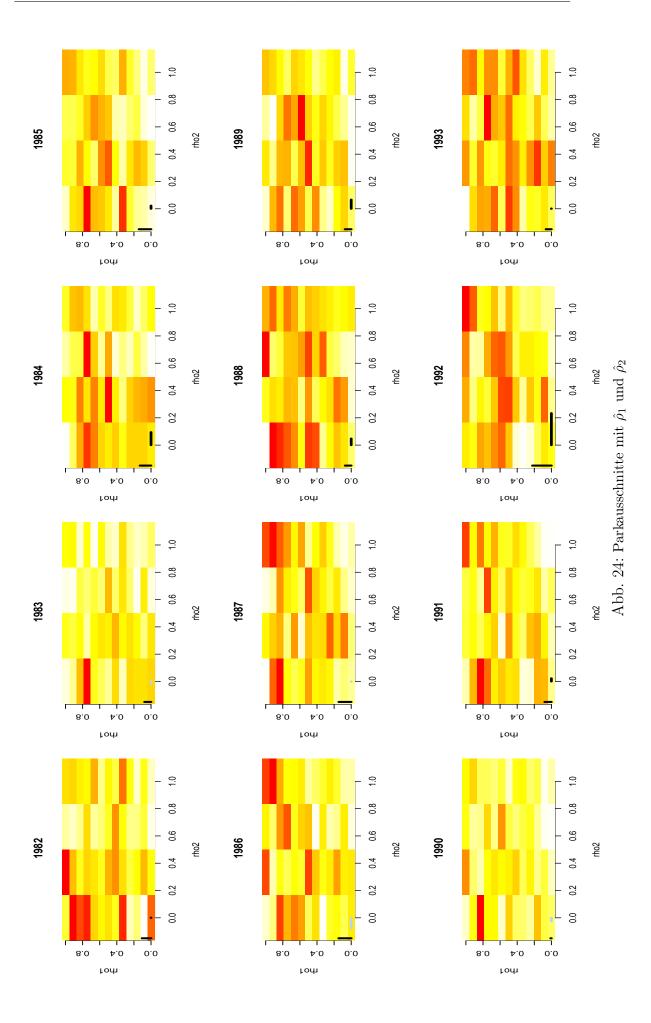
Abb. 23: Ergebnisse der Schätzungen für ρ_1 und ρ_2 über alle Jahre

sche Kovarianz $s_{(X_{ij},X_{i(j+1)})}$ negativ ist, der $\hat{\rho}_2$ -Wert ebenfalls als negativ geschätzt

wurde. In den Abbildungen 24 sind diese negativen Werte in grau eingezeichnet. Derlei niedrige Werte können dadurch zustande gekommen sein, dass in dieser waagerechten Richtung nur vier Zellen betrachtet worden sind. Für das Jahr 1992 wird ρ_2 mit 0.234 am höchsten geschätzt. In der Abbildung des Jahres 1992 sind die Abstufungen in benachbarten Zellen in der Horizontalen schön zu erkennen. Für dieses Jahr wird der ρ_1 -Wert auf 0.222 geschätzt und ist somit auch der am höchsten geschätzte ρ_1 -Wert. Dies spiegelt sich wiederum in der dazugehörigen Hitzekarte wieder. Somit besteht im Jahr 1992 der größte Zusammenhang in beiden Richtungen. Die Giraffen breiten sich gleichmäßig in die Höhe und Breite des Krüger-Nationalparks aus. Im Jahr 1984 sind ebenfalls beide Richtungsparameter als groß geschätzt worden. Bei der Betrachtung des räumlichen Modells wurde der Richtungsparameter für dieses Jahr am höchsten geschätzt.

Aus den Zeitreihen der beiden Parameter geht hervor, dass ab dem Jahr 1988 die beiden geschätzten Parameterwerte jeweils relativ nah beieinander liegen. Die Giraffen bevorzugten in diesen Jahren keine der beiden Richtungen. Für die Jahre 1984 bis 1987 gibt es für den Parameter ρ_1 keine großen Veränderungen.

Die geschätzten Werte für den Parameter ρ_1 sind bis auf eine Ausnahme in jedem Jahren größer als für den Parameter ρ_2 . Da der betrachtete Parkausschnitt die obere und untere Parkhälfte verbindet, müssen die Tiere, um von der einen zur andern Hälfte zu gelangen, diesen Ausschnitt vertikal durchqueren.



Zwei Richtungen mit Zeit

Zum Schluss wird der Parkausschnitt in drei Richtungen betrachtet. Dazu wurden die 12 Parkausschnitte übereinandergelegt und gleichzeitig die vertikale, horizontale und die Zeitrichtung betrachtet. Die berechneten empirischen Kennwerte für die Schätzung der Parameter lauten: $\bar{x}=45.64423,\ s^2=1010.955,\ s_{(X_{ijk},X_{(i+1)jk})}=60.71512,\ s_{(X_{ijk},X_{i(j+1)k})}=221.9839$ und $s_{(X_{ijk},X_{ij(k+1)})}=453.6085.$ Die Schätzungen der fünf Parameter mit der Momentenschätzung fanden diesmal mit dem Softwarepaket R statt, da das Softwareprogramm MATLAB in diesem Fall auch nach längerer Wartezeit noch keine Lösung ausgegeben hatte. Die geschätzten Parameter lauten: $\hat{a}=0.50110826,\ \hat{b}=40.63961453,\ \hat{\rho}_1=0.03147924,\ \hat{\rho}_2=0.11654161$ und $\hat{\rho}_3=0.24958875.$

Der Parameter ρ_3 wird in diesem Fall am größten geschätzt, was bedeutet, dass für die drei Richtungen die Zeitrichtung die Korrelation der Zellen am größten ist. Bei Betrachtung dieser Richtungen ergibt sich für die Giraffen, dass sie insgesamt in den 12 Jahren nicht viel umher wandern und wenn dann über die Breite des Parkausschnitts ziehen.

5. Zusammenfassung

Diese Diplomarbeit befasst sich mit dem verallgemeinerten Gamma-Poisson-Mischungsmodell. Es wird überprüft, ob sich das Modell dazu eignet Korrelationen zwischen benachbarten Zellen in einem Gitternetz mit dem Parameter ρ wiederzugeben. Dazu werden Daten auf Grundlage dieses Modells simuliert und die Parameter aus den simulierten Daten geschätzt. Anschließend werden die geschätzten und vorgegebenen Parameter miteinander verglichen.

Das verallgemeinerte Gamma-Poisson-Mischungsmodell wird auf verschiedenen Hintergründen betrachtet. Das Modell soll die räumliche Korrelation, die Korrelation in zwei und drei Richtungen und die Korrelation von einer Zelle über mehrere Zeitpunkte beschreiben.

Nachdem die vier verschiedenen Modelle auf ihre Anwendbarkeit geprüft wurden, sollen sie auf einem biometrischen Datensatz angewendet werden. Dieser enthält Daten über Tiere, die in den Jahren 1977-1997 im Krüger-Nationalpark in Südafrika lebten. Zur Auswertung wird nur die Anzahl der Giraffen näher betrachtet.

Die Anwendbarkeit des verallgemeinerten Gamma-Poisson-Mischungsmodell zur Beschreibung von Zusammenhängen von benachbarten Zellen in Gitternetzen konnte bestätigt werden. Allerdings gibt es Einschränkungen für jedes der betrachteten Modelle. Das Modell, das eine einzelne Zelle über mehrere Zeitpunkte betrachtet, eignet sich nur, um relativ große Korrelationen wiederzugeben. Alle anderen drei Modelle eignen sich dazu, relativ kleine Korrelationen zu beschreiben.

Bei der Betrachtung jeder einzelnen Zelle über die 12 Jahre kann keine eindeutige Aussage über das Verhalten der Giraffen getroffen werden. Falls keine bestimmte Richtung betrachtet wird, verteilen sich die Giraffen gleichmäßig in diesem Parkausschnitt. Für den Fall, dass der Parkausschnitt mit zwei Richtungen betrachtet wird, wandern die Giraffen über die Länge des Parks. Wenn noch die Zeitrichtung

mit betrachtet wird, dann halten sich immer ähnlich viele Giraffen im Park auf und sie wandern leicht über die Breite des Parkausschnittes. Mit dem verallgemeinerten Gamma-Poisson-Mischungsmodell könnten ebenfalls noch die Anzahl der anderen 18 Tierarten, die in den Datensatz aufgeführt sind, bearbeitet werden. Außerdem könnte der kleinere Parkausschnitt, der in Kapitel 4.2 vorgestellt wird, zur Auswertung betrachtet werden.

Der Datensatz eignet sich für weitere statistische Fragestellungen aus der Biometrie.

Für das verallgemeinerten Gamma-Poisson-Mischungsmodell könnte die Anwendbarkeit in weiteren Simulationen überprüft werden. Vor allem für das Modell, bei dem drei Richtungen betrachtet werden. Somit könnten noch weitere Aussagen über dieses Modell getroffen werden.

Das verallgemeinerten Gamma-Poisson-Mischungsmodell lässt sich auch in anderen Studien anwenden. Überall dort wo ein Gitternetz über ein Bereich gelegt werden kann, ist die Anwendung möglich. Somit könnten Risse, die in einer Stahlplatte entstehen, auf ihre Ausbreitungrichtung hin überprüft werden.

A. R-Code

Abbildungen 1 & 2: Problem bei der Simulation aller Daten

```
eindimen <- function(n,a,b,rho){
lambdaber <- function(rho1,t1,a1,b1){</pre>
gamm <- rgamma(t1,a1,scale=b1)</pre>
1<-rep(0,t1)</pre>
for(x in (1:t1)) {
z \leftarrow rep(0,t1)
for(r in ((1):(t1))){
z[r] <- rho1^(abs(x-r))*gamm[r] }</pre>
l[x] \leftarrow sum(z)
rpois(t1,1)
zahlen <- lambdaber(rho,n,a,b)
##Bild
x < -(seq(1,n,1))
reihe <-cbind(x,zahlen)
plot(reihe,type="1",xlab="Zeitpunkt",ylab="Anzahl")
eindimen(500,50,0.5,0.9)
eindimen(500, 0.5, 50, 0.9)
Abbildungen 3: Simulation Teildaten (1-dim)
eindimen <- function(n,a2,b2,rho){
lambdaber <- function(rho1,t1,a1,b1){</pre>
gamm <- rgamma(t1,a1,scale=b1)</pre>
l<-rep(0,t1)</pre>
for(x in (1:t1)) {
z \leftarrow rep(0,t1)
for(r in ((1):(t1))){
z[r] \leftarrow (rho1^(abs(x-r)))*gamm[r] 
l[x] \leftarrow sum(z)
rpois(t1,1)
```

```
}
allezahlen <- lambdaber(rho,n,a2,b2)</pre>
zahlen \leftarrow allezahlen[(n*33/100):(n*67/100)]
#Berechnungen
xquer <-mean(zahlen)</pre>
empVar <- var(zahlen)</pre>
covdach1 \leftarrow cov(zahlen[-(n*33/100)], zahlen[-1])
#Berechnungen der Werte mit den Parametern, die angegeben wurden
xquerwirklich<- (1+rho)/(1-rho)*a2*b2
empVarwirklich<- (1+rho)/(1-rho)*a2*b2+(1+rho^2)/(1-rho^2)*a2*b2^2
covdach1wirklich<- 2*rho/(1-rho^2)*a2*b2^2</pre>
print(zahlen)
print(xquer)
print(xquerwirklich)
print(empVar)
print(empVarwirklich)
print(covdach1)
print(covdach1wirklich)
momentfunc <-
function(xquer,empVar,covdach1) {
func<- function(x){</pre>
      a <- x[1]
      b < -x[2]
      w < -x[3]
      (xquer-(((1+w)/(1-w))*a*b))^2+
      (empVar-(((1+w^2)/(1-w^2))*a*b^2 + ((1+w)/(1-w))*a*b))^2+
      (covdach1-((w*2/(1-w^2))*a*b^2))^2
                   }
optim(c(15,0,0), func, method="L-BFGS-B", lower=c(0,0,0),
upper=c(Inf,Inf,1))
}
moment <- try(momentfunc(xquer,empVar,covdach1))</pre>
####Bilder
par(mfrow=c(1,2))
```

```
x<-(seq((n*33/100),(n*67/100),1))
reihe <-cbind(x,zahlen)</pre>
plot(reihe,type="l",xlab="Zeitpunkt",ylab="Anzahl")
image(t(zahlen),col=rev(heat.colors(15)),ylab="rho",oldstyle=T,
frame.plot = FALSE,xaxt = "n")
lines(c(-0.998, -0.998), c(moment*par[3], 0))
moment
eindimen(2000, 0.5, 50, 0.6)
Abbildung 4: Darstellung einer zu minimierenden Funktion
par(xaxs="i")
par(yaxs="i")
w<-0.9
u \leftarrow seq(1,5,length=10)
v \leftarrow seq(1,5,length=50)
f \leftarrow function(u,v) \{ r \leftarrow (473.0164 - (((1+0.9)/(1-0.9))*u*v))^2 + ((1+0.9)/(1-0.9))*u*v) \}
(770.8164 - (((1+0.9^2)/(1-0.9^2))*u*v^2 + ((1+0.9)/(1-0.9))*u*v))^2 +
(469.9944-((0.9*2/(1-0.9^2))*u*v^2))^2
```

Abbildung 12: Histogramme von Simulationen mit 2 Richtungen

xlab = "a", ylab = "b" , zlab = "Funktionswert", las=3

persp(u, v, z, theta = 45, phi = 20, expand = 0.6,

,main="a, b variabel, festes rho=0.9"

z <- outer(u, v, f)

col = "red3",

ticktype = "detailed",

```
par(mfrow=c(1,2))
zweidim0909<-read.table("Dateipfad/zweirich(50,50,0.2,50,0.9,0.9).txt",
header=F,sep=";")
#par(mfrow=c(2,2))
dat3<-zweidim0909[-c(1:4),8]
hist(dat3,freq=FALSE,main="rho1=0.9, rho2=0.9",xlab="rho1 geschätzt")
xx <- seq(min(dat3),max(dat3), length=100)</pre>
```

```
lines(xx, dnorm(xx, mean(dat3), sd(dat3)), col="red")
zweidim0805<-read.table("Dateipfad/zweirich(50,50,0.2,50,0.8,0.5).txt",</pre>
header=F,sep=";")
\#par(mfrow=c(2,2))
dat2 < -zweidim0805[-c(1:4),9]
hist(dat2,freq=FALSE,main="rho1=0.8, rho2=0.5",xlab="rho2 geschätzt")
xx <- seq(min(dat2), max(dat2), length=100)</pre>
lines(xx, dnorm(xx, mean(dat2), sd(dat2)), col="red")
Abbildung 18: Plot mit Giraffenanzahlen in "Extrem"-Zellen
giraffenzr <- read.table ("Dateipfad/Giraffendaten_Zeitreihe.txt", sep=";",
header=FALSE)
plot(giraffenzr$V10,type="o",col=1,axes=T,xaxt = "n",xlab="Jahr",
ylab="Giraffenanzahl",ylim=c(0,160),main="Giraffenanzahlen in den
Zellen über die Jahre")
par(new=T)
plot(giraffenzr$V28,type="o",col=2,axes=F,xlab="Jahr",
ylab="Giraffenanzahl",ylim=c(0,160))
par(new=T)
plot(giraffenzr$V38,type="o",col=3,axes=F,xlab="Jahr",
ylab="Giraffenanzahl",ylim=c(0,160))
par(new=T)
plot(giraffenzr$V45,type="o",col=4,axes=F,xlab="Jahr",
ylab="Giraffenanzahl",ylim=c(0,160))
par(new=T)
plot(giraffenzr$V48,type="o",col=5,axes=F,xlab="Jahr",
ylab="Giraffenanzahl",ylim=c(0,160))
legend("topright", "(x,y)", lty=1, pch=1,col=c(1,2,3,4,5),
       title="Zelle",legend= c("(3,1)","(7,3)","(10,1)","(11,4)",
       "(12,3)"),inset = .02)
axis(1,1:12,c(82,83,84,85,86,87,88,89,90,91,92,93))
par(new=F)
```

Abbildungen 21 & 23: Zeitreihen für die geschätzten ρ -Werte

```
#Räumlich
rho<- c(0.499,0.418,0.525,0.373,0.288,0.256,0.335,0.444,0.341,0.406,
0.309, 0.238)
plot(rho, xaxt="n", xlab="Jahr", ylab="rho geschätzt", type="b",
main="Geschätzte rho-Werte")
axis(1,1:12,c(82,83,84,85,86,87,88,89,90,91,92,93))
#Zwei Richtungen
rho1<-c(0.1050,0.0760,0.1290,0.1410,0.1460,0.1450,0.0710,0.0720,0.0090,
0.0780,0.2220,0.0630)
rho2<-c(0.0010,-0.017,0.0950,0.0230,-0.078,-0.002,0.0480,0.0670,-0.025,
0.0230,0.2340,0.0020)
plot(rho1,xaxt="n",xlab="Jahr",ylab="rho geschätzt",type="b",
main="Geschätzte rho-Werte", ylim=c(-0.1,0.25))
par(new=T)
plot(rho2,xaxt="n",type="b",xlab="Jahr",ylab="rho geschätzt",
ylim=c(-0.1,0.25),col=2)
par(new=F)
axis(1,1:12,c(82,83,84,85,86,87,88,89,90,91,92,93))
legend("topleft", "(x,y)", lty=1, pch=1,col=c(1,2),
legend= c("rho1","rho2"),inset = .02)
Simulation des eindimensionalen Modells
#######anzahl = Anzahl der Wiederholungen
####### = Anzahl der Daten, die simuliert werden sollen
#######a2,b2,rho = festzulegende Parameter
eindimen <- function(anzahl,n,a2,b2,rho){
ergebnis <- matrix(rep(0,3*anzahl+3), ncol=3, byrow=TRUE)
##Berechnungen der Kennzahlen mit den vorgegebenen Parametern
ergebnis[1,1] \leftarrow (1+rho)/(1-rho)*a2*b2
ergebnis[1,2]<- (1+rho)/(1-rho)*a2*b2+(1+rho^2)/(1-rho^2)*a2*b2^2
ergebnis[1,3]<- 2*rho/(1-rho^2)*a2*b2^2
```

###Datensimulation:

```
lambdaber <- function(rho1,t1,a1,b1){</pre>
gamm <- rgamma(t1,a1,scale=b1)</pre>
1 < -rep(0,t1)
for(x in (1:t1)) {
z \leftarrow rep(0,t1)
for(r in (1:t1)){
z[r] \leftarrow (rho1^(abs(x-r)))*gamm[r] 
l[x] \leftarrow sum(z)
rpois(t1,1)
}
for (i in (1:anzahl)){
allezahlen <- lambdaber(rho,n,a2,b2) #Daten werden erzeugt
zahlen \leftarrow allezahlen[(n*33/100):(n*67/100)] #33% der mittleren Daten
#Berechnungen
ergebnis[i+1,1] <-mean(zahlen)
ergebnis[i+1,2] <- var(zahlen)
ergebnis[i+1,3] <- cov(zahlen[-(n*33/100)],zahlen[-1])
print(ergebnis)
eindimen(50,2000,0.5,50,0.9)
Simulation des räumlichen Modells
######anzahl = Anzahl der Wiederholungen
####### = Anzahl der Zeilen und Spalten
#######a2,b2,rho = festzulegende Parameter
raeumlich<-function(anzahl,n,a,b,rho){
ergebnis <- matrix(rep(0,3*anzahl+3), ncol=3, byrow=TRUE)
##Berechnungen der Kennzahlen mit den vorgegebenen Parametern
ergebnis[1,1] < ((1+rho)/(1-rho))^2*a*b
ergebnis[1,2]<- ((1+rho)/(1-rho))^2*a*b+((1+rho^2)/(1-rho^2))^2*a*b^2
ergebnis[1,3]<- ((2*rho)/(1-rho^2))^2*a*b^2
###Datensimulation:
lambdaber <- function(rho11,n1,a1,b1){</pre>
```

```
#Gammazufallszahlen erzeugen
gamm <- matrix(rgamma(n1^2,a1,scale=b1),ncol=n1,byrow=TRUE)</pre>
l<-matrix(rep(0,n1^2),ncol=n1,byrow=TRUE)</pre>
for(i in (1:n1)){
for(j in (1:n1)){
z <- matrix(rep(0,n1^2),ncol=n1,byrow=TRUE)</pre>
for(t1 in (1:n1)){
for(s1 in (1:n1)){
z[s1,t1] \leftarrow (rho11^(abs(i-s1)+abs(j-t1)))*gamm[s1,t1] 
l[i,j] < -sum(z)
    }
for (i1 in 1:anzahl){
12<-as.vector(1)
poi<-rpois(n1^2,12)</pre>
matrix(test,ncol=n1,byrow=FALSE)
}
gesmatrix<-lambdaber(rho,n,a,b)</pre>
#33% der mitte der Matrix für weitere berechnungen
#betrachtung33 <- round(sqrt(n^2*33/100)+0.49)
f < -round(n/2+0.4)
st<-f/2
matrixber<-gesmatrix[-c(seq(1:round(st+0.4)),seq(round(st+0.4)+f,n)),</pre>
-c(seq(1:round(st+0.4)), seq(round(st+0.4)+f,n))]
########
#Berechnungen
I<-n-2*st-1
J<-n-2*st-1
xquer <- 1/(I*J)*sum(matrixber)</pre>
empVar <- 1/(I*J-1)*sum((matrixber-xquer)^2)</pre>
covdach <- {
xquer1 \leftarrow 1/((I-1)*(J-1))*(sum(matrixber)-sum(matrixber[I,J]))
xquer2 \leftarrow 1/((I-1)*(J-1))*(sum(matrixber)-sum(matrixber[1,1]))
f <- matrixber-xquer1</pre>
```

```
g <- matrixber-xquer2
d \leftarrow \{p \leftarrow (I-1)*(J-1)\}, nrow=I-1, ncol=J-1, byrow=F\}
for(i in (1:(I-1))) {
for(j in (1:(J-1))) {
p[i,j] \leftarrow (f[i,j]*g[i+1,j+1])
sum(p)
1/((I-1)*(J-1)-1)*d
ergebnis[i1+1,1] <- xquer
ergebnis[i1+1,2] <- empVar
ergebnis[i1+1,3] <- covdach
try(print(ergebnis))
raeumlich(50,50,0.2,50,0.9)
Beispiel für den Einstichproben-t-Test
raeumlich06 <- read.table("G:/raeumlich(50,50,0.2,50,0.6).txt",</pre>
header=F,sep=";")
t.test(raeumlich06[-1,5],mu=0.2)
t.test(raeumlich06[-1,6],mu=50)
t.test(raeumlich06[-1,7],mu=0.6)
Simulation des Modells mit zwei Richtungen
######anzahl = Anzahl der Wiederholungen
####### = Anzahl der Zeilen und Spalten
#######a2,b2,rho1,rho2 = festzulegende Parameter
zweirich<-function(anzahl,n,a,b,rho1,rho2){</pre>
ergebnis <- matrix(rep(0,4*anzahl+4), ncol=4, byrow=TRUE)
##Berechnungen der Kennzahlen mit den vorgegebenen Parametern
ergebnis[1,1] \leftarrow (1+rho1)/(1-rho1)*(1+rho2)/(1-rho2)*a*b
ergebnis[1,2] \leftarrow (1+rho1)/(1-rho1)*(1+rho2)/(1-rho2)*a*b+
(1+rho1^2)/(1-rho1^2)*(1+rho2^2)/(1-rho2^2)*a*b^2
ergebnis[1,3]<- (1+rho2^2)/(1-rho2^2)*(2*rho1)/(1-rho1^2)*a*b^2
```

```
ergebnis[1,4]<- (1+rho1^2)/(1-rho1^2)*(2*rho2)/(1-rho2^2)*a*b^2
###Datensimulation:
lambdaber <- function(rho11,rho22,n1,a1,b1){</pre>
#Gammazufallsvariablen
gamm <- matrix(rgamma(n1^2,a1,scale=b1),ncol=n1,byrow=TRUE)</pre>
l<-matrix(rep(0,n1^2),ncol=n1,byrow=TRUE)</pre>
for(i in (1:n1)){
for(j in (1:n1)){
z <- matrix(rep(0,n1^2),ncol=n1,byrow=TRUE)</pre>
for(t1 in (1:n1)){
for(s1 in (1:n1)){
z[s1,t1] \leftarrow (rho11^(abs(i-s1)))*(rho22^(abs(j-t1)))*gamm[s1,t1] 
l[i,j] < -sum(z)
     }
12<-as.vector(1)
poi<-rpois(n1^2,12)</pre>
matrix(poi,ncol=n1,byrow=FALSE)
}
for (i1 in 1:anzahl){
gesmatrix<-lambdaber(rho1,rho2,n,a,b)</pre>
#33% der mitte der Matrix für weitere Berechnungen
f < -round(n/2+0.4)
st<-f/2
matrixber<-gesmatrix[-c(seq(1:round(st+0.4)),seq(round(st+0.4)+f,n)),</pre>
-c(seq(1:round(st+0.4)), seq(round(st+0.4)+f,n))]
#######
#Berechnungen
I < -n-2 * st-1
J<-n-2*st-1
xquer <- 1/(I*J)*sum(matrixber)</pre>
empVar <- 1/(I*J-1)*sum((matrixber-xquer)^2)</pre>
covdach1 <- {
xquer1 <- 1/((I-1)*J)*(sum(matrixber)-sum(matrixber[I,]))</pre>
```

```
xquer2 <- 1/((I-1)*J)*(sum(matrixber)-sum(matrixber[1,]))</pre>
f <- matrixber-xquer1</pre>
g <- matrixber-xquer2
d \leftarrow sum(g[-1,,]*f[-I,,])
1/((I-1)*J-1)*d
}
covdach2 <- {
xquer3 <- 1/(I*(J-1))*(sum(matrixber)-sum(matrixber[,J]))</pre>
xquer4 <- 1/(I*(J-1))*(sum(matrixber)-sum(matrixber[,1]))</pre>
k <- matrixber-xquer3
1 <- matrixber-xquer4</pre>
o <- sum(k[,-1]*l[,-J])
 1/(I*(J-1)-1)*o
}
ergebnis[i1+1,1] <- xquer
ergebnis[i1+1,2] <- empVar
ergebnis[i1+1,3] <- covdach1</pre>
ergebnis[i1+1,4] <- covdach2
}
try(print(ergebnis))
}
zweirich(50,50,0.2,50,0.8,0.7)
Simulation des Modells mit drei Richtungen
######anzahl = Anzahl der Wiederholungen
####### = Anzahl der Zeilen und Spalten
#######a2,b2,rho1,rho2,rho3 = festzulegende Parameter
dreirich<-function(anzahl,n,a,b,rho1,rho2,rho3){</pre>
ergebnis <- matrix(rep(0,5*anzahl+5), ncol=5, byrow=TRUE)
##Berechnungen der Werte mit den Parametern, die angegeben wurden
ergebnis[1,1] \leftarrow (1+rho1)/(1-rho1)*(1+rho2)/(1-rho2)*(1+rho3)/(1-rho3)*a*b
ergebnis[1,2] < (1+rho1)/(1-rho1)*(1+rho2)/(1-rho2)*(1+rho3)/(1-rho3)*a*b+
(1+rho1^2)/(1-rho1^2)*(1+rho2^2)/(1-rho2^2)*(1+rho3^2)/(1-rho3^2)*a*b^2
ergebnis[1,3]<- (1+rho2^2)/(1-rho2^2)*(2*rho1)/(1-rho1^2)*
```

```
(1+rho3^2)/(1-rho3^2)*a*b^2
ergebnis[1,4]<- (1+rho1^2)/(1-rho1^2)*(2*rho2)/(1-rho2^2)*
(1+rho3^2)/(1-rho3^2)*a*b^2
ergebnis[1,5]<- (1+rho1^2)/(1-rho1^2)*(1+rho2^2)/(1-rho2^2)*
(2*rho3)/(1-rho3^2)*a*b^2
#Berechnungen der poissonwerte für lambda_ij
lambdaber <- function(rho11,rho22,rho33,n1,a1,b1){</pre>
#Gammazufallsvariablen
gamm <- array(rgamma(n1^3,a1,scale=b1),dim=c(n1,n1,n1))</pre>
l<-array(rep(0,n1^3),dim=c(n1,n1,n1))</pre>
for(i in (1:n1)){
for(j in (1:n1)){
for(k in (1:n1)){
z <- array(rep(0,n1^3),dim=c(n1,n1,n1))</pre>
for(u1 in (1:n1)){
for(t1 in (1:n1)){
for(s1 in (1:n1)){
z[s1,t1,u1] < -(rho11^(abs(i-s1)))*(rho22^(abs(j-t1)))*(rho33^(abs(k-u1)))*
gamm[s1,t1,u1] }
}}
l[i,j,k] < -sum(z)
}}}
poi <- rpois(n1^3,1)</pre>
array(poi,dim=c(n1,n1,n1))
for (i1 in 1:anzahl){
gesarray<-lambdaber(rho1,rho2,rho3,n,a,b)</pre>
#33% der mitte der Datenmatrix für weitere Berechnungen
f < -round(n/2+0.4)
st<-f/2
matrixber <- gesarray [-c(seq(1:round(st+0.4)), seq(round(st+0.4)+f,n)),
-c(seq(1:round(st+0.4)), seq(round(st+0.4)+f,n)),
-c(seq(1:round(st+0.4)), seq(round(st+0.4)+f,n))]
########
```

```
#Berechnungen
I<-n-2*st-1
J<-n-2*st-1
K<-n-2*st-1
xquer <- 1/(I*J*K)*sum(matrixber)</pre>
empVar <- 1/(I*J*K-1)*sum((matrixber-xquer)^2)</pre>
covdach1 <- {
xquer1 <- 1/((I-1)*J*K)*(sum(matrixber)-sum(matrixber[I,,]))</pre>
xquer2 <- 1/((I-1)*J*K)*(sum(matrixber)-sum(matrixber[1,,]))</pre>
f <- matrixber-xquer1</pre>
g <- matrixber-xquer2
d \leftarrow sum(g[-1,,]*f[-I,,])
1/((I-1)*J*K-1)*d
covdach2 <- {
xquer3 <- 1/(I*(J-1)*K)*(sum(matrixber)-sum(matrixber[,J,]))</pre>
xquer4 \leftarrow 1/(I*(J-1)*K)*(sum(matrixber)-sum(matrixber[,1,]))
k <- matrixber-xquer3
1 <- matrixber-xquer4</pre>
o \leftarrow sum(l[,-1,]*k[,-J,])
1/(I*(J-1)*K-1)*o
covdach3 <- {
xquer5 <- 1/(I*J*(K-1))*(sum(matrixber)-sum(matrixber[,,K]))</pre>
xquer6 <- 1/(I*J*(K-1))*(sum(matrixber)-sum(matrixber[,,1]))</pre>
v <- matrixber-xquer5
w <- matrixber-xquer6
z \leftarrow sum(w[,,-1]*v[,,-K])
1/(I*(J)*(K-1)-1)*z
}
ergebnis[i1+1,1] <- xquer
ergebnis[i1+1,2] <- empVar
ergebnis[i1+1,3] <- covdach1</pre>
ergebnis[i1+1,4] <- covdach2
```

```
ergebnis[i1+1,5] <- covdach3
try(print(ergebnis))
dreirich(45,14,0.2,50,0.5,0.2,0.1)
Schätzungen der Parameter beim Modell mit drei Richtungen
datei <- read.table("Dateiname.txt",sep="")</pre>
momentfunc <-
function(xquer,empVar,covdach1,covdach2,covdach3) {
func<- function(x){</pre>
 a <- x[1]
                 #a
b < -x[2]
                 #b
 w < -x[3]
                 #rho1
v \leftarrow x[4]
                 #rho2
 z < -x[5]
                 #rho3
 (xquer-(((1+w)/(1-w))*((1+v)/(1-v))*((1+z)/(1-z))*a*b))^2+
 (empVar - (((1+w^2)/(1-w^2))*((1+v^2)/(1-v^2))*((1+z^2)/(1-z^2))*a*b^2+
     ((1+w)/(1-w))*((1+v)/(1-v))*((1+z)/(1-z))*a*b))^2+
 (covdach1-((w*2/(1-w^2))*((1+v^2)/(1-v^2))*((1+z^2)/(1-z^2))*a*b^2))^2+
 (covdach2-((v*2/(1-v^2))*((1+w^2)/(1-w^2))*((1+z^2)/(1-z^2))*a*b^2))^2+
 (covdach3-((z*2/(1-z^2))*((1+w^2)/(1-w^2))*((1+v^2)/(1-v^2))*a*b^2))^2
optim(c(1,0,0,0,0), func, method="L-BFGS-B", lower=c(0,0,-1,-1,-1),
upper=c(50000,50000,1,1,1))
}
ergebnis <- matrix(rep(0,6*nrow(datei)), ncol=6, byrow=TRUE)
for (i in 1:(nrow(datei))){
moment <- try(momentfunc(datei[i,1],datei[i,2],datei[i,3],datei[i,4],</pre>
datei[i,5]),silent=TRUE)
if (class(moment) == "try-error"){
ergebnis[i,1]<-8.888888
}else
{
```

```
ergebnis[i,1]<-moment$par[1]</pre>
ergebnis[i,2]<-moment$par[2]
ergebnis[i,3]<-moment$par[3]</pre>
ergebnis[i,4]<-moment$par[4]</pre>
ergebnis[i,5]<-moment$par[5]
ergebnis[i,6]<-moment$value
} }
print(ergebnis)
KNP: Auswertung der Giraffenanzahl
daten2 <- read.table("G:/datenab81_84.txt",sep=",",header=TRUE)</pre>
daten3 <- read.table("G:/datenab85_97.txt",sep=",",header=TRUE)</pre>
giraffed2 <-daten2[daten2$SP=="kp",] #Giraffen herausrausfiltern
jahr1981 <- giraffed2[giraffed2$YEAR=="1981",]</pre>
jahr1982 <- giraffed2[giraffed2$YEAR=="1982",]</pre>
jahr1983 <- giraffed2[giraffed2$YEAR=="1983",]</pre>
jahr1984 <- giraffed2[giraffed2$YEAR=="1984",]</pre>
giraffed3 <-daten3[daten3$SP=="kp",]</pre>
jahr1985 <- giraffed3[giraffed3$YEAR=="1985",]</pre>
jahr1986 <- giraffed3[giraffed3$YEAR=="1986",]</pre>
#Rechteckgröße bestimmen
I1<-4
        #Abschnitt x-Achse
J<-13
        #Abschnitt y-Achse
K<-0.1 #Länge der Abschnitte
#Abschnitte bestimmmen
###Längengrad
ir1 < - seq(31.6, 32.0 - K + 0.01, by = K)
ir2 <- seq(31.6+K-0.01,32.0,by=K)
mi <- cbind(ir1,ir2)</pre>
###Breitengrad
jr1 < -seq(-24.2, -25.4-K+0.01, by=-K)
```

```
jr2 < -seq(-24.2+K-0.01, -25.4, by=-K)
mj <- cbind(jr2,jr1)</pre>
x <- rep(ir1, each=J)
y<- mj[,1]
 A \leftarrow cbind(x,y)
 ####################################
 ###Wieviele Tiere leben in jedem der Abschnitte
 ###Jahr 1982
n82<-{
 I <- length(mi[,1])</pre>
 J \leftarrow length(mj[,2])
v < -rep(0, I * J)
n <- matrix(v,ncol=J)</pre>
for(i in 1:I) {
 for(j in 1:J) {
for(k in (1:sum(jahr1982$SP=="kp"))){
 if (sum(mi[i,1] \le jahr1982[k,5] \& jahr1982[k,5] \le mi[i,2]\&mj[j,2] \le jahr1982[k,5] \le mi[i,2]\&mj[j,2] \le jahr1982[k,5] \le jahr19
 jahr1982[k,4] & jahr1982[k,4] <= mj[j,1])>0) {
        n[i,j] < -n[i,j] + jahr 1982[k,3]
}}}
n
 }
###Plot in schwarz-weiß für das Jahr 1982####
x <- rep(jr1, each=I1)
y<- mi[,1]
A \leftarrow cbind(y,x)
z <- cbind(rep(1,I1*J),rep(1,I1*J))</pre>
M < - A + (K/2) * z
E \leftarrow A+(K-1)*z
matvec <- as.vector(n82)</pre>
nullvec <- rep(0,I1*J)</pre>
yachse <- cbind(nullvec,matvec/1000)</pre>
 a \leftarrow pi/3
```

```
drehmat <- matrix(c(sin(a),-cos(a),cos(a),sin(a)),byrow=TRUE,nrow=2)</pre>
neuendpkt <- yachse %*% drehmat
gedreht <- neuendpkt+M</pre>
punkt \leftarrow c(0,0)
plot(punkt,xlim=c(31.6,32.02),ylim=c(-25.5,-24.0),xlab="Längengrad",
ylab="Breitengrad",type="n",main="1982")
for(i in 1:(I1*J)){
segments(M[i,1],M[i,2],gedreht[i,1],gedreht[i,2],col=c(if(0==matvec[i])
{"white"}, if (1<=matvec[i]) {"black"}))}
#######Hitzekarte für das Jahr 1982####
nn82<-matrix(rev(n82),nrow=4,ncol=13, byrow=F)
image(nn82,col=rev(heat.colors(100)),oldstyle=T,main=" 1982",
frame.plot = FALSE,ylab="rho1",xlab="rho2")
lines(c(-0.15,-0.15),c(0.105,0),lwd=3)
                                           #Linie von rho1
lines(c(0.0010,0),c(0,0),lwd=3)
                                           #Linie von rho2
```

B. MATLAB-Code

```
%%/für Schätzungen mit 3 Parametern (räumlich):
Matrix = xlsread('Dateiname.xls');
size(Matrix)
syms a b p;
format longG
lsga=zeros(3,1);
lsgb=zeros(3,1);
lsgp=zeros(3,1);
for i=1:size(Matrix,1)
coeff1=Matrix(i,1);
coeff2=Matrix(i,2);
coeff3=Matrix(i,3);
syms a b p;
g1 = (coeff1-((1+p)/(1-p))^2*a*b)^2;
g2 = (coeff2-(((1+p)/(1-p))^2*a*b+((1+p^2)/(1-p^2))^2*a*b^2))^2 ;
g3 = (coeff3-(2*p/(1-p^2)^2)^2*a*b^2)^2;
[L1 L2 L3] = solve(g1,g2,g3);
 if imag(L3(2,1))^{-}=0
                     %Bestrafungswert, falls Schätzung komplexe
 lsga(i,1)=8888;
 lsgb(i,1)=8888;
                     %Werte annimmt
 lsgp(i,1)=8888;
 else
lsga(i,1)= double(round(L1(2,1)*1000)/1000);
lsgb(i,1)= double(round(L2(2,1)*1000)/1000);
lsgp(i,1)= double(round(L3(2,1)*1000)/1000);
end
end
lsga lsgb lsgp
```

```
%%%für Schätzungen mit 3 Parametern (1-dim mit Zeit):
Matrix = xlsread('Dateiname.xls');
size(Matrix)
syms a b p;
format longG
lsga=zeros(3,1);
lsgb=zeros(3,1);
lsgp=zeros(3,1);
for i=1:size(Matrix,1)
coeff1=Matrix(i,1);
coeff2=Matrix(i,2);
coeff3=Matrix(i,3);
syms a b p;
g1 = (coeff1-((1+p)/(1-p))*a*b)^2;
g2 = (coeff2 - (((1+p)/(1-p))*a*b+((1+p^2)/(1-p^2))*a*b^2))^2;
g3 = (coeff3-(2*p/(1-p^2))*a*b^2)^2;
[L1 L2 L3] = solve(g1,g2,g3);
if imag(L1(4,1))^{-}=0
 lsga(i,1)=8888;
 lsgb(i,1)=8888;
 lsgp(i,1)=8888;
else
lsga(i,1)= double(round(L1(1,1)*1000)/1000);
lsgb(i,1) = double(round(L2(1,1)*1000)/1000);
lsgp(i,1) = double(round(L3(1,1)*1000)/1000);
end
end
lsga
lsgb
lsgp
```

%%für Schätzungen mit 4 Parametern (2 Richtungen):

```
Matrix = xlsread('Dateiname.xls');
size(Matrix)
syms a b p1 p2;
format longG
lsga=zeros(4,1);
lsgb=zeros(4,1);
lsgp1=zeros(4,1);
lsgp2=zeros(4,1);
for i=1:size(Matrix,1)
coeff1=Matrix(i,1);
coeff2=Matrix(i,2);
coeff3=Matrix(i,3);
coeff4=Matrix(i,4);
syms a b p1 p2;
g1 = (coeff1-((1+p1)/(1-p1))*((1+p2)/(1-p2))*a*b)^2;
g2 = (coeff2-(((1+p1)/(1-p1))*((1+p2)/(1-p2))*a*b+((1+p1^2)/(1-p1^2))
      *((1+p2^2)/(1-p2^2))*a*b^2))^2;
g3 = (coeff3-((1+p2^2)/(1-p2^2))*(2*p1/(1-p1^2))*a*b^2)^2;
g4 = (coeff4-((1+p1^2)/(1-p1^2))*(2*p2/(1-p2^2))*a*b^2)^2;
[L1 L2 L3 L4] = solve(g1,g2,g3,g4);
 if imag(L1(16,1))^{-}=0
  lsga(i,1)=8888;
  lsgb(i,1)=8888;
  lsgp1(i,1)=8888;
  lsgp2(i,1)=8888;
 else
lsga(i,1)= double(round(L1(4,1)*1000)/1000);
lsgb(i,1)= double(round(L2(4,1)*1000)/1000);
lsgp1(i,1)= double(round(L3(4,1)*1000)/1000);
lsgp2(i,1) = double(round(L4(4,1)*1000)/1000);
end
end
lsga lsgb lsgp1 lsgp2
```

C. Tabellen und Grafiken

C.1. Simulationen

Anzahl Durchläufe: 50, Anzahl Daten: 2000								
$(\mathrm{a,b},\!\rho)$	n	p-Wert \hat{a}	p-Wert \hat{b}	p-Wert $\hat{\rho}$				
(0.5,50,0.9)	46	0.6958	0.01368	0.001379				
(0.5, 50, 0.8)	50	0.2634	0.5674	0.1650				
(0.5, 50, 0.7)	50	0.6782	0.522	0.01487				
(0.5, 50, 0.6)	50	0.03574	0.02738	0.2199				
(0.5, 50, 0.5)	50	0.04805	0.2306	0.7813				
(0.5, 50, 0.4)	50	0.5931	0.5951	0.05175				
(0.5, 50, 0.3)	50	0.3195	0.2626	0.3054				
(0.5, 50, 0.2)	50	0.9534	0.8136	0.00044				
(0.5, 50, 0.1)	50	0.9555	0.9079	0.002144				
(0.2,150,0.9)	50	0.1744	0.9755	0.2996				
(0.2, 150, 0.8)	50	0.1443	0.4366	0.2514				
(0.2, 150, 0.7)	50	0.08001	0.3063	0.4386				
(0.2, 150, 0.6)	50	0.9563	0.3010	0.003789				
(0.2, 150, 0.5)	50	0.5502	0.7407	0.0543				
(0.2, 150, 0.4)	50	0.4378	0.9845	0.1266				
(0.2, 150, 0.3)	50	0.4512	0.7128	0.04788				
(0.2, 150, 0.2)	50	0.1762	0.2176	0.02496				
(0.2, 150, 0.1)	50	0.9513	0.9787	0.08496				
(0.2,50,0.9)	45	0.08366	0.04581	0.2165				
(0.2, 50, 0.8)	50	0.2634	0.13	0.01150				
(0.2, 50, 0.7)	50	0.1018	0.3399	0.272				
(0.2, 50, 0.6)	50	0.09797	0.3981	0.5006				
(0.2, 50, 0.5)	50	0.2210	0.04707	0.02113				
(0.2, 50, 0.4)	50	0.06823	0.1179	0.3206				
(0.2, 50, 0.3)	50	0.4577	0.8337	0.009072				
(0.2, 50, 0.2)	50	0.593	0.6985	1.269e-05				
(0.2, 50, 0.1)	50	0.9352	0.3343	0.001161				

Tab. 2: Schätzungen Ergebnistabelle Simulation (1-dim mit Zeit) - Teil 1

Anzahl Du	Anzahl Durchläufe: 50, Anzahl Daten: 2000								
$({\bf a},\!{\bf b},\!\rho)$	n	p-Wert \hat{a}	p-Wert \hat{b}	p-Wert $\hat{\rho}$					
(0.7,93,0.9)	50	0.3096	0.5712	0.03262					
(0.7, 93, 0.8)	50	0.007827	0.007366	0.6491					
(0.7, 93, 0.7)	50	0.7956	0.6984	0.07338					
(0.7, 93, 0.6)	50	0.6144	0.6644	0.2808					
(0.7, 93, 0.5)	50	0.8145	0.239	0.09012					
(0.7, 93, 0.4)	50	0.07813	0.3718	0.006706					
(0.7, 93, 0.3)	50	0.5637	0.9647	0.0001651					
(0.7, 93, 0.2)	50	0.6138	0.7618	0.0009773					
(0.7, 93, 0.1)	50	0.1134	0.7626	5.434e-06					
(2,20,0.9)	33	0.2236	0.04543	0.6652					
(2,20,0.8)	50	0.827	0.0588	0.002253					
(2,20,0.7)	50	0.3332	0.5136	0.09027					
(2,20,0.6)	50	0.06771	0.09015	0.2081					
(2,20,0.5)	50	0.6308	0.5204	0.0137					
(2,20,0.4)	50	0.826	0.6612	0.03951					
(2,20,0.3)	50	0.8456	0.5641	0.005336					
(2,20,0.2)	50	0.5921	0.2969	0.003328					
(2,20,0.1)	50	0.1056	0.08715	0.0003266					

Tab. 3: Schätzungen Ergebnistabelle Simulation (1-dim mit Zeit) - Teil 2

Anzahl Du	Anzahl Durchläufe: 50									
Anzahl Da	Anzahl Daten: 2500 (50 $ imes$ 50-Matrix)									
$(\mathrm{a,b,}\rho)$	n	p-Wert \hat{a}	p-Wert \hat{b}	p-Wert $\hat{\rho}$						
(0.2, 50, 0.9)	0	NA	NA	NA						
(0.2, 50, 0.8)	0	NA	NA	NA						
(0.2, 50, 0.7)	1	NA	NA	NA						
(0.2, 50, 0.6)	50	3.103 e-07	0.08407	2.563e-16						
(0.2, 50, 0.5)	50	0.5985	0.8301	6.719e-05						
(0.2, 50, 0.4)	50	0.04003	0.1082	0.8358						
(0.2, 50, 0.3)	50	0.3221	0.6423	0.8926						
(0.2, 50, 0.2)	50	0.1857	0.8221	0.6081						
(0.2, 50, 0.1)	38	0.2303	0.4628	0.6625						
(0.1,25,0.9)	0	NA	NA	NA						
(0.1, 25, 0.8)	0	NA	NA	NA						
(0.1, 25, 0.7)	45	2.757e-05	0.3703	7.457e-12						
(0.1, 25, 0.6)	50	0.4048	0.09542	0.0001965						
(0.1, 25, 0.5)	50	0.007581	0.003793	0.4049						
(0.1, 25, 0.4)	50	0.006658	0.4215	0.3277						
(0.1, 25, 0.3)	50	0.06278	0.5698	0.5998						
(0.1, 25, 0.2)	50	0.0004341	0.1968	0.00638						
(0.1, 25, 0.1)	36	0.3702	0.7331	0.338						

Tab. 4: Schätzungen Ergebnistabelle Simulation (räumlich)

Anzahl Durchläufe: 50, Anzahl Daten: 2500 (50×50 -Matrix)								
(a,b,ρ_1,ρ_2)	n	p-Wert \hat{a}	p-Wert \hat{b}	p-Wert $\hat{\rho}_1$	p-Wert $\hat{\rho}_2$			
(0.2,50,0.9,0.9)	47	2.915e-09	5.071e-09	<2.2e-16	<2.2e-16			
(0.2, 50, 0.9, 0.8)	45	5.518e-08	2.743e-11	6.73e-16	9.07e-13			
(0.2, 50, 0.9, 0.7)	44	2.913e-05	0.001858	6.416e-11	7.534e-08			
(0.2, 50, 0.9, 0.6)	43	3.652 e-07	0.0004604	4.73e-10	6.107e-08			
(0.2, 50, 0.9, 0.5)	45	0.0003682	0.007885	2.291e-10	5.748e-07			
(0.2, 50, 0.9, 0.4)	41	5.114e-07	0.001458	8.589e-09	2.402 e-06			
(0.2, 50, 0.9, 0.3)	39	0.001275	0.1559	6.898e-05	0.001097			
(0.2, 50, 0.9, 0.2)	37	0.0001705	0.199	0.0004788	0.005377			
(0.2, 50, 0.9, 0.1)	40	0.0003231	0.02391	0.0001309	0.01769			
(0.2, 50, 0.8, 0.9)	47	4.813e-10	7.43e-10	1.024e-11	< 2.2 e-16			
(0.2, 50, 0.8, 0.8)	47	1.186e-07	3.787e-06	2.757e-08	2.325 e-06			
(0.2, 50, 0.8, 0.7)	47	1.703 e-06	4.862 e-06	3.362e-06	9.632 e-06			
(0.2, 50, 0.8, 0.6)	48	1.708e-06	4.102 e-05	0.003816	8.121e-07			
(0.2, 50, 0.8, 0.5)	49	6.784 e - 06	8.625 e-06	0.0207	2.827e-05			
(0.2, 50, 0.8, 0.4)	49	3.393 e-06	2.871 e-05	0.06298	4.023e-06			
(0.2, 50, 0.8, 0.3)	50	0.0008534	0.4707	0.01519	0.449			
(0.2, 50, 0.8, 0.2)	50	0.000646	0.01534	0.1169	0.01485			
(0.2, 50, 0.8, 0.1)	50	0.01873	0.5211	0.8955	0.1201			
(0.2, 50, 0.7, 0.9)	45	5.959 e-05	1.163 e-06	1.055e-07	3.601 e-12			
(0.2, 50, 0.7, 0.8)	50	4.455 e - 06	5.007e-09	3.788e-08	5.465 e - 08			
(0.2, 50, 0.7, 0.7)	50	9.56 e-07	0.0002209	0.0003856	0.0003204			
(0.2, 50, 0.7, 0.6)	50	0.0004254	0.1932	0.0003004	0.006125			
(0.2, 50, 0.7, 0.5)	50	1.218e-05	0.001795	1.362e-06	0.0002458			
(0.2, 50, 0.7, 0.4)	50	3.418e-05	0.0002641	0.004544	1.897 e - 05			
(0.2, 50, 0.7, 0.3)	50	0.0006627	0.07943	0.1856	0.1538			
(0.2, 50, 0.7, 0.2)	50	0.0001055	0.003895	0.07997	0.092			
(0.2, 50, 0.7, 0.1)	50	0.009273	0.8229	0.1240	0.7204			
(0.2, 50, 0.6, 0.9)	41	2.761e-06	0.006182	9.032e-09	1.264 e-11			
(0.2, 50, 0.6, 0.8)	50	5.116e-07	1.151e-05	2.14e-08	0.0001073			
(0.2, 50, 0.6, 0.7)	50	1.997 e-05	0.001078	0.001052	0.01921			
(0.2, 50, 0.6, 0.6)	50	8.183e-06	0.0003153	0.01508	2.273e-06			
(0.2, 50, 0.6, 0.5)	50	0.001089	0.2827	0.1728	0.06147			
(0.2, 50, 0.6, 0.4)	50	0.06405	0.8775	0.4338	0.953			
(0.2, 50, 0.6, 0.3)	50	2.730e-05	0.0003993	0.04234	0.004337			
(0.2, 50, 0.6, 0.2)	50	0.09677	0.8822	0.7826	0.7927			
(0.2, 50, 0.6, 0.1)	50	0.01189	0.4136	0.2221	0.06291			

Tab. 5: Schätzungen Ergebnistabelle Simulation Matrix (2 Richtungen) - Teil 1

Anzahl Durch	Anzahl Durchläufe: 50, Anzahl Daten: 2500 (50 $ imes$ 50-Matrix)								
(a,b,ρ_1,ρ_2)	n	p-Wert \hat{a}	p-Wert \hat{b}	p-Wert $\hat{\rho}_1$	p-Wert $\hat{\rho}_2$				
(0.2,50,0.5,0.9)	35	2.986e-06	1.837e-06	1.502e-07	9.157e-07				
(0.2, 50, 0.5, 0.8)	48	3.658 e - 05	0.001854	3.947e-06	0.003256				
(0.2, 50, 0.5, 0.7)	50	0.0009488	0.06452	0.006169	0.1671				
(0.2, 50, 0.5, 0.6)	50	2.701 e-05	0.01976	0.0002238	0.006933				
(0.2, 50, 0.5, 0.5)	50	4.881 e-05	0.1367	0.00093	0.05382				
(0.2, 50, 0.5, 0.4)	50	0.0001917	0.04419	0.004803	0.01652				
(0.2, 50, 0.5, 0.3)	50	0.0007249	0.0402	0.1690	0.1022				
(0.2, 50, 0.5, 0.2)	50	0.02373	0.02478	0.5639	0.6964				
(0.2, 50, 0.5, 0.1)	50	0.005561	0.2826	0.9062	0.2641				
(0.2,50,0.4,0.9)	35	9.616e-05	0.2045	0.004676	8.037e-06				
(0.2, 50, 0.4, 0.8)	50	0.003752	0.3634	0.1414	0.1321				
(0.2, 50, 0.4, 0.7)	50	0.001313	0.5216	0.2047	0.03061				
(0.2, 50, 0.4, 0.6)	50	0.004131	0.9192	0.002753	0.2370				
(0.2, 50, 0.4, 0.5)	50	0.000339	0.001451	0.00903	0.1019				
(0.2, 50, 0.4, 0.4)	50	0.1246	0.7478	0.8711	0.3975				
(0.2, 50, 0.4, 0.3)	50	0.02002	0.5519	0.04582	0.2955				
(0.2, 50, 0.4, 0.2)	50	0.0029	0.1953	0.02049	0.05282				
(0.2, 50, 0.4, 0.1)	50	0.003102	0.07435	0.1795	0.5771				
(0.2,50,0.3,0.9)	43	0.0001785	0.2097	0.05884	0.0001484				
(0.2, 50, 0.3, 0.8)	50	0.005275	0.2196	0.09394	0.6961				
(0.2, 50, 0.3, 0.7)	50	0.0005838	0.0898	0.01191	0.3958				
(0.2, 50, 0.3, 0.6)	50	0.002151	0.1158	0.03856	0.1450				
(0.2, 50, 0.3, 0.5)	50	0.0005402	0.03836	0.04671	0.0718				
(0.2, 50, 0.3, 0.4)	50	0.2272	0.6801	0.3938	0.6362				
(0.2, 50, 0.3, 0.3)	50	0.004862	0.1400	0.01175	0.3227				
(0.2, 50, 0.3, 0.2)	50	0.008638	0.1431	0.7661	0.08446				
(0.2, 50, 0.3, 0.1)	50	0.1354	0.7872	0.2359	0.1811				

Tab. 6: Schätzungen Ergebnistabelle Simulation Matrix (2 Richtungen) - Teil 2

Anzahl Durchläufe: 50, Anzahl Daten: 2500 (50 $ imes$ 50-Matrix)								
(a,b,ρ_1,ρ_2)	n	p-Wert \hat{a}	p-Wert \hat{b}	p-Wert $\hat{\rho}_1$	p-Wert $\hat{\rho}_2$			
(0.2, 50, 0.2, 0.9)	40	6.223e-05	0.008616	0.0007922	5.395e-06			
(0.2, 50, 0.2, 0.8)	50	0.00206	0.009798	0.0008689	0.6858			
(0.2, 50, 0.2, 0.7)	50	0.003533	0.6065	0.0106	0.1463			
(0.2, 50, 0.2, 0.6)	50	0.06446	0.7322	0.3307	0.2589			
(0.2, 50, 0.2, 0.5)	50	0.000246	0.1118	0.04338	0.001156			
(0.2, 50, 0.2, 0.4)	50	0.001514	0.07572	0.04857	0.2493			
(0.2, 50, 0.2, 0.3)	50	0.002950	0.05637	0.3854	0.637			
(0.2, 50, 0.2, 0.2)	50	0.1010	0.3204	0.2245	0.5788			
(0.2, 50, 0.2, 0.1)	50	0.03147	0.6428	0.2802	0.0729			
(0.2, 50, 0.1, 0.9)	40	8.187e-05	0.03157	0.003056	0.0003957			
(0.2, 50, 0.1, 0.8)	49	0.006581	0.01357	0.2887	0.8706			
(0.2, 50, 0.1, 0.7)	50	0.009024	0.7373	0.3216	0.06347			
(0.2, 50, 0.1, 0.6)	50	0.005663	0.5683	0.06399	2.136e-12			
(0.2, 50, 0.1, 0.5)	50	0.2590	0.7061	0.832	0.5212			
(0.2, 50, 0.1, 0.4)	50	0.01137	0.333	0.1327	0.2249			
(0.2, 50, 0.1, 0.3)	50	0.01340	0.7172	0.6677	$\boldsymbol{0.6682}$			
(0.2, 50, 0.1, 0.2)	50	0.003017	0.06762	0.5866	0.2109			
(0.2, 50, 0.1, 0.1)	50	0.00372	0.4131	0.1259	0.1478			

Tab. 7: Schätzungen Ergebnistabelle Simulation Matrix (2 Richtungen) - Teil 3

C.2. Anwendung

bs	Bushbuck (Tragelaphus scriptus)
d	Duiker (Sylvicapra grimmia)
e	Eland (Taurotragus oryx)
k	Zebra (Equus burchelli)
ko	Kudu (Tragelaphus strepsiceros)
kp	Giraffe (Giraffa camelopardalis)
ks	Klipspringer (Oreotragus oreotragus)
n	Nyala (Tragelaphus angasii)
r	Impala (Aepyceros melampus)
rb	Reedbuck (Redunca fulvorufula)
ro	Roan (Hippotragus equinus)
s	Sable (Hippotragus niger)
sr	Black rhino (Diceros bicornis)
st	Steenbuck
t	Tsessebe (Damaliscus lunatus)
v	Warthog (Phacochoerus africanus)
w	Blue wildebeest (Connochaetes taurinus)
wb	Waterbuck (Kobus ellipsiprymnus)
wr	White rhino (Ceratotherium simum)

Tab. 8: Aufschlüsselung der Tiere in den Datensätzen

Jahr	# Giraffen	# Gruppen	Breitengrad	Längengrad
1982	5162	1774	-25.4422.64	31.01-32.03
1983	4726	1473	-25.4422.64	30.99 - 32.03
1984	5341	1662	-25.4422.64	31.03 - 32.03
1985	5332	1699	-25.5122.47	31.01 - 32.02
1986	5802	1728	-25.4822.64	30.99 - 32.03
1987	4990	1721	-25.4422.51	30.97 - 32.03
1988	5067	1666	-25.5022.47	30.97 - 32.03
1989	4877	1620	-25.4922.48	30.96 - 32.02
1990	4812	1752	-25.5022.62	30.92 - 32.03
1991	4616	1718	-25.5022.49	30.91 - 32.03
1992	4610	1794	-25.5222.51	31.00 - 32.02
1993	4923	1833	-25.5222.41	31.03 - 32.02
1994	2763	1279	-24.7822.46	31.01 - 32.01
1995	2729	1028	-24.7022.45	30.98 - 32.01
1996	2297	747	-24.6222.63	31.16 - 32.00
1997	771	299	-25.4024.97	31.21-32.02

Tab. 9: Kennwerte über die Anzahl der Giraffen und Lokalisation im Krüger-Nationalpark

Jahr	# Giraffen	\overline{x}
1982	2534	48.73077
1983	2395	46.05769
1984	2829	54.40385
1985	2533	48.71154
1986	2790	53.65385
1987	2228	42.84615
1988	2340	45.00000
1989	2400	46.15385
1990	2302	44.26923
1991	1911	36.75000
1992	2013	38.71154
1993	2207	42.44231

Tab. 10: Kennwerte über die Anzahl der Giraffen im Teilausschnitt

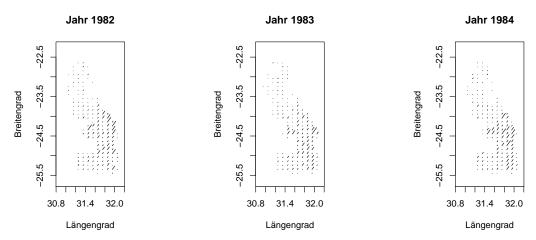


Abb. 25: Giraffenanzahlen in den Zellen in den Jahren 1982-84

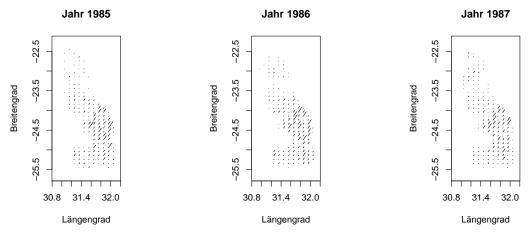


Abb. 26: Giraffenanzahlen in den Zellen in den Jahren 1985-87

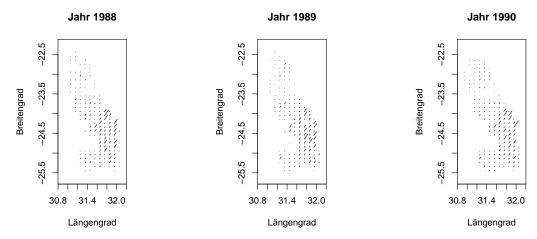


Abb. 27: Giraffenanzahlen in den Zellen in den Jahren 1988-90

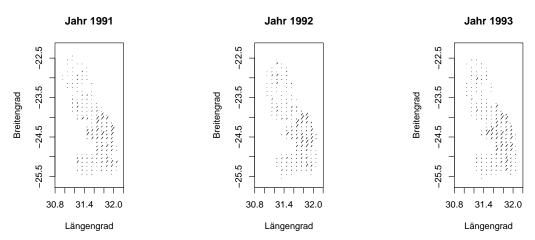


Abb. 28: Giraffenanzahlen in den Zellen in den Jahren 1991-93

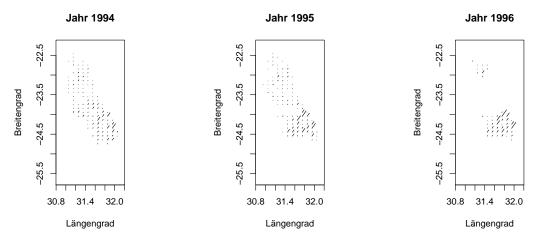


Abb. 29: Giraffenanzahlen in den Zellen in den Jahren 1994-96

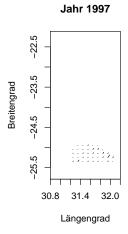


Abb. 30: Giraffenanzahlen in den Zellen im Jahr 1997

			1		2		3	}	4	Į.
	\overline{x}	\hat{a}	75.33	3.358	35.75	4.95	63.5	4.672	11.167	1.269
1	s^2	\hat{b}	911.33	15.208	903.841	12.36	1763.727	18.618	115.061	9.048
	s_{xy}	$\hat{ ho}$	309.664	0.192	-426.227	-0.262	-517.964	-0.156	-2.855	-0.014
	\overline{x}	\hat{a}	78.417	2.299	17.083	2.569	33.083	4.575	59.41167	2.476
2	s^2	\hat{b}	1259.902	21.542	100.084	5.646	236.265	6.652	1423.537	23.469
	s_{xy}	$\hat{ ho}$	507.8	0.226	13.455	0.082	16.9	0.042	30.327	0.011
	\overline{x}	\hat{a}	50.67	4.169	37	1.154	50.083	2.565	103	4.624
3	s^2	\hat{b}	653.515	12.025	625.636	21.773	643.902	14.949	1617.818	17.884
	s_{xy}	$\hat{ ho}$	6.436	0.005	216.982	0.191	154.864	0.133	327.264	0.109
	\overline{x}	\hat{a}	45.583	7.979	79.25	14.044	39.75	4.625	89.33	0.907
4	s^2	\hat{b}	248.811	5.026	1477.295	9.378	287.114	7.262	2925.152	50.342
	s_{xy}	$\hat{ ho}$	25.882	0.064	-654.8	-0.249	41.3	0.084	1661.391	0.324
	\overline{x}	\hat{a}	44	7.679	50.25	4.959	49.583	11.682	67.083	12.021
5	s^2	\hat{b}	347.455	6.274	581.659	10.35	469.72	5.847	707.72	7.186
	s_{xy}	$\hat{ ho}$	-27.418	-0.045	-11.3	-0.011	-130.2	-0.159	-158.59	-0.126
	\overline{x}	\hat{a}	38.67	4.772	59.167	1.531	45.583	3.739	49.583	8.382
6	s^2	\hat{b}	298.788	7.366	1689.424	32.374	885.902	14.848	672.083	8.368
	s_{xy}	$\hat{ ho}$	24.709	0.048	285.227	0.088	-163.482	-0.098	-207.609	-0.172
	\overline{x}	\hat{a}	31.417	3.784	67.67	7.487	90.417	6.615	54.083	13.79
7	s^2	\hat{b}	207.174	6.743	409.879	6.598	1325.356	13.663	561.174	5.835
	s_{xy}	$\hat{ ho}$	36.073	0.104	104.255	0.156	0.445	0	-191.5	-0.196
	\overline{x}	\hat{a}	37.167	11.293	32	1.948	53.167	4.1	49.917	4.872
8	s^2	\hat{b}	345.788	5.008	421.091	14.048	533.97	10.733	682.992	11.368
	s_{xy}	$\hat{ ho}$	-122.473	-0.207	60.445	0.078	89.827	0.094	-65.627	-0.052
	\overline{x}	\hat{a}	48.83	2.546	40.83	4.103	34.75	3.629	37.33	0.754
9	s^2	\hat{b}	552.88	13.681	375.788	9.014	252.023	7.64	2315.151	54.84
	s_{xy}	$\hat{ ho}$	164.055	0.167	33.1	0.05	48.227	0.112	-230.445	-0.051
	\overline{x}	\hat{a}	31.25	4.935	24.417	6.198	54	24.704	29.083	0.811
10	s^2	\hat{b}	86.386	2.913	100.629	3.419	270.545	2.903	456.083	21.564
	s_{xy}	$\hat{ ho}$	35.882	0.37	16.1	0.11	-59.809	-0.141	200.145	0.249
	\overline{x}	\hat{a}	21.33	2.82	27.583	13.978	60	11.709	41.417	3.495
11	s^2	\hat{b}	138.424	6.398	121.72	2.553	496.545	6.063	547.902	12.038
	s_{xy}	$\hat{ ho}$	19.409	0.083	-23.718	-0.128	-72.764	-0.084	-7.9	-0.008
	\overline{x}	\hat{a}	16.25	1.821	22.417	1.588	58	44.126	37.83	5.529
12	s^2	\hat{b}	169.295	9.165	321.72	13.726	322	2.276	386.879	7.905
	s_{xy}	$\hat{ ho}$	-4.118	-0.013	8.391	0.014	-131.973	-0.268	-49.991	-0.072
	\overline{x}	\hat{a}	21.17	1.598	7.583	2.765	33.5	0.635	35.75	0.768
13	s^2	\hat{b}	151.061	8.621	29.538	2.817	841.727	34.062	814.205	30.487
	s_{xy}	$\hat{ ho}$	52.591	0.211	-0.591	-0.013	332.845	0.215	311.455	0.209

Tab. 11: Kennzahlen & Schätzungen für die Zellen des Ausschnitts über die Zeit

		Jahr											
Zelle	$\hat{ ho}$	82	83	84	85	86	87	88	89	90	91	92	93
(12,3)	-0.268	67	29	81	67	50	80	61	65	41	35	77	43
(11,4)	-0.08	22	67	62	18	42	12	47	38	87	58	20	24
(7,3)	0	70	60	155	125	137	55	62	133	59	69	89	71
(3,1)	0.005	44	15	75	56	83	92	35	39	34	79	26	30
(10,1)	0.37	38	28	36	36	50	29	33	37	31	21	18	18

Tab. 12: Giraffenanzahlen in den Zellen

Jahr	\overline{x}	s^2	$S(X_{ij}, X_{(i+1)(j+1)})$	\hat{a}	\hat{b}	$\hat{ ho}$
1982	48.73077	1221.573	749.6278	0.07	77.934	0.499
1983	46.05769	935.8986	450.6341	0.137	56.602	0.418
1984	54.40385	1156.991	747.7913	0.078	67.393	0.525
1985	48.71154	1283.621	530.4643	0.146	69.472	0.373
1986	53.65385	1588.701	434.3976	0.244	67.146	0.288
1987	42.84615	953.466	210.0143	0.323	46.568	0.256
1988	45	810.549	277.9675	0.255	43.691	0.335
1989	46.15385	1074.682	566.0437	0.101	67.613	0.444
1990	44.26923	1084.397	388.7619	0.175	61.006	0.341
1991	36.75	692.5833	318.754	0.128	51.402	0.406
1992	38.71154	653.4642	195.7333	0.278	38.858	0.309
1993	42.44231	579.7417	108.9667	0.604	26.629	0.238

Tab. 13: Kennzahlen & Schätzungen für die Ausschnitte für jedes Jahr räumlich

Jahr	\overline{x}	s^2	$s_{(X_{ij},X_{(i+1)j})}$	$s_{(X_{ij},X_{i(j+1)})}$	\hat{a}	\hat{b}	$\hat{ ho}_1$	$\hat{ ho}_2$
1982	48.73077	1221.573	243.953	3.052632	1.350	29.150	0.1050	0.0010
1983	46.05769	935.8986	134.8644	-30.26113	1.903	21.492	0.0760	-0.017
1984	54.40385	1156.991	279.7114	207.1201	1.150	30.181	0.1290	0.0950
1985	48.71154	1283.621	342.5106	56.53239	1.033	33.865	0.1410	0.0230
1986	53.65385	1588.701	437.5461	-237.747	1.506	31.071	0.1460	-0.078
1987	42.84615	953.466	259.0957	-3.020243	1.179	27.210	0.1450	-0.002
1988	45	810.549	107.6662	72.74291	1.671	21.249	0.0710	0.0480
1989	46.15385	1074.682	147.1924	137.3394	1.210	28.875	0.0720	0.0670
1990	44.26923	1084.397	19.48404	-52.44332	2.011	22.729	0.0090	-0.025
1991	36.75	692.5833	101.5824	30.14777	1.393	21.557	0.0780	0.0230
1992	38.71154	653.4642	260.3387	273.0769	0.468	32.647	0.2220	0.2340
1993	42.44231	579.7417	67.52305	1.920378	2.606	14.301	0.0630	0.0020

Tab. 14: Kennzahlen & Schätzungen für die Ausschnitte für jedes Jahr für zwei Richtungen

Literaturverzeichnis

- Braack, Dr L. E. O. (2011). Kruger National Park. New Holland, London
- Burrell, Q. L. (1990). Using the Gamma-Poisson Model to Predict Library Circulations. Journal of the American society for information science 41(3), 164-170
- Fahrmeir, L., Künstler, R., Pigeot, I. und Tutz, G. (2007). Statistik. Der Weg zur Datenanalyse. Springer, Berlin
- Genschel, U. und Becker, C. (2005). Schließende Statistik. Grundlegende Methoden. Springer, Berlin
- Hwang, W.-H. und He, F. (2011). Estimating abundance from presence/absence maps. *Methods in Ecology and Evolution* **2**(5), 550–559
- MATLAB 7.9.0 (R2009b). http://www.mathworks.de/products/matlab/
- Müller, Ch. H., Huggins, R. und Hwang, W.-H. (2011). Consistent estimation of species abundance from a presence-absence map. Statistics & Probability Letters 81(9), 1449-1457
- R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://cran.r-project.org/
- R Development Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://cran.r-project.org/
- Sachs, L. und Hedderich, J. (2009). Angewandte Statistik: Methodensammlung in R. Springer, Berlin
- Stuart, Ch. und Stuart, T. (2001). Field guide to mammals of Southern Africa. Struik, Kapstadt
- Toft, N., Innocent, G. T., Mellor, D. J. und Reid, S. W. J. (2006). The Gamma-Poisson model as a statistical method to determine if micro-organisms are randomly distributed in a food matrix. *Food Microbiology* **23**(1), 90-94

- Tosh, Ch. A., Reyers, B. und van Jaarsveld, A. S. (2004). Estimating the abundances of large herbivores in the Kruger National Park using presencabsence data. *Animal Conservation* **7**(1), 55-61
- Karte des Krüger-Nationalparks: http://www.gotravel24.com/files/media/ Places/Limpopo/KrugerPark_map_goafrikasafari.jpg (Datum: 6.2.2012)
- Quelle für die Datensätze: http://dataknp.sanparks.org/sanparks/metacat? action=read&qformat=sanparks&sessionid=0&docid=judithk.739.17 (Datum: 3.7.2011)

Eidesstattliche Erklärung des Urhebers

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbststä anderen als die angegebenen Quellen und Hilfsmittel benur sinngemäße Zitate kenntlich gemacht habe.	
Dortmund, den	
	Tanja K. Rausch
Einverständniserklärung des Urhebers	
Ich erkläre mich hiermit einverstanden, dass eine Diplom	arbeit nach §6 (1) des
URG der Öffentlichkeit durch die Übernahme in die Bereic	chsbibliotheken zugäng-
lich gemacht wird. Damit können Leser der Bibliothek die	Arbeit einsehen und zu
persönlichen wissenschaftlichen Zwecken Kopien aus dieser	Arbeit anfertigen.
Weitere Urheberrechte werden nicht berührt.	
Dortmund, den	

Tanja K. Rausch