Einführung zu Markov Chain Monte Carlo Verfahren mit Anwendung auf Gesamtschadenmodelle *

Claudia Czado
Technische Universität München, Zentrum Mathematik
email: cczado@ma.tum.de

1 Einführung

Um komplexe Modelle der Versicherungs- und Finanzmathematik schätzen zu können, gewinnen neuere computerintensive statistische Verfahren immer größere Bedeutung. In diesem Artikel wird eine solche Klasse von Verfahren, nämlich die Markov Chain Monte Carlo (MCMC) Verfahren, eingeführt und diskutiert. Damit wollen wir diese Methoden einem größeren Kreis von Anwendern zugänglich zu machen.

Bevor wir MCMC Verfahren allgemein einführen, betrachten wir zunächst den klassischen Ansatz der Statistik. Es seien die Daten $\boldsymbol{x} = (x_1, \cdots, x_n)$ als Realisierungen der Zufallsvariablen $X_1, \cdots, X_n$ mit gemeinsamer Dichte $f(\boldsymbol{x}|\boldsymbol{\theta})$ gegeben. Beim klassischen Ansatz wird angenommen, dass der Parameter $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_d)$ fest und unbekannt ist. Dieser wird dann häufig durch die Maximum Likelihood (ML) Methode geschätzt. Im Gegensatz dazu wird beim Bayesianischen Ansatz der Parameter $\boldsymbol{\theta}$ als zufällig mit a-priori Dichte $\pi(\boldsymbol{\theta})$ angenommen. Nach der Beobachtung der Daten $\boldsymbol{x}$ wird das Wissen über die Verteilung von $\boldsymbol{\theta}$ angepasst. Insbesondere ist man an der a-posteriori Verteilung von $\boldsymbol{\theta}$ interessiert, deren Dichte durch

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}$$

gegeben ist. Beim Bayesianischen Ansatz (siehe zum Beispiel die Monografien von Lee (1997) und Gelman et al. (1995)) wird $\boldsymbol{\theta}$ durch den Erwartungswert oder Modus der a-posteriori Verteilung geschätzt. Im Gegensatz zum klassischen Ansatz hat man nicht nur Punktschätzer für die Inferenz zur Verfügung, sondern die gesamte a-posteriori Verteilung. Daneben kann man direkte Wahrscheinlichkeitsaussagen über $\boldsymbol{\theta}$ treffen, was im klassischen Ansatz nicht möglich ist. Die a-posteriori Verteilung erlaubt damit die Bestimmung von Intervallen oder Regionen, in denen $\boldsymbol{\theta}$ mit a-posteriori Wahrscheinlichkeit $1 - \alpha$ liegt. Ein weiterer Vorteil eines Bayesianischen Ansatzes ist die Möglichkeit sowohl Parameterunsicherheit als auch Modellunsicherheit zu berücksichtigen. Die Berücksichtigung der Parameterunsicherheit wird später in dem Anwendungsbeispiel illustriert, während die Modellunsicherheit mit Hilfe von

*basiert auf einem eingeladenen Vortrag gehalten anlässlich der *Scientific Conference on Insurance and Finance* der Deutschen Gesellschaft für Versicherungs- und Finanzmathematik am 28. April 2003 in Bonn

Bayes factors untersucht werden kann (siehe z.B. Hoeting, Madigan, Raftery, und Volinsky (1999)).

Falls man eine nicht informative a-priori Verteilung für $\boldsymbol{\theta}$ wählen kann, dann stimmen der ML Schätzer und der a-posteriori Modusschätzer überein. Unter Regularitätsbedingungen ist die asymptotische a-posteriori Verteilung normalverteilt mit Erwartungwert $\hat{\boldsymbol{\theta}}$ und Kovarianzmatrix gegeben durch die Inverse der Fisher Informationsmatrix (Berger 1985, S. 224). Hier bezeichnet $\hat{\boldsymbol{\theta}}$ den ML Schätzer von $\boldsymbol{\theta}$. Dies bedeutet, dass der Einfluß der a-priori Verteilung auf die Bayesianische Schätzung für $n \to \infty$ verschwindet.

In komplexen statistischen Modellen insbesondere mit zeitlichen und/oder räumlichen Abhängigkeitsstrukturen kann sowohl die Berechnung des ML Schätzers als auch der a-posteriori Verteilung nicht nur analytisch sondern auch numerisch nicht mehr berechenbar sein. Insbesondere wird dies der Fall sein, wenn man hoch dimensionale Integrale berechnen muss. Diese treten bei a-posteriori Verteilungen im Nenner und bei Modellen mit latenten Variablen bei der Berechnung des Likelihoods auf. Mit Hilfe von MCMC Verfahren kann man dort jedoch noch gute Approximationen der a-posteriori Verteilung erhalten.

Im Abschnitt 2 werden wichtige Grundlagen der MCMC Verfahren zusammengefasst und in dem Abschnitt 3 die speziellen MCMC Verfahren Gibbs Sampler und Metropolis Hastings Algorithmus dargestellt. Der Abschnitt 4 beschreibt statistische Softwarepakete zur Implementierung von MCMC Verfahren. Im letzten Abschnitt werden zur Illustration zwei klassische Gesamtschadenmodelle mit Hilfe von MCMC Verfahren geschätzt. Dabei wird ein bekannter Datensatz aus der Extremwerttheorie, nämlich Daten einer dänischen Feuerversicherung verwendet.

2 Theoretische Grundlagen der Markov Chain Monte Carlo Verfahren

MCMC Verfahren sind Simulationsverfahren bei denen Markovketten mit vorgegebener Grenzverteilung π konstruiert werden. Solche Verfahren werden notwendig, wenn die Grenzverteilung π so komplex ist, dass sie direkt nicht berechenbar ist. Mit Hilfe der Markovkette werden nun approximativ Beobachtungen von der Grenzverteilung π erzeugt. Im Zusammenhang mit Parameterschätzungen in komplexen statistischen Modellen wird die gewünschte a-posteriori Verteilung als Grenzverteilung gewählt.

Im folgenden werden einige notwendige Begriffe aus der Theorie der Markovketten in Erinnerung gerufen. Eine ausführlichere Darstellung kann man z.B. in den Monografien von Guttorp (1995) und Resnick (1992) finden. Eine diskrete Markovkette ist ein stochastischer Prozess $\{\theta^{(n)}, n \geq 0\}$ mit Werten im Zustandsraum S, der die Markoveigenschaft

$$P(\theta^{(n+1)} \in A | \theta^{(n)} = x, \theta^{(n-1)} \in A_{n-1}, .., \theta^{(0)} \in A_0) = P(\theta^{(n+1)} \in A | \theta^{(n)} = x)$$

für alle $n, A_0, .., A_{n-1}, A \subset S$ und $x \in S$ erfüllt. Hängt $P(\theta^{(n+1)} \in A | \theta^{(n)} = x)$ nicht von n ab, so spricht man von einer homogenen Markovkette.

Falls $S = \{x_1, x_2, ...\}$ diskret und die Markovkette homogen ist, dann reicht es aus, die Übergangswahrscheinlichkeiten

$$P(x, y) := P(\theta^{(n+1)} = y | \theta^{(n)} = x)$$

und deren Übergangsmatrix 1. Ordnung $P := (P(x_i, x_j))_{i,j \in S}$ zu studieren. Man kann mit Hilfe der Chapman-Kolmogorov Gleichungen die Übergangsmatrix n-ter Ordnung $P^n := (P^n(x_i, x_j)_{i,j \in S})$, mit $P^n(x, y) := P(\theta^{(n)} = y | \theta^{(1)} = x)$, als Matrixprodukt von

$$\underbrace{P \cdot \ldots \cdot P}_{n-mal}$$

berechnen. Insbesondere ist man an dem Grenzverhalten von P^n für $n \to \infty$ interessiert. Mit Hilfe von stationären Verteilungen ist es möglich, die Grenzverteilung von P^n zu charakterisieren. Eine Verteilung π heißt stationär falls

$$\pi = \pi P \text{ für } \quad \pi = (\pi(x_1), \pi(x_2), \ldots)$$

gilt. Dies bedeutet, dass die Verteilung von $\theta^{(n)}$ gleich π ist für alle n, falls die Markovkette mit Anfangsverteilung π startet. Im Falle einer diskreten Markovkette mit diskretem Zustandsraum S existiert unter Regularitätsbedingungen wie Irreduzibilität, positive Rekurrenz und Aperiodizität eine eindeutige stationäre Verteilung π, und diese ist auch die Grenzverteilung, d.h. es gilt

$$\lim_{n \to \infty} P^n(x, y) = \pi(y) \quad \forall x, y \in S.$$

Wie im Fall von unabhängigen Zufallsvariablen gilt für jede Funktion $t(\theta)$ mit $E_\pi(t(\theta)) < \infty$

$$\bar{t}_n := \frac{1}{n} \sum_{i=1}^{n} t(\theta^{(i)}) \to E_\pi(t(\theta)) \text{ mit W-keit 1.}$$

Unter weiteren Regularitätsbedingungen gilt auch ein zentraler Grenzwertsatz für Markovketten, d.h. dass $\bar{t}_n$ asymptotisch normalverteilt ist. Für stetige Zustandsräume kann man ähnliche Resultate herleiten (siehe z.B. Meyn und Tweedie (1993)). Man betrachtet hier den Übergangskern, der durch

$$p(x, y) := P(\theta^{(n+1)} \leq y | \theta^{(n)} = x)$$

gegeben ist anstatt Übergangswahrscheinlichkeiten. Es sei anzumerken, dass es in komplexen statistischen Modellen schwierig ist, die Regularitätsbedingungen für Stationärität, die Existenz der Grenzverteilung und dem zentralen Grenwertsatz für Markovketten nachzuweisen. Für diskrete Zustandsräume kann man auch direkt Übergangskerne verwenden.

Um MCMC Verfahren praktisch durchzuführen, muss man in der Lage sein Realisierungen einer Markovkette zu erzeugen. Im Falle einer Markovkette $\{\boldsymbol{\theta}^{(n)}, n \geq 0\}$ mit stetigen Zustandsraum $S \subset \mathbb{R}^d$, Übergangskern $p(x, y)$ und Startverteilung $\pi^{(0)}$ geschieht dies wie folgt:

- Erzeuge $\boldsymbol{\theta}^{(0)} \sim \pi^{(0)}$.
- Erzeuge $\boldsymbol{\theta}^{(1)} \sim p(\theta^{(0)}, \cdot)$.

  $\vdots$
- Erzeuge $\boldsymbol{\theta}^{(n)} \sim p(\theta^{(n-1)}, \cdot)$.

Falls die Regularitätsbedingungen erfüllt sind, so dass eine Grenzverteilung π existiert, werden die so erzeugten Realisierungen für große n eine Verteilung nahe der Grenzverteilung π besitzen und können daher als eine approximative Stichprobe von π betrachtet werden. In der Praxis bedeutet dies, dass man für $m > 1$ geeignet gewählt die Werte $(\boldsymbol{\theta}^{(m)}, \ldots, \boldsymbol{\theta}^{(n)})$ als Stichprobe der Grenzverteilung betrachtet. Der Wert m wird als burnin bezeichnet.

3 Spezielle Markov Chain Monte Carlo Verfahren

Die beiden häufigsten MCMC Verfahren sind der Gibbs Sampler und der Metropolis Hastings Algorithmus, die wir nun vorstellen. Roberts und Smith (1994) geben einfache Bedingungen an, unter denen diese Verfahren konvergieren.

3.1 Gibbs Sampler

Der Gibbs Sampler war das erste Simulationsverfahren für die Erzeugung von Realisierungen von einer Markovkette bei gegebener Grenzverteilung. Es wurde zunächst zur Simulation der Gibbsverteilung aus der Bildanalyse verwendet, daher sein Name. Geman und Geman (1984) beschäftigten sich mit diesem Problem, und Gelfand und Smith (1990) waren die ersten die bemerkten, dass der Gibbs Sampler auch für Realisierungen anderer Verteilungen verwendet werden kann.

Der Gibbs Sampler ist ein MCMC Verfahren, wobei der Übergangskern mit Hilfe der vollständigen bedingten Verteilungen, im Englischen 'full conditionals', gebildet wird. Für $\boldsymbol{\theta} = (\theta_1, .., \theta_d)^t$ definiere folgende Vektoren aus $\mathbb{R}^{d-1}$, bei denen jeweils eine Komponente weggelassen wurde:

$$\boldsymbol{\theta}_{-i} := (\theta_1, .., \theta_{i-1}, \theta_{i+1}, .., \theta_d)^t \text{ für } i = 1, \cdots, d.$$

Bezeichne die interessierende Verteilung von $\boldsymbol{\theta}$ mit π. Die vollständigen bedingten Verteilungen sind nun durch die bedingten Dichten von θ_i gegeben $\boldsymbol{\theta}_{-i}$, die mit

$$\pi(\theta_i | \boldsymbol{\theta}_{-i}) \qquad i = 1, .., d$$

bezeichnet werden, definiert. Damit kann der Gibbs Sampler wie folgt beschrieben werden:

Gibbs Sampler Algorithmus:

1. Setze $j = 1$ und wähle Startwerte $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, .., \theta_d^{(0)})^t$

2. Erzeuge $\boldsymbol{\theta}^{(j)} = (\theta_1^{(j)}, .., \theta_d^{(j)})^t$ durch

$$\begin{aligned}
\theta_1^{(j)} &\sim \pi\left(\theta_1 | \theta_2^{(j-1)}, .., \theta_d^{(j-1)}\right) \\
\theta_2^{(j)} &\sim \pi\left(\theta_2 | \theta_1^{(j)}, \theta_3^{(j-1)}, .., \theta_d^{(j-1)}\right) \\
&\vdots \\
\theta_d^{(j)} &\sim \pi\left(\theta_d | \theta_1^{(j)}, .., \theta_{d-1}^{(j)}\right)
\end{aligned}$$

3. Setze $j = j + 1$ und wiederhole Schritt 2 bis die erzeugte Markovkette sich stationär verhält.

Zum Beispiel in Gamerman (1997) wird gezeigt, dass dieses Verfahren unter Regularitätsbedingungen eine Markovkette mit stationärer Grenzverteilung π erzeugt. Um den Gibbs Sampler anwenden zu können, muss man in der Lage sein, von allen $\pi(\theta_i|\boldsymbol{\theta}_{-i})$ zufällige Realisierungen ziehen zu können. Oft sind einige dieser bedingten Dichten nur bis auf eine Proportionalitätskonstante bekannt. In diesem Fall benötigt man den Metropolis Hastings Algorithmus.

3.2 Metropolis Hastings Algorithmus

Chib und Greenberg (1995) geben eine verständliche Einführung in den Metropolis Hastings Algorithmus, der von Metropolis et al. (1953) und Hastings (1970) entwickelt wurde. Dieser Algorithmus erzeugt eine reversible Markovkette, d.h. es gilt

$$\pi(\boldsymbol{\theta})p(\boldsymbol{\theta}, \boldsymbol{\phi}) = \pi(\boldsymbol{\phi})p(\boldsymbol{\phi}, \boldsymbol{\theta}) \quad \forall \boldsymbol{\theta}, \boldsymbol{\phi} \ \in S.$$

Es ist leicht zu sehen, dass dann $\pi(\boldsymbol{\theta})$ stationär ist. Um Reversibilität zu erreichen, benötigt man eine

- Vorschlagsdichte $q(\boldsymbol{\theta}, \cdot)$ und die
- Akzeptanzwahrscheinlichkeit

$$\alpha(\boldsymbol{\theta}, \boldsymbol{\phi}) := \min\Big\{1, \frac{\pi(\boldsymbol{\phi})q(\boldsymbol{\phi}, \boldsymbol{\theta})}{\pi(\boldsymbol{\theta})q(\boldsymbol{\theta}, \boldsymbol{\phi})}\Big\}.$$

Mit diesen Bezeichnungen ist dann der Metropolis Hastings Algorithmus gegeben durch

Metropolis Hastings (MH) Algorithmus

1. Setze $j = 1$ und wähle Startwerte $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, .., \theta_d^{(0)})^t$
2. Erzeuge eine Realisierung $\boldsymbol{\phi}$ der Vorschlagsdichte $q(\boldsymbol{\theta}^{(j-1)}, \cdot)$.
3. Berechne die Akzeptanzwahrscheinlichkeit $\alpha(\boldsymbol{\theta}^{(j-1)}, \boldsymbol{\phi})$.
   Mit dieser Wahrscheinlichkeit wird $\boldsymbol{\phi}$ akzeptiert, in diesem Fall setze

$$\boldsymbol{\theta}^{(j)} = \boldsymbol{\phi},$$

   ansonsten

$$\boldsymbol{\theta}^{(j)} = \boldsymbol{\theta}^{(j-1)}.$$

Es reicht aus $\dfrac{\pi(\boldsymbol{\phi})}{\pi(\boldsymbol{\theta})}$ zu kennen, um die Akzeptanzwahrscheinlichkeit berechnen zu können. Dies bedeutet, dass man die Grenzverteilung π nur bis auf eine Konstante kennen muss. Das ist insbesondere nützlich für die Simulation von a-posteriori Verteilungen, die häufig wegen der Nichtberechenbarkeit des Integrals im Nenner nur bis auf Proportionalität bekannt sind.

Das *Acceptance-Rejection* Verfahren (siehe z.B. Gamerman (1997) S.24-26) benötigt zwar auch keine Normierungskonstante, dafür braucht man jedoch eine Vorschlagsdichte, die die Grenzverteilung π so gering überdeckt, dass diese nicht zu kleinen Akzeptanzraten führt.

Folgende zwei Spezialfälle werden in der Praxis häufig verwendet. Falls man die Vorschlagsdichte symmetrisch wählt, d.h. $q(\boldsymbol{\theta}, \boldsymbol{\phi}) = q(\boldsymbol{\phi}\ , \boldsymbol{\theta})$, dann spricht man von einem symmetrischen MH. Für einen symmetrischen MH verwendet man häufig eine symmetrische Verteilung mit Erwartungswert $\boldsymbol{\phi}$ als Vorschlagsdichte $q(\boldsymbol{\theta}, \boldsymbol{\phi})$. Beim Unabhängigkeits MH Algorithmus wählt man die Vorschlagsdichte unabhängig vom momentanen Zustand der Markovkette, d.h. es gilt $q(\boldsymbol{\theta}, \cdot) = f(\cdot)$.

Nach Konstruktion ist es bei dem MH Algorithmus möglich, dass die erzeugte Markovkette sich nicht bewegt. Daher sollte man immer die Akzeptanzrate, d.h. den prozentuale Anteil aller Iterationen bei denen sich die Markovkette bewegt, bestimmen. Man ist an hohen Akzeptanzraten interessiert. Diese kann man z.B. erreichen, falls man als Vorschlag in der j-ten Iteration eine multivariate Normalverteilung mit Erwartungswertvektor $\boldsymbol{\theta}^{(j-1)}$ und Kovarianzmatrix mit kleinen diagonalen Einträgen verwendet. Diese Wahl führt zu kleinen Bewegungen der Markovkette und daher zu einer Akzeptanzrate nahe 1. Anderseits müssen die Realisierungen der Markovkette jedoch den gesamten Parameterraum durchlaufen können, damit diese konvergiert. Kleine Bewegungen führen jedoch zu einer langsamen Konvergenz der Markovkette. Auf der anderen Seite fallen große Bewegungen in die Schwänze der Verteilung, was wiederum zu einer niedrigen Akzeptanzrate führt. Daher sollte die Vorschlagdichte so gewählt sein, dass grössere Bewegungen vom momentanen Wert mit grösserer Akzeptanzwahrscheinlichkeit möglich sind. Bennett et al. (1996) und Besag et al. (1995) raten Vorschlagsdichten so zu wählen, dass man eine Akzeptanzrate von 20-50% erhält.

Die Monografie von Gamerman (1997) gibt eine allgemeine Einführung zu MCMC Verfahren, während W.R. Gilks et al. (1996) viele Anwendungen enthält. Ein Übersichtsartikel ist in Chib (2001) enthalten.

4 Statistische Software für MCMC Verfahren

Es gibt momentan noch keine kommerzielle statistische Software für den Entwickler von MCMC Verfahren für komplexe Modelle. Solche Entwickler schreiben daher in der Regel ihre eigenen Programme in C(++) und/oder MATLAB. Für den Anwender von bereits entwickelten MCMC Verfahren bieten (win)bugs und BayesX eine reiche Auswahl von Modellen. Beide Softwareprogramme sind free share Programme.

Das Paket (win)bugs wurde gemeinsam am MRC Biostatistics und Imperial College of Medicine at St. Mary's, London entwickelt und ist unter

`http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml`

kostenlos erhältlich. David Scollnik hat unter

`http://www.math.ucalgary.ca/~scollnik/abcd/`

(win)bugs Materialien für den Bereich Aktuarwissenschaften zusammengestellt und einen Übersichtsartikel (siehe Scollnik (2001)) geschrieben.

Viele interessante Beispiele, die (win)bugs verwenden, findet man in den Monografien von Congdon (2001) und Congdon (2003).

Das Paket BayesX wurde am Institut für Statistik an der Universität München entwickelt und ist ein Windows basiertes Anwendungstool. Es ist unter

`http://www.stat.uni-muenchen.de/~lang/bayesx/bayesx.html`

kostenlos erhältlich.

Für den Anwender von MCMC Methoden ist es auch wichtig den burnin und das Konverengenzverhalten der MCMC Iterationen zu beurteilen. Ein Plot der Iterationen und die Betrachtung der geschätzten Autokorrelationen geben dabei einen ersten Einblick. Mathematische Methoden basieren auf der Theorie der Markovketten und sind momentan in der Praxis nur beschränkt anwendbar. Cowles und Carlin (1996) und Brooks und Roberts (1998) geben einen Überblick über Methoden zur Konvergenzdiagnostik, die direkt aus den beobachteten Iterationswerten bestimmt werden können. Insbesondere ist die Gelman-Rubin Statistik (Gelman und Rubin 1992) direkt in (win)bugs implementiert. Weitere Konvergenzdiagnostiken sind in der Splus Library CODA() implementiert. Diese Library ist unter der gleichen Internetadresse wie (win)bugs erhältlich.

5 Anwendung auf Gesamtschadenmodelle

Wir werden nun die Bayesianischen Verfahren mit Hilfe von MCMC Methoden an einem einfachen Gesamtschadenmodell illustrieren. Ein wichtiges Problem der Versicherungswirtschaft ist die Vorhersage der Gesamtschadenverteilung. Der Gesamtschaden S_t in der Zeitperiode t ergibt sich als Summe der Einzelschäden, d.h.

$$S_t = Y_{t,1} + \cdots + Y_{t,N_t}, \tag{5.1}$$

wobei $Y_{t,1}, .., Y_{t,N_t}$ die Einzelschäden und N_t die Anzahl der Schäden in der Zeitperiode t bezeichnen. Für erste einfache Modelle wird angenommen, dass die Einzelschäden unabhängig und identisch verteilt sind. Eine weitere Annahme ist, dass die Einzelschäden unabhängig von der Anzahl der Schadensfälle auftreten. In dem klassischen Ansatz der mathematischen Statistik erfolgte die Parameterschätzung für die Einzelschadenverteilung und die Verteilung der Anzahl der Schäden durch Maximum Likelihood. Rekursive Methoden zur Bestimmung der Gesamtschadenverteilung müssen eingesetzt werden (siehe z.B. Klugman et al. (1998)) Hierbei wird angenommen, dass die geschätzten Parameter die wahren Parameter sind. Dies vernachlässigt die Parameterunsicherheit, und Dickson et al. (1998) zeigen z.B., dass die Vorhersage der Gesamtschadenverteilung stark fehlerbehaftet sein kann. Dieses Problem kann durch einen Bayesianischen Ansatz gelöst werden. Panjer und Willmot (1983) schreiben:

> The operational actuarial interpretation is that the risk is first selected from the whole set of risks in accordance with the risk distribution, and the performance of the selected risk is then monitored. The statistical interpretation is essentially Bayesian.

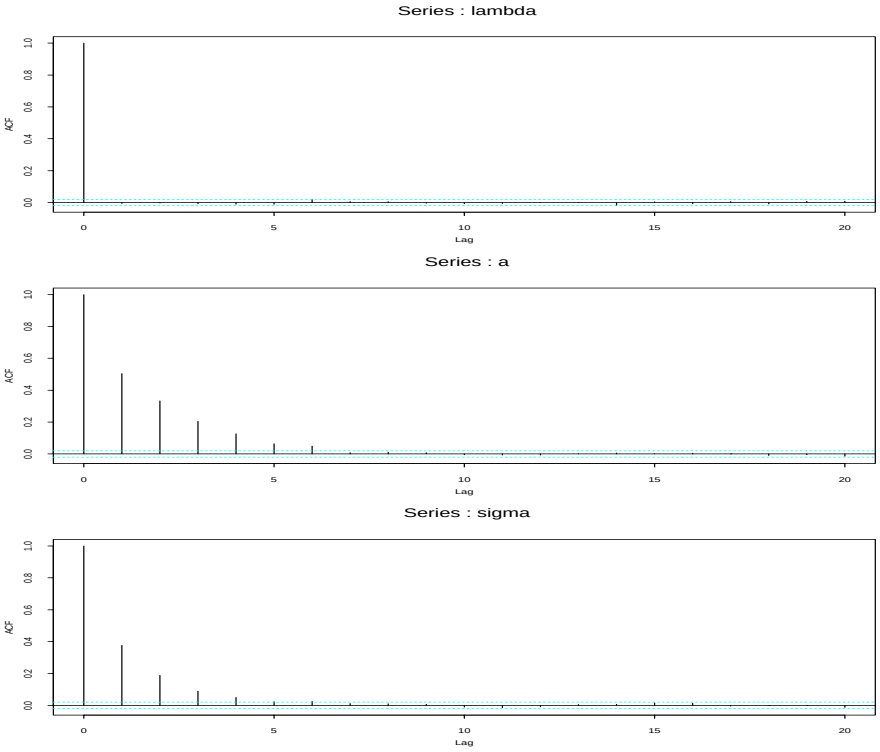

Abbildung 1: Autokorrelationen für das GPD-Poisson Modell

Wir stellen nun einen Bayesianischen Ansatz für das Gesamtschadenmodell (5.1) mit den vereinfachenden Annahmen vor und zeigen wie (win)bugs genutzt werden kann, um die Parameterunsicherheit in der Gesamtschadenverteilung zu berücksichtigen. Wir lassen dabei zunächst folgende Einzelschadenverteilungen zu:

- **Pareto Verteilung**
  $Y_{t,i} \sim Pareto(a, b)$ mit Dichte

$$f(y) = \frac{ab^a}{y^{a+1}},\ y \geq b > 0, a > 0. \tag{5.2}$$

- **Generalisierte Pareto Verteilung (GPD) mit** $a > 0$
  $Y_{t,i} \sim GPD(a, \mu, \sigma)$ mit Dichte

$$f(y) = \frac{1}{\sigma\left[1 + \frac{1}{a\sigma}(y-\mu)\right]^{1+a}},\ y \geq \mu \in \mathbb{R}, \sigma > 0, a > 0. \tag{5.3}$$

Die Pareto Verteilung ist eine Standardschadenverteilung, während die GPD Verteilung in der Extremwerttheorie auftaucht (siehe z.B. die Monografie von Embrechts, Klüppelberg, und Mikosch (1997)). Die GPD Verteilung tritt als Grenzverteilung für Überschreitungen eines Schwellenwert, falls der Schwellenwert gegen unendlich konvergiert, (siehe Balkema und de Haan (1974) und Pickands (1975)). Insbesondere kann man die GPD Verteilung mit $a > 0$, als die Verteilung von $Y = \mu + a \times \sigma \times (X - 1)$, wobei $X \sim Pareto(a, 1)$ identifizieren.

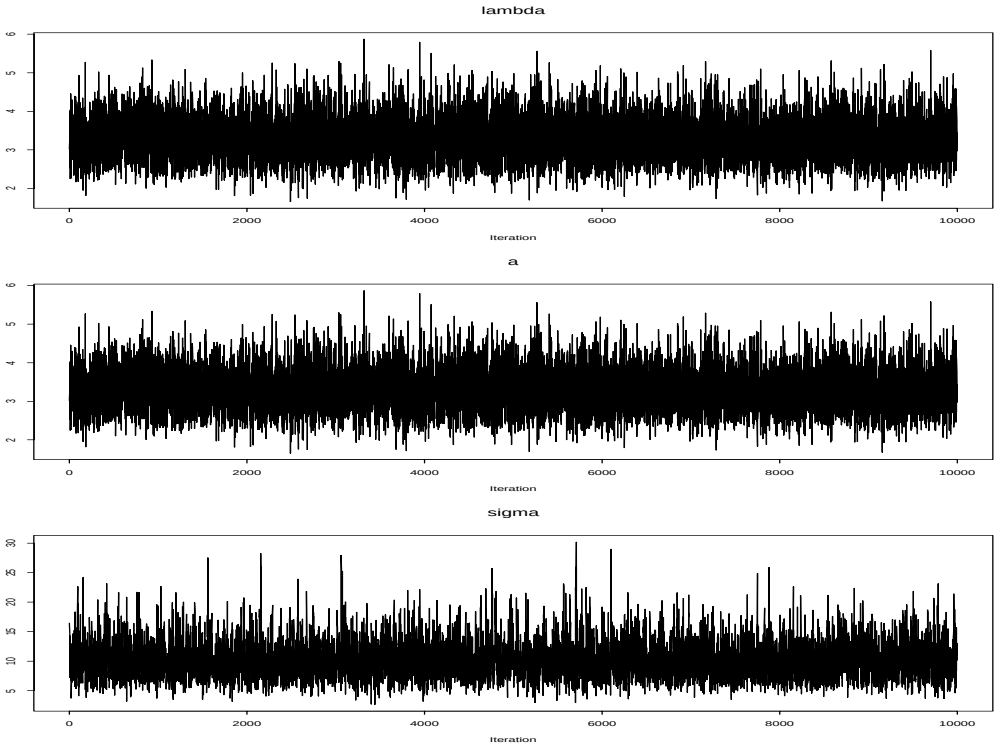

Abbildung 2: MCMC Iterationen für das GPD-Poisson Modell mit $a > 0$

Wir betrachten hier den Schwellenwert b in der Pareto Verteilung als bekannt. David Scollnik lässt für Pareto verteilte Schäden auch einen unbekannten Schwellenwert zu.

Für die Verteilung der Anzahl der Schäden nehmen wir folgendes Mischungsmodell an:

$$N_t|\lambda \sim Poisson(\lambda) \text{ mit } \lambda \sim Gamma(\alpha, \beta) \qquad \alpha > 0, \beta > 0 \tag{5.4}$$

Dies erlaubt die Modellierung von Überdispersion in der Anzahl der Schäden, d.h. $Var(N_t) > E(N_t)$, insbesondere gilt, dass die marginale Verteilung von N_t negativ binomial verteilt ist.

Wir wenden nun diese Modelle auf einen bekannten Datensatz aus der Extremwerttheorie an, nämlich auf Daten einer dänischen Feuerversicherung, die schon in McNeil (1997) und Resnik (1997) untersucht worden sind. Der Datensatz ist auf der Homepage von EVIS (`http://www.math.ethz.ch/~mcneil/software.html`) erhältlich. Der Gesamtdatensatz enthält 2157 Einzelschäden mit einer Schadensumme von über 1 Million Dänischen Kronen (DKK) aus den Jahren 1980 bis einschließlich 1990. Wir beschränken uns auf die Anpassung der Schäden, deren Schadenhöhe 20 Millionen DKK überschreiten. Dies war bei 36 Schäden der Fall. McNeil (1997) zeigt in einer explorativen Datenanalyse, dass es keine Hinweise gibt, dass die Einzelschäden in diesem Datensatz nicht unabhängig und identisch verteilt sind. Im Anhang sind die entsprechenden (win)bugs Programme zu finden, die Pareto und GPD verteilte Einzelschäden in Kombination mit dem hierarchischen Poisson Mischungsmodell (5.4) für die Anzahl der Schäden schätzen. Die MCMC Iterationen wurden daher in (win)bugs durchgeführt und die Abbildungen wurden mit der Statistiksoftware Splus erzeugt.

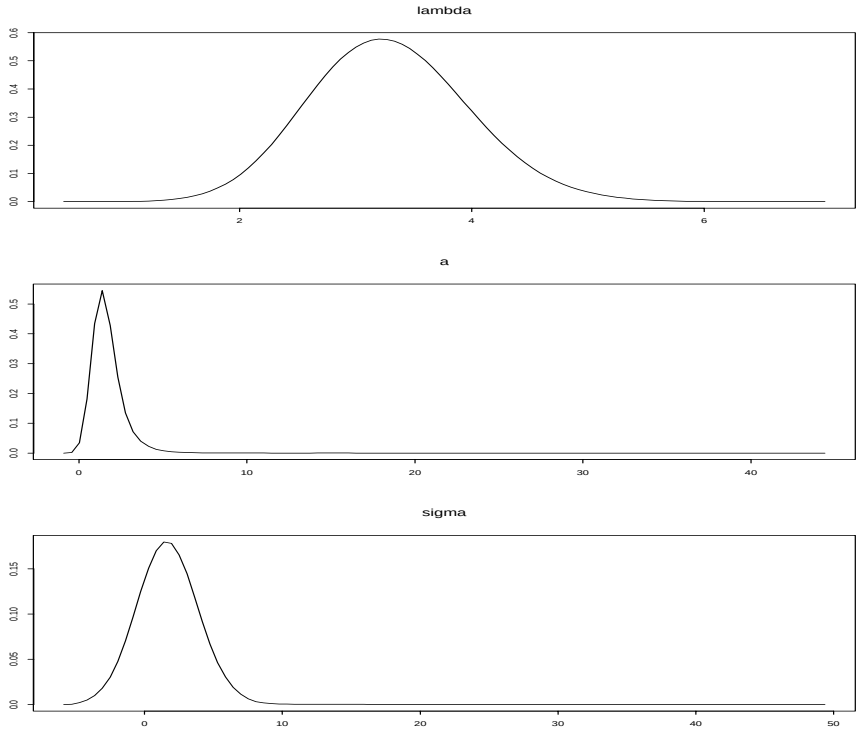

Abbildung 3: a-posteriori Dichteschätzungen für das GPD-Poisson Modell mit $a > 0$

Für alle Modelle wurden 10000 MCMC Iterationen durchgeführt und ein burnin von 500 Iterationen hat sich nach Ansicht der MCMC Iterationen als ausreichend erwiesen. Insbesondere sind auch die beobachteten Autokorrelationen zwischen den Iterationen recht niedrig (siehe Abbildung 1). Dies deutet auf eine schnelle Konvergenz des MCMC Algorithmus hin. Für das GPD-Poisson Modell zeigt Abbildung 2 die Iterationswerte und Abbildung 3 die dazugehörigen a-posteriori Dichteschätzungen.

Auch die hier nicht gezeigten MCMC Iterationen und Autokorrelationen des Pareto-Poisson Modells zeigen schnelle Konvergenz auf. In Tabelle 1 kann man für beide Modelle die a-posteriori Punkt- und Intervalschätzungen basierend auf den letzten 9500 Iterationen finden.

Um die Anpassungsgüte der verschiedenen Modelle zu beurteilen, wurde die Verteilungsfunktion der Einzelschäden durch die empirische Verteilung, die geschätzte Verteilungfunktion basierend auf ML Schätzung und durch die geschätzte a-posteriori Verteilungfunktion angepasst. Bei der Bayesianischen Schätzung der Verteilungsfunktion wurden a-posteriori Mediane verwendet. Punktweise 95 % Bayesianische Vertrauensintervalle wurden auch bestimmt. Die Ergebnisse sind in Abbildung 4 zu sehen. Insbesondere sieht man, dass die ML Schätzung die Verteilung unterschätzt, wenn ein Pareto Modell für die Einzelschäden verwendet wird. Dagegen überschätzt der ML Schätzer für das GPD Modell die Verteilung. Die Bayesianische Schätzung stimmt gut mit der ML Schätzung im Fall des Pareto und GPD Modells überein. Die Bayesianischen Intervallschätzungen sind präziser für das Pareto als für das GPD Modell. Dies ist zu erwarten, da das GPD Modell einen Parameter mehr als das Pareto Modell verwendet. Für beide Modelle überdecken die Intervallschätzer die empi-

GPD-Poisson Modell mit $a > 0$ (siehe (5.3))					
node	mean	sd	2.5 %	median	97.5 %
a	1.708	1.117	.708	1.484	3.893
σ	10.270	3.211	5.229	9.871	17.640
λ	3.295	.574	2.294	3.258	4.519
Pareto-Poisson Modell (siehe (5.2))					
node	mean	sd	2.5 %	median	97.5 %
a	1.810	.298	1.270	1.796	2.422
λ	3.305	.577	2.273	3.267	4.550

Tabelle 1: A-posteriori Punkt- und Intervallschätzungen

rischen Werte. Da die Überschätzung im Pareto Modell geringer ist als für das GPD Modell und die Intervallschätzer genauer, scheint das Pareto Modell geeigneter zu sein für diesen Datensatz. Mehr formale Bayesianische Modellwahlansätze wie Bayesfaktoren (siehe (Kass und Raftery 1995) und (Han und Carlin 2001)) für nicht geschachtelte Modelle und das DIC Kriterium (Spiegelhalter et al. (2002)) sind hier auch denkbar.

Wir wenden uns nun den Bayesianischen Vorhersagen zu. Eine zukünftige Beobachtung X_{n+1} basierend auf den Daten $\mathbf{X} = (X_1, \ldots, X_n)$ mit a-posterori Verteilung $\pi(\boldsymbol{\theta}|\mathbf{X})$ wird mit Hilfe der prediktiven Dichte

$$f(x_{n+1}|x_1, .., x_n) := \int f(x_{n+1}|\boldsymbol{\theta}) \cdot \pi(\boldsymbol{\theta}|\mathbf{X}) d\boldsymbol{\theta}$$

vorhergesagt. Diese Dichte kann approximiert werden durch

$$f(x_{n+1}|x_1, .., x_n) \approx \frac{1}{R-B} \sum_{i=B+1}^{R} f(x_{n+1}|\boldsymbol{\theta}^{(i)})$$

wobei $\{\boldsymbol{\theta}^{(i)}, i = 1, .., R\}$ der Output eines MCMC Verfahrens mit burnin$= B$ ist. Für unser Modell bedeutet dies für die Anzahl der Schadenfälle im Jahr 1991, dass die prediktiven Wahrscheinlichkeiten durch

$$\begin{aligned} P(N_{91} = n|Daten) &\approx \frac{1}{R-B} \sum_{i=B+1}^{R} P(N_{91} = n|\lambda^{(i)}) \\ &= \frac{1}{R-B} \sum_{i=B+1}^{R} \frac{e^{-\lambda^{(i)}}[\lambda^{(i)}]^n}{n!} \end{aligned}$$

geschätzt werden können. Hierbei ist $\lambda^{(i)}$ der i-te Iterationswert für λ. Für unseren Datensatz sind diese prediktiven Wahrscheinlichkeiten in Tabelle 2 enthalten. Wir sehen, dass die Wahl des Modells für die Einzelschadenverteilung keinen Einfluss hat. Dies ist zu erwarten, da angenommen wurde, dass die Anzahl der Schäden unabhängig von der Höhe der Einzelschäden ist.

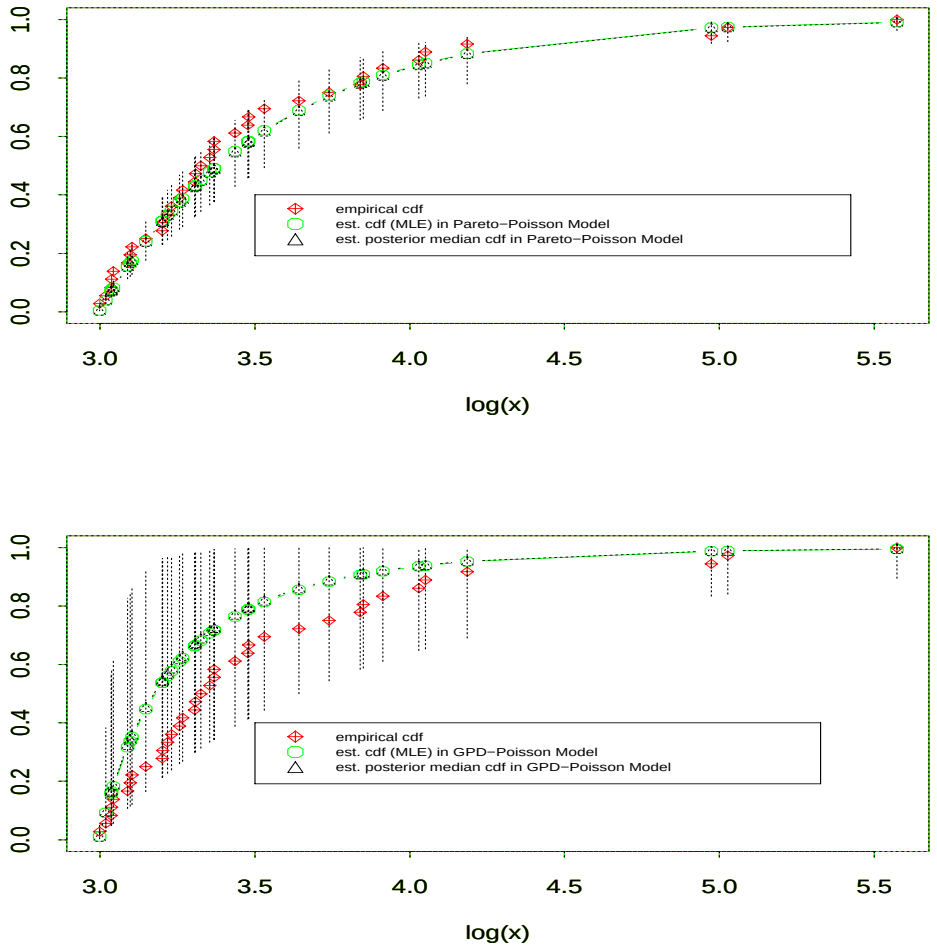

Abbildung 4: Schätzungen der Verteilungsfunktion der Einzelschäden (Empirisch, ML und Bayesianisch mit 95 % Vertrauensintervalle) für das Pareto und GPD Einzelschadenmodell

Für den Gesamtschaden im Jahr 1991

$$S_{91} = \begin{cases} Y_{91,1} + .. + Y_{91,N_{91}} & \text{falls } N_{91} > 0 \\ 0 & \text{sonst} \end{cases}$$

ist die prediktive Dichte für das Pareto-Poisson Modell gegeben durch

$$f(S_{91}|Daten) = \int_0^\infty \int_0^\infty f(S_{91}|\lambda, a) f(\lambda, a|Daten) d\lambda\, da.$$

Hierbei ergibt sich das Problem, dass $f(S_{91}|\lambda, a)$ rekursiv berechnet werden muss. Daher ist es einfacher die prediktive Dichte zu simulieren. Hierzu wird für $i = B+1, \cdots, R$ folgende Simulation durchgeführt

1. Erzeuge $N_{91}^{(i)} \sim Poisson(\lambda^{(i)})$

n	**GPD-Poisson Modell** $\hat{P}(N_{91} = n \mid Daten)$	**Pareto-Poisson Modell** $\hat{P}(N_{91} = n \mid Daten)$
0	0.044	0.043
1	0.130	0.129
2	0.201	0.200
3	0.212	0.212
4	0.173	0.173
5	0.116	0.117
6	0.066	0.067
7	0.033	0.034
8	0.015	0.015
9	0.006	0.006
10	0.002	0.002
11	0.001	0.001
12	0.000	0.000

Tabelle 2: A-posteriori Vorhersage der Anzahl der Schäden im Jahr 1991

2. Falls $N_{91}^{(i)} = 0$, dann setze $S_{91}^{(i)} = 0$

3. Falls $N_{91}^{(i)} > 0$, dann erzeuge

$$Y_{91,k}^{(i)} \sim Pareto(a^{(i)}, b) \quad \text{für}\, k = 1, .., N_{91}^{(i)}$$

und setze
$$S_{91}^{(i)} = Y_{91,1}^{(i)} + .. + Y_{91,N_{91}^{(i)}}^{(i)},$$

falls das Pareto-Poisson Modell verwendet wird. Für das GPD-Poisson Modell wird die Pareto Verteilung im Schritt 3 durch die GPD Verteilung ersetzt. Die prediktive Dichte wird nun durch die empirische Verteilung von $\{S_{91}^{(i)}, i = B+1, .., R\}$ approximiert. Dafür wurden die letzten 5000 MCMC Iterationen verwendet. Die entsprechenden geschätzten prediktiven Dichten für einen von Null verschiedenen Gesamtschaden im Jahr 1991 sind in Abbildung 5 enthalten. Die ML Schätzungen sind wie erwartet konzentrierter als die Bayesianischen Schätzungen. Das Pareto-Poisson Modell sagt einen größeren Gesamtschaden für Schäden über 20 Millionen DKK vorher als das GPD-Poisson Modell. Dies ist zu erwarten, da nach Abbildung 4 die erwartete Höhe eines Einzelschadens im Pareto-Poisson Modell größer ist als die erwartete Höhe eines Einzelschaden im GPD-Poisson Modell. Die geschätzte Wahrscheinlichkeit für keinen Schaden über 20 Millionen DKK in 1991 ist .0358 (.0398) für das GPD-Poisson Modell wenn ML (Bayesianische) Schätzung verwendet wird. Für das Pareto-Poisson Modell ist diese Wahrscheinlichkeit .0386 (.0426). Zusammenfassend kann man sagen, dass bei diesem Datensatz die Parameterunsicherheit nicht zu gross ist und dass das Pareto-Poisson Modell eine gute Anpassung der Daten erlaubt.

Zielsetzung dieses Artikels ist es, MCMC Verfahren einem neuen Kreis von Anwendern zugänglich zu machen. Deswegen wurde ein einfaches Beispiel zur Illustration gewählt, was

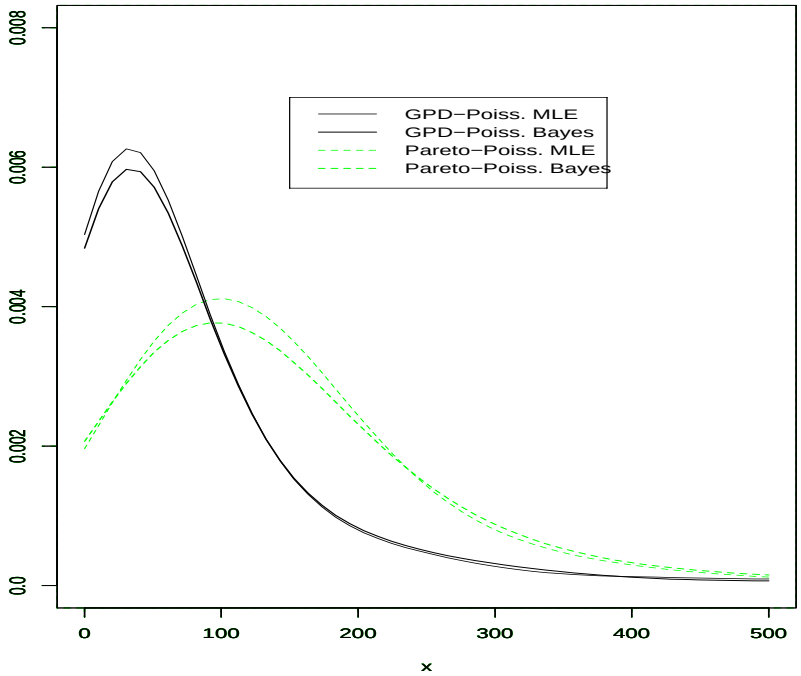

Abbildung 5: A-posteriori und ML Dichteschätzungen für einen von Null verschiedenen Gesamtschaden im Jahr 1991

jedoch nur beschränkt die Möglichkeit der Parameterschätzung in komplexen Modellen aufzeigen kann. Abschliessend möchten wir deswegen auf Literatur verweisen, die diese Möglichkeiten mehr ausschöpfen.

Im Bereich von Versicherungsdaten verwenden Bottolo et al. (2003) Reversible Jump MCMC Verfahren, um Heterogenität bei extremen Ereignissen zu berücksichtigen. Diese Heterogenität wird durch einen Mischansatz modelliert, wobei die Anzahl der Komponenten der Mischverteilung gemeinsam mit den anderen Parametern geschätzt wird. Für die Schätzung der durchschnittlichen Schadenhöhe in einem Portfolio verwenden Dimakos und di Rattalma (2002) einen Poisson/Gamma Ansatz im Modell (5.1). Da alle Schäden betrachtet werden, wird hier die Gamma Verteilung als Verteilung der Einzelschäden herangezogen. Sowohl auf der Ebene Schadenanzahl als auch auf der Ebene der Schadenhöhe werden fixe Effekte von Kovariablen und zufällige Effekte zugelassen. Mit Hilfe von MCMC schätzen Dimakos und di Rattalma (2002) alle Parameter gemeinsam und verwenden diese, um die Risikoprämie zu bestimmen. Hardy (2002) zeigt wie MCMC verwendet werden kann, um Parameterunsicherheit im Bereich von Risikomanagement für eigenkapitalgebundene Versicherungen zu berücksichtigen. Dabei werden sowohl *regime-switching* als auch *GARCH* Modelle unter mehreren Risikostrategien untersucht.

Im Bereich von Finanzdaten untersucht Eraker (2001) Diffusionsmodelle mit Hilfe von MCMC, während Hedgingstrategien von Wechselkursen mit *GARCH* Modellen und stochastische Volatilitätsmodelle (SV) von Bos et al. (2001) betrachtet werden. Statistische Modelle für multivariate Finanzzeitreihen und deren Schätzung mit MCMC Verfahren werden z.B.

von Pitt und Shephard (1999) und Vrontos et al. (2001) entwickelt. Ein Modell für ordinal skalierte Zeitreihen wie sie z.B. bei Preisveränderungen bei hochfrequenten Finanzdaten auftreten wurde von Müller und Czado (2002) untersucht.

Zusammenfassung

Modellierung von realistischen Zusammenhängen erfordern komplexe statistische Modelle. Standard Schätzmethoden wie Maximum Likelihood sind dann oft auch numerisch nicht mehr verwendbar, während eine Bayesianische Schätzung noch durchführbar ist. Dies ist möglich durch das Einsetzen von Markov Chain Monte Carlo (MCMC) Methoden. Dieser Artikel beschreibt in Grundzügen wie MCMC Verfahren arbeiten und diskutiert zwei Standardverfahren, nämlich den Gibbs Sampler und den Metropolis Hastings Algorithmus. Praktische Hinweise zur Implementierung und Auswertung werden gegeben. Dabei wird die Software (win)bugs eingesetzt. Zum Abschluss wird die Implementierung von MCMC Algorithmen und deren Bayesianische Auswertung an einem einfachen statistischen Modells illustriert. Dabei wurden Standardmodelle wie das Pareto-Poisson und das GPD-Poisson Modell für die Modellierung des Gesamtschaden in der Versicherungswirtschaft eingesetzt und die Parameterunsicherheit bei der Vorhersage berücksichtigt.

Introduction to Markov Chain Monte Carlo Methods with Application to Aggregate Insurance Loss

Modeling of realistic dependency structures requires the usage of complex statistical models. In this case standard estimating methods such as maximum likelihood are not even numerically tractable in contrast to Bayesian approaches. This is facilitated by utilising Markov Chain Monte Carlo (MCMC) methods. This paper introduces MCMC approaches such as the Gibbs sampler and the Metropolis Hastings algorithm. Practical advice with regard to implementation and evaluation will be given. The statistical software (win)bugs will be used. Finally a data example requiring modeling of aggregate insurance loss will be discussed in detail when parameter uncertainty is accounted for. In particular standard models such as the Pareto-Poisson and the GPD-Poisson models were used to illustrate the Bayesian approach with MCMC methods.

Appendix

(win)bugs Programm für das Pareto-Poisson Modell

```
Model
model;
{
   # Model structure for Pareto modeling.
   for(i in 1:L){
         Y[i]~dpar(a,20)
     }
   a~dgamma(0.001,0.001);
   # Model structure for Poisson Modelling
   N91 ~ dpois( lambda )
   for( i in 1 : T ){
       N[i] ~ dpois( lambda )
    }
   lambda ~ dgamma( 0.001, 0.001 )
}

Data
list( N  = c(3, 4, 5, 0,0,3, 1, 4, 8, 5, 3),
 Y  = c(26.214641, 21.961933, 263.250366, 34.141547, 20.969856, 56.225426,
        50.065531, 24.970273, 20.049941, 65.707491, 27.262595, 22.258226,
22.137567, 46.500000, 57.410636, 29.026037, 23.283859, 32.467532,
        29.037106, 27.829314, 38.154392, 27.338066, 25.288376, 20.452529,
        47.019521, 24.578527, 25.95386, 31.055901, 24.555461, 42.091448,
20.863675, 152.413209, 32.387807, 20.826733, 144.657591, 28.630363),
 N91  = NA,  T  = 10,  L  = 36)

Inits
list( a = 4, lambda=15)
```

(win)bugs Programm für das GPD-Poisson Modell

```
Model
model;
{
   # Model structure for GPD modelling.
   for(i in 1:L){
         ones[i]<-1
         ones[i]~dbern(p[i])
         p[i]<-pow(1/(1+((Y[i]-20)/(sigma*a))),a+1)/(bignum*sigma)
     }
   bignum<-100
```

```
   sigma~dgamma(0.001,0.001);
   a~dgamma( 0.001, 0.001 ) ;
   # Model structure for Poisson Modelling
   N91 ~ dpois( lambda )
   for( i in 1 : T ){
       N[i] ~ dpois( lambda )
    }
   lambda ~ dgamma( 0.001, 0.001 )
}

Data
list(... wie bei Pareto-Poisson Modell)

Inits
list(a = 4, lambda=15,sigma=10)
```

Literatur

Balkema, A. und L. de Haan (1974). Residual life time at great age. *Annals of Probability 2*, 792–804.

Bennett, J., A. Racine-Poon, und J. Wakefield (1996). MCMC for nonlinear hierarchical models. In *Markov Chain Monte Carlo in Practice (eds. W.R. Gilks, S. Richardson and D. J. Spiegelhalter)*, London, pp. 339–357. Chapman & Hall.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis (Second edition)*. Berlin: Springer.

Besag, J., P. Green, D. Higdon, und K. Mengersen (1995). Bayesian computation and stochastic systems (with discussion). *Stat. Science 10*, 3–66.

Bos, C. S., R. J. Mahieu, und H. K. van Dijk (2001). Daily exchange rate behaviour and hedging of currency risk. *J. Appl. Econometrics 15*, 671–696.

Bottolo, L., G. Consonni, P. Dellaportas, und A. Lijoi (2003). Bayesian analysis of extreme values by mixture modeling. *Extremes 6*, 25–47.

Brooks, S. P. und G. O. Roberts (1998). Assessing convergence of Markov Chain Monte Carlo algorithms. *Statistics and Computing 8*, 319–335.

Chib, S. (2001). Markov Chain Monte Carlo methods: Computation and inference. In *Handbook of Econometrics, Volume 5 (eds. J.J. Heckman and E. Leamer)*, pp. 3569–3649. Elsivier Science B.V.

Chib, S. und E. Greenberg (1995). Understanding the Metropolis-Hastings algorithm. *Amer. Statistician 49*, 327–335.

Congdon, P. (2001). *Bayesian Statistical Modelling*. New York: Wiley & Sons.

Congdon, P. (2003). *Applied Bayesian Modelling*. New York: Wiley & Sons.

Cowles, M. K. und B. P. Carlin (1996). Markov Chain Monte Carlo convergence diagnostics: A comparative review. *J. Amer. Statist. Assoc. 91*, 883– 904.

Dickson, D., L. Tedesco, und B. Zehnwirth (1998). Predictive aggregate claims distributions. *The Journal of Risk and Insurance 65*, 689–709.

Dimakos, X. K. und A. F. di Rattalma (2002). Bayesian premium rating with latent structure. *Scand. Actuarial J. 3*, 162–184.

Embrechts, P., C. Klüppelberg, und T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer.

Eraker, B. (2001). MCMC analysis of diffusion models with application to finance. *Amer. Stat. Assoc. J. of Business and Econ. Statistics 19*, 177–191.

Gamerman, D. (1997). *Stochastic simulation for Bayesian inference*. London: Chapman & Hall.

Gelfand, A. E. und A. F. M. Smith (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc. 85*, 398–409.

Gelman, A., J. B. Carlin, H. S. Stern, und D. B. Rubin (1995). *Bayesian data analysis*. London: Chapman & Hall.

Gelman, A. und D. Rubin (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science 7*, 457–511.

Geman, S. und D. Geman (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transaction on Pattern Analysis and Machine Intelligence 6*, 721–741.

Guttorp, P. (1995). *Stochastic Modeling of Scientific Data*. Chapman & Hall.

Han, C. und B. Carlin (2001). MCMC methods for computing Bayes factors: a comparative review. *Journal of the American Statistical Association 96*, 1122–1132.

Hardy, M. (2002). Bayesian risk management for equity-linked insurance. *Scand. Actuarial J. 3*, 185–211.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika 57*, 97–109.

Hoeting, J. A., D. Madigan, A. E. Raftery, und C. T. Volinsky (1999). Bayesian model averaging: A tutorial (disc: P401-417). *Statistical Science 14*, 382–401. Corrected version available at www.stat.washington.edu/www/research/online/hoeting1999.pdf.

Kass, R. E. und A. E. Raftery (1995, June). Bayes factors. *Journal of the American Statistical Association 90*, 773–795.

Klugman, S. A., H. Panjer, und G. Willmot (1998). *Loss models: from data to decisions*. New York: Wiley & Sons.

Lee, P. M. (1997). *Bayesian Statistics: An Introduction*. Arnold.

McNeil, A. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin 27*, 117–137.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, und E. Teller (1953). Equation of state calculations by fast computing machines. *J. of Chemical Physics 21*, 1087–1092.

Meyn, S. und R. Tweedie (1993). *Markov Chains and stochastic stability*. New York: Springer.

Müller, G. und C. Czado (2002). Regression models for ordinal valued time series with application to high frequency financial data. *SFB 386 Statistische Analyse diskreter Strukturen , Discussion Paper 301 ,http://www.stat.uni-muenchen.de/sfb386/*.

Panjer, H. und G. Willmot (1983). Compound poisson models in actuarial risk theory. *Econometrics 23*, 63–76.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics 3*, 119–131.

Pitt, M. K. und N. Shephard (1999). Time-varying covariances: A factor stochastic volatility approach. In *Bayesian Statistics 6 (eds. J.M. Bernardo et.al.)*, pp. 547–570. Oxford University Press.

Resnick, S. I. (1992). *Adventures in Stochastic Processes*. Birkhäuser.

Resnik, S. (1997). Discussion of the danish data on large fire insurance losses. *ASTIN Bulletin 27*, 139–151.

Roberts, G. O. und A. F. M. Smith (1994). Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. *Stoch. Proc. Appl. 49*, 207–216.

Scollnik, D. (2001). Actuarial modeling with MCMC and BUGS. *North American Actuarial J. 5*, 96–124.

Spiegelhalter, D., N. Best, B. Carlin, und A. van der Linde (2002). Bayesian measures of model complexity and fit. *J. Roy. Stat. Soc. B. 64*, 583–639.

Vrontos, I., S. Giakoumatos, P. Dellaportas, und D. Politis (2001). An application of three bivariate time-varying volatility models. *App. Stoch. Model. Bus. Ind. 17*, 121–133.

W.R. Gilks, W., S. Richardson, und D. Spiegelhalter (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall.