

Praktische Erfahrungen mit Grossschadenverteilungen

Autor(en): **Schnieper, René**

Objektyp: **Article**

Zeitschrift: **Mitteilungen / Schweizerische Vereinigung der
Versicherungsmathematiker = Bulletin / Association Suisse des
Actuaires = Bulletin / Swiss Association of Actuaries**

Band (Jahr): - **(1993)**

Heft 2

PDF erstellt am: **23.07.2024**

Persistenter Link: <https://doi.org/10.5169/seals-551141>

Nutzungsbedingungen

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.

Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.

Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

Haftungsausschluss

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

Praktische Erfahrungen mit Grossschadenverteilungen

1. Einführung

In ihrem Buch «Loss Distributions» analysieren *Hogg* und *Klugmann* die von der American Insurance Association während der Periode 1949–1980 registrierten Wirbelstürme, die in den USA einen Marktschaden von mindestens \$ 1 000 000 verursacht haben. Es wurden 38 solche Ereignisse registriert. Die Schäden werden mit Hilfe des Baukostenindex in das Jahr 1981 indexiert. Die Autoren erhalten somit 35 Schäden, die eine Schwelle von \$ 5 000 000 übersteigen.

Diese Schäden werden statistisch analysiert. Die Autoren passen fünf verschiedene Verteilungen an: die Pareto-Verteilung, eine verallgemeinerte Pareto-Verteilung, die Burr-Verteilung, die Weibull- und die Lognormal-Verteilung. Die verallgemeinerte Pareto- und die Burr-Verteilung sind Verallgemeinerungen der Pareto-Verteilung. Im vorliegenden Fall werden sie aufgrund eines Likelihood-Ratio-Testes zugunsten der Pareto-Verteilung verworfen. Die Wahl zwischen der Pareto-, Weibull- und Lognormal-Verteilung erfolgt aufgrund einer Ad-hoc-Statistik, der «Limited Expected Value Function». Die Autoren kommen zum Schluss, dass die Weibull-Verteilung den besten Fit liefert.

Die Höhe des Einzelschadens über \$ 5 000 000 ist somit durch folgende Verteilungsfunktion gegeben:

$$P(X \leq x) = 1 - e^{-cx^\tau}$$

wobei die geschätzten Parameter die folgenden sind:

$$\tau = 0.52$$

$$c = 7.5 \cdot 10^{-5}.$$

Die erwartete Anzahl Schäden pro Jahr über \$ 5 000 000 ist gleich der beobachteten Häufigkeit dieser Ereignisse:

$$n_{5''} = \frac{35}{32} = 1.094$$

Im August 1992 hat Hurricane Andrew in Florida und Louisiana einen Marktschaden verursacht, der auf ca. $\$ 15.5 \cdot 10^9$ geschätzt wird (Quelle: Property Claims Services Division of the American Insurance Services Group, gemäss Business Insurance, March 1, 1993). Wir können diesen Schaden benutzen, um das Modell zu testen. Dazu müssen wir den Schadenaufwand bestimmen, den der Wirbelsturm verursacht hätte, wenn er sich 1981 ereignet hätte. Wir gehen dabei analog vor wie *Hogg* und *Klugmann*.

Zwischen 1981 und 1992 ist der Konsumentenpreisindex in USA von 100 auf 153 gestiegen (Quelle: International Financial Statistics, International Monetary Fund). In 1981-Dollar ausgedrückt, beträgt der Andrew-Schaden somit ca $\$ 10.1 \cdot 10^9$. Gemäss Modell ist die erwartete Anzahl derartiger Grossschäden pro Jahr:

$$\begin{aligned} n_{10.1'''} &= n_{5.1''} P(X > 10.1 \cdot 10^9 \mid X > 5 \cdot 10^6) \\ &= 1.094 \cdot \frac{e^{-c(10.1 \cdot 10^9)^\tau}}{e^{-c(5 \cdot 10^6)^\tau}} \\ &= 8.9 \cdot 10^{-6} \end{aligned}$$

und die Wiederkehrperiode zwischen zwei solchen Ereignissen beträgt:

$$\text{Wiederkehrperiode} \simeq (n_{10.1'''})^{-1} \simeq 110\,000 \text{ Jahre.}$$

Mit anderen Worten: Hurricane Andrew ist gemäss Modell so gut wie unmöglich. Da das Ereignis aber trotzdem eingetreten ist, muss am Modell gezweifelt werden. Das Modell von *Hogg* und *Klugmann* ist eine gute Illustration der folgenden zwei Thesen, die für die praktische Analyse von Schadendaten von grosser Bedeutung sind.

Schadendaten aus der Vergangenheit müssen richtig interpretiert und korrigiert werden. Im vorliegenden Fall wurden die alten Schäden zwar indexiert, es wurde aber nicht berücksichtigt, dass sich die versicherten Werte in den exponierten Gebieten viel stärker erhöht haben als die Inflation. Diese Unterlassung führt zu einer groben Unterschätzung des Grossschadenpotentials.

Bei der Bestimmung von Grossschadenverteilungen reicht es in der Regel nicht, wenn man sich auf die Schadenerfahrung abstützt, da diese naturgemäss sehr spärlich ist.

Nebst der Schadenerfahrung muss auch die vorhandene A-priori-Information ausgewertet werden. Im vorliegenden Fall kann aufgrund der geographischen Verteilung der versicherten Werte, der physikalischen Eigenschaften von Wirbelstürmen,

des Zusammenhangs zwischen Windgeschwindigkeit und Zerstörungsgrad mittels Simulationen eine A-priori-Verteilungsfunktion des Schadenaufwandes ermittelt werden. Dies ist auch die Art und Weise, wie in der Praxis der maximale Schaden ermittelt wird, mit dem «realistischerweise» für ein bestimmtes Portefeuille gerechnet werden muss (Probable Maximum Loss). Eine Bestimmung dieser Grösse aufgrund der Schadenerfahrung allein wäre mit zu grossen Unsicherheiten behaftet. Für die praktische Durchführung einer solchen Analyse siehe z. B. *K. M. Clark* (1986).

2. Die Schätzung des Parameters einer Verteilungsfunktion

Anhand eines einfach Beispiels wollen wir nun illustrieren, wie in der Praxis die Daten mit der A-priori-Information kombiniert werden. Wir betrachten das Problem der Tarifierung einer Schadenexzedentendeckung; bei einer solchen Deckung übernimmt der Rückversicherer denjenigen Teil des Schadens, der eine bestimmte Priorität übersteigt. In einem solchen Fall ist nur der rechte Schwanz der Verteilungsfunktion von Interesse.

Oft wird angenommen, dass diese Grossschäden paretoverteilt sind:

$$P(X > x) = \left(\frac{x_0}{x}\right)^\alpha \quad x \geq x_0$$

wobei der Parameter α von Portefeuille zu Portefeuille verschieden ist und geschätzt werden muss.

Falls x_1, x_2, \dots, x_n , Grossschäden sind (d.h. Schäden, die x_0 übersteigen), so ist

$$\hat{\alpha} = \frac{n-1}{\sum_{i=1}^n \log\left(\frac{x_i}{x_0}\right)}$$

ein erwartungstreuer Schätzer von α mit minimaler Varianz. Dies sieht man folgendermassen ein: Die Pareto-Verteilung gehört zur Exponentialfamilie, $\sum_{i=1}^n \log\left(\frac{x_i}{x_0}\right)$ ist eine suffiziente Statistik für α , somit ist jede Funktion von $\sum_{i=1}^n \log\left(\frac{x_i}{x_0}\right)$ ein Schätzer des Erwartungswertes dieser Funktion mit minimaler Varianz; andererseits gilt $E(\hat{\alpha}) = \alpha$. Letzteres ergibt sich aus der Tatsache, dass $\log\left(\frac{x_i}{x_0}\right)$ exponentialverteilt ist mit Parameter α und somit $\sum_{i=1}^n \log\left(\frac{x_i}{x_0}\right)$ gammaverteilt ist. Daraus ergibt sich auch unmittelbar $\text{Var}(\hat{\alpha}) = \frac{\alpha^2}{n-2}$.

Bei der in der Praxis auftretenden Stichprobengrößen (typischerweise zwischen 2 und 5 Grossschäden) ist die Varianz des Schätzers sehr gross, seine Präzision unbefriedigend. Andererseits weiss ein Tarifierer aus Erfahrungen mit ähnlichen Portefeuilles, in welcher Größenordnung der unbekannte Parameter α zu erwarten ist. Dies lässt sich am besten formalisieren, indem man den unbekannt Parameter α als Zufallsgrösse betrachtet und die A-priori-Kenntnisse über α in die A-priori Verteilungsfunktion fliessen lässt.

Bei einer Pareto-Likelihood erweist es sich als zweckmässig, eine Gamma-A-priori-Verteilung zu wählen:

$$f(\alpha) = \frac{c^\gamma}{\Gamma(\alpha)} \alpha^{\gamma-1} e^{-c \cdot \alpha}$$

Die mit Hilfe des Satzes von Bayes erhaltene A-posteriori-Dichte ist dann:

$$\begin{aligned} f(\alpha | x) &= k \cdot f(x | \alpha) f(\alpha) \\ &= k \alpha^n e^{-\alpha \sum_{i=1}^n \log\left(\frac{x_i}{x_0}\right)} \cdot \alpha^{\gamma-1} \cdot e^{-c \cdot \alpha} \\ &= k \alpha^{(\gamma+n)-1} e^{-\left(c + \sum_{i=1}^n \log\left(\frac{x_i}{x_0}\right)\right) \cdot \alpha} \end{aligned}$$

d.h. die A-posteriori-Dichte ist wiederum eine Gamma-Dichte mit den neuen Parametern

$$\begin{aligned} \gamma' &= \gamma + n \\ c' &= c + \sum_{i=1}^n \log\left(\frac{x_i}{x_0}\right). \end{aligned}$$

(N. B.: Die Dichten werden durch Ihre Argumente indexiert; insbesondere bezeichnen $f(\alpha)$ und $f(x)$ i.d.R. verschiedene Dichten.)

Der beste – im Sinn der mittleren quadratischen Abweichung – Schätzer für α ist die A-posteriori-Erwartung

$$E(\alpha | x) = \frac{\gamma'}{c'} = \frac{\gamma + n}{c + \sum_{i=1}^n \log\left(\frac{x_i}{x_0}\right)}$$

Und der mittlere quadratische Fehler des Schätzers ist

$$\begin{aligned}\text{Var}(E[\alpha | \underline{x}]) &= \text{Var}(\alpha) - E[\text{Var}(\alpha | \underline{x})] \\ &= \frac{\gamma}{c^2} - E\left[\frac{\gamma + n}{\left(c + \sum \log\left(\frac{x_i}{x_0}\right)\right)^2}\right] \\ &= \frac{1}{\gamma} \left(\frac{\gamma}{c}\right)^2 - \frac{1}{\gamma + n} E\left[\left[\frac{\gamma + n}{c + \sum \log\left(\frac{x_i}{x_0}\right)}\right]^2\right] \\ \text{Var}(E[\alpha | \underline{x}]) &\simeq \alpha^2 \frac{n}{\gamma(\gamma + n)}\end{aligned}$$

Numerisches Beispiel

Die Parameter der A-priori-Verteilung sind

$$\gamma = 11.39 \quad \text{und} \quad c = 8.44;$$

dies entspricht der folgenden Wahl der Momente von α

$$E(\alpha) = \frac{\gamma}{c} = 1.35 \quad \text{und} \quad \text{Var}(\alpha) = \frac{\gamma}{c^2} = 0.16.$$

Wir haben eine Realisation des Parameters α simuliert und dabei $\alpha = 1.40$ erhalten.

Danach haben wir der Reihe nach drei Realisationen, von Grossschäden stammend, aus einer Verteilung mit $\alpha = 1.40$ (und $x_0 = 1$) simuliert; dabei haben wir erhalten:

$$x_1 = 1.06, \quad x_2 = 1.16, \quad x_3 = 25.03$$

somit

$$\begin{aligned}T &= \sum_{i=1}^n \log\left(\frac{x_i}{x_0}\right) = 3.43 \\ \hat{\alpha} &= \frac{n-1}{T} = 0.58\end{aligned}$$

und

$$E(\alpha | \underline{x}) = \frac{\gamma + n}{c + T} = 1.21$$

Dabei gilt für die Varianzen:

$$\text{Var}(\hat{\alpha}) = \frac{\alpha^2}{n-2} = 1.96 = 1.40^2$$

$$\text{Var}(E[\alpha | \underline{x}]) \simeq \alpha^2 \frac{n}{\gamma(\gamma + n)} = 0.04 = 0.19^2$$

Die Genauigkeit des Bayesischen Schätzers ist viel grösser als diejenige des klassischen Schätzers.

3. Anpassung einer Verteilungsfunktion

Wir betrachten ein Beispiel aus der Feuerversicherung. Es soll eine Schadenexzedenten-Deckung tarifiert werden. Zu diesem Zweck werden die Grossschäden der letzten 10 Jahre geliefert. Es sind dies 17 Schäden, die eine Schwelle von NOK 10 Millionen übersteigen. Die Rohdaten müssen zunächst bereinigt werden: Die Schadenfrequenz muss um die Erhöhung der Anzahl Risiken korrigiert werden, die individuellen Schadenhöhen müssen um die Inflation und allenfalls darüber hinaus um das reelle Wachstum der Zeichnungslimiten korrigiert werden. Die Ermittlung dieser «as if»-Statistik, d.h. der Schäden, wie sie angefallen wären, wenn sie im Deckungsjahr eingetreten wären, ist eine sehr komplexe Aufgabe, die in praktischen Anwendungen mindestens so aufwendig ist wie die eigentliche statistische Analyse. Wir wollen aber nicht näher darauf eingehen und gehen von den bereinigten Daten aus. Es sind dies 17 Grossschäden, die eine indexierte Schwelle von NOK 22,0 Mio übersteigen.

In chronologischer Reihenfolge erhalten wir folgende Schadenstatistik:

Jahr	Schadenhöhe (Mio NOK)	Jahr	Schadenhöhe (Mio NOK)
1983	42.719	1989	25.590
1984	105.860		24.130
1985	—		23.208
1986	29.172	1990	37.772
	22.654		34.126
1987	61.992		27.990
	35.000	1991	—
1988	26.891	1992	53.472
			36.269
			31.088
			25.907

Aufgrund dieser Statistik können wir die empirische Verteilungsfunktion ermitteln. Sie ist im Anhang 1 graphisch dargestellt. In einem nächsten Schritt wird eine ganze Serie von analytischen Verteilungsfunktionen an die empirische Verteilung angepasst, d.h. es werden die Parameter der verschiedenen Verteilungen

mittels der Maximum-Likelihood-Methode geschätzt, und die so erhaltenen Verteilungsfunktionen werden mit der empirischen Verteilung verglichen. Im vorliegenden Fall haben wir die Pareto-, die verallgemeinerte Pareto-, die Lognormal-, die Gamma- und die Loggamma-Verteilung angepasst. Sowohl die Pareto- wie die Loggamma-Verteilung liefern einen guten Fit, wie aus Anhang 1 ersichtlich ist. Wir haben somit zwei statistische Modelle, die im Bereich zwischen NOK 20 Mio und 100 Mio eine sehr gute Anpassung an die Daten liefern, aufgrund der Schäden allein ist es jedoch unmöglich zu beurteilen, wie gut diese statistischen Modelle für die Haftstrecke zwischen NOK 100 Mio und dem Zeichnungslimit NOK 350 Mio sind. Auch im vorliegenden Fall ist jedoch A-priori-Information vorhanden: Bei der Tarifierung von Deckungen wie der vorliegenden liefert der Zedent die Verteilung der Versicherungssummen des Portefeuilles. Die Verteilung des Schadengrades (d.h. des Quotienten des Schadens durch die Versicherungssumme) für das Kollektiv aller Industrie-Feuerversicherungsportefeuilles eines gewissen Marktes ist bekannt und ist in der Form von Franchisen-Rabattkurven gegeben. Somit lässt sich auch die A-priori-Schadenverteilung ermitteln (a priori in dem Sinn, dass sie auf der Schadengradverteilung des Kollektivs beruht).

Diese A-priori-Schadenverteilung wurde in Form einer Treppenfunktion ermittelt:

Schadenhöhe	Gewicht	Schadenhöhe	Gewicht
25.0	0.3110	116.6	0.0223
31.1	0.2066	145.2	0.0108
38.8	0.1589	180.9	0.0039
48.4	0.1063	225.4	0.0011
60.3	0.0793	280.9	0.0005
75.1	0.0603	350.0	0.0004
93.5	0.0386		

Die Anzahl Punkte ist so gewählt, dass die Approximation an die zugrunde liegende A-priori-Verteilung befriedigend ist; die Punkte sind so gewählt, dass sie im logarithmischen Massstab äquidistant sind. Die Verteilungsfunktion ist im Anhang 2 graphisch dargestellt. Durch Auswerten der vorhandenen A-priori-Information, die im Risikoprofil der Gesellschaft und in der Schadengradverteilung enthalten ist, haben wir nun einen Anhaltspunkt über den Verlauf der Schadenverteilungsfunktion im Bereich zwischen NOK 100 Mio und NOK 350 Mio, eine

Information, die wir aus der reinen statistischen Auswertung der Schadendaten nicht erhalten können, da der grösste «as if»-Schaden NOK 100 Mio nicht wesentlich übersteigt. Es stellt sich nun die Frage der Kombination zwischen der A-priori-Information und der Schadenstatistik. Eine sehr einfache Vorgehensweise besteht darin, dass man von der an die empirische Verteilung angepassten Pareto-Verteilung ($\alpha = 2.219$) und von der an die A-priori-Verteilung angepassten Pareto-Verteilung ($\alpha = 1.825$) ausgeht und als A-posteriori-Verteilung eine Pareto-Verteilung wählt mit einem Parameter, der ein gewichtetes Mittel zwischen dem A-priori und dem empirischen Parameter ist. Die Wahl der Gewichte ist dabei willkürlich und bringt zum Ausdruck, wie stark man die A-priori und die empirische Information relativ zueinander gewichtet. Dieses Vorgehen lässt sich auch auf die angepassten Loggamma-Verteilungen anwenden.

Geht man vom Modell aus, das im Abschnitt 2 präsentiert wurde (Pareto-Likelihood- und Gamma-A-priori-Verteilung), so drängt sich ein anderes Verfahren auf. Im Rahmen dieses Modells ist die A-priori-Verteilung

$$F_0(x) = \int_0^{\infty} P(X \leq x \mid \alpha) f(\alpha) d\alpha$$

$$F_0(x) = 1 - \left(\frac{c}{c + \log\left(\frac{x}{x_0}\right)} \right)^{\gamma}$$

und die A-posteriori-Verteilungsfunktion ist

$$F(x \mid \underline{x}) = P(X \leq x \mid \underline{x}) = \int_0^{\infty} P(X \leq x \mid \alpha) f(\alpha \mid \underline{x}) d\alpha$$

$$F(x \mid \underline{x}) = 1 - \left(\frac{c + T}{c + T + \log\left(\frac{x}{x_0}\right)} \right)^{\gamma+n}$$

wobei

$$T = \sum_{i=1}^n \log\left(\frac{x_i}{x_0}\right) \quad (x_0 = 22.4)$$

Als Parameter der Verteilung von α wählen wir

$$\gamma = 30 \quad \text{und} \quad c = 16.$$

Sie entsprechen den folgenden Annahmen über die Momente der Verteilung von α

$$E[\alpha] = \frac{\gamma}{c} = 1.88 \quad \text{und} \quad \text{Var}(\alpha) = \frac{\gamma}{c^2} = 0.12.$$

Andererseits gilt

$$T = \sum_{i=1}^n \log\left(\frac{x_i}{x_0}\right) = 7.36 \quad (x_0 = 22.4, \quad n = 17)$$

Die A-posteriori-Verteilung ist somit

$$F(x | \underline{x}) = 1 - \left(\frac{23.36}{23.36 + \log\left(\frac{x}{22.4}\right)} \right)^{47}$$

Wir entscheiden uns für ein drittes Verfahren, ein Credibility-gewichtetes Mittel zwischen der A-priori-Verteilung $F_0(x)$, so wie sie aufgrund des Risikoprofils und der Schadensgradverteilung ermittelt wurde, und der empirischen Verteilung $F_n(x)$.

$$F(x) = z \cdot F_n(x) + (1 - z) \cdot F_0(x)$$

Dabei gilt für den Credibility-Faktor

$$z = z(x) = \frac{n}{n + N(x)}$$

mit

$$N(x) = \frac{E_{\alpha}[P(x | \alpha)(1 - P(x | \alpha))]}{\text{Var}_{\alpha}[P(x | \alpha)]}$$

Für eine Herleitung dieser Ergebnisse siehe *Jewell* (1974).

Im Rahmen des statistischen Modells aus Abschnitt 2 erhält man

$$N(x) = \frac{\left[\frac{c}{c + \log\left(\frac{x_0}{x}\right)} \right]^{\gamma} - \left[\frac{c}{c + 2 \cdot \log\left(\frac{x_0}{x}\right)} \right]^{\gamma}}{\left[\frac{c}{c + 2 \cdot \log\left(\frac{x_0}{x}\right)} \right]^{\gamma} - \left[\frac{c}{c + \log\left(\frac{x_0}{x}\right)} \right]^{\gamma}}$$

wobei wir für c und γ die oben ermittelten Werte einsetzen ($\gamma = 30$, $c = 16$). Eine graphische Darstellung der Zeitkonstante $N(x)$ ist im Anhang 3 gegeben. Man sieht, dass die Funktion $N(x)$ für grosse Werte von x eine wachsende

Funktion ihres Argumentes ist, somit ist der Credibility-Faktor für grosse Werte von x «klein». Für solche Werte stützt man sich hauptsächlich auf die A-priori-Verteilung ab. Es ist zu bemerken, dass im vorhergehenden Verfahren die Modellannahmen nur einen Einfluss auf den Credibility-Faktor $z(x)$ haben; $F_0(x)$ und $F_n(x)$ sind unabhängig von den Modellannahmen. Das Verfahren ist somit robust gegenüber Abweichungen von den Modellannahmen.

Wir geben eine tabellarische Darstellung der A-posteriori-Verteilung $F(x)$ an den Sprungstellen an (vgl. Tabelle auf nächster Seite).

Für praktische Zwecke nehmen wir an, dass die Verteilungsfunktion zwischen je 2 Sprungstellen konstant ist. Das entspricht nicht genau der Credibility-Formel, da $z(x)$ und somit $F(x)$ auch zwischen zwei Sprungstellen variiert; die Annahme hat aber den Vorteil, dass $F(x)$ monoton wachsend ist, was sonst im allgemeinen nicht der Fall ist.

4. Wahl einer analytischen Verteilungsfunktion

Die folgenden zwei Ansätze stammen von *G. Benktander* (1970). Sie entsprechen dem Wunsch, ein in bezug auf seine Grossschadenträchtigkeit geordnetes System von Verteilungsfunktionen zu definieren. Beide Ansätze liefern als Nebenprodukt je ein Hilfsmittel zur Beurteilung der Anpassung von analytischen Verteilungen an empirische Verteilungen. Der erste Ansatz besteht in der Modellierung der mittleren Exzessschadenfunktion.

$$m(x) = E[X - x \mid X > x].$$

Es ist leicht einzusehen, dass die Verteilungsfunktion durch die mittlere Exzessschadenfunktion charakterisiert ist. Aus der Definition des mittleren Exzessschadens

$$m(x) = \frac{\int_x^\infty (y - x) dF(y)}{1 - F(x)}$$

folgt nämlich

$$F(x) = 1 - k_0 \cdot e^{-\int_{x_0}^x \frac{1+m'(y)}{m(y)} dy}$$

x	$F(x)$	$z(x)$	$F_n(x)$	$F_0(x)$
22.5	0.0000	0.0047	0.0000	0.0000
22.7	0.0008	0.0138	0.0588	0.0000
23.2	0.0041	0.0348	0.1176	0.0000
24.1	0.0119	0.0675	0.1765	0.0000
25.0	0.2982	0.0948	0.1765	0.3110
25.6	0.3062	0.1106	0.2353	0.3110
25.9	0.3090	0.1179	0.2941	0.3110
26.9	0.3169	0.1396	0.3529	0.3110
28.0	0.3271	0.1596	0.4118	0.3110
29.2	0.3394	0.1778	0.4706	0.3110
31.1	0.5200	0.2005	0.5294	0.5176
34.1	0.5336	0.2258	0.5882	0.5176
35.0	0.5476	0.2316	0.6471	0.5176
36.3	0.5626	0.2388	0.7059	0.5176
37.8	0.5783	0.2456	0.7647	0.5176
38.8	0.6985	0.2495	0.7647	0.6765
42.7	0.7148	0.2604	0.8235	0.6765
48.4	0.7937	0.2683	0.8235	0.7828
53.5	0.8097	0.2705	0.8824	0.7828
60.3	0.8676	0.2695	0.8824	0.8621
62.0	0.8834	0.2688	0.9412	0.8621
75.1	0.9273	0.2602	0.9412	0.9224
93.5	0.9562	0.2443	0.9412	0.9610
105.9	0.9701	0.2334	1.0000	0.9610
116.6	0.9870	0.2245	1.0000	0.9833
145.2	0.9953	0.2029	1.0000	0.9941
180.9	0.9984	0.1808	1.0000	0.9980
225.4	0.9992	0.1592	1.0000	0.9991
280.9	0.9997	0.1389	1.0000	0.9996
350.0	1.0000	0.1202	1.0000	1.0000

Die praktische Bedeutung des mittleren Exzessschadens ist offensichtlich: $m(x)$ ist die erwartete Schadenhöhe im unlimitierten Layer über der Priorität x ; falls $f(x)$ die Schadenfrequenz über der Priorität x bezeichnet, so ist $f(x) \cdot m(x)$ der erwartete Schadenaufwand im unlimitierten Layer über x .

Benktander geht von der Feststellung aus, dass die in der Praxis auftretenden Exzessschadenfunktionen wachsende Funktionen von x sind und dass ihre Zuwachsschnelligkeit eine abnehmende Funktion von x ist. Ein analytisches Modell muss somit die folgenden Bedingungen erfüllen:

$$m'(x) \geq 0, \quad m''(x) \leq 0$$

Die Pareto- und die Exponential-Verteilung sind zwei Grenzfälle unter den zulässigen Verteilungen

$$m(x) = ax \quad (\text{Pareto-Verteilung})$$

$$m(x) = b \quad (\text{Exponential-Verteilung})$$

Bei der Modellierung von Grossschadenverteilungen für praktische Zwecke ist die Exponentialverteilung i.d.R. zu wenig grossschadenträchtig; die Pareto-Verteilung hingegen ist i.d.R. zu grossschadenträchtig. Zur Modellierung von realistischen Grossschadenverteilungen schlägt *Benktander* eine Schar von Verteilungsfunktionen vor, die zwischen diesen beiden Extremen liegen. Die mittlere Exzessschadenfunktion dieser Schar ist von der Form

$$m(x) = \frac{x^b}{a} \quad 0 \leq b \leq 1, \quad a > 0$$

Somit sind die obigen Bedingungen erfüllt. Die Pareto- und die Exponential-Verteilung sind Spezialfälle des Modells ($b = 1$ bzw. $b = 0$).

Für $b \neq 1, 2$ erhält man eine Verteilungsfunktion der Form

$$F(x) = 1 - k \cdot x^{-b} \cdot e^{-\frac{a}{1-b}x^{1-b}}$$

Diese Verteilungsfunktion ist in der versicherungsmathematischen Literatur unter dem Namen «Benktander II» bekannt. Asymptotisch verhält sie sich wie eine Weibull-Verteilung. Umgekehrt gilt für die mittlere Exzessschadenfunktion einer Weibull-Verteilung

$$m(x) \simeq \frac{x^b}{a} \quad (x \rightarrow \infty)$$

Für einen Beweis siehe *J. Beirland* und *J. Teugels* (1992). Somit kann die Weibull-Verteilung stellvertretend für die Benktander II-Verteilung verwendet werden. Die empirische mittlere Exzessschadenfunktion

$$\widehat{m}(x) = \frac{1}{\#\{i \mid x_i > x\}} \sum_{i=1}^n (x_i - x)^+$$

ist ein wertvolles Hilfsmittel bei der Wahl einer analytischen Verteilungsfunktion. Ist sie konstant, so liegt eine Exponentialverteilung vor; wächst sie linear, so liegt eine Pareto-Verteilung vor. Verläuft $\widehat{m}(x)$ wie x^b , so liegt eine Benktander II-Verteilung vor; diese wiederum kann für praktische Zwecke durch eine Weibull-Verteilung ersetzt werden. In Anhang 4 ist $\widehat{m}(x)$ für das Beispiel aus Abschnitt 3 graphisch dargestellt. Die empirische mittlere Exzessschadenfunktion verläuft ungefähr wie eine Gerade, dies ist mit der Pareto-Annahme verträglich. Der zweite Ansatz, der auch von *Benktander* (1970) stammt, besteht in der Modellierung der Sterbeintensität des Schadens

$$\lambda(x) = \frac{f(x)}{1 - F(x)}$$

Dabei bezeichnet $f(x)$ die Dichte der Verteilungsfunktion. Die Sterbeintensität charakterisiert die Verteilungsfunktion

$$F(x) = 1 - k_0 e^{-\int_{x_0}^x \lambda(y) dy}$$

Die Interpretation der Sterbeintensität ergibt sich aus der folgenden Beziehung:

$$\lambda(x)dx = P(X \in (x, x + dx) \mid X > x)$$

Die Bedeutung der obigen Wahrscheinlichkeit ist offensichtlich, falls X eine Lebensdauer ist. Falls X eine Schadenhöhe ist, ist diese Bedeutung weniger offensichtlich. Es kann jedoch festgehalten werden: je stärker $\lambda(x)$ abnimmt, desto grosschadenträchtiger ist die Verteilungsfunktion. *Benktander* postuliert nun eine Sterbeintensität der folgenden Form:

$$\lambda(x) = k \cdot x^p \quad \text{mit } p \in [-1, 0]$$

Für $p = 0$ erhält man die Exponential-Verteilung, für $p = -1$ die Pareto-Verteilung und für $p \in (-1, 0)$ eine Weibull-Verteilung. Der zweite Ansatz liefert also dasselbe System von Verteilungen wie der erste.

Die kumulative Sterbeintensitätsfunktion

$$\Lambda(x) = \int_{x_0}^x \lambda(y) dy$$

kann ähnlich wie die mittlere Exzessschadenfunktion zur Diskriminierung zwischen analytischen Verteilungsfunktionen verwendet werden. Unter der obigen Annahme über die Form von $\lambda(x)$ gilt nämlich

$$\Lambda(x) = \begin{cases} a \log(x) + b & p = -1 \\ ax^{p+1} + b & p \in (1, 0] \end{cases}$$

Der Verlauf der empirischen kumulativen Sterbeintensitätsfunktion

$$\Lambda_n(x) = -\log(1 - F_n(x))$$

ist somit eine Hilfsmittel bei der Bestimmung der zugrundeliegenden Verteilungsfunktion.

In Anhang 5 ist die Kurve $(\log(x), \Lambda(x))$ für das Beispiel aus Abschnitt 3 graphisch dargestellt. Sie verläuft ungefähr wie eine Gerade, dies bestätigt die Pareto-Annahme.

In den Fällen, wo die Schadenstatistik umfangreich genug ist, (wie etwa im Beispiel aus Abschnitt 3), können die mittlere Exzessschadenfunktion und die kumulative Sterbeintensität zur Modellidentifikation verwendet werden. In der Regel ist aber die Anzahl Schäden eines Portefeuilles nicht gross genug, um eine solche Analyse zu erlauben. Sofern man jedoch die mittlere Exzessschadenfunktion oder die kumulative Sterbeintensität einer bestimmten Sparte in einem bestimmten Land kennt, kann man aufgrund der obigen Überlegungen die Schadenverteilungsfunktion bestimmen. Zu lösen bleibt dann das einfachere Problem der Schätzung der Parameter einer bekannten Verteilungsfunktion.

In gewissen Fällen kann die Modellierung des Schaden verursachenden Prozesses zur Bestimmung der Schadenverteilungsfunktion benützt werden. Ein Beispiel eines solchen Vorgehens liefert *D. Shpilberg* (1977). Die Dauer eines Feuers ist eine Zufallsgrösse T , deren Verteilungsfunktion durch die Sterbeintensität $\lambda(t)$ charakterisiert ist:

$$P(T \leq t) = 1 - e^{-\int_0^t \lambda(s) ds}$$

Der durch das Feuer verursachte Schaden X ist eine exponentielle Funktion der Dauer des Feuers

$$X = x_0 e^{k_0 T}$$

Die Verteilungsfunktion von X ergibt sich somit aus der Wahl von $\lambda(t)$. *Sphilberg* betrachtet folgende Spezialfälle:

- (1) Falls $\lambda(t) = \alpha$, so ist T exponentialverteilt und X paretoverteilt

$$P(X \leq x) = 1 - \left(\frac{x_0}{x}\right)^{\frac{\alpha}{k}}, \quad x \geq x_0$$

- (2) Falls $\lambda(t) = \alpha + \beta t$, so gilt

$$P(X \leq x) = 1 + \left(\frac{x_0}{x}\right)^{\left(\frac{\alpha}{k} + \frac{1}{2} \frac{\beta}{k^2} \log\left(\frac{x}{x_0}\right)\right)}, \quad x \geq x_0$$

- (3) Falls $\lambda(t) = e^{\alpha t}$, so ist T Weibull-verteilt

$$P(X \leq x) = 1 - e^{-\frac{1}{\alpha} \left(\frac{x}{x_0}\right)^{\frac{\alpha}{k}}}, \quad x \geq x_0$$

Man erhält somit wiederum die Pareto- und die Weibull-Verteilung als Spezialfälle.

5. Abschliessende Bemerkung

Die Bestimmung von Grossschadenverteilungen ist in verschiedenen praktischen Situationen (Tarifierung von Schadenexzedentendeckungen, Ermittlung von PMLs) von eminenter Bedeutung. In praktischen Anwendungen ist die richtige Interpretation und entsprechende Korrektur der Schäden aus der Vergangenheit der erste und oft wichtigste Schritt in der Datenanalyse. Da es andererseits in der Natur der Sache liegt, dass Grossschäden selten eintreten, ist i.d.R. auch nach Korrektur der Daten das statistische Material äusserst spärlich, und die Bestimmung der Verteilung der Grossschäden kann nicht aufgrund der Daten allein erfolgen, sondern es muss noch die vorhandene A-priori-Information berücksichtigt werden.

Die Art und Weise, wie diese A-priori-Information ausgewertet wird, hängt vom jeweiligen praktischen Problem ab. In der Feuerversicherung lässt sich aufgrund

der Verteilung der Versicherungssummen und der Schadensgradverteilung die A-priori-Verteilung der Schäden ermitteln, wie in Abschnitt 3 gezeigt wurde. Ein ähnliches Verfahren lässt sich etwa in der allgemeinen Haftpflichtversicherung anwenden. In der Kraftfahrzeughaftpflichtversicherung ist in den meisten europäischen Ländern die überwiegende Mehrzahl der Deckungssummen unbeschränkt, somit sind auch die einzelnen Portefeuilles in einem gegebenen Land austauschbar, und die individuellen Schadensätze lassen sich mit Hilfe der Credibility-Theorie miteinander kombinieren (siehe *Bühlmann* und *Straub* [1970]). Das Vorgehen bei der Wirbelsturmversicherung in den USA wurde unter Abschnitt 1 kurz besprochen. Typischerweise wird die Erdbebenversicherung ähnlich behandelt. Sowohl die PML-Bestimmung wie auch die Tarifierung von Schadenexzedentendeckungen erfolgt aufgrund der geographischen Verteilung der versicherten Werte, der A-priori-Kenntnissen über die Wiederkehrperiode und Intensität von Erdbeben an bestimmten Orten und des bekannten Zusammenhanges zwischen Erdbebenintensität und Schadensgrad. Im Fall von Erdbebenversicherung ist die Schadenerfahrung so gut wie wertlos, da wegen des seltenen Eintretens von Erdbeben die letzten Ereignisse i.d.R. zu weit zurückliegen und somit für Versicherungszwecke nicht aussagekräftig sind.

Die Erfahrungen des Tarifierers in einer bestimmten Sparte, in einem bestimmten Land fließen bei der Wahl einer Schadenverteilungsfunktion ein; die mittlere Exzessschadenfunktion und die kumulative Sterbeintensität bilden dabei wichtige Hilfsmittel. Auch bei der Schätzung von Verteilungsparametern werden die A-priori-Kenntnisse des Tarifierers berücksichtigt. Ein Beispiel wurde in Abschnitt 2 gegeben.

6. Benutzte Software

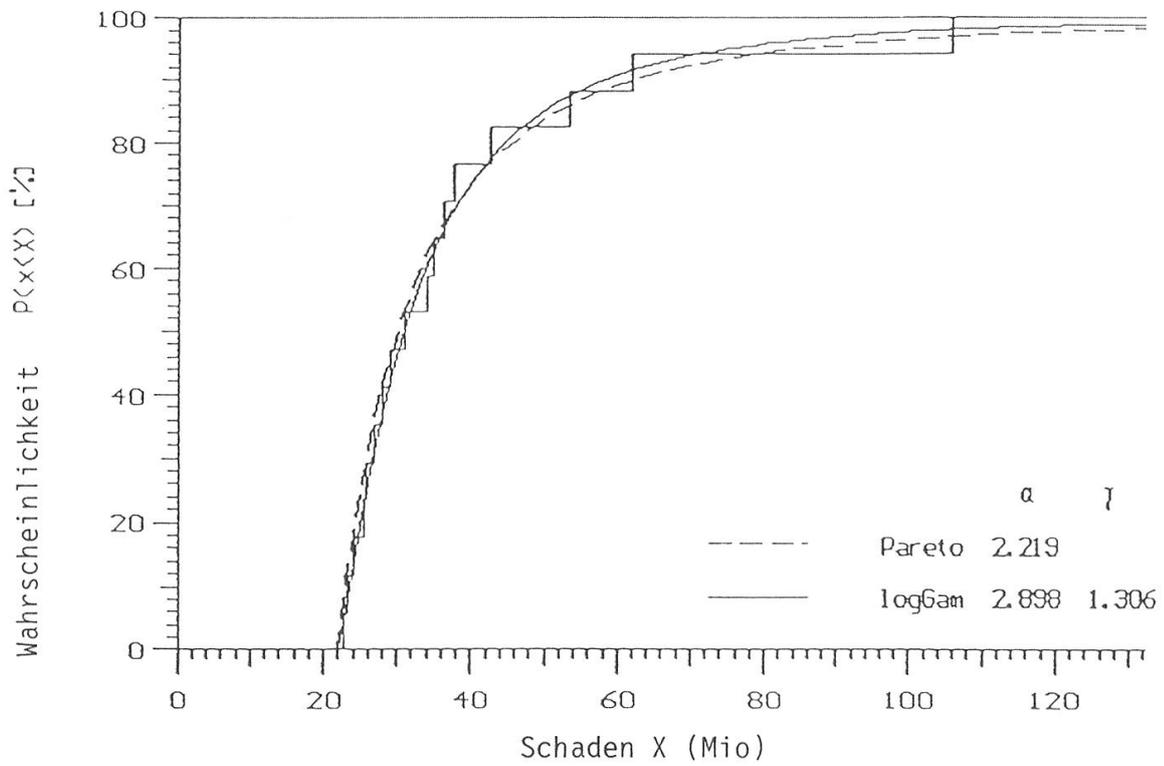
Die Graphiken von Anhang 3, 4 und 5 wurden mit einem Standard Software Paket erstellt. Die Graphiken von Anhang 1 und 2 wurden mit einem spezialisierten Tarifierungssoftware-Produkt erstellt, das von *S. Bernegger* entwickelt wurde.

René Schnieper
Winterthur-Versicherungen
Postfach
8401 Winterthur

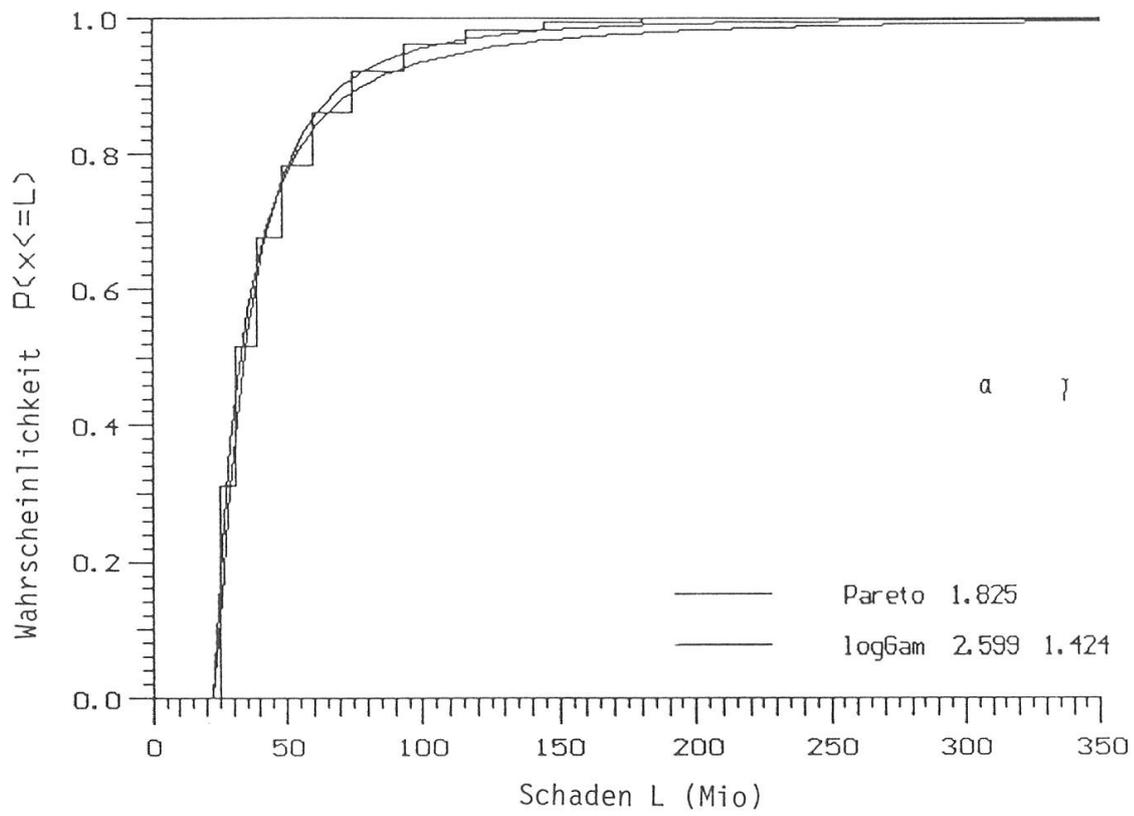
Literatur

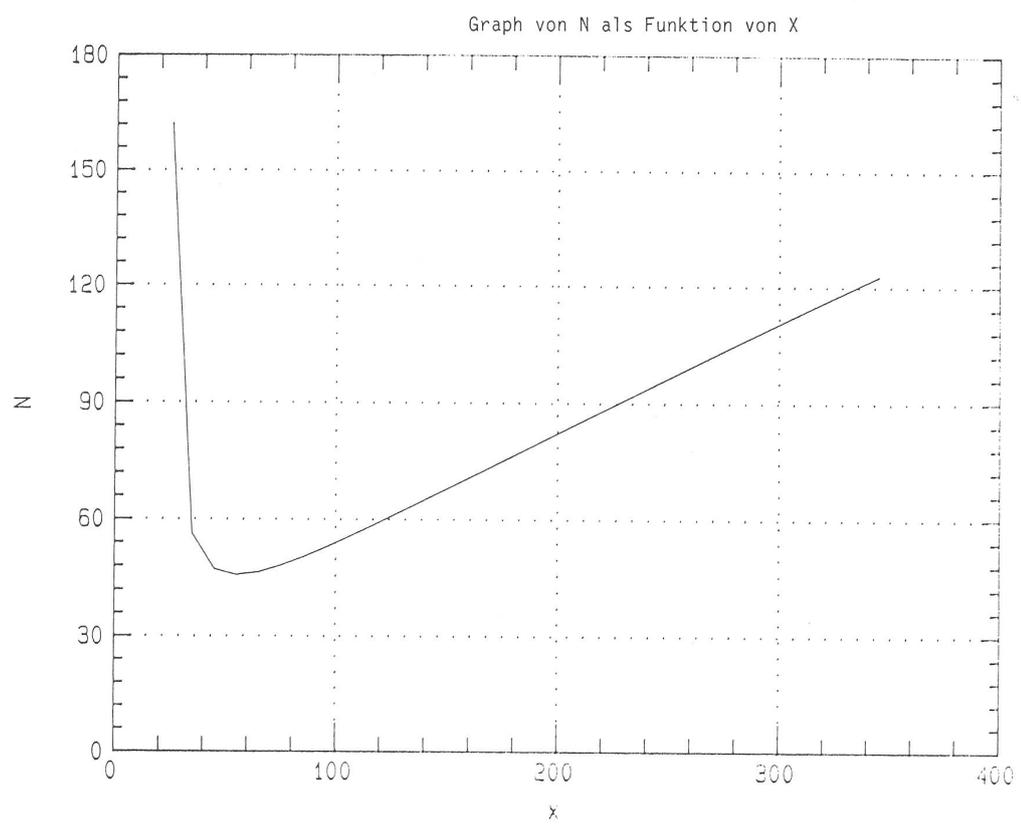
- Beirland J., Teugels J.L.:* Modeling Large Claims in Non-Life Insurance. Insurance: Mathematics and Economics 11 (1992).
- Benktander G.:* Schadenverteilung nach Grösse in der Nicht-Leben-Versicherung. MVSVM, Heft 2, 1970.
- Bühlmann H., Straub E.:* Glaubwürdigkeit für Schadensätze. MVSVM, Heft 1, 1970.
- Clark K.M.:* A Formal Approach to Catastrophe Risk Assessment and Management. PCAS, Vol 73, Part 2, 1986.
- Hogg R.V., Klugmann S.A.:* Loss Distributions. Wiley, 1984.
- Jewell W.S.:* The Credible Distribution. ASTIN-Bulletin 7:3, 1974.
- Shpilberg D.C.:* The Probability Distribution of Fire Loss Amount. The Journal of Risk and Insurance, Vol. XIIV, No. 1, 1977.

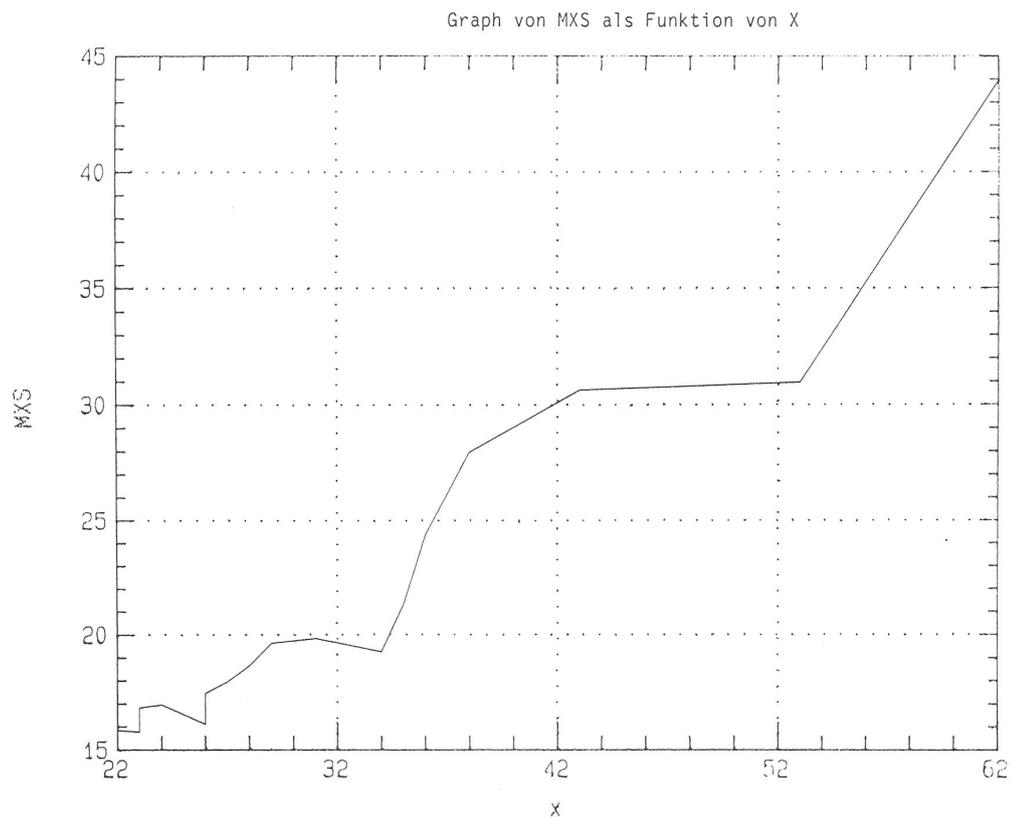
Anhang 1

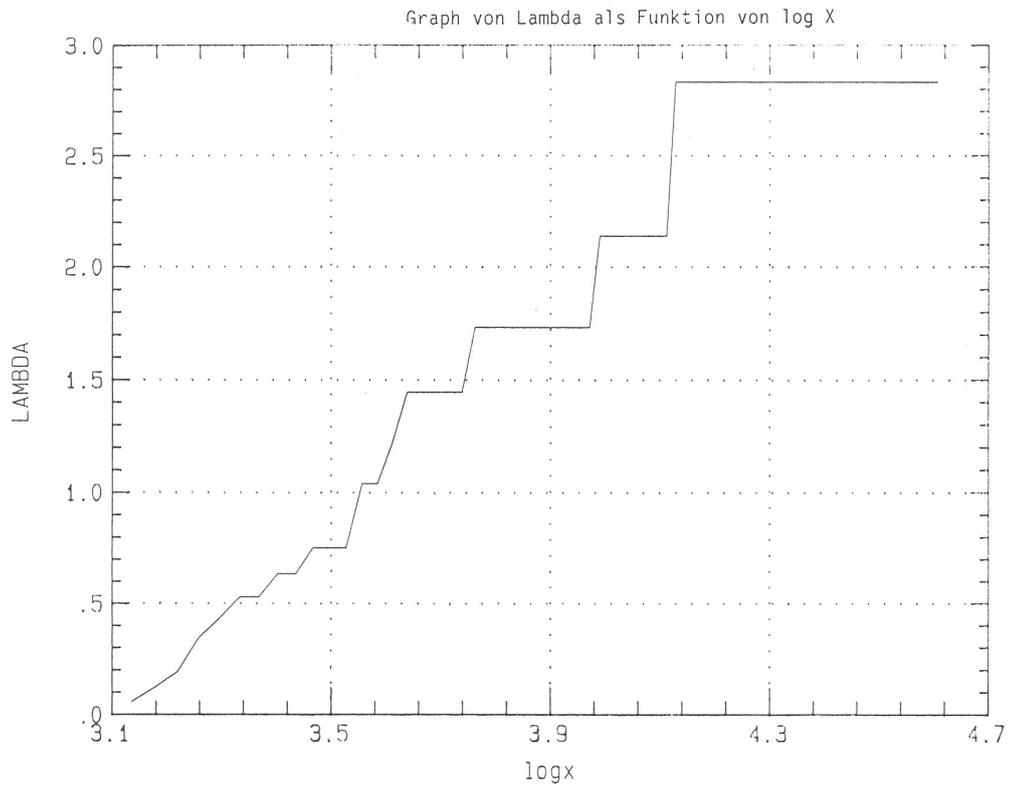


Anhang 2









Zusammenfassung

Anhand von Beispielen aus der Praxis wird gezeigt, dass es bei der Bestimmung von Grossschadenverteilungen sehr auf die richtige Interpretation der Schäden aus der Vergangenheit ankommt. Da Grossschäden selten eintreten, ist das Datenmaterial i.d.R. spärlich. Eine klassische statistische Analyse führt somit zu Schätzungen, die mit grossen Unsicherheiten behaftet sind. Es werden verschiedene Bayes'sche Verfahren präsentiert, die es erlauben, die Daten mit der Information über das Kollektiv zu kombinieren. Darüber hinaus werden verschiedene Ansätze zur Wahl eines Systems von Verteilungsfunktionen bei der Modellierung von Grossschäden präsentiert. Nebst der mittleren Exzessschadensfunktion erweist sich die in der Risikotheorie weniger bekannte kumulative Sterbeintensität als gutes Hilfsmittel zur Wahl zwischen verschiedenen Grossschaden-Verteilungsfunktionen.

Summary

Based on practical examples it is shown that a correct interpretation of past claims data is essential to determine the distribution of large claims. Since large claims do not occur often there is generally not much data available. A classical statistical analysis therefore, leads to estimates which are unprecise. Different bayesian procedures are presented, which allow a combination of data with collateral information. In addition we present different systems of distributions for the modelisation of large claims. The mean excess claim function and the cumulative death intensity function which is less well known in risk theory are seen to be helpful tools to choose between different distribution functions for large claims.

Résumé

Sur la base d'exemples pratiques il est démontré qu'une interprétation correcte des sinistres du passé est essentielle pour déterminer la distribution des grands sinistres. Les grands sinistres étant peu fréquents, il n'y a généralement que peu de données disponibles. Pour cette raison une analyse statistique classique produit des estimations imprécises. Différents procédés bayesiens sont présentés qui permettent de combiner les données avec les informations sur le collectif. De plus, différents systèmes de distributions pour la modélisation des grands sinistres sont présentés. La fonction du sinistre d'excédent espéré ainsi que la fonction de l'intensité cumulative de décès, moins connue en théorie de risque, apparaissent comme des instruments, utiles pour le choix d'une fonction de distribution des grands sinistres.

