FFT versus Panjer Rekursion für zusammengesetzte Verteilungen

Jens Röhrig

28.06.2010

1 Einleitung

Dieses Handout dient dem Zweck, den Inhalten des gleichnamigen Seminarvortrags besser folgen zu können.

2 Grundlegendes zur Risikotheorie

Wir werden zunächst auf das sogenannte "Kollektive Modell" der Risikotheorie eingehen, um den Vortrag anwendungsbezogen zu halten und die darauf folgenden mathematischen Erwägungen zu Begründen.
Das "Individuelle Modell" ist ein sehr intuitives Modell, es betrachtet ein Risiko in Form eines abgeschlossenen Versicherungsvertrages als eine nichtnegative Zufallsvariable X. Nimmt diese einen Wert an, ist ein Schaden entstanden. Um alle Policen (Versicherungsverträge) abbilden zu können werden diese in einem Portfolio $P = \{X_k : k = 1, ..., n\}$ zusammengefasst. Dabei ist n die Anzahl der abgeschlossenen Policen. Betrachtet wird nun der Gesamtschaden als die „zusammengesetzte Summenvariable" $S_n = \sum_{k=1}^{n} X_k$. Von praktischen Interesse, zum Beispiel zur Berechnung der der nötigen Rückstellungen einer Versicherung für das Portfolio ist nun die Verteilung von S_n, sowie deren Mittelwert und Varianz.
Das "Kollektive Modell" läßt nun n nicht mehr fest, sondern betrachtet auch die Anzahl der eingetretenen Schäden als zufällig und ersetzt es durch die Zufallsvariable N. Somit muss sich auch die Interpretation der Risiken X_k ändern. Nimmt X_k nun einen Wert an, wird das nun nicht mehr als in Anspruchnahme der k-ten Police gewertet, sondern als der k-te wirklich aufgetretene Schaden.

Zusammenfassend:

X_k Risiko mit Schwereverteilung F_k
N Schadenzahl mit Häufigkeitverteilung Q
S_N Gesamtschaden mit Zusammengesetzter Verteilung G

Voraussetzungen:
Die Schadenzahl N sein auf dem W-Raum $(\aleph_0, \wp(\aleph_0))$ definiert, somit müssen wir, um Definitionsschwierigkeiten zu vermeiden von einem unendlichen Portfolio $P = \{X_k : k \in \aleph\}$ ausgehen. Alle Risiken X_k seien auf demselben W-Raum $(\Omega, \Im, P)$ definiert und gleich verteilt, setze $F := F_1$. Desweiteren seien $N, X_1, X_2, ...$ stochastisch unabhängig.

Satz 1 Verteilungs der Gesamtschadens
Der Gesamtschaden S_N (zusammengesetzte Summenvariable) eines oben beschriebenen Portfolios P hat die folgende Verteilungsfunktion

$$G = \sum_{k=0}^{\infty} P(N = k) F^{*k} \tag{1}$$

Satz 2 Fouriertransformierte des Gesamtschadens
Es sei P ein oben beschriebenes Portfolio, dann ist die Fouriertransformierte des Gesamtschadens:

$$\Phi_{S_N}(t) = \Theta_N(\Phi_{X_1}(t)) \quad \forall t \geq 0 \tag{2}$$

Mit $\Theta_N := E(t^N)$ momenterzeugende Funktion von N

ausserdem gilt:

$$EW(S_N) = EW(N)EW(X_1)$$

und

$$VAR(S_N) = VAR(N)(EW(X_1))^2 + EW(N)VAR(X_1)$$

3 Numerische Ansätze

Wir werden nun drei Methoden diskutieren, wie man mittels Formeln 1 bzw 2 die Verteilung G des Gesamtschadens S_N bestimmen kann.

3.1 Ein "naiver" Ansatz

Die Idee dieses Ansatzes ist es, die Verteilung des Gesamtschadens mit Formel 1 zu berechnen, indem man die Schwereverteilung diskretisiert und dadurch die k-te Faltungspotenz rekursiv berechnen kann.
Die Schwereverteilung F soll auf dem Gitter $h \cdot \aleph_0 = \{0, h, 2h, 3h, ...\}$ diskretisiert werden mit Dichte $h > 0$. Das erreicht man, indem F durch die diskrete Verteilung $F^h = \{f_{h,j}\}_{j \in \aleph_0}$ ersetzt wird. Für die Wahl von f gibt es verschiedene Ansätze, hier zwei Beispiele:

Möglichkeiten für die Diskretisierungsfunktion:

Rundungs-Methode

$f_{h,j} = F(jh + \frac{h}{2}) - F(jh - \frac{h}{2})$

Einschliessungs-Methode

$f^+_{h,j} = F((j+1)h) - F(jh)$
$f^-_{h,j} = F(jh) - F((j-1)h)$

Es ist klar, dass wenn man das Gitter immer enger macht ($h \to 0$), die diskrete Schwereverteilung F^h schwach gegen F und die diskrete Verteilung des Gesamtschadens G^h schwach gegen G konvergieren.

Berechnung der diskreten k-ten Faltungspotenz

$$f_n^{*k} = \begin{cases} 1, & \text{wenn } j = 0 \text{ und } n = 0 \\ 0, & \text{wenn } j = 0 \text{ und } n \in \aleph \\ \sum_{i=0}^{n} f_{n-i}^{*j-1} f_i, & \text{wenn } j > 0 \end{cases}$$

$\Rightarrow P(S_N = n) = \sum_{k=0}^{\infty} P(N = k) f_n^{*k}$ (Formel 1) berechenbar.

3.2 Die Panjer Rekursion

Eine zweite Möglichkeit den Gesamtschaden zu bestimmen ist mittels der Panjer Rekursion. Dazu verwenden wir folgende, weit verbreitete Notation:

$q_k := P(N = k)$ (Häufigkeitsverteilung)
$f_k := P(X_i = k)$ (Schwereverteilung)
$g_k := P(S_N = k)$ (Verteilung des Gesamtschadens)

Die Panjer Rekursion ist allerdings nur bei ganz bestimmten Schadenszahlen N anwendbar. Häufigkeitsverteilungen, für die die folgende Rekursion gilt sind in der Panjer Klasse.

Satz 3 Anwendbarkeit der Panjer Rekursion
Für eine Häufigkeitsverteilung N gilt:

N ist entweder poissonverteilt oder binomialverteilt oder negativbinomialverteilt oder P(N=0)=1.
$\Leftrightarrow$
$\exists a, b \in \Re$ s.d $q_k = (a + \frac{b}{k})q_{k-1}$ für $k \geq 1$

(benutze den Startwert $p_0 = \Theta_N(0)$)

Satz 4 Der Panjer Algorithmus
Falls die Häufigkeitsverteilung in der Panjer Klasse liegt und das Portfolio P nur Werte auf einem Gitter $h * \aleph_0$ hat, gilt zur Berechnung der Verteilung des Gesamtschadens die Rekursion:

$$g_n = \frac{1}{1 - af_0} \sum_{j=1}^{n} (a + \frac{bj}{n}) f_j g_{n-j} \; n \geq 1$$

(mit Startwert $g_0 = \Theta_N(f_0)$)

3.3 Ein FFT basierter Algorithmus

Im folgenden werden wir die Verteilung des Gesamtschadens durch Formel 2 berechnen:

$$\Phi_{S_N}(t) := EW(e^{itS_N}) = \Theta_N(\Phi_{X_1}(t)) \;\; \forall t \geq 0$$

Allerdings ist die explizite Berechnung der Verteilung des Gesamtschadens von S_N mittels Fouriertransformation nur möglich, wenn die Verteilungs Funktion von X_1 sehr einfach aussieht (z.B. stückweise linear). Im Folgenden betarchten wir deshalb wieder das diskretisierte Problem.

Diskrete Fourier Transformation (DFT)
Nehme einen Vektor $f := (f_0, ..., f_{m-1}) \in \Re^m$ mit $m \in \aleph$
Berechne deren diskrete Fouriertransformierte $\widehat{f} = (\widehat{f}_0, ..., \widehat{f}_{m-1})$
diese ist

$$\widehat{f_j} := \sum_{k=0}^{m-1} f_k e^{\frac{i2\pi jk}{m}}$$

umgekehrt ist es möglich die f_j wieder zurück zu gewinnen durch:

$$f_j := \frac{1}{m} \sum_{k=0}^{m-1} \widehat{f}_k e^{\frac{-i2\pi jk}{m}}$$

Grundgerüst des Algorithmus

1. Stelle aus den gegebenen Daten den Vektor $f \in \Re^m$ auf
2. Berechne die DFT $\widehat{f} \in \Re^m$
3. Wende die momenterzeugende Funktion von N darauf an und interpretiere sie als DFT $\Theta_N(\widehat{f}) =: \widehat{g}$
4. Berechne die Inverse DFT um g^m als eine Approximation der Verteilung des Gesamtschadens zu erhalten

Fehleranalyse
Für die Approximation gilt die folgende Fehlerschranke:

$$\sum_{n=0}^{m-1} |g_n^m - g_n| = \sum_{n=m}^{\infty} g_n - \Theta'_N(1) \sum_{n=m}^{\infty} f_n + O((\sum_{n=m}^{\infty} f_n)^2)$$

Somit ist der Gesamtfehler klein, wenn m gross ist.
Der Fehler ist Null, wenn $\sum_{j=0}^{m-1} f_j = 1$
Um den sogenannten aliasing Fehler zu reduzieren. wendet man einen Kipp-Operator (tilting Operator) auf f an bevor man den Algorithmus startet:

$$E_\tau(f) := (e^{-\tau j} f_j)_{j=0,\dots,m-1}$$

dies kann man am Schluß durch Anwendung von $E_{-\tau}(f)$ einfach wieder zurücktransformieren, weil gilt:

$$G(F) = E_{-\tau}(G(E_\tau(F))$$

da der Kippoperator mit der Faltungsoperation verträglich ist.

3.4 Vergleich der Algorithmen

Tabelle 1: Algorithmenvergleich

	Naiver Algorithmus	Panjer Algorithmus	FFT
Lauftzeit	O(n^3)	O(n^2)	O($n \log(n)$)
Komplexität	komplex	leicht	leicht
Fehlerschranke	nein	ja	ja
Erweiterbarkeit	-	ja	-
Anwendbarkeit	universell	beschränkt	universell
Effizienz	niedrig	hoch	sehr hoch

4 Anhang

Implementiertes Beispiel: Quellcode

```
function [ g1,g2,g3 ] = FP( lambda, alpha,beta,h,m,teta )
%Berechnet die diskrete Verteilung des Gesamtschadens bei einer
%poissonverteilten Schadenszahl und paretoverteilten Risiken.
%g1 ist durch FFT bestimmt
%g2 durch FFT mit tilt Operator teta
%g3 durch Panjer Rekursion

hold off

% FFT ohne tilting Parameter
f=pareto(alpha, beta,h,m);
fh=fft(f);
gh=exp(lambda*(fh-ones(1,m)'));
g1=(m^(-1))*ifft(gh);

% FFT mit tilting Parameter
f=pareto(alpha, beta,h,m);
f=exp(-teta*(0:m-1)').*f;
fh=fft(f);
gh=exp(lambda*(fh-ones(1,m)'));
g=(m^(-1))*ifft(gh);
g2=exp(teta*(0:m-1)').*g;

% Panjer Algorithmus a=0 b=lambda
f=pareto(alpha, beta,h,m);
g3=ones(m,1);
g3(1,1)=exp(-lambda);
for n=1:m-1
  g3(n+1,1)=sum(((lambda*(1:n)')./(ones(n,1)*n)).*f(2:n+1,1).*g3(n:-1:1,1)
end
```

Übersicht: Panjer Rekursion

Verteilung	$P[N=k]$	a	b	p_0	$W_N(x)$	$E[N]$	$Var(N)$
Binomial	$\binom{n}{k} p^k (1-p)^{n-k}$	$\frac{p}{p-1}$	$\frac{p(n+1)}{1-p}$	$(1-p)^n$	$(px+(1-p))^n$	np	$np(1-p)$
Poisson	$e^{-\lambda}\frac{\lambda^k}{k!}$	0	λ	$e^{-\lambda}$	$e^{\lambda(x-1)}$	λ	λ
Negative Binomial	$\frac{\Gamma(r+k)}{k!\,\Gamma(r)} p^r (1-p)^k$	$1-p$	$(1-p)(r-1)$	p^r	$\left(\frac{p}{1-x(1-p)}\right)^r$	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$

Quelle:Wikipedia