COMPUTATION OF COMPOUND DISTRIBUTIONS I: ALIASING ERRORS AND EXPONENTIAL TILTING

BY

RUDOLF GRÜBEL and RENATE HERMESMEIER

Universität Hannover

ABSTRACT

Numerical evaluation of compound distributions is one of the central numerical tasks in insurance mathematics. Two widely used techniques are Panjer recursion and transform methods. Many authors have pointed out that aliasing errors imply the need to consider the whole distribution if transform methods are used, a potential drawback especially for heavy-tailed distributions. We investigate the magnitude of aliasing errors and show that this problem can be solved by a suitable change of measure.

KEYWORDS AND PHRASES

Total claim size distribution, ruin probabilities, random sums, Fourier transformation, aliasing, change of measure.

1. INTRODUCTION

The starting point for stochastic modelling in insurance mathematics is the assumption that claims made to an insurance company are of random size and arrive at random time. A widely used stochastic model for this situation stipulates that the claim arrival times constitute a Poisson process of constant rate, that the claim sizes are independent and identically distributed, and that claim sizes are independent of the arrival process. This classical risk model is discussed in many textbooks; see e.g. Beard, Pentikäinen and Personen (1984), Grandell (1991) or Hipp and Michel (1990).

As a consequence of the structural model assumptions the quantities of interest depend on the rate λ of the Poisson arrival process and the distribution μ of the individual claims only. One such quantity is the distribution ν of the total claim amount over a given period of time, say $[0, t]$. If p_k is the probability of exactly k claims in $[0, t]$ and with "$*$"

denoting convolution, ν can be written in the form

$$\nu = \sum_{k=0}^{\infty} p_k \mu^{*k}, \tag{1.1}$$

i.e. ν is a *compound distribution*. In the classical model the number of claims has a Poisson distribution so that

$$p_k = \exp(-\lambda t)\frac{(\lambda t)^k}{k!} \tag{1.2}$$

and ν is a *compound Poisson distribution*. The *compound geometric* case, with $p_k = (1-\rho)^k\rho$ for some $\rho \in (0,1)$, is important in connection with ruin probabilities; see Section 4 below. For general compound distributions $(p_k)_{k\in\mathbb{N}_0}$ can be any sequence of non-negative real numbers with $\sum_{k=0}^{\infty} p_k = 1$. If $X_1, X_2, \ldots$ are independent random variables with distribution μ and if τ is another random variable, independent of the X-variables and with $P(\tau = k) = p_k$ for all $k \in \mathbb{N}_0$, then ν is the distribution of the random sum $S_\tau := X_1 + X_2 + \ldots + X_\tau$; we take this sum to be zero if $\tau = 0$. In the context of insurance modelling, the X-values are the individual claims and S_τ is the total claim amount.

The question of how to calculate ν from μ and $(p_k)_{k\in\mathbb{N}_0}$ has attracted much interest over the last decades and continues to be an active topic of research; see e.g. Embrechts, Grübel and Pitts (1993), where transform methods are explained and reviewed, and Asmussen and Binswanger (1997), who advocate a simulation approach in the case of compound geometric distributions. These methods can both be applied for general compound distributions. In the classical risk model the special cases of compound Poisson and compound geometric distributions are of main interest. For these and μ of lattice type there exists a recursive scheme due to Panjer; see Panjer (1981) or Kapitel 3 in Hipp and Michel (1990). Further methods exist, and several authors have investigated the relative merits and potential drawbacks of the various approaches; see e.g. Bühlmann (1984), Buchwalder, Chevallier and Klüppelberg (1993) and Schröter (1995).

The situation described above is typical for stochastic modelling insofar as cases in which ν can be given explicitly are few and far between, so various approximations have to be made which all entail some inevitable associated error. These errors are random if the Monte Carlo method is employed. Both Panjer recursion and the transform approach depend on an initial discretization of the claim size distribution μ: For a given discretization parameter $h > 0$ let μ_h be the distribution concentrated on $h\mathbb{N}_0 = \{nh : n \in \mathbb{N}_0\}$ with

$$\mu_h(\{nh\}) := \mu(I_{h,n}), \quad I_{h,n} := \left(\left(n-\frac{1}{2}\right)h, \left(n+\frac{1}{2}\right)h\right).$$

In words, intervals of length h are lumped to their respective centres. If h is small enough then the compound distribution ν_h associated with μ_h is close to the compound distribution ν associated with μ (if e.g. "close" refers to the supremum distance of the distribution functions). In a companion paper, Grübel and Hermesmeier (1999), we discuss the discretization error and show that it can often be reduced dramatically with the help of extrapolation techniques; see also Embrechts, Grübel and Pitts (1993) for a "weak" result in this direction. This is of interest for transform *and* recursion methods. In the present paper we ignore the discretization error and assume that the claim size distribution is of lattice type, i.e. concentrated on some hN_0, $h > 0$. A simple scaling argument shows that we may then assume that $h = 1$, so that we are dealing with claim size distributions concentrated on the non-negative integers. Of course, both recursion and transform methods require an additional truncation step, i.e. we replace the infinite set $\mathbb{N}_0$ by $\{0, 1, \ldots, N-1\}$ with some $N \in \mathbb{N}$. Panjer recursion, if it applies and if we ignore errors from floating point representation etc., obtains the exact values $\nu(\{n\})$, $n = 0, \ldots, N-1$, from $\mu(\{n\})$, $n = 0, \ldots, N-1$. Transform methods introduce an additional error, the *aliasing error*, essentially a wrap-around effect due to the replacement of the usual summation of integers by summation modulo the truncation point N, but their operation count grows as $N \log N$ only if the fast Fourier transform (FFT) algorithm is used whereas recursion needs an operation count of order N^2. Therefore, for the comparison of these methods, the order of magnitude of the aliasing error and techniques for its reduction are of theoretical and practical importance. In this connection heavy-tailed claim size distributions are of special interest, see Asmussen and Binswanger (1997) and the recent monograph by Embrechts, Klüppelberg and Mikosch (1997).

In the present paper we obtain a simple general bound for the aliasing error. We further show that the local behaviour of the functional that maps μ to ν can be used to investigate the asymptotic behaviour of this error as the truncation point N tends to infinity. From this analysis it follows that the general bound is asymptotically sharp in cases where the tail of the individual claim size distribution decreases at a faster rate than the tail of the compound distribution. We also discuss the case where these tails are of the same order of magnitude. In our view this approach contributes to the theoretical understanding of transform methods. From a practical point of view, and especially for heavy-tailed claim size distribution, our second finding appears to be of some immediate use: We show that aliasing errors can be eliminated for all practical purposes by a suitable change of measure. This technique is of considerable importance in many areas of probability and statistics and is also known as *exponential tilting* and closely related to the *Esscher transform*.

The paper is organized as follows. In Section 2 we first give a summary of the transform method. This has been done previously by other authors, see e.g. Bühlmann (1984) or Embrechts, Grübel and Pitts (1993), so we keep this brief. Our treatment will be somewhat abstract, which enables us to expose

the simplicity of the underlying ideas once a bit of notation is introduced. We then discuss the aliasing error and explain the exponential change of the measure. Section 2 ends with the description of an algorithm designed to reduce such errors. In Section 3 we look at a particular numerical example, taken from Embrechts, Grübel and Pitts (1993), which was chosen to exhibit the aliasing error in an extreme case. The algorithm introduced in Section 2 is shown to work well. Section 4 explains two related applications, to the calculation of ruin probabilities and the mean function in the Sparre-Andersen model of risk theory. Some concluding remarks are collected in the final section.

2. Results

Distributions μ that are concentrated on $\mathbb{N}_0$ can be described by the sequence $a = (a_n)_{n\in\mathbb{N}_0}$ of their atoms $a_n := \mu(\{n\})$. Such sequences are elements of the space

$$\ell_1 := \left\{ (a_n)_{n\in\mathbb{N}_0} \in \mathbb{C}^{\mathbb{N}_0} : \sum_{n=0}^{\infty} |a_n| < \infty \right\}$$

of absolutely summable complex sequences, which, when endowed with the norm

$$\|\cdot\|_1 : \ell_1 \to \mathbb{R}, \quad \|a\|_1 := \sum_{n=0}^{\infty} |a_n|,$$

becomes a Banach space. Conversely, every $a \in \ell_1$ defines a complex valued finite measure on $\mathbb{N}_0$. We can similarly interpret the elements $a^N = (a_0^N, \ldots, a_{N-1}^N)$ of the standard N-dimensional unitary vector space $\mathbb{C}^N$ as complex-valued measures on the cyclic group $\mathbb{G}_N$ of order N. This group may be identified with $\{0, 1, \ldots, N-1\}$ if the usual arithmetic operations are carried out modulo N. We will use the letters $a, b, c, \ldots$ for elements of ℓ_1 and $a^N, b^N, c^N, \ldots$ for elements of $\mathbb{C}^N$and we write a_n or $(a^N)_n$ for the n^{th} component of a sequence a or a vector a^N; note that indices start at $n = 0$ in both cases. On $\mathbb{C}^N$ we will use the norm $\|a^N\|_1 := \sum_{n=0}^{N-1} |a_n^N|$ (the "city block norm"; it should be clear from the context whether $\|\cdot\|_1$ refers to ℓ_1 or $\mathbb{C}^N$).

Both $\mathbb{N}_0$ and $\mathbb{G}_N$ have an additive structure, which leads to the notion of convolution for measures on these sets. The convolution product $c = a * b$ of two sequences $a = (a_n)_{n\in\mathbb{N}_0}$, $b = (b_n)_{n\in\mathbb{N}_0} \in \ell_1$ is defined by

$$c_n := \sum_{m=0}^{n} a_m b_{n-m} \quad \text{for all} \quad n \in \mathbb{N}_0.$$

For elements a^N, b^N of $\mathbb{C}^N$ we use the same symbol "$*$" and put

$$(a^N * b^N)_n := \sum_{m=0}^{N-1} a_m^N b_{n-m}^N \quad \text{for all} \quad n = 0, 1, \ldots, N-1,$$

where the b^N-index is obtained on subtracting in $\mathbb{G}_N$, i.e. modulo N. The *norm inequality* $\|a * b\|_1 \leq \|a\|_1 \|b\|_1$ holds for all $a, b \in \ell_1$, and similarly $\|a^N * b^N\|_1 \leq \|a^N\|_1 \|b^N\|_1$ for all $a^N, b^N \in \mathbb{C}^N$; we even have $\|a * b\|_1 = \|a\|_1 \|b\|_1$ if all entries of a and b are non-negative. We write a^{*k} for the k^{th} convolution power of $a \in \ell_1$ and use the convention $a^{*0} = \delta_0$, where $\delta_{00} = 1$ and $\delta_{0n} = 0$ for $n > 0$. The sequence δ_0 is the unit element with respect to convolution; the corresponding definitions for $\mathbb{C}^N$ should be obvious.

In this formal framework the relationship between the distribution of individual claims and the distribution of the total claim amount is given by a nonlinear operator (functional) Ψ,

$$\Psi : \{a \in \ell_1 : \|a\|_1 \leq 1\} \to \ell_1, \quad \Psi(a) := \sum_{k=0}^{\infty} p_k a^{*k}.$$

Here and in the following we regard the sequence $(p_k)_{k \in \mathbb{N}_0}$ as fixed; p_k is the probability of exactly k claims. For some special p-sequences the elements of $b = \Psi(a)$ can be obtained recursively: If e.g. $p_k = e^{-\alpha}\alpha^k/k!$ for all $k \in \mathbb{N}_0$, then

$$b_0 = e^{-\alpha(1-a_0)}, \quad b_n = \frac{\alpha}{n} \sum_{m=1}^{n} m a_m b_{n-m} \quad \text{for all} \quad n \in \mathbb{N}. \tag{2.1}$$

This formula arises in the context of discrete infinite divisibility, see e.g. Johnson, Kotz and Kemp (1992), p. 352. It provides the basis for Panjer's recursive algorithm for the computation of compound Poisson distributions.

Since convolution can also be done in $\mathbb{C}^N$ we have an analogue of Ψ for measures on $\mathbb{G}_N$,

$$\Psi_N : \left\{a^N \in \mathbb{C}^N : \|a^N\|_1 \leq 1\right\} \to \mathbb{C}^N, \quad \Psi_N(a^N) := \sum_{k=0}^{\infty} p_k (a^N)^{*k}.$$

This is the total claim size distribution function if the aggregate claims are readjusted by subtracting a suitable multiple of N whenever the sum overshoots the threshold N.

We now connect the sequence and vector spaces by three bounded linear operators, which represent truncation, zero padding and aliasing respectively:

$$T_N : \ell_1 \to \mathbb{C}^N, \quad (T_N(a))_n := a_n \quad \text{for} \quad n = 0, \ldots, N-1,$$

$$U_N : \mathbb{C}^N \to \ell_1, \quad \left(U_N(a^N)\right)_n := \begin{cases} a_n^N, & \text{for } n = 0, \ldots, N-1, \\ 0, & \text{otherwise,} \end{cases}$$

$$V_N : \ell_1 \to \mathbb{C}^N, \quad \left(V_N(a)\right)_n := \sum_{j=0}^{\infty} a_{n+jN} \quad \text{for } n = 0, \ldots, N-1.$$

With "Id" denoting the identity operator on $\mathbb{C}^N$ we obviously have

$$T_N \circ U_N = V_N \circ U_N = \mathrm{Id}, \tag{2.2}$$

and, due to the rules of addition modulo N,

$$V_N(a * b) = V_N(a) * V_N(b) \quad \text{for all} \quad a, b \in \ell_1$$

(note that "$*$" refers to different spaces on the two sides of this formula). This, together with the continuity of V_N, implies

$$\Psi_N \circ V_N = V_N \circ \Psi. \tag{2.3}$$

To complete our notational round up we require Fourier transformation. For $a \in \ell_1$ the Fourier transform $\hat{a}$ is given by

$$\hat{a} : [0, 2\pi) \to \mathbb{C}, \quad \hat{a}(\theta) := \sum_{n=0}^{\infty} a_n e^{in\theta} \quad \text{for all} \quad \theta \in [0, 2\pi). \tag{2.4}$$

Let $w_N := \exp(2\pi i/N)$ be the canonical N^{th} root of unity. The Fourier transform $(a^N)^\wedge$ of some $a^N \in \mathbb{C}^N$ is given by

$$(a^N)^\wedge \in \mathbb{C}^N, \quad (a^N)_n^\wedge := \sum_{m=0}^{N-1} a_m^N w_N^{m \cdot n} \quad \text{for} \quad n = 0, \ldots, N-1. \tag{2.5}$$

If we define an $N \times N$-matrix $W = (w_{kl})_{k,l=1}^N$ by $w_{kl} := w_N^{(k-1)(l-1)}$ then, in matrix notation, $(a^N)^\wedge = Wa^N$. Writing $\overline{W}$ for the complex conjugate of W we further have $W^{-1} = N^{-1}\overline{W}$ and consequently

$$a_n^N = \frac{1}{N} \sum_{m=0}^{N-1} (a^N)_m^\wedge w_N^{-m \cdot n} \quad \text{for} \quad n = 0, \ldots, N-1, \tag{2.6}$$

i.e. there is a simple inversion formula for Fourier transformation on $\mathbb{G}_N$. Note that Fourier transformation on $\mathbb{G}_N$ is a numerically stable operation: apart from a constant factor both W and W^{-1} are unitary matrices so that blow-up of approximation errors due to rounding etc. need not be feared.

It is well-known that convolution becomes pointwise multiplication on the transform side, i.e.

$$(a * b)^\wedge = \hat{a} \cdot \hat{b}, \quad (a^N * b^N)^\wedge = (a^N)^\wedge \cdot (b^N)^\wedge \tag{2.7}$$

for all $a, b \in \ell_1$, $a^N, b^N \in \mathbb{C}^N$ (multiplication of vectors is understood to be componentwise–we regard the elements of $\mathbb{C}^N$ as complex functions on the set $\{0, \ldots, N-1\}$). Let H be the probability generating function of the p-sequence, i.e.

$$H : \{z \in \mathbb{C} : |z| \leq 1\} \to \mathbb{C}, \quad H(z) := \sum_{k=0}^{\infty} p_k z^k.$$

Then the following identities, which are fundamental for our purposes and again are well-known, follow easily from (2.7),

$$\Psi(a)^{\wedge} = H \circ \hat{a}, \quad \Psi_N(a^N)^{\wedge} = H \circ (a^N)^{\wedge}, \tag{2.8}$$

for all a, a^N in the respective range of definition. We mention in passing that the relation

$$(V_N(a))_n^{\wedge} = \hat{a}(2\pi n/N) \quad \text{for} \quad n = 0, \ldots, N-1$$

between the transforms on ℓ_1 and $\mathbb{C}^N$ leads to an immediate proof of relations such as (2.3).

Putting pieces together we arrive at the following *transform algorithm* for the computation of an approximation b^N to the compound distribution $b := \Psi(a)$ for a given claim size distribution a.

Algorithm 1
(i) *truncate the input sequence a at some threshold* $N : a \to a^N := T_N(a)$;
(ii) *apply* (2.5) *to the result of step* (i): $a^N \to (a^N)^{\wedge}$;
(iii) *apply H to the result of step* (ii): $(a^N)^{\wedge} \to H \circ (a^N)^{\wedge} = (\Psi_N(a^N))^{\wedge}$;
(iv) *apply* (2.6) *to the result of step* (iii): $(\Psi_N(a^N))^{\wedge} \to b^N := \Psi_N \circ T_N(a)$.

Note that we used (2.8) in Step (iii); for many cases of interest H can be given explicitly. Apart from errors such as those induced by the floating point representation of real numbers on a computer steps (ii)-(iv) of this algorithm evaluate compound distributions on $\mathbb{G}_N$ exactly.

It is clear that things might go badly wrong if there is substantial mass near N, as this mass will simply be wrapped around the threshold N and reappear at 0 (a year 2000 problem, so to speak). This can seriously distort the outcome, especially for heavy-tailed distributions. The following theorem gives bounds for the *aliasing error* $b^N - T_N(b)$, $b := \Psi(a)$.

Theorem 2
With b and b^N *as above,*

$$b_n \leq b_n^N \leq b_n + \sum_{j=1}^{\infty} b_{n+jN} \quad \text{for} \quad n = 0, \ldots, N-1.$$

In particular, $\|b^N - T_N(b)\|_1 \leq \sum_{n=N}^{\infty} b_n$.

Proof: On ℓ_1 and $\mathbb{C}^N$ we consider the componentwise ordering, e.g. we write $c \geq 0$ if all components of c are non-negative. In particular, this holds for a as its entries are probabilities. Note that Ψ is monotone in the sense of

$$a, b \in \ell_1,\ 0 \leq a \leq b \quad \Rightarrow \quad \Psi(a) \leq \Psi(b),$$

and similarly with Ψ_N. For the truncation operator we have

$$T_N(a * b) \leq T_N(a) * T_N(b) \quad \text{for all} \quad a, b \in \ell_1,\ a, b \geq 0.$$

This implies

$$\begin{aligned} b^N = \Psi_N \circ T_N(a) &= \sum_{k=0}^{\infty} p_k (T_N(a))^{*k} \\ &\geq \sum_{k=0}^{\infty} p_k T_N(a^{*k}) \\ &= T_N\left(\sum_{k=0}^{\infty} p_k a^{*k}\right) \\ &= T_N \circ \Psi(a) = T_N(b), \end{aligned}$$

which is the left hand side of the inequality. Further, $T_N(a) \leq V_N(a)$ because of $a \geq 0$, hence monotonicity of Ψ_N and (2.3) together give

$$\Psi_N \circ T_N(a) \leq \Psi_N \circ V_N(a) = V_N \circ \Psi(a) = V_N(b),$$

which yields the right hand side. □

Note that the upper bound $\sum_{n=N}^{\infty} b_n$ has a simple interpretation in the risk model, it is the probability that the sum of the claims exceeds the threshold N. In particular, we can draw the qualitative conclusion that the aliasing error becomes negligible with increasing N. For practical purposes, however, a more quantitative statement would be of interest. Using the componentwise ordering introduced in the above proof we can rewrite the first statement of the theorem as

$$T_N(b) \leq b^N \leq V_N(b), \tag{2.9}$$

with the central term the result of the algorithm and the left term the target value. Is the general upper bound $V_N(b)$ too conservative? To deal with questions of this type we investigate the asymptotic behaviour of b^N as $N \to \infty$, using a differentiability property of the functional Ψ. We assume that the sequence $(p_k)_{k \in \mathbb{N}_0}$ defining Ψ satisfies the condition

$$\sum_{k=1}^{\infty} k^2 p_k < \infty, \tag{2.10}$$

and may then define another functional Φ by

$$\Phi : \{a \in \ell_1 : \|a\|_1 \leq 1\} \to \ell_1, \quad \Phi(a) := \sum_{k=1}^{\infty} k p_k a^{*(k-1)}.$$

Lemma 3
If $a(N), N \in \mathbb{N}$ *and* a *are elements of* $\{a \in \ell_1 : \|a\|_1 \leq 1\}$ *with* $\|a(N) - a\|_1 \to 0$ *as* $N \to \infty$, *then*

$$\|\Psi(a(N)) - \Psi(a) - \Phi(a) * (a(N) - a)\|_1 = O\left(\|a(N) - a\|_1^2\right).$$

Proof: Simple algebra shows that, for all $k \in \mathbb{N}$,

$$a(N)^{*k} - a^{*k} = (a(N) - a) * \sum_{j=0}^{k-1} a(N)^{*(k-1-j)} * a^{*j}.$$

In particular, it follows with the norm inequality that

$$\left\|a(N)^{*k} - a^{*k}\right\|_1 \leq k \cdot \|a(N) - a\|_1 \quad \text{for all} \quad k \in \mathbb{N}.$$

We can now write

$$\begin{aligned}
&\Psi(a(N)) - \Psi(a) - \Phi(a) * (a(N) - a) \\
&\qquad = (a(N) - a) * \sum_{k=2}^{\infty} p_k \sum_{j=0}^{k-2} \left(a(N)^{*(k-1-j)} * a^{*j} - a^{*(k-1)}\right) \\
&\qquad = (a(N) - a)^{*2} * \sum_{k=2}^{\infty} p_k \sum_{j=0}^{k-2} a^{*j} * \sum_{l=0}^{k-2-j} a(N)^{*(k-2-j-l)} * a^{*l}.
\end{aligned}$$

The j-sum in the last term is bounded in norm by $O(k^2)$ so that the assertion follows on using (2.10). □

We can now obtain an expansion of the difference between the result of Algorithm 1 and the upper bound in Theorem 2 in terms of $\sum_{n=N}^{\infty} a_n$, i.e. the tails of the distribution of individual claims.

Theorem 4
Assume that (2.10) *holds; let* $a(N) := U_N \circ T_N(a)$. Then, as $N \to \infty$,

$$\left\| V_N(b) - b^N - V_N(\Phi(a)) * V_N(a - a(N)) \right\|_1 = O\left(\left(\sum_{n=N}^{\infty} a_n\right)^2\right).$$

Proof: Using (2.2) and (2.3) we obtain

$$\begin{aligned} V_N(b) - b^N &= V_N \circ \Psi(a) - \Psi_N \circ T_N(a) \\ &= V_N \circ \Psi(a) - \Psi_N \circ V_N \circ U_N \circ T_N(a) \\ &= V_N(\Psi(a) - \Psi(a(N))). \end{aligned} \tag{2.11}$$

Lemma 3 shows that

$$\|\Psi(a(N)) - \Psi(a) - \Phi(a) * (a(N) - a)\|_1 = O\left(\|a(N) - a\|_1^2\right).$$

Clearly, applying V_N does not increase the norm, and $\|a(N) - a\|_1 = \sum_{n=N}^{\infty} a_n$. □

To use this theorem in connection with the asymptotic behaviour of aliasing errors we note that (2.9) implies

$$\left\|b^N - T_N(b)\right\|_1 + \left\|V_N(b) - b^N\right\|_1 = \|V_N(b) - T_N(b)\|_1.$$

The first term is the aliasing error $\sum_{n=0}^{N-1} |b_n^N - b_n|$, the right hand side is the tail $\sum_{n=N}^{\infty} b_n$ of the compound distribution. For the middle term we obtain from Theorem 4,

$$\left\|V_N(b) - b^N\right\|_1 = \|V_N(\Phi(a)) * V_N(a - a(N))\|_1 + O\left(\left(\sum_{n=N}^{\infty} a_n\right)^2\right).$$

Using the fact that all entries of $\Phi(a)$ and $a - a(N)$ are nonnegative we obtain

$$\begin{aligned} \|V_N(\Phi(a)) * V_N(a - a(N))\|_1 &= \|\Phi(a) * (a - a(N))\|_1 \\ &= \|\Phi(a)\|_1 \|a - a(N)\|_1 \\ &= \left(\sum_{k=1}^{\infty} k p_k\right) \|a - a(N)\|_1 \\ &= H'(1) \sum_{n=N}^{\infty} a_n, \end{aligned}$$

so we have the following relation between the aliasing error, the tail of the individual claims and the tail of the compound distribution,

$$\sum_{n=0}^{N-1} |b_n^N - b_n| = \sum_{n=N}^{\infty} b_n - H'(1) \sum_{n=N}^{\infty} a_n + O\left(\left(\sum_{n=N}^{\infty} a_n\right)^2\right). \tag{2.12}$$

As a first consequence we see that under certain circumstances the upper bound in Theorem 2 is asymptotically tight in the sense that the output b^N of the algorithm is asymptotically closer to $V_N(b)$ than to $T_N(b)$: If $\sum_{n=N}^{\infty} a_n = o\left(\sum_{n=N}^{\infty} b_n\right)$ then we obtain

$$\lim_{N\to\infty} \frac{\sum_{n=0}^{N-1} |b_n^N - b_n|}{\sum_{n=N}^{\infty} b_n} = 1. \tag{2.13}$$

This condition means that the tail of the distribution of individual claims is asymptotically negligible in comparison to the tail of the total claim size distribution. It is obviously satisfied if e.g. claim sizes are bounded. Indeed, in this case we obtain from (2.11) that $b^N = V_N(b)$ for all $N \geq N_0$, $N_0 := \min\{N \in \mathbb{N} : \sum_{n=N}^{\infty} a_n = 0\}$, hence the upper bound in Theorem 2 cannot be improved without further conditions on the claim size distribution.

Of course, if the individual tails $\sum_{n=N}^{\infty} a_n$ are of the same order as the compound tails $\sum_{n=N}^{\infty} b_n$ then (2.13) may fail to hold. A notable example for this situation is the case where $(p_k)_{k\in\mathbb{N}}$ decreases at an exponential rate (as in the compound Poisson and compound geometric case) and where the claim size distribution is of subexponential type, i.e. (in the present discrete setup)

$$\sum_{n=N}^{\infty} (a^{*2})_n \sim 2 \sum_{n=N}^{\infty} a_n,$$

where the tilde means that the ratio of the two quantities tends to 1 as $N \to \infty$. The class $\mathcal{S}$ of subexponential distributions has been the object of much research as it provides the natural setting for many limit theorems; see e.g. Embrechts, Goldie and Veraverbeke (1979), Embrechts and Veraverbeke (1982) and the references therein. In particular, under the above assumptions,

$$\sum_{n=N}^{\infty} b_n \sim H'(1) \sum_{n=N}^{\infty} a_n, \tag{2.14}$$

so (2.12) leads to

$$\lim_{N\to\infty} \frac{\sum_{n=0}^{N-1} |b_n^N - b_n|}{\sum_{n=N}^{\infty} b_n} = \lim_{N\to\infty} \frac{\sum_{n=0}^{N-1} |b_n^N - b_n|}{\sum_{n=N}^{\infty} a_n} = 0. \tag{2.15}$$

Note, however, that the fact that in this case the outcome of the algorithm is asymptotically closer to the left than to the right boundary of the interval provided by Theorem 2 is of somewhat marginal significance because of the large size of this interval. Nevertheless, it would be interesting to know whether a second order refinement of (2.14) together with the use of higher

order derivatives of Ψ would lead to more precise statements on the asymptotic behaviour of the aliasing error for subexponential claim size distributions.

Subexponential distributions are heavy-tailed. Related classes $\mathcal{S}(\gamma)$, $\gamma > 0$, with exponential rate of tail decrease can be defined by the requirements

$$\sum_{n=0}^{\infty} e^{\gamma n} a_n < \infty, \quad \sum_{n=N+1}^{\infty} a_n \sim e^{-\gamma} \sum_{n=N}^{\infty} a_n, \quad \sum_{n=N}^{\infty} \left(a^{*2}\right)_n \sim 2\kappa \sum_{n=N}^{\infty} a_n,$$

in which case we necessarily have $\kappa = \sum_{n=0}^{\infty} e^{\gamma n} a_n (> 1)$. Again, this is a well-studied class; see Chover, Ney and Wainger (1973) in addition to the above references for $\mathcal{S} = \mathcal{S}(0)$. For such distributions it holds that

$$\sum_{n=N}^{\infty} b_n \sim H'(\kappa) \sum_{n=N}^{\infty} a_n,$$

provided that $p_k = O\left(\rho^{-k}\right)$ for some $\rho > \kappa$. The H-functions that are of interest to us have a strictly monotone first derivative so that $H'(1) < H'(\kappa)$. Hence (2.12) implies that b^N is asymptotically strictly between the two bounds in Theorem 2, i.e. we have a behaviour intermediate between (2.13) and (2.15):

$$0 < \lim_{N \to \infty} \frac{\sum_{n=0}^{N-1} \left| b_n^N - b_n \right|}{\sum_{n=N}^{\infty} b_n} = \frac{H'(\kappa) - H'(1)}{H'(\kappa)} < 1.$$

The above considerations show that the local behaviour of the functional Ψ can be used to relate the asymptotics of the aliasing error to the tail behaviour of the distributions involved. From a purely practical point of view the main conclusion remains, however, that for the purposes of Algorithm 1 it would be desirable to have rapidly decreasing compound tails. The main idea of the alternative algorithm that we propose now is a change of measure that forces the tails of the compound distribution to decrease at an exponentially rate.

The mathematical basis for our aliasing error reduction method is the observation that "an exponential change of measure commutes with the compound distribution functional". To make this precise we require one final bit of notation: For $a \in \ell_1$ and $\theta \in \mathbb{R}$, let $E_\theta a$ be the complex sequence defined by

$$\left(E_\theta a\right)_n := e^{-n\theta} a_n \quad \text{for all} \quad n \in \mathbb{N}_0.$$

The operator E_θ *tilts* the input sequence, θ is the *tilting parameter*. For $\theta \geq 0$ E_θ maps ℓ_1 into ℓ_1 and $\|E_\theta a\|_1 \leq \|a\|_1$. We further have for all $a, b \in \ell_1$,

$$\begin{aligned}((E_\theta a) * (E_\theta b))_n &= \sum_{m=0}^{n} (E_\theta a)_m (E_\theta b)_{n-m} \\ &= e^{-n\theta} \sum_{m=0}^{n} a_m b_{n-m} \\ &= (E_\theta(a * b))_n,\end{aligned}$$

so that $E_\theta(a * b) = E_\theta(a) * E_\theta(b)$. From this we easily obtain $\Psi \circ E_\theta = E_\theta \circ \Psi$, which is the idea behind the following algorithm.

Algorithm 5

(i) *Tilt the input sequence a with some suitable* $\theta > 0 : a \to a_\theta := E_\theta(a)$;
(ii) *truncate the result of step* (i) *at some threshold* $N : a_\theta \to a_\theta^N := T_N(a_\theta)$;
(iii) *apply* (2.5) *to the result of step* (ii): $a_\theta^N \to (a_\theta^N)^\wedge$;
(iv) *apply* H *to the result of step* (iii): $(a_\theta^N)^\wedge \to H \circ (a_\theta^N)^\wedge = (\Psi_N(a_\theta^N))^\wedge$;
(v) *apply* (2.6) *to the result of step* (iv): $(\Psi_N(a_\theta^N))^\wedge \to b_\theta^N := \Psi_N \circ T_N(a_\theta)$;
(vi) *undo the tilting:* $b_\theta^N \to b^N := E_{-\theta} b_\theta^N$.

The exponential tilting in step (i) produces rapidly decreasing tails, hence the aliasing error introduced by steps (ii)-(v) will be small by Theorem 2 if N is large. At least for small indices n the inevitable blow-up of errors introduced by the multiplication with potentially large factors in step (vi) is negligible in comparison with the overall improvement.

3. A NUMERICAL EXAMPLE

In Embrechts, Grübel and Pitts (1993) the difficulties arising from aliasing (and discretization) errors were illustrated in a special case, where the claim size distribution was taken to be the stable distribution with index 1/2. This distribution is concentrated on the non-negative real numbers and has an additional scale parameter α, the corresponding density is

$$f_\alpha(x) := \frac{\alpha}{\sqrt{2\pi x^3}} \exp\left(-\frac{\alpha^2}{2x}\right), \quad x > 0.$$

Since $f_\alpha * f_\beta = f_{\alpha+\beta}$ it is easy to obtain a corresponding compound distribution numerically to any desired degree of precision directly from the definition (1.1). In the above paper the Fourier transform based approximation for the distribution function of the associated compound Poisson distribution with $\lambda t = 20$ in (1.2), i.e. the numerical approximation resulting from discretization and subsequent application of Algorithm 1, was

compared to the true distribution function and it was found that, even with $N = 16384$ discretization intervals, the supremum distance would never be less than 0.034, irrespective of the truncation point.

Here we ignore the discretization error, which we will investigate in Part II of the present paper, and we compare the results obtained with Panjer recursion and transform methods. We also compare individual probabilities instead of distribution functions: By design, the supremum distance between the distribution functions cannot be smaller than the probability that the compound distribution exceeds the threshold value, if the supremum is taken over the whole real line. This elementary observation applies to both Panjer recursion and the transform based algorithm.

In practice the parameter θ involved in the exponential tilting cannot be chosen arbitrarily large as this might result in under- or overflow errors. A rough guideline would take the truncation point $x_0 = Nh$ into account. The minimal and maximal factors arising in steps (i) and (vi) of Algorithm 5 would be $\exp(\pm\theta x_0)$; a value of about 20 for the product θx_0 will in general not lead to numerical difficulties.

Table 1 gives some numerical values for the special case described above; $N = 1024$ discretization intervals of length 1 were used. The second column gives the true values for $\nu([x - 0.5,\ x + 0.5])$, the third the corresponding approximations obtained after discretization and use of (2.1). The column labeled "Alg1" shows the results of Algorithm 1, the unadorned Fourier transform algorithm. The alias or "wrap-around" effect is easily seen: the results for small x-values are at least an order of magnitude always from the true values. The remaining columns were obtained with Algorithm 5, with tilting parameters $\theta = 0.001$, 0.0049 and 0.0244 respectively, corresponding to the values 1, 5 and 25 of the product of tilting parameter and truncation threshold discussed above.

TABLE 1

COMPOUND PROBABILITIES AND APPROXIMATIONS

x	*true*	*Panjer*	*Alg1*	*Alg5a*	*Alg5b*	*Alg5c*
1	1.078E-07	2.462E-07	2.064E-04	7.346E-05	1.560E-06	2.462E-07
10	3.075E-05	3.432E-05	2.380E-04	1.067E-04	3.562E-05	3.432E-05
100	1.156E-03	1.156E-03	1.321E-03	1.215E-03	1.157E-03	1.156E-03
1000	2.013E-04	2.012E-04	2.134E-04	2.056E-04	2.013E-04	2.012E-04

Obviously, the accuracy is improved considerably by the change of measure. This is perhaps easier to see in Table 2, where the logarithm of the ratio of the approximation and the value obtained by recursion is displayed. The "large" value for $x = 1000$ in the last column can be explained by the occurrence of large factors in the last step of Algorithm 5: with tilting

parameter $\theta = 0.001$ factors of the order 10^{10} appear and rounding errors begin to be important. The sum of all absolute differences between the probabilities obtained with Panjer's algorithm and the transform based algorithms were 0.0714, 0.0255, 0.000459 and 0.0000003121 respectively, which is further support for the overall conclusion that for practical purposes the aliasing error can be eliminated by a judicious use of Algorithm 5. The upper bound $\sum_{n=N}^{\infty} b_n$ from Theorem 2 takes the value 0.4641 in this example, much larger than the actual value 0.0714. Note that this is an example with a subexponential claim size distribution, so this is in accordance with (2.15). We finally note that a comparison of the values in the second and third column of Table 1 shows that there is a pronounced discretization error, again most notable for small x-values.

TABLE 2

LOGARITHMIC DIFFERENCES

x	*Alg1*	*Alg5a*	*Alg5b*	*Alg5c*
1	6.732	5.698	1.84659	0.0000000110
10	1.936	1.134	0.03715	0.0000000001
100	0.134	0.050	0.00092	0.0000000000
1000	0.059	0.022	0.00040	0.0000247729

4. RELATED APPLICATIONS

In this section we briefly mention two further applications of the above methods in the context of risk theory, the calculation of ruin probabilities in the classical model with Poisson arrivals and the calculation of the mean aggregate claim size as a function of time in the Sparre-Andersen model. Details, including numerical examples, can be found in Hermesmeier (1997).

The probability of ruin as a function of the initial capital (risk reserve), given a specific premium income mechanism, is another central quantity of interest in risk theory. Let Y_t be the insurance surplus at time t. When a claim of size X_i arrives at time T_i then the stochastic process $Y = (Y_t)_{t \geq 0}$ has a corresponding downward jump, i.e. $Y_{T_i} - Y_{T_i-} = -X_i$. The classical model assumes that Y increases linearly between claims with rate c, c is the *premium income rate*. Let

$$\psi(u) = P(Y_t < 0 \quad \text{for some} \quad t \geq 0 | Y_0 = u)$$

be the probability that ruin will eventually occur if the initial risk reserve is u. We assume that the claim size distribution has finite mean m_C and that the *relative safety loading*

$$\theta := \frac{c}{\lambda m_C} - 1$$

is strictly greater than 0; ruin is certain if $\theta \leq 0$. Let μ_1 be the *tail measure* associated with the claim size distribution μ, μ_1 is the distribution with density function $x \to \mu((x,\infty))/m_C$. Ruin probabilities can be related to the distribution of the maximum of a random walk with negative drift, and classical random walk theory leads to the representation

$$\psi(u) = \sum_{k=0}^{\infty} (1-\rho)^k \rho\, \mu_1^{*k}((u,\infty)), \quad \text{with} \quad \rho := \frac{\theta}{1+\theta} \tag{4.1}$$

(see e.g. Asmussen (1987), Chapter XIII). Hence ψ is the tail function of a distribution ν of the form (1.1) with $p_n = (1-\rho)^n \rho$, i.e. the tail function of a compound geometric distribution.

This situation is essentially the same as the one treated in Section 2, and the same arguments apply. Again, for the weights in (4.1) a recursive method is available (here too the first step is discretization), transform methods apply, aliasing errors arise and can be handled as in Section 2. There is a considerable literature relating the asymptotics of ψ to the tails of the claim size distribution; see e.g. Embrechts and Veraverbeke (1982).

In the Sparre-Andersen model the assumption that the claim arrival times form a Poisson process is generalized to a situation where the times between claim arrivals are independent and identically distributed random variables. Let μ_0 be the distribution of these interarrival times. From a technical point of view, the arrival rate λ in Section 1 is replaced by the distribution μ_0. Let S_t be the aggregate claim size at time t. Then, as explained in Embrechts, Grübel and Pitts (1993), the stochastic process $S = (S_t)_{t \geq 0}$ is a renewal reward process, and for the calculation of e.g. the mean function $t \to ES_t$ of S the renewal measure $\sum_{k=0}^{\infty} \mu_0^{*k}$ associated with μ_0 is important. This corresponds to a situation of type (1.1) with $p_k = 1$, but the renewal measure is not finite as these weights are not summable. One possibility to overcome the resulting difficulties is to rewrite the renewal equation in such a way that one arrives safely back in ℓ_1 (see Embrechts, Grübel and Pitts (1993)), a different one is to use the ideas of Section 2 and to consider tilted renewal sequences. Finally, the ruin probability function in this more general model continues to be of the form (4.1), but μ_1 depends on μ_0 and the claim size distribution μ in a more complicated way. We still have the connection to the supremum of a random walk, which also occurs in connection with the stationary waiting time in a G/G/1 queue, so that the transform based algorithm in Grübel (1991) can be used.

5. CONCLUSIONS AND COMMENTS

We have already mentioned in Section 1 that the operation count grows quadratically in N if $b_0, \ldots, b_{N-1}$ are evaluated with Panjer recursion. On first sight the matrix multiplications in (2.5) and (2.6) seem to be no different, but if N is e.g. a power of 2, then the fast Fourier transform (FFT) algorithm can be used, resulting in an operation count of the order $N \log N$ only. Bühlmann (1984) conducted some explicit comparisons and found a timing advantage for the FFT-based algorithm from $N = 256$ onwards. The 15 years which elapsed since then have seen an enormous increase in computing power available on the desktop–we have used transform based algorithms with $N = 1048576 (= 2^{20})$ and more. Recursion with N of this magnitude is not feasible. The significance of large N-values is especially important if high quantiles of the distribution of the total claim amount are to be calculated (or, correspondingly, risk reserves with very small probability of ruin). In the situation explained in Section 3, for example, the 0.95-quantile of the total claim size distribution is about 10^6, and with discretization intervals of the length considered in our numerical example we would still have a noticeable discretization error (see Table 1) and already require the value 2^{20} for N.

Algorithm efficiency can also be decisive in connection with statistical analyses. Given that the structural assumptions are accepted as a sensible approximation to the real situation, practical usage of the model would need to begin with inference on λ and μ.We refer the reader to Pitts (1994) and the references given there for inference on compound distributions. Modern statistical techniques such as bootstrap confidence regions require the numerical evaluation of the estimator in a great many cases (resamples), which is feasible only if an efficient algorithm is available.

Finally, shifting from a given measure to a new one that has an exponential density with respect to the original measure is an important technique in many areas of theoretical and applied probability and statistics. It underlies saddle point approximations, it is a standard technique in large deviation theory, it can be found in the stochastic analysis treatment of the Black-Scholes formula, and the concept of exponential families in traditional mathematical statistics makes use of this idea. Exponential tilting also connects the classes $\mathcal{S}$ and $\mathcal{S}(\gamma)$ that were used in Section 2. On the transform side this shift corresponds to a shift of the integration range in the complex domain, which we could do here in view of the fact that claim size distributions are concentrated on the non-negative half-line. Transform methods *per se* can handle two-sided distributions (in contrast to recursion methods), but in the context of exponential tilting this advantage is lost. The gain is in the effective elimination of aliasing errors, which makes it possible to use transform algorithms if e.g. only low order quantiles are of interest, a situation which up to now was considered to be the exclusive domain of recursion algorithms.

ACKNOWLEDGEMENT

We thank Dr. Binswanger for sending us a copy of Asmussen and Binswanger (1997). We also thank the two anonymous referees for their valuable comments and suggestions, which led to a considerable improvement of the paper.

REFERENCES

ASMUSSEN, S. (1987). *Applied Probability and Queues*. Wiley, Chichester.

ASMUSSEN, S. and BINSWANGER, K. (1997). Simulation of ruin probabilities for subexponential claims. *ASTIN Bull.* **27**, 297-318.

BEARD, R.E., PENTIKÄINEN, T. and PERSONEN, E. (1984). *Risk Theory: The Stochastic Basis of Insurance*. (3rd ed.) Chapman and Hall, London.

BUCHWALDER, M., CHEVALIER, E. and KLÜPPELBERG, C. (1993). Approximation methods for the total claimsize distribution–an algorithmic and graphical presentation. *Mitteilungen der Schweiz. Vereinigung der Versicherungsmathematiker*, Heft 2/1993.

BÜHLMANN, H. (1984). Numerical evaluation of the compound Poisson distribution: recursion or fast Fourier transform? *Scand. Actuarial J.* **2**, 116-126.

CHOVER, J., NEY, P. and WAINGER, S. (1973). Functions of probability measures. *J. Analyse Math.* **26**, 255-302.

EMBRECHTS, P., GOLDIE, C.M. and VERAVERBEKE, N. (1979). Subexponentiality and infinite divisibility. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **49**, 335-347.

EMBRECHTS, P., GRÜBEL, R. and PITTS, S.M. (1993). Some applications of the fast Fourier transform algorithm in insurance mathematics. *Statistica Neerlandica* **47**, 59-75.

EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, Th. (1997). *Modelling Extremal Events*. Springer, Berlin.

EMBRECHTS, P. and VERAVERBEKE, N. (1982). Estimates of the probability of ruin with special emphasis on the possibility of large claims. *Insurance: Mathematics and Economics* **1**, 55-72.

GRANDELL, J. (1991). *Aspects of Risk Theory*. Springer, New York.

GRÜBEL, R. (1991). Algorithm AS 265: G/G/1 via Fast Fourier Transform. *Applied Statistics (J. R. Stat. Soc. C)* **40**, 355-365.

GRÜBEL, R. and HERMESMEIER, R. (1999). Computation of compound distributions II: discretization errors and Richardson extrapolation. In preparation.

HERMESMEIER, R. (1997). *Transformationsmethoden und Diskretisierungseffekte bei stochastischen Modellen in der Versicherungsmathematik*. Doctoral thesis, University of Hannover.

HIPP, C. and MICHEL, R. (1990). *Risikotheorie: Stochastische Modelle und Statistische Methoden*, Verlag Versicherungswirtschaft e.V., Karlsruhe.

JOHNSON, N.L., KOTZ, S. and KEMP, A.W. (1992). *Univariate Discrete Distributions*, 2nd edition. Wiley, New York.

PANJER, H.H. (1981). Resursive evaluation of a family of compound distributions. *ASTIN Bulletin* **12**, 22-26.

PITTS, S.M. (1994). Nonparametric estimation of compound distributions with applications in insurance. *Ann. Inst. Stat. Math.* **46**, 537-555.

SCHRÖTER, K.J. (1995). *Verfahren zur Approximation der Gesamtschadenverteilung: Systematisierung, Techniken und Vergleiche*. Verlag Versicherungswirtschaft e.V., Karlsruhe.

RUDOLF GRÜBEL and RENATE HERMESMEIER
Institut für Mathematische Stochastik
Universität Hannover
Postfach 60 09
D-30060 Hannover
e-mail: rgrubel@stochastik.uni-hannover.de