Decompounding random sums: A nonparametric approach

Hansen, Martin Bøgsted; Pitts, Susan M.

Publication date:
2006

Document Version
Publisher's PDF, also known as Version of record

AALBORG UNIVERSITY

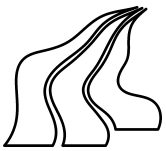
Decompounding random sums: A nonparametric approach

by

Martin B. Hansen and Susan M. Pitts

R-2006-20 May 2006

Department of Mathematical Sciences
Aalborg University
Fredrik Bajers Vej 7 G ▪ DK - 9220 Aalborg Øst ▪ Denmark
Phone: +45 96 35 80 80 ▪ Telefax: +45 98 15 81 29
URL: http://www.math.aau.dk



ISSN 1399–2503 ▪ On-line version ISSN 1601–7811

Decompounding random sums: A nonparametric approach

MARTIN B. HANSEN* and SUSAN M. PITTS**

*Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg Ø, Denmark. Email: mbh@math.aau.dk

**Statistical Laboratory, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WB, UK. Email: S.Pitts@statslab.cam.ac.uk

Summary. Observations from sums of random variables with a random number of summands, known as *random*, *compound* or *stopped sums* arise within many areas of engineering and science. Quite often it is desirable to infer properties of the distribution of the terms in the random sum. In the present paper we review a number of applications and consider the non-linear inverse problem of inferring the cumulative distribution function of the components in the random sum. We review the existing literature on non-parametric approaches to the problem. The models amenable to the analysis are generalized considerably and the properties of the proposed estimators are studied. Bootstrap methods are suggested to provide confidence bounds. Finally a number of algorithms are suggested to make the methods operational and tested on simulated data. In particular we show how Panjer recursion in general can be inverted for the Panjer as well as the Willmot class.

Keywords: Asymptotic normality; Compound distributions; Compound Poisson; Decompounding; Empirical processes; Functional central limit; Infinite-dimensional delta-method; Inverse problems; Sampled network traffic; Workload.

1. Introduction

Let N be an integer-valued discrete random variable with probability function

$$p_k = \mathrm{P}(N = k), \text{ for } k = 0, 1, 2, \ldots,$$

and $X_1, X_2, X_3, \ldots$ a series of strictly positive independent and identically distributed (iid) random variables with cumulative distribution function (cdf) F independent of N. Following the tradition of insurance mathematics N will be denoted the *claim number* and F the *claim size distribution*.

Now the cdf of the random sum $Y = X_1 + \cdots + X_N$ is given by

$$G = \Psi(F) \text{ with } \Psi(F) = \sum_{k=0}^{\infty} p_k F^{\star k}, \tag{1}$$

where $\star k$ denotes k-fold convolution. The main objective of the present paper is to infer F given an iid sample $Y_1, \ldots, Y_n$ from G.

This problem arises in a number of applications within engineering and science. A prominent example is the compound Poisson process which has many applications - within insurance mathematics and queueing theory - as G denotes the distribution of the amount

of a random number of randomly sized insurance claims or workload requirements over a given time interval, see Grandell (1997) for a comprehensive overview of the literature on this subject. Buchmann and Grübel (2003) considered this decompounding problem in detail and describe motivations from insurance mathematics and queueing theory. In their work the problem was solved by proposing a non-parametric estimator of the cdf F and its properties (in weak convergence sense) were investigated.

Another motivation for the set-up presented above arises when one intends to sample (or probe) the status of communication systems by test traffic. This is a standard tool for performance evaluation. It is e.g. used in call-center evaluations, where the distribution of the waiting time to service is estimated by the empirical distribution function of the call answer time for repeated phone calls. Another application is when a call admission controller in an ATM network decides whether there are sufficient resources to allow a new connection to be established, based on information obtained by sampling the workload at neighbouring nodes. This problem was considered by Sharma and Mazumdar (1998) and Sharma (1999). They consider an $M/G/1$ queue which is probed by Poisson traffic from which moments of the service time distribution are inferred by the well-known moment relations which can be derived from the Pollaczeck-Khintchine formula (Asmussen, 2003, Theorem VIII.5.7 (5.6) and (5.7)). In Hansen and Pitts (2006) the even harder problem of inferring the whole distribution function of the service time was solved. Hansen and Pitts (2006) extended the work by Buchmann and Grübel (2003) to the geometric case and proposed a semi-parametric approach to estimate the parameter of the geometric distribution and the service time cdf. Moreover, the paper showed one can deal with the problem for regenerative samples from G, a problem which is especially relevant dealing with queuing systems.

Another motivation arises in an infinite-capacity storage model, where inputs $S_1, S_2, S_3, \ldots$ to the storage facility arrive as a Poisson process with rate λ, where $S_1, S_2, S_3 \ldots$ are independent identically distributed random variables, with distribution function F independent of the Poisson process. The total amount in the storage facility at time t has the same distribution as the workload distribution in the $M/G/1$ queueing model above. Suppose that observations on the sampled total amounts at times $t = 1, 2, 3 \ldots$ are available, and that interest lies in inference for the distribution F of the inputs to the facility. This inference problem is exactly analogous to that described above for the queueing model and the methods of that paper apply.

Another motivation arises in network traffic. According to Hohn and Veitch (2003) many routers have the ability to output statistics about packets and flows that traverse them. However as the generation of detailed traffic statistics does not scale well with link speed, many manufactures have implemented sampling strategies at the packet level. Hohn and Veitch (2003) describe a solution to this binomial deconvolution problem and derivation of the statistical properties of this estimator is work in progress.

The Neyman-Scott cluster point process, originally developed in Neyman and Scott (1958) to describe the distribution of galaxies in space has become an important representation for a broad range of phenomena in the physical, biological, and social sciences. Here the decompounding problem arises if one assumes a random number (distributed according to p_k) of entities per observation field each of which produces a random number (distributed according to F) of offspring, and one asks for an estimate of F, based on observations of the number of offspring only.

Finally, one could also consider examples where some of the p_k's are zero, e.g. $p_2 = 1$ and

$p_0, p_1, p_3, \cdots = 0$, then we are dealing with the following non-linear deconvolution problem

$$G(y) = F \star F(y) = \int F(y-z)F(dz) \tag{2}$$

which resembles the classical linear deconvolution problem

$$G(y) = F \star H(y) = \int F(y-z)H(dz)$$

where F is an unknown distribution function and H is a known error function. This is an ill-posed and highly studied inverse problem, see e.g. Zhang (1990). However, the non-linearity in the present problem requires new methods to be developed.

In what follows a cdf of a non-negative random variable is denoted by a capital letter, A, say. The Laplace-Stieltjes transform $\mathcal{L}(A)$ or $\widetilde{A}$ is denoted by

$$\mathcal{L}(A)(\theta) = \int_{[0,\infty)} e^{-\theta x} A(dx)$$

for those $\theta \in \mathbb{R}$ for which the integral is defined. The Fourier-Stieltjes transform $\mathcal{F}(A)$ or $\widehat{A}$ is denoted by

$$\mathcal{F}(A)(\theta) = \int e^{ix\theta} A(dx), \quad \theta \in \mathbb{R}.$$

We suggest the empirical cumulative distribution function (ecdf) as an estimator for G

$$G_n(x) = n^{-1} \sum_{i=1}^{n} 1(Y_i \leq x), \quad 0 \leq x < \infty. \tag{3}$$

Its n'th *empirical process* counterpart is defined as

$$\mathbb{G}_n(x) = n^{1/2}(G_n(x) - G(x)), \quad 0 \leq x < \infty. \tag{4}$$

Assume we want to make statistical inference about F. Formula (1) can be inverted under conditions in the following way

$$F(x) = \sum_{k=1}^{\infty} \pi_k (G^o)^{\star k}(x), \tag{5}$$

where the π_k's can be expressed in terms of the p_k's. For any function h we write h^o for the function $x \mapsto h(x) - h(0)$ and the convergence in (5) is in a suitable weighted Banach space, see Theorem 1 below. This leads to the following plug-in estimator of F

$$F_n(x) = \sum_{k=1}^{\infty} \pi_k (G_n^o)^{\star k}(x). \tag{6}$$

In (5), F is given in terms of G, so that F is determined by G, and we can write F as a functional of G. The proposed estimator F_n is then the result of applying the same functional to the estimator G_n of G respectively. Given an appropriate asymptotic normality result for G_n, the infinite-dimensional delta method can be used to derive asymptotic normality of F_n, provided that the functional in question satisfies a particular differentiability result.

For general descriptions of the infinite-dimensional delta method, see Gill (1989) and van der Vaart (1998).

From (1), we see that G is a compound distribution function, based on F, and so the inverse functional that takes G onto F is a *decompounding* functional that "decompounds" the compound distribution. As such, this functional is closely related to that in Buchmann and Grübel (2003), where the notion of decompounding is introduced in the context of decompounding for a compound Poisson distribution. The set-up and proofs for the definition and differentiability of our functional follow those in Buchmann and Grübel (2003), making adaptations for the general case as necessary. Moreover, if data do not consist of independent identically distributed observations, but rather exhibit regenerative structure, then asymptotic normality results for the input estimator can be obtained using the approach of Hansen and Pitts (2006), whose approach was obtained using an empirical central limit theorem for regenerative data, see Tsai (1998) and Levental (1988).

The infinite-dimensional delta method has been used for stochastic models in previous work, and this paper follows the set-up and approach developed in, for example, Grübel and Pitts (1993), Hansen and Pitts (2006), Bingham and Pitts (1999b) and Bingham and Pitts (1999a). These last two papers study inference for service-time distributions given data on busy and idle periods for the $M/G/1$ and the $M/G/\infty$ queues respectively, and can be regarded as tackling inverse problems, in the same way that inference for decompounding can be regarded as an inverse problem.

In this paper we estimate the distribution function and in our work (Hansen and Pitts, 2006) we have, inspired by recent interests in queueing and insurance applications been much concerned with tail behaviour. However, in many informal investigations of the properties of a given set of data a density estimator is popular as it can give valuable indications of such features as skewness and multi-modality. It is indeed possible to make a kernel-estimator of the density function based on the methods presented in this paper

$$f_n(x) = \int K(x-y)F_n(dy),$$

where K is a smoothing-kernel, see e.g. Wand and Jones (1995) for more details. We are not going to discuss any properties of such estimators whatsoever, but merely refer to detailed papers on this matter. More specifically Hall and Park (2004) gives an approach to estimation of the density of F from busy period data presented in Bingham and Pitts (1999a), and van Es et al. (2005) for a density approach to Buchmann and Grübel (2003).

The paper is organised as follows. In Section 2 some necessary background from complex analysis as well as (sub-) probability generating functions are reviewed. This insight is used to define the estimator and show it is well-defined. The section is concluded with statements of the asymptotic normality for F_n as well as stating asymptotic properties of the proposed bootstrap confidence region. Section 3 contains a comparisons of necessary bounds under the developed general theory and the bounds which can be derived using the explicit parametric structure of the compounding under study. Various algorithms for dealing with the general approach to decompounding are introduced in Section 4. In this section an example of the estimator in action is also given and a discussion and conclusion are provided in Section 5. Proofs related to the properties of the estimator are given in Section 6.

2. Estimation

2.1. Background on sub-probability generating functions

In order to invert (1) we rephrase the problem as reversing (in line with the tradition of complex analysis we use the word reverse for finding the power series of the inverse with respect to composition of a power series, see Henrici (1974, p. 47)) the following power series

$$g(z) = \sum_{k=1}^{\infty} p_k z^k, \ z \in \mathbb{C}. \tag{7}$$

Notice, that g is actually a sub-probability generating function (if $p_0 > 0$, otherwise a probability generating function) and will therefore always have a radius of convergence $r(g)$ greater than or equal to 1 (Johnson et al., 2005, Section 1.2.11).

A standard way to calculate candidates for the coefficients π_k of the reversion $f(w) = \sum_{k=1}^{\infty} \pi_k w^k$ is formally to solve for π_k in the equation

$$z = \sum_{k=1}^{\infty} \pi_k \left[\sum_{l=1}^{\infty} p_l z^l \right]^k \tag{8}$$

by equating powers of z. This leads to the recursion schemes in Henrici (1974, (1.7-2)), which indeed have a unique solution, if $p_1 \neq 0$, see Henrici (1974, Theorem 1.7a).

However, the procedure to obtain the coefficients π_k of f is usually difficult to implement in practice. This issue will be treated in Section 4.1

The next step is to prove that the formal power series is a reverse of g. We need to give a precise statement of results from complex analysis about reversion of power series, as they apply to our power series g. The next proposition follows directly from Theorem 2.4b and the proof of Theorem 2.4c in Henrici (1974).

PROPOSITION 1. *Let $g(z) = \sum_{k=1}^{\infty} p_k z^k$ where $p_k = \mathrm{P}(N = k)$, $p_1 \neq 0$, and $r(g)$ is the radius of convergence.*

(a) Then the reversion of g

$$f(w) = \sum_{k=1}^{\infty} \pi_k w^k$$

(as defined above) has positive radius of convergence $r(f)$.

(b) There exist σ_g, $0 < \sigma_g \leq r(g)$, and σ_f, $0 < \sigma_f \leq r(f)$, such that $|w| < \sigma_f$ implies that there is a unique z in $|z| < \sigma_g$ such that $g(z) = w$, and this z is given by $z = f(w)$.

(c) We have $|w| < \sigma_f$ implies that

$$z = f(w) = g^{-1}(w) = \sum_{k=1}^{\infty} \pi_k w^k \text{ and } |f(w)| < \sigma_g. \tag{9}$$

(d) There exists ρ_g with $0 < \rho_g \leq \sigma_g$ such that

$$|z| < \rho_g \text{ implies that } |g(z)| < \sigma_f, \tag{10}$$

and so for z such that $|z| < \rho_g$, we have $z = f(g(z)) = \sum_{k=1}^{\infty} \pi_k \big(g(z)\big)^k$.

It is also possible to give lower bounds on those ρ_g and σ_f for which Proposition 1(d) holds. The next lemma follows directly from Copson (1935, p. 123) (according to Copson this result was first derived by Laudau)

LEMMA 1 (LANDAU). *Under the assumptions of Proposition 1, let* $0 < R < r(g)$ *and* $|g(z)| \leq M$ *for* $|z| \leq R$. *Then if* $|w| < (Rp_1)^2/(6M)$ *we have that there is a unique* z *in* $|z| < R^2 p_1/(4M)$ *such that* $g(z) = w$, *and this* z *is given by* $z = f(w)$.

In this case it is possible to choose

$$\rho_g = R^2 p_1/(4M)$$

and

$$\sigma_f = (Rp_1)^2/(6M).$$

Actually, Redheffer (1962) improved upon the lower bound for the radius of convergence for the reversed series. The following result follows directly by an adaption of his proof.

LEMMA 2 (REDHEFFER). *Under the assumptions of Proposition 1, let* $0 < R < r(g)$ *and*

$$\frac{1}{2\pi}\int_0^{2\pi} |g(re^{i\theta})|^2 d\theta \leq M^2$$

for $|z| < R$, *and define* $A = Rp_1/M$. *Then if* $|w| < (Rp_1)^2(1-(3/4)A^2)^{-1/2}/(4M)$ *we have that there is a unique* z *in* $|z| < R\sin(\arctan(A(1-A^2)^{-1/2}/2))$ *such that* $g(z) = w$, *and this* z *is given by* $z = f(w)$.

In this case it is possible to choose

$$\rho_g = R\sin(\arctan(A(1-A^2)^{-1/2}/2))$$

and

$$\sigma_f = (RA)^2(1-(3/4)A^2)^{-1/2}/(4M).$$

REMARK 1. *As the sub-probability generating function always has a radius of convergence greater than one, we immediately obtain, the following "universal" bounds on* z *and* $g(z)$ *for reverting the series in (7) (take* $R = 1$ *and* $M = 1$*)*

$$\rho_g = \sin(\arctan((p_1(1-p_1^2)^{-1/2}/2)) \quad (11)$$

$$\sigma_f = p_1^2(1-(3/4)p_1^2)^{-1/2}/4 \quad (12)$$

To our knowledge there do not exist any systematic studies of improvements on the Redheffer bound for power series. Although very interesting and relevant for the present paper we leave this as an open question.

2.2. Definition of the estimator

We avoid trivialities by requiring throughout the paper that the random variable N is not concentrated at 0. We also rule out the case where F is concentrated at one point and N is concentrated at 1 (so that G is not concentrated at one point).

Turning now to the general decompounding problem, we define appropriate weighted spaces for our estimator. We follow the approach and methodology of Buchmann and

Grübel (2003), and in particular we use the weighted spaces defined there, which we now describe. Let $D[0,\infty)$ be the Banach space of cadlag functions h on $[0,\infty)$. For τ in $\mathbb{R}$, let $D_\tau[0,\infty)$ be the space of all functions $h:[0,\infty)\to\mathbb{R}$, such that the function $x\mapsto e^{-\tau x}h(x)$, is in $D[0,\infty)$. For h in $D_\tau[0,\infty)$, let $\|h\|_{\infty,\tau}=\sup_{x\geq 0}e^{-\tau x}|h(x)|$, so that $(D_\tau[0,\infty),\|\cdot\|_{\infty,\tau})$ is a Banach space. Theorem 1 below states that, under conditions on G, the right-hand side of (5) is in $D_\tau[0,\infty)$. Recall that σ_f, σ_g and ρ_g are defined in Proposition 1.

THEOREM 1. *Let $\tau>0$ and assume that G is a distribution function on $[0,\infty)$ with $\widetilde{G}^o(\tau)<\sigma_f$. Then the series*

$$\Lambda(G) \;=\; \sum_{k=1}^{\infty}\pi_k(G^o)^{\star k}$$

converges in $D_\tau[0,\infty)$.

Suppose in addition that $\widetilde{F}(\tau)<\sigma_g$. Then

$$G=\Psi(F) \text{ implies that } F=\Lambda(G).$$

Note that, with ρ_g defined as in (10), $\widetilde{F}(\tau)<\rho_g$ $(\leq\sigma_g)$ implies that $\widetilde{G}^o(\tau)<\sigma_f$ $(\leq r(f))$, so that in this case $\Lambda(G^o)$ is defined and is equal to F. This means that the conditions on $\widetilde{F}(\tau)$ and $\widetilde{G}^o(\tau)$ in the statement of the theorem may be replaced by the single condition $\widetilde{F}(\tau)<\rho_g$.

REMARK 2. *As argued in Remark 1 it is possible to obtain universal bounds on ρ_g and σ_f which are strictly greater than 0. As $\widetilde{G}^o(\tau)\to 0$ and $\widetilde{F}(\tau)\to 0$ as $\tau\to\infty$, it always possible to find a space $D_\tau[0,\infty)$ such that (1) is invertible!*

In (5), our quantity of interest F is given in terms of G, and this inverse representation was made precise in Theorem 1 above. Thus the the plug-in estimator given in (6) is then

$$F_n=\Lambda(G_n).$$

2.2.1. Properties of the estimator

In this section, we state two results, one giving asymptotic normality of the proposed estimator, in terms of convergence in distribution to a Gaussian process in a D_τ-space, and the second giving asymptotic validity of bootstrap confidence regions. Throughout the paper, weak convergence in Banach spaces refers to σ-algebras generated by the open balls in the respective norms, see Pollard (1984), page 199. We can now formulate the main result on weak convergence of the inverse estimator of the distribution function F.

THEOREM 2. *Assume that $\tau>0$ is such that $\widetilde{F}(\tau)<\sigma_g$ and $\widetilde{G}^o(\tau)<\sigma_f$. Then*

$$\sqrt{n}(F_n-F)\;\to_{\mathcal{D}}\; A \text{ as } n\to\infty,$$

in $(D_\tau[0,\infty),\|\cdot\|_{\infty,\tau})$, where A is a centered Gaussian process with

$$cov\big(A(s),A(t)\big)\;=\;\int\int G^o\big((s-x)\wedge(t-y)\big)H(dx)H(dy)-G^o\star H(s)G^o\star(H)(t),\quad s,t\geq 0,$$

where

$$H \;=\; \sum_{k=1}^{\infty} k\pi_k \, (G^o)^{\star(k-1)} \, .$$

Turning to the question of obtaining confidence regions in $D_\tau[0,\infty)$ for the unknown F, we define, for $z \geq 0$,

$$R_n(z) \;=\; \mathrm{P}\big(\sqrt{n}\|F_n - F\|_{\infty,\tau} \leq z\big),$$

and

$$R(z) \;=\; \mathrm{P}\big(\|A\|_{\infty,\tau} \leq z\big).$$

From the convergence in distribution proved in Theorem 2, we know that

$$R_n(z) \to R(z)$$

at all continuity points of R. If the distribution function R were known, then an asymptotic $100\alpha\%$ confidence region for F would be the set of all distribution functions F such that

$$\|F_n - F\|_{\infty,\tau} \;\leq\; \frac{q(\alpha)}{\sqrt{n}},$$

where $q(\alpha)$ is a continuity point of R with $R\big(q(\alpha)\big) \geq \alpha$. However, R is not known, and moreover, it depends on G in a rather complicated way.

This motivates the use of the bootstrap to obtain confidence regions for F. We give below a precise statement of its asymptotic validity, in a similar manner to that of Grübel and Pitts (1993). In the present paper, we give sufficient explanation of the notation and ideas needed to formulate the precise results.

The bootstrap arises by resampling with replacement from the empirical distribution function G_n. Following Grübel and Pitts (1993), we capture the resampling mechanism via the map $\mathbb{F}_n : \mathbb{R}^n \to D[0,\infty)$, where, for $x = (x_1, \ldots, x_n)$,

$$\mathbb{F}_n(x) \;=\; \frac{1}{n}\sum_{i=1}^{n} 1_{[x_i,\infty)},$$

where $1_B(z) = 1$ if z is in the set B and is zero otherwise. If $Y_1, \ldots, Y_n$ are iid observations from G, then the empirical distribution function based on $Y_1, \ldots, Y_n$ is given by $G_n = \mathbb{F}_n(Y_1, \ldots, Y_n)$. With this notation, the distribution function $R_n(z)$ is given by

$$R_n(z) \;=\; \int_{\mathbb{R}^n} 1_{[0,z]} \left(\sqrt{n}\|\Lambda\big(\mathbb{F}_n(x)\big) - \Lambda(G)\|_{\infty,\tau}\right) G^{\otimes n}(dx),$$

where $G^{\otimes n}$ refers to product measure. The bootstrap estimator $\hat{R}_n$ of R_n is obtained by replacing G by G_n in the above formula for R_n. Since G_n corresponds to a discrete measure that puts mass n^{-1} at each of $Y_1, \ldots, Y_n$, it is easy to see that the integral becomes a sum, and that

$$\hat{R}_n(z) \;=\; \frac{1}{n^n} \sum_{(i_1,\ldots,i_n)\in\mathcal{I}_n} 1_{[0,z]} \left(\sqrt{n}\|\Lambda\big(\mathbb{F}_n(Y_{i_1}, \ldots, Y_{i_n})\big) - \Lambda\big(G_n\big)\|_{\infty,\tau}\right),$$

where $\mathcal{I}_n = \{1, \ldots, n\}^n$. Let $\hat{q}_n(\alpha)$ be the α-quantile of $\hat{R}_n$.

Table 1. Comparisons of upper bounds on $\widetilde{F}\tau$

Compound	Landau (Lemma 1)	Redheffer (11)	Direct
Geometric	$\frac{\rho(1-\rho)}{4}$	$\sin(\arctan(\rho(1-\rho)(1-(\rho(1-\rho))^2)^{-1/2}/2))$	$\frac{1}{\rho}\frac{1}{2}$
Poisson	$\frac{\lambda e^{-\lambda}}{4}$	$\sin(\arctan(\lambda e^{-\lambda}(1-(\lambda e^{-\lambda})^2)^{-1/2}/2))$	$\frac{1}{\lambda}\log(2)$

THEOREM 3 (BOOTSTRAP). *Let G_n, G, F_n, F, $\hat{q}_n(\alpha)$ be as above. Assume that τ is such that $\widetilde{F}(\tau) < \sigma_g$ and $\widetilde{G}^o(\tau) < \sigma_f$. Then*

$$\lim_{n\to\infty} \mathrm{P}\left(\sqrt{n}\|F_n - F\|_{\infty,\tau} \le \hat{q}_n(\alpha)\right) = \alpha.$$

This result gives rise to a confidence region for F given by the set of all distribution functions F' such that $\|F' - F_n\|_{\infty,\tau} \le \hat{q}_n(\alpha)/\sqrt{n}$, and the theorem tells us that the probability that this region covers the true F approaches α as n tends to infinity.

In practice, an approximation to the bootstrap estimator $\hat{R}_n$ is obtained via Monte Carlo methods. This means that a large number B of (re)samples $Y_j^* = \{Y_{ji_1}, \ldots, Y_{ji_n}\}$, $j = 1, \ldots, B$, are simulated with replacement from the original data $Y_1, \ldots, Y_n$. For each Y_j^*, we calculate the corresponding estimator $F_j^* = \Lambda\big(\mathbb{F}_n(Y_{ji_1}, \ldots, Y_{ji_n})\big)$. The empirical distribution function of the B real numbers $\sqrt{n}\|F_j^* - F_n\|_{\infty,\tau}$, $j = 1, \ldots, B$, is an approximation to $\hat{R}_n$, and the α quantile of this provides an approximation q_{approx} to $\hat{q}_n(\alpha)$. Using this approximating quantile then leads to an approximate $100\alpha\%$ confidence region for F given by the set of distribution functions such that $\|F' - F_n\|_{\infty,\tau} \le q_{\text{approx}}/\sqrt{n}$,

3. Comparisons

In Theorem 2 it is required that there exists a τ such that $\widetilde{F}(\tau)$ is less than σ_g and $\widetilde{G}^o(\tau)$ is less than σ_f, which (according to Proposition 1, Lemma 1 and 2) is certainly the case if ρ_g is either the Landau or Redheffer bound.

The interpretation is – the smaller it is necessary to choose τ the less one has to regularize the supremum-norm and the bigger function classes the decompounding works for. As $\widetilde{F}(\tau)$ is a decreasing function in τ this means the larger bound one can get on $\widetilde{F}(\tau)$ for which the reversion certainly exists, the larger function classes one can ensure the decompounding works for, see Figure 1 for an illustration of this. Actually, this resembles usual smoothness conditions put on function classes in inverse problems.

Another interpretation is that the lighter tails the p_k's have, the lesser summands will be convolved in the random sum and consequently the inverse problem will be easier to solve. Henceforth, the intuition tells us that we would expect heavier tails of the p_k's will imply stricter function classes and thereby smaller bounds on $\widetilde{F}(\tau)$ for which we can assure the decompounding to work.

This point is most easily investigated in a parametric family. We here investigate in detail the compound geometric and Poisson cases. Calculations are summarized in Table 1.

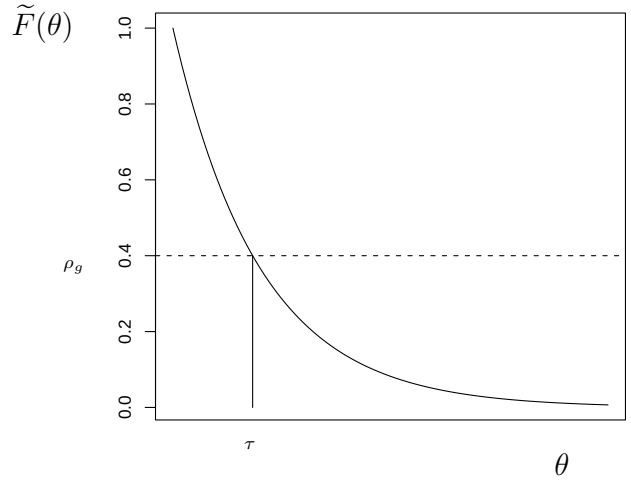


Fig. 1. Choose the smallest τ such that $\widetilde{F}(\tau) \leq \rho_g$

3.1. The geometric case

Decompounding the compound geometric distribution amounts to study the following convolution series

$$G = \sum_{k=0}^{\infty} p_k(\rho) F^{\star k},$$

where $0 \leq \rho \leq 1$ and $p_k(\rho) = (1-\rho)\rho^k$ for all $k = 0, 1, \ldots$. For the Landau bound choose τ such that:

$$\begin{aligned} \widetilde{F}(\tau) &\leq \frac{p_1(\rho)}{4} \\ &= \frac{\rho(1-\rho)}{4}. \end{aligned}$$

For the Redheffer bound, recall $A = Rp_1(\rho)$ and choose τ such that

$$\begin{aligned} \widetilde{F}(\tau) &\leq \sin(\arctan(A(1-A^2)^{-1/2}/2)) \\ &= \sin(\arctan(\rho(1-\rho)(1-(\rho(1-\rho))^2)^{-1/2}/2)). \end{aligned}$$

Now, as ρ gets bigger the more mass is put in the tail of the geometric distribution and intuition tells us the decompounding problem should be worse. This is indeed reflected in the Landau and Redheffer bounds above as we have to choose larger τ's to ensure the decompounding works.

An "efficiency" comparison can be done by dividing the bound provided by Landau (B_L) with the bound provided by Redheffer (B_R), and view it as a function of the parameter of the geometric distribution, see Figure 2, and in the heavy tail case

$$\lim_{\rho \to 1} = \frac{B_L}{B_R} = \frac{1}{2}.$$

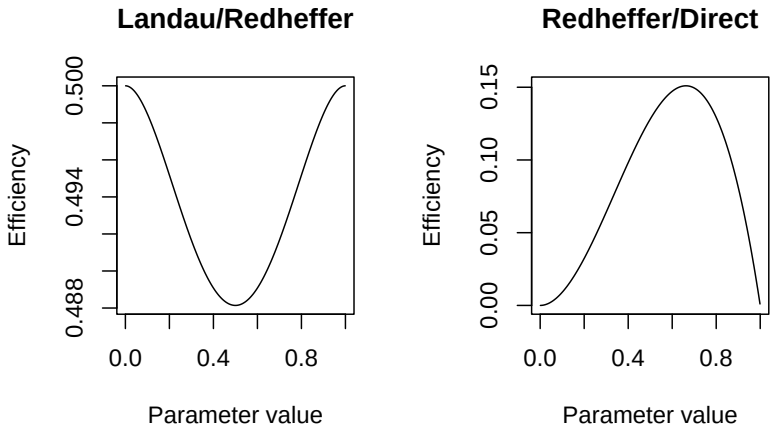


Fig. 2. For the compound geometric case. Left-hand side: Comparison of Laudau bound versus Redheffer bound. Right-hand side: Comparison of Redheffer bound versus the direct bound.

We notice that in the the heavy tail limit the Redheffer bound is 2 times "better" than the Landau bound.

In the geometric case by careful investigation of the derivatives of the $\pi_k(\rho)$'s as done in Hansen and Pitts (2006, Theorem 2) it is possible to obtain a direct bound of

$$\widetilde{F}(\tau) \leq \frac{1}{2\rho}.$$

We would expect this bound to better as it is more careful studying the properties of the geometric random sum. This is indeed so and it it most easily seen by an "efficiency" calculation where we devide the bound provided by Redheffer (B_R) with the bound provided by Hansen and Pitts (2006, Theorem 2) (B_D), see Figure 2 and in the heavy tail case

$$\lim_{\rho \to 1} \frac{B_R}{B_D} = 0.$$

3.2. The Poisson case

Decompounding the compound Poisson distribution amounts to studying the following convolution series

$$G = \sum_{k=0}^{\infty} p_k(\lambda) F^{\star k},$$

where $\lambda \geq 1$ and $p_k(\lambda) = e^{-\lambda}\lambda^k/k$ for all $k = 0, 1, \ldots$. The calculations and conclusions from the geometric case are easily carried over and summarized in Table 1 and Figure 3. A similar carefully derived bound as existing in the geometric case was derived by Buchmann and Grübel (2003, Theorem) and is given by

$$\widetilde{F}(\tau) \leq \frac{\log(2)}{\lambda}.$$

Henceforth the efficiency comparisons can also be carried over and similar conclusions to the geometric case exists.

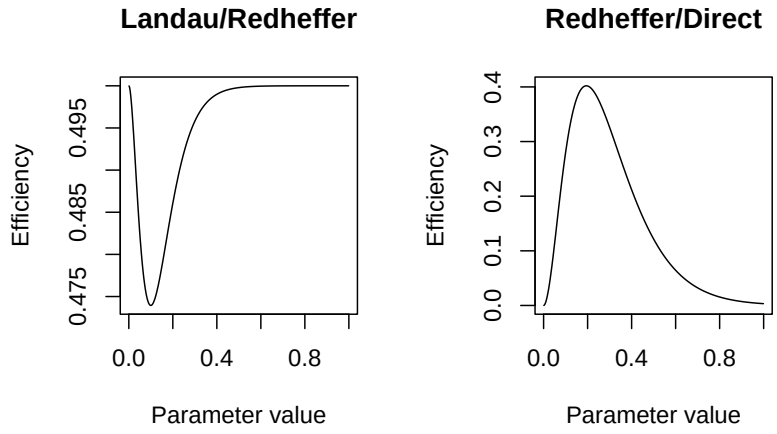


Fig. 3. For the compound Poisson case. Left-hand side: Comparison of Laudau bound versus Redheffer bound. Right-hand side: Comparison of Redheffer bound versus the direct bound.

4. Applications

Although, complex analysis provides a solution to the inverse problem, we will follow (Henrici, 1974)

> ... The algorithmic attitude to toward mathematics - not to consider a problem solved unless an algorithm for constructing the solution has been found ...

and continue in this section with various algorithmic aspects of solving the reversion in practice.

This section will discuss two ways of solving the inverse problem. In Section 4.1 we base our methods on the actual π_k's and suggest either an approach based on Fourier methods or a method based on the the direct convolutions. If it is possible to find a recursion scheme for a discretized version of the direct problem it is straightforward to reverse this scheme and provide a very powerful inversion method - this will be dealt with in Section 4.2. Finally, in Section 4.3 we conclude with a simulated example where we invert the compound Delaporte distribution and provide bootstrapped confidence bounds for the estimated claim size distribution.

4.1. Direct and Fourier methods

As mentioned in Section 2.1, the procedure to obtain the actual series can be difficult to implement in practice. We shall only give a brief account on the problems involved. The two standard methods to compute the coefficients π_k are reversion of series and the Bürmann-Lagrange Theorem. The first one requires to solve for π_k in equation (8) as described in Section 2.1.

For efficient ways to implement the recursions we refer to the comprehensive treatment in Section 1.9 of Henrici (1974). Especially Theorem 1.9(a) in Henrici (1974), states a theorem by Schur and Jabotinski which reformulates the problem of finding the reverse of a power series to a problem of calculating the reciprocal of the series. This in turn can be done efficiently by a recursion, which is a modification of the Miller algorithm see Problem

5, Section 1.8, in Henrici (1974). For a recent overview of the literature and an approach based on nested derivatives we refer to Dominici (2003).

However in statistical applications the practitioners would not bother dealing with implementing algorithms, but rather, if possible use a mathematical computation system like Maple. Fortunately it is indeed possible to reverse power series of the type (7) in Maple, by using the `solve` command. For illustration purposes let U be a shifted gamma-distribution with probability density function (pdf) given by

$$f_U(u) = \frac{\beta^\alpha}{\Gamma(\alpha)}(u-\gamma)^{\alpha-1}e^{-\beta(u-\gamma)},\ u \geq \gamma > 0.$$

Now assume $N|U$ is $pois(U)$-distributed whereby the unconditional N becomes a mixed Poisson distribution with pgf (Grandell, 1997, Example 2.2)

$$g(z) = e^{-\gamma(1-z)}\left(1+\frac{1-z}{\beta}\right)^{-\alpha} \tag{13}$$

and probability function (pf)

$$p_k = \sum_{n=0}^{k}\frac{\gamma^{k-n}}{(k-n)!}e^{-\gamma}\begin{pmatrix}\alpha+n-1\\ n\end{pmatrix}\left(\frac{\beta}{\beta+1}\right)^\alpha\left(\frac{1}{1+\beta}\right)^n. \tag{14}$$

This distribution is known in the actuarial literature as the Delaporte distribution, (Delaporte, 1959). It has been suggested as an alternative to the more usual assumption of a two-parameter gamma mixture in the theory of insurance claims. The distribution has also been fitted to several data in the literature (Ruohonen, 1988) and we take it as an example, throughout the paper. In that case we arrive at the following compound Delaporte cdf

$$G = \sum_{k=0}^{\infty}\left(\sum_{n=0}^{k}\frac{\gamma^{k-n}}{(k-n)!}e^{-\gamma}\begin{pmatrix}\alpha+n-1\\ n\end{pmatrix}\left(\frac{\beta}{\beta+1}\right)^\alpha\left(\frac{1}{1+\beta}\right)^n\right)F^{\star k}.$$

Here follows a Maple-script yielding the first 5 p_k's for the compound Delaporte distribution (discussed in Section 4.3, below):

$$> g := series\left(e^{-\gamma\,(1-z)}\left(1+\frac{1-z}{\beta}\right)^{-\alpha} - e^{-\gamma}\left(\frac{\beta+1}{\beta}\right)^{-\alpha}, z=0,6\right)$$

To limit space we have left out the actual Maple output. The result can be reversed to get the first 5 π_k's for the inverse compound Delaporte distribution

$$> solve(series(g, z=0,6) = w, z)$$

which can be simplified for the actual parameters used in Section 4.3 below.

4.1.1. Fourier method

A computationally efficient way to implement the convolutions in the direct method is to use Fourier transforms. This is motivated by noticing that one can express F as

$$F = \sum_{k=1}^{\infty}\pi_k\mathcal{F}^{-1}\left[\left(\widehat{G}^o\right)^k\right].$$

Although, a number of warnings have been reported using transform methods (see Section 3 of Buchmann and Grübel (2003) and references therein) some authors also report that using carefully designed algorithms *it is* possible to get stable inversion by transform methods (Abate and Whitt, 1992). Without resorting to these advanced algorithms we got satisfactory results using a direct approach.

Assume for now that that the π_k's have been found as described above. First choose the number of terms in the inverse expansion, K, say. Secondly, choose a discretization level $h > 0$. Thirdly, let Γ_k denote the discretized values of G_n^o

$$\Gamma_l = G_n^o((l + 0.5)h),\ l = 0, \ldots, L$$

and γ_l the discretized probability function

$$\gamma_l = \Gamma_{l+1} - \Gamma_l,\ l = 0, \ldots, L - 1$$

for some number L.

Now, an estimate of the probability function ϕ for the discretized version of F can be calculated by the fast Fourier transform (`fft`) and the inverse fast Fourier transform (`ifft`), see e.g. Bracewell (1999), as provided in most mathematical software packages,

$$\phi_l = \texttt{ifft}\left(\sum_{k=1}^{K} \pi_k \,\texttt{fft}\,(\gamma)^k\right)_l, \quad l = 0, \ldots, L.$$

This procedure is general in nature but have a number of disadvantages, as one has to choose number of terms in the expansion K and discretization level h. Theoretical as well as practical guidance needs to be developed.

For particular cases, the relationship (1) may be expressed neatly in terms of Fourier transforms. For example, if N has a Poisson distribution with mean λ, then

$$\mathcal{F}(G)(\theta) = \exp\big(\lambda((\mathcal{F}(F)(\theta) - 1\big).$$

Buchmann and Grübel (2003) refer to inherent difficulties in using transform methods to invert such an expression directly, and we have encountered such difficulties in practice. Our approach above, where the convolution series is truncated and the fast Fourier transform algorithm is used to calculate convolution powers, avoids these difficulties, but involves the choice of truncation parameter K.

4.1.2. Direct method

It should be mentioned that, although convolution can be based on fast transform algorithms as described above it is much more straightforward to code the convolutions directly, and sometimes even computationally faster. As described above we first choose the number of terms in the inverse expansion, say K. Secondly, choose a discretization level $h > 0$. Thirdly, by recalling the definition of self-convolution of empirical cumulative (sub)-distribution functions given in (2), let Γ_k denote the discretized values of G_n^o

$$\Gamma_l = G_n^o((l + 0.5)h),\ l = 0, \ldots, L$$

and γ_l the discretized probability function

$$\gamma_l = \Gamma_{l+1} - \Gamma_l,\ l = 0, \ldots, L - 1$$

for some number L. Then define the discrete convolutions

$$\Gamma_l^{\star(k+1)} = h \sum_{j=0}^{l} \Gamma_{k-j}^{\star k} \gamma_j,\ l = 0, \ldots, L.$$

Now, we have the following discrete approximation to F

$$\Phi_l = \sum_{k=1}^{K} \pi_k \Gamma_l^{\star k},\ l = 0, \ldots, L.$$

As noted above this procedure is general in nature but have the same disadvantages as the Fourier method as one has to choose number of terms in the expansion K and discretization level h. Theoretical as well as practical guidance needs further exploration.

4.2. Recursion method

As we mentioned the direct and Fourier methods have some clear disadvantages, but fortunately in some cases an alternative exists. The idea is to discretize the claim size distribution, whereby the compound distribution also becomes discrete. First, choose a disretization level h. Secondly, let f_k denote the mass given by F to the interval $((k-0.5)h, (k+0.5)h]$, in the following way

$$f_k = F((k+0.5)h) - F((k-0.5)k),$$

and let the discrete distribution that gives mass f_k to the point kh, $k = 0, 1, 2, \ldots$, be an approximation to the distribution F. Assume now that the random variables $Y, X_1, X_2, X_3, \ldots$ from the introduction satisfies

$$\begin{aligned} f_k &= \mathrm{P}(X_i = k), \\ g_k &= \mathrm{P}(Y = k) \end{aligned}$$

and $f_k^m = \mathrm{P}(X_1 + \cdots + X_m = k)$ or

$$f_k^{\star,m+1} = \sum_{j=0}^{k} f_j f_{k-j}^{\star,m}.$$

In many cases g_k can be expressed as a linear combination of $g_1, \ldots, g_{k-1}$ and $f_0, \ldots, f_{k-1}$ whereby providing a recursive algorithm for calculating the discrete compound distribution g_k. In the seminal work of Panjer (1981) the following general algorithm

$$g_k = \frac{1}{1 - a f_0} \sum_{j=1}^{k} \left(a + \frac{bj}{k} \right) f_j g_{k-j}$$

were derived for situations where the following assumption holds

$$p_{m+1} = \left(a + \frac{b}{m+1} \right) p_m,\ m = 0, 1, 2, \ldots.$$

Actually, the Panjer recursion only applies to the Poisson, the negative binomial and the binomial distribution. However, there have since Panjer's work been an intensive interest in

this and various extensions have been developed for distributions belonging to the so-called Willmot class.

All, these algorithms can easily be inverted, and provides thereby what we could call inverse Panjer recursion, and subsequently be used for inverse decompounding. For the Panjer class we easily arrive at

$$g_0 = g(z) + p_0 \tag{15}$$

$$f_k = \frac{1}{(a+b)g_0}\left((1 - af_0 - \sum_{i=1}^{k-1}\left(a + \frac{ib}{k}\right) f_i g_{k-i}\right), \tag{16}$$

where it might be necessary to implement a numerical routine to invert (15). Actually, the Poisson case was used in Buchmann and Grübel (2003) and the geometric case in Hansen and Pitts (2006) to invert Poisson and geometric random sums, respectively.

For a comprehensive overview of recursive algorithms for discrete compound distributions see Section 8.4 of Grandell (1997). In particular for the discrete compound Delaporte distribution Schröter (1990) (see also Grandell (1997), Example 8.6) derived the following recursion

$$(1+\beta)g_k = \sum_{j=0}^{k}\left(\left(1 + \frac{(\alpha - 1 + \gamma\beta + \gamma)j}{k}\right) f_j - \frac{\gamma j}{2k} f_j^{\star 2}\right) g_{k-j}, \text{ for } k = 1, 2, 3, \ldots, \tag{17}$$

which in turn can be inverted to give

$$\begin{aligned} g_0 &= e^{-\gamma(1-f_0)}\left(1 + \frac{1-f_0}{\beta}\right)^{-\alpha} \\ f_k &= \frac{1}{(\alpha + \gamma\beta + \gamma - \gamma f_0)g_0}\left[(1 + \beta - f_0)g_k + \frac{\gamma g_0}{2}\sum_{i=1}^{k-1} f_i f_{k-i}\right. \\ &\quad \left. - \sum_{j=1}^{k-1}\left(\left(1 + \frac{\alpha - 1 + \gamma\beta + \gamma}{k} j\right) f_j - \frac{\gamma j}{2k} f_j^{\star 2}\right) g_{k-j}\right] \text{ for } k = 1, 2, 3, \ldots \end{aligned}$$

4.3. Example: Inverse Compound Delaporte distribution

All calculations in the example have been carried out in the statistical software package **R**, except the π_k's, which have been calculated in Maple. To save space the actual scripts are left out, but can be obtained from `http://www.math.aau.dk/~mbh/Mysoftware`

To check performance of the proposed procedure, we applied it to various simulated data sets. Results from a typical case is summarized in this section. As mentioned in Section 3 it is possible in many standard cases to get explicit expressions for the π_k's and develop a theory along the lines of Buchmann and Grübel (2003) and Hansen and Pitts (2006). Let us now develop an example where the approach with explicit expressions are not obvious.

In what follows the discretization level was chosen to be $h = 0.01$ and number of discretization points $L = 2^{16}$. Furthermore, an iid sample of size 500 was drawn from a compound Delaporte distribution with parameters $\alpha = 1$, $\beta = 1$, $\gamma = 0.5$ and exponentially distributed claim sizes with mean 1.

We now consider the parameters α and β as fixed at 1 and γ at 0.5, but we have to estimate nonparametrically the claim size distribution. This corresponds to the interpretation (Ruohonen, 1988) that the risk is composed of an underlying population risk which is

Table 2. Estimated π_k's.

k	1	2	3	4	5	6	7
π_k	3.3	-6.8	16.1	-41.2	111.1	-310.3	888.1
k	8	9	10	11	12	13	14
π_k	-2587.7	7647.1	-22851.8	68912.5	-209391.2	640306.7	-1968726.2

Poisson distributed with a parameter γ and a time-varying risk which is known and gamma distributed with parameters α and β.

Remark 2 ensures that there exists a τ such that the asymptotic results of Theorem 1, 2 and 3 holds. Maple provided the first 14 p_k's by the command:

$$> g \quad := \quad series(exp(-.5 * (1+z)/(2-z) - .5 * exp(-.5), z = 0, 15);$$

which again can be reversed to get the first 14 π_k's for the inverse compound Delaporte distribution

$$> solve(series(g, z = 0, 6) = w, z, 15);$$

The result is summarized in Table 2

The full-drawn line at the left-hand side of Figure 4 shows the ecdf for an iid sample of size 500 from the compound Delaporte distribution with parameters $\alpha = \beta = 1$, $\gamma = 0.5$ and exponentially distributed claim sizes with mean 1. The dashed line shows the underlying theoretical cdf, calculated via the direct Panjer recursion provided in (17) and the discretization parameters given in the previous paragraph. At the left-hand side of Figure 4 the inversion has been carried out by the Fourier method with $K = 14$ and the above-mentioned discretization parameters. The dashed line shows the exponential distribution with parameter 1.

In Figure 5 (left-hand side) the full-drawn grey line shows an estimator based on inverse Panjer recursion, with discretization parameters chosen as above.

Theorem 3 refers to confidence regions in the space $D_\tau[0,\infty)$, and calculating these in practice require the user to make a choice of a suitable τ. The bounds in Section 3 provide some help with this choice. For the example here, we have taken a pragmatic approach, and constructed a bootstrap confidence band using the $\|\cdot\|_\infty$-norm, as this is easy to construct in practice (although not supported by our theorems).

Figure 5 (right-hand side) shows, the 90% bootstrapped confidence bounds (with respect to the sup-norm) with $N = 100$. In both plots the dashed line shows the exponential distribution with mean 1.

5. Discussion

In the discussion of Buchmann and Grübel (2003) they claim their basic problem (decompounding Poisson random sums) is amenable to generalizations within stochastic processes. We believe to have done so in the previous sections by indicating the problem can be solved in great generality. Moreover, we have directed the attention to a number of real-life applications in e.g. control of queuing systems, infinite capacity storage models, insurance mathematics, network statistics as well as biology. We also showed how Panjer recursion can be inverted for the Panjer as well as the Willmot class, hereby providing a very strong tool to discretize and solve the inverse problem at hand. Finally, we developed by means of Maple and R effective tools to decompound general random sums.

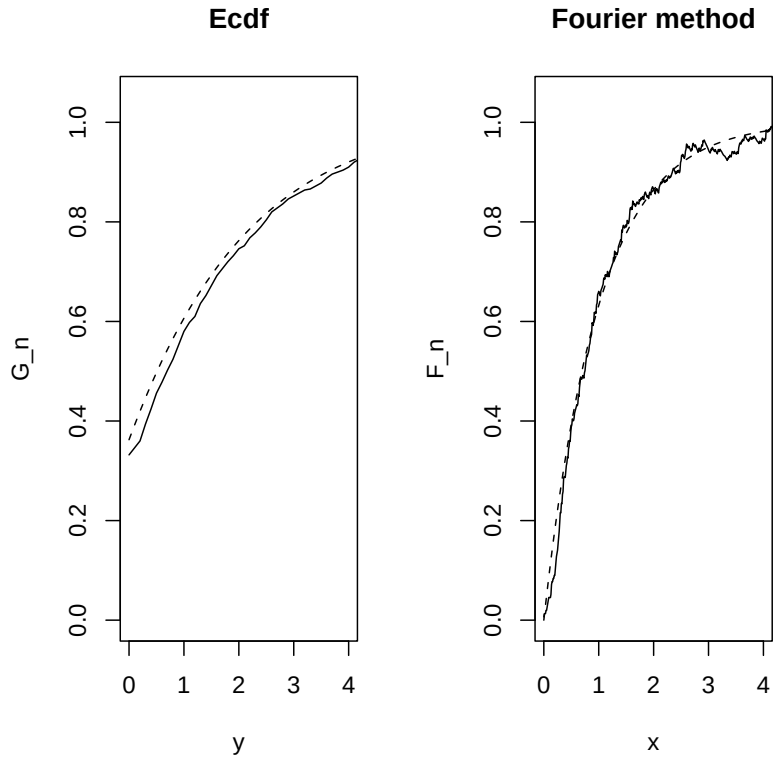


Fig. 4. Left-hand side: Estimate of the compound claim size distribution with exponentially distributed claim sizes with mean one, $n = 500$, $\alpha = \beta = 1$ and $\gamma = 0.5$. Right-hand side: Inversion by the Fourier method. The dashed lines show the compound Delaporte and the exponential distribution with mean 1, respectively.

Despite the general set-up there are still a number of unsolved problems for future research. One may e.g. wonder what happens when the first $n-1$ derivatives of g vanish at 0, i.e. $p_i = 0$ for $i \leq n-1$ and $p_n \neq 0$. In this case, the equation $w = g(z)$ has a solution

$$z = f(w) = \sum_{k=1}^{\infty} \pi_k w^{k/n}, \tag{18}$$

where the series of powers of $w^{1/n}$ converges in a neighbourhood of 0. Henceforth, f is n-valued function of w, having a branch-point at 0 (Copson, 1935, p. 123).

This corresponds to solving the inverse problem

$$G = p_n F^{\star n} + p_{n+1} F^{\star(n+1)} + \cdots \tag{19}$$

for $p_n \neq 0$. Hence the plug-in estimator is given by

$$F_n = \sum_{k=1}^{\infty} \pi_k (G_n)^{\star k/n}.$$

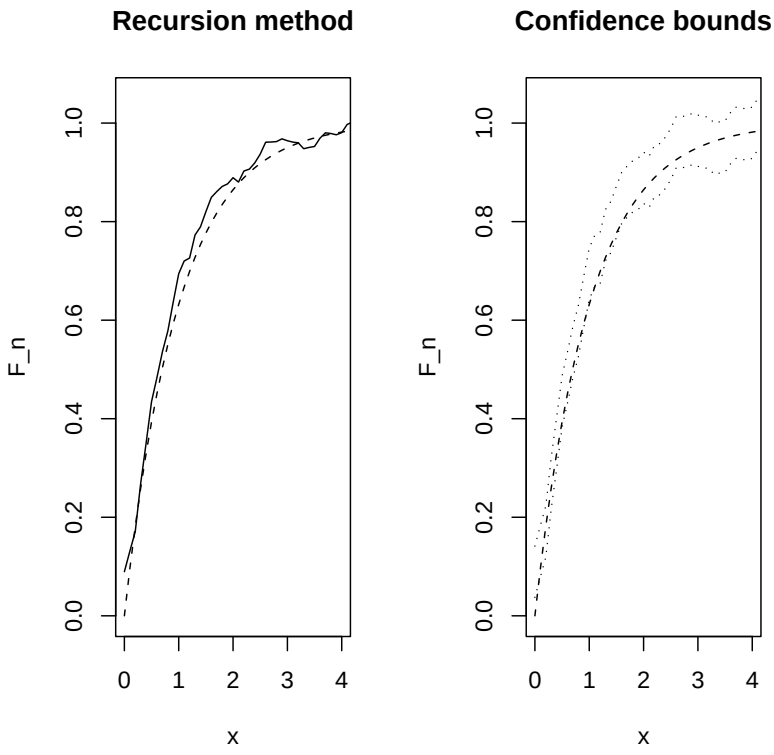


Fig. 5. Left-hand side: Inversion by the recursion method. Right-hand side: 90% confidence bounds. The dashed line shows the exponential distribution with mean 1.

This makes sense, if we define fractional convolution as

$$f^{\star\alpha} = \mathcal{L}^{-1}(\mathcal{L}(f)^{\alpha}),\ \alpha \in \mathbb{R},$$

whenever the Laplace transforms are well-defined, as the inverse of (19) can be derived by Laplace transformation

$$\widetilde{G} \quad = \quad p_n \widetilde{F}^n + p_{n+1}\widetilde{F}^{(n+1)} + \cdots$$

hence by means of (18) we arrive at

$$\begin{aligned} F &= \sum_{k=1}^{\infty} \pi_k \mathcal{L}^{-1}(\widetilde{G}^{k/n}). \\ &= \sum_{k=1}^{\infty} \pi_k G^{\star k/n}. \end{aligned}$$

However, in this case we do not know of any lower bound on the radius of convergence of f, and we are forced to either develop these or do a case by case analysis.

In passing it may be worth noting, that we have defined fractional convolution for any $\alpha \in \mathbb{R}$ as

$$H = F^{\star\alpha} \;\Leftrightarrow\; \mathcal{L}(H) = \mathcal{L}(F)^{\alpha}$$

whenever the Laplace transforms are well-defined. In fact such dualities are the basis for deriving fractional operations in general. In recent years, there has been put an enormous effort in the field of *fractional analysis*. We will not pursue this subject further, but refer the interested reader to Bultheel and Martínez-Sulbaran (2003), where a comprehensive overview of the field and its applications in signal processing, geometry, optics, mechanics, stochastic processes etc. is provided.

6. Proofs

Proof of Theorem 1. The proof of Theorem 1 uses the approach of Buchmann and Grübel (2003), see also Hansen and Pitts (2006) and applies it to reversion of power series. In common with the above papers, our functional Λ involves convolutions of distribution functions, and so we summarise here the spaces introduced in Buchmann and Grübel (2003) for functions that may be used as integrators in convolution integrals. Let $D(\infty) = \cup_{\tau>0} D_\tau[0,\infty)$ and let $D_{\mathrm{m}}(\infty)$ be the subset of $D(\infty)$ consisting of functions having finite variation on $[0,x]$ for all $x > 0$. Then a function H in $D_{\mathrm{m}}(\infty)$ corresponds to a signed measure μ_H given by $\mu_H\big([0,x]\big) = H(x)$. For H in $D_{\mathrm{m}}(\infty)$ and g in $D(\infty)$ we may define the convolution $g \star H(x) = \int g(x-y)H(dy)$ for $x \geq 0$. As observed in Buchmann and Grübel (2003), members of $D_{\mathrm{m}}(\infty)$ are identified via their Laplace transforms. Any H in $D_{\mathrm{m}}(\infty)$ is also in $D(\infty)$ and hence is in $D_\tau[0,\infty)$ for some $\tau > 0$. We note that the Laplace transform $\widetilde{H}(\theta)$ is then defined for $\theta > \tau$.

We need a technical result about convolutions from Lemma 6 of Buchmann and Grübel (2003), which we state without proof. Let $D^+_{\mathrm{m}}(\infty)$ be the subset of $D_{\mathrm{m}}(\infty)$ consisting of functions H such that the measure μ_H is non-negative. Then

$$\|g \star H\|_{\infty,\tau} \;\leq\; \|g\|_\infty \widetilde{H}(\tau) \text{ for all } \tau > 0. \tag{20}$$

Turning to the convolution series given by $\Lambda(G)$, we first note that, for any distribution function G on $[0,\infty)$, we have $\sum_{k=1}^m \pi_k \big(G^o\big)^{\star k}$ is in $D_\tau[0,\infty)$ for each finite m. Using (20) we have for $k \geq 1$,

$$\big\|(G^o)^{\star k}\big\|_{\infty,\tau} \;\leq\; \|G^o\|_{\infty,\tau} \widetilde{G}^o(\tau)^{k-1} \;\leq\; \widetilde{G}^o(\tau)^{k-1}.$$

Then it follows that

$$\sum_{k=1}^{\infty} |\pi_k| \big\|(G^o)^{\star k}\big\|_{\infty,\tau} \;\leq\; \sum_{k=1}^{\infty} |\pi_k| \widetilde{G}^o(\tau)^{k-1}.$$

It is assumed in the statement of the theorem that $\widetilde{G}^o(\tau) < \sigma_f \; (\leq r(f))$, so that

$$\sum_{k=1}^{\infty} |\pi_k| \widetilde{G}^o(\tau)^{k-1}$$

converges. Thus $\Lambda(G)$ is in $D_\tau[0,\infty)$.

To obtain the relationship between F and G, first note that the series $\sum_{k=1}^{\infty} \pi_k (G^o)^{\star k}$ can be written as the difference of two non-decreasing functions:

$$\sum_{k=1}^{\infty} \pi_k (G^o)^{\star k} \;=\; \sum_{k:\pi_k \geq 0} \pi_k (G^o)^{\star k} - \sum_{k:\pi_k < 0} |\pi_k| (G^o)^{\star k},$$

and so $\Lambda(G)$ is in $D_{\mathrm{m}}(\infty)$. This means that $\Lambda(G)$ is identified by its Laplace transform.

From the definition of G in terms of F, we have, for $\theta > \tau$,

$$\widetilde{G}^o(\theta) \;=\; \sum_{k=1}^{\infty} p_k \widetilde{F}(\theta) \;=\; g\big(\widetilde{F}(\theta)\big).$$

We have $\widetilde{G}^o(\theta) < \sigma_f$ and thus, by (1) there exists a unique z with $|z| < \sigma_g$ such that $g(z) = \widetilde{G}^o(\theta)$, and this z is given by the power series reversion, ie $z = \sum_{k=1}^{\infty} \pi_k \big(\widetilde{G}^o(\theta)\big)^k$. Since $\widetilde{F}(\theta) < \sigma_g$, this implies that $z = \widetilde{F}(\theta)$ and that

$$\widetilde{F}(\theta) \;=\; \sum_{k=1}^{\infty} \pi_k \big(\widetilde{G}^o(\theta)\big)^k.$$

The right-hand side of this last expression is $\widetilde{\Lambda(G)}(\theta)$, from which we conclude that $F = \Lambda(G)$, and Theorem 1 is proved.

Proof of Theorem 2. This is an application of the infinite-dimensional delta method, which combines an appropriate differentiability property of the map Λ with the relevant asymptotic normality result for the empirical process associated with the input estimators G_n, $n \geq 1$ (see, Gill (1989), van der Vaart (1998)). The proposition below gives an appropriate differentiability result for Λ.

Proposition 2. *Suppose that G_n, $n \in \mathbb{N}$, and G are distribution functions with $G_n(0) = G(0) = 0$ for all $n \in \mathbb{N}$. Let $\tau > 0$ be such that $\widetilde{G}(\tau) < \sigma_f$ and $\widetilde{F}(\tau) < \sigma_g$, and suppose that*

$$\|\sqrt{n}\big(G_n - G\big) - h\|_{\infty} \to 0 \text{ as } n \to \infty$$

for some h in $D[0,\infty)$. Then

$$\|\sqrt{n}\big(\Lambda(G_n) - \Lambda(G)\big) - h \star H\|_{\infty,\tau} \to 0 \text{ as } n \to \infty,$$

where $H = \sum_{k=1}^{\infty} k\pi_k G^{\star(k-1)}$.

Comparing this proposition with Proposition 8 in Buchmann and Grübel (2003) and Proposition 4 in Hansen and Pitts (2006), it is clear that it is the relevant generalisation of those differentiability results to the general decompounding case considered here.

The proof of Proposition 2 follows similar methods to the proofs of the corresponding propositions in the above earlier papers, with η in those proofs replaced here by $\eta \in (0,1)$ chosen such that $\widetilde{G}(\tau) < \eta\sigma_f$, and then using the fact that the series $\sum_{k=1}^{\infty} \pi_k (\eta\sigma_f)^k$ is absolutely convergent. We refer the reader to the earlier papers for further details.

The second ingredient for the infinite-dimensional delta method is an asymptotic normality result for the input estimators, which is given by the empirical central limit theorem (Pollard (1984) V.2.11). This states that

$$\sqrt{n}\big(G_n - G\big) \rightarrow_{\mathcal{D}} B \circ G \text{ in } (D[0,\infty), \|\cdot\|_\infty),$$

where B is a standard Brownian bridge and $B \circ G$ is a rescaled Brownian bridge given by $B \circ G(t) = B(G(t))$, $t \geq 0$. The limit process $B \circ G$ is thus a zero mean Gaussian process with

$$\text{cov}(B \circ G(s), B \circ G(t)) \;=\; G(s \wedge t) - G(s)G(t).$$

The empirical central limit result is then combined with Proposition 2 to yield Theorem 2, see Buchmann and Grübel (2003) for details. The covariance structure of the limiting Gaussian process A in Theorem 2 arises easily, as in Buchmann and Grübel (2003).

Proof of Theorem 3. The proof proceeds by first showing that $\hat{R}_n$ converges in distribution to R in $D[0,\infty)$, and then using this to infer the result of the theorem. The technicalities are the same as those in the proof of Theorem 2.3 in Grübel and Pitts (1993), and the interested reader is referred to that paper.

Acknowledgment

The authors thank Nick Bingham, University of Sheffield, for a most helpful discussion.

References

Abate, J. and W. Whitt (1992). The Fourier-series method for inverting transforms of probability distributions. *Queueing Systems 10*, 5–88.

Asmussen, S. (2003). *Applied Probability and Queues* (2 ed.). Springer Verlag, New York, NY.

Bingham, N. H. and S. M. Pitts (1999a). Nonparametric estimation for the $M/G/\infty$ queue. *Annals of the Institute of Statistical Mathematics 51*, 71–97.

Bingham, N. H. and S. M. Pitts (1999b). Nonparametric inference from $M/G/1$ busy periods. *Communications in Statistics and Stochastic Models 15*, 247–272.

Bracewell, R. N. (1999). *The Fourier Transform and its Applications* (3 ed.). New York: McGraw-Hill.

Buchmann, B. and R. Grübel (2003). Decompounding: An estimation problem for Poisson random sums. *Annals of Statistics 31*, 1054–1074.

Bultheel, A. and H. Martínez-Sulbaran (2003). Recent developments in the theory of the fractional Fourier transforms and linear canonical transforms. *Bulletin of the Belgian mathematical Society- Simon Stevin*. Accepted.

Copson, E. T. (1935). *An Introduction to the Theory of Functions of a Complex Variable*. Oxford: The Clarendon Press.

Delaporte, P. (1959). Quelques problemés de statistique mathématique posés par l'assurance automobile et le bonus pour non sinistre. *Bulletin Trimestriel de l'Institut des Actuaires Francais 227*, 87–102.

Dominici, D. (2003). Nested derivatives: A simple method for computing series expansions of inverse functions. *International Journal of Mathematics and Mathematical Sciences 58*, 3699–3715.

Gill, R. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method (Part 1). *Scandinavian Journal of Statistics 16*, 97–128.

Grandell, J. (1997). *Mixed Poisson Processes*. London: Chapman and Hall.

Grübel, R. and S. M. Pitts (1993). Non-parametric estimation in renewal theory, I: the empirical renewal function. *Annals of Statistics 21*, 1431–1451.

Hall, P. and J. Park (2004). Nonparametric inference about service time distribution from indirect measurements. *Journal of the Royal Statistical Society 66*, 861–875.

Hansen, M. B. and S. M. Pitts (2006). Nonparametric inference from the $M/G/1$ workload. Accepted for publication in Bernoulli.

Henrici, P. (1974). *Applied and Computational Complex Analysis, Vol 1: Power Series - Integration - Conformal Mapping - Location of Zeros*. New York: John Wiley and Sons.

Hohn, N. and D. Veitch (2003). Inverting sampled traffic. In *Proceedings of the 3rd ACM SIGCOMM conference on Internet measurement*, Miami Beach, FL, USA, pp. 222–233. ACM Press New York, NY, USA.

Johnson, N. L., A. W. Kemp, and S. Kotz (2005). *Univariate Discrete Distributions* (3 ed.). New York: John Wiley and Sons.

Levental, S. (1988). Uniform limit theorems for Harris recurrent Markov chains. *Probability Theory and Related Fields 80*, 101–118.

Neyman, J. and E. L. Scott (1958). A statistical approach to problems of cosmology. *Journal of the Royal Statistical Society, Series B 20*, l–43.

Panjer, H. H. (1981). Recursive evaluation of a family of compound distributions. *Astin Bulletin 12*, 22–26.

Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer Verlag.

Redheffer, R. M. (1962). Reversion of power series. *American Mathematical Monthly 69*, 423–425.

Ruohonen, M. (1988). On a model for the claim number process. *Astin Bulletin 18*, 57–68.

Schröter, K. (1990). On a family of counting distributions and reccursions for related compound distributions. *Scandinavian Actuarial Journal 304*, 161–175.

Sharma, V. (1999). Estimating traffic intensities at different nodes in networks via a probing stream. In *GLOBECOM 1999: Seamless Interconnection for Universal Services*, pp. 374–380. Piscataway, NJ: IEEE Operations Center.

Sharma, V. and R. Mazumdar (1998). Estimating traffic parameters in queueing systems with local information. *Performance Evaluation 32*, 217–230.

Tsai, T. H. (1998). *The uniform CLT and LIL for Markov chains*. Ph. D. thesis, University of Wisconsin, Madison, WI.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

van Es, B., S. Gugushvili, and P. Spreij (2005). A kernel type nonparametric density estimator for decompounding. URL: front.math.ucdavis.edu/math.ST/0505355.

Wand, M. P. and M. Jones (1995). *Kernel Smoothing*, Volume 60 of *Monographs on Statistics and Applied Probability*. Chapman and Hall.

Zhang, C. (1990). Fourier methods for estimating mixing densities and distributions. *Annals of Statistics 18*, 806–831.