Ann Inst Stat Math (2010) 62:855–872
DOI 10.1007/s10463-008-0200-6

Decompounding random sums: a nonparametric approach

Martin Bøgsted · Susan M. Pitts

Received: 16 February 2007 / Revised: 23 May 2008 / Published online: 6 November 2008

Abstract A compound distribution is the distribution of a random sum, which consists of a random number N of independent identically distributed summands, independent of N. Buchmann and Grübel (Ann Stat 31:1054–1074, 2003) considered decompounding a compound Poisson distribution, i.e. given observations on a random sum when N has a Poisson distribution, they constructed a nonparametric plug-in estimator of the underlying summand distribution. This approach is extended here to that of general (but known) distributions for N. Asymptotic normality of the proposed estimator is established, and bootstrap methods are used to provide confidence bounds. Finally, practical implementation is discussed, and tested on simulated data. In particular we show how recursion formulae can be inverted for the Panjer class in general, as well as for an example drawn from the Willmot class.

Keywords Asymptotic normality · Compound distributions · Decompounding · Empirical processes · Functional central limit theorem · Infinite-dimensional delta method · Inverse problems

1 Introduction

This paper deals with nonparametric estimation in the context of decompounding of random sums. Decompounding is an example of an inverse problem, and so a natural

M. Bøgsted
Department of Haematology and Center for Cardiovascular Research, Aalborg Hospital, Aarhus University Hospital, Sdr. Skovvej 15, 9000 Aalborg, Denmark
e-mail: martin.boegsted@rn.dk

S. M. Pitts (✉)
Statistical Laboratory, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WB, UK
e-mail: S.Pitts@statslab.cam.ac.uk

starting point is to describe the corresponding "direct" problem of compounding. A compound distribution is the distribution of a random sum $Y = X_1 + \cdots + X_N$, where $X_1, X_2, X_3, \ldots$ are independent identically distributed (iid) random variables with distribution function F, say, and N is a discrete random variable, independent of the X_i's, taking values in $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$. If we let $p_k = \mathrm{P}(N = k)$, $k \in \mathbb{N}_0$, then the distribution function of Y is $G = \sum_{k=0}^{\infty} p_k F^{\star k}$, where $\star k$ denotes k-fold convolution. We assume throughout this paper that the p_k's are known. In this case, once F is known, then G is completely determined, and so we may write

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

where Ψ is the functional that maps a distribution function F onto the corresponding compound distribution G. Decompounding is the inverse problem to this, i.e., supposing that G is given, can we determine F, and if so, how? From (1), decompounding requires the inverse functional, Λ, to Ψ. We shall see that, under conditions to be made precise later, (1) may be inverted in the following way,

$$F = \Lambda(G) = \sum_{k=1}^{\infty} \pi_k (G^o)^{\star k}, \tag{2}$$

where the π_k's can be expressed in terms of the p_k's, and where, for any function h, we write h^o for the function $x \mapsto h(x) - h(0)$. Clearly, convergence of the convolution series in (2) needs to be established formally, and we do this in an appropriate function space below.

For the statistical problem, suppose that F and G are both unknown, that interest lies in inference about F, and that observations $Y_1, \ldots, Y_n$ are available from G. We use these observations to construct an estimator G_n of G, and our proposed estimator F_n of F is obtained by plugging G_n into the decompounding functional so that $F_n = \Lambda(G_n)$.

Our main result is an asymptotic normality result for F_n in terms of convergence in distribution to a Gaussian process. This is proved by combining together two separate components. The first is an appropriate asymptotic normality result for the estimator G_n, and the second is a differentiability property for Λ, so that, roughly, Λ may be approximated locally by a linear map. This means that the asymptotic normality of the input estimator G_n is carried over to a corresponding asymptotic normality property for the output quantity $F_n = \Lambda(G_n)$. This is the well-known infinite-dimensional delta method (see for example Gill 1989 and van der Vaart 1998), and the necessary details and proofs for the particular functional Λ are given in later sections of this paper.

We note in passing that, once the differentiability of Λ is established, then we will automatically have asymptotic normality for F_n constructed for any choice of estimator G_n for which an appropriate asymptotic normality result in the relevant function space holds. In subsequent sections of this paper we assume for simplicity of exposition that $Y_1, \ldots, Y_n$ are iid observations of G and we let G_n be the empirical distribution function,

$$G_n(x) = n^{-1} \sum_{i=1}^{n} 1_{[Y_i,\infty)}(x), \quad 0 \le x < \infty,$$

where $1_B(x) = 1$ if x is in the set B and is zero otherwise. However, we see below that asymptotic normality may hold for other choices, including estimators constructed from non-iid data.

The above approach follows, in a more general context, that of Buchmann and Grübel (2003), who coined the term "decompounding." They consider the decompounding problem when N has a Poisson distribution and the observations on G are iid, and they introduce an appropriate functional set-up. Hansen and Pitts (2006) tackle the case where N has a geometric distribution and where the observations on G exhibit regenerative structure.

Decompounding arises in various situations. We briefly mention some applications here, and refer the reader to the references for further details. Hansen and Pitts (2006) consider decompounding for a geometric random sum in a setup motivated by a standard tool in performance evaluation of communication networks. This involves probing the system to obtain observations of the workload and using the resulting data to make inferences about the job-size distribution. The results in Hansen and Pitts (2006) are immediately transferrable to the infinite storage model that corresponds to the $M/G/1$ queue (see Asmussen 2003, Sect. XIV.1), where the service-times and workload correspond to the storage amount and the inputs to the storage facility respectively. Random sums also arise in Neyman–Scott cluster processes (see Neyman and Scott 1958), giving rise to decompounding problems. Another application, concerns a binomial deconvolution problem arising in a network traffic setting (see Hohn and Veitch 2003).

The infinite-dimensional delta method has been used for other inverse estimation problems in queueing theory in Bingham and Pitts (1999a,b), where data on busy and idle periods are used to make inferences about the service time distribution. Alternative approaches are given in Hall and Park (2004), where kernel densities are used in estimation of the service time density from busy period data, and van Es et al. (2007), which has a density approach to Poisson decompounding.

The paper is organised as follows. In Sect. 2 (sub-)probability generating functions and some necessary background from complex analysis are reviewed. This insight is used to define the estimator and show it is well defined. The section concludes with statements of the asymptotic normality for F_n as well as stating asymptotic properties of the proposed bootstrap confidence region. Section 3 contains comparisons of various bounds. Some algorithms for dealing with the general approach to decompounding are introduced in Sect. 4, where an example of the estimator in action is also given. Discussion is provided in Sect. 5. Outlines of proofs related to the properties of the estimator are given in Sect. 6. These broadly follow those in Buchmann and Grübel (2003) and Hansen and Pitts (2006), making adaptations for the general case as appropriate.

2 Estimation

2.1 Background on sub-probability generating functions

We write $\widetilde{A}$ for the Laplace–Stieltjes transform of a distribution function A, where $\widetilde{A}(\theta) = \int_{[0,\infty)} e^{-\theta x} A(dx)$ for those $\theta \in \mathbb{R}$ for which the integral is defined. Then, from (1), we typically have

$$\widetilde{G}(\theta) = \sum_{k=0}^{\infty} p_k (\widetilde{F}(\theta))^k = g_N\left(\widetilde{F}(\theta)\right),$$

where $g_N(z) = \sum_{k=0}^{\infty} p_k z^k$ is the probability generating function of N. Thus the compounding functional is related to power series, and inversion of (1) corresponds to reversion of power series, where in line with the traditions of complex analysis we use the term “reversion” for finding the power series of the inverse with respect to composition of power series see Henrici (1974, p. 47). It is standard to consider reversion of power series with $p_0 = 0$ and so we consider reversion of the power series

$$g(z) = \sum_{k=1}^{\infty} p_k z^k, \quad z \in \mathbb{C}. \tag{3}$$

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, Sect. 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 $z = f(g(z))$, i.e.

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

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).

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 that* $|w| < \sigma_f$ *implies that* $z = f(w) = g^{-1}(w) = \sum_{k=1}^{\infty} \pi_k w^k$ *and* $|f(w)| < \sigma_g$.
(d) *There exists* ρ_g *with* $0 < \rho_g \leq \sigma_g$ *such that*

$$|z| < \rho_g \quad \text{implies that} \quad |g(z)| < \sigma_f, \tag{4}$$

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

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 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*

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

for $0 < r < R$, *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 $\sigma_f = (p_1 R)^2(1-(3/4)A^2)^{-1/2}/(4M)$.

Remark 1 If the sub-probability generating function 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 (3) (take $R = 1$ and $M = 1$) $\rho_g = \sin(\arctan(p_1(1-p_1^2)^{-1/2}/2))$ and $\sigma_f = p_1^2(1-(3/4)p_1^2)^{-1/2}/4$.

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). We assume that $p_1 > 0$ (see Proposition 1 and the discussion in Sect. 5).

Turning now to the general decompounding problem, observe that (1) implies that $G^o = G - p_0 = \sum_{k=1}^{\infty} p_k F^{\star k}$, so that, given G, we might expect that F is given by the reversed convolution power series $\sum_{k=1}^{\infty} \pi_k (G^o)^{\star k}$. This must be made precise, and in particular, function spaces and norms for G^o must be defined in order to ensure convergence of the reversed convolution power series. Since we are extending the approach of Buchmann and Grübel (2003) and Hansen and Pitts (2006) to general $\{p_k\}$, we use the same weighted spaces used in these two papers. Let $D[0, \infty)$ be the Banach space of cadlag functions h on $[0, \infty)$ such that h has a finite limit at infinity. We define a family of exponentially weighted D-spaces indexed by τ in $\mathbb{R}$, which have the property that the associated weighted norm of G^o decreases as τ tends to infinity, so that appropriate choice of τ allows for the weighted norm to be small enough for convergence of the reversed convolution power series in the weighted space. To define these spaces, 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 $x \mapsto \mathrm{e}^{-\tau x} h(x)$ is in $D[0, \infty)$. For h in $D_\tau[0, \infty)$, let $\|h\|_{\infty,\tau} = \sup_{x \geq 0} \mathrm{e}^{-\tau x} |h(x)|$, 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 (2) 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 $\widetilde{G}^o(\tau) \left(= \int_{[0,\infty)} \mathrm{e}^{-\tau x} G^o(dx)\right) < \sigma_f$. 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 $G = \Psi(F)$ implies that $F = \Lambda(G)$.

Note that, with ρ_g defined as in (4), $\widetilde{F}(\tau) < \rho_g$ $(\leq \sigma_g)$ 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. Thus the conditions on $\widetilde{F}(\tau)$ and $\widetilde{G}^o(\tau)$ in the statement of the theorem may be replaced by the condition $\widetilde{F}(\tau) < \rho_g$.

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

Theorem 1 makes precise the informal representation of the decompounding functional in (2). Since $\widetilde{G}_n^o(\tau) \to \widetilde{G}^o(\tau)$ as $n \to \infty$ with probability one, then, the plug-in estimator $F_n = \Lambda(G_n)$ is defined with probability tending to 1, when n tends to infinity. We observe here that the π_k's are not necessarily postive and so, in common with the estimators in Buchmann and Grübel (2003) and Hansen and Pitts (2006), the proposed estimator need not be a distribution function itself (see also Sect. 5).

2.3 Properties of the estimator

In this section, we state two results, the first 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 closed balls in the respective norms, see Pollard (1984), p. 199. There are other ways of dealing with measurability issues with empirical processes, in particular the technique of using outer probabilities (see e.g. van der Vaart and Wellner 1996). However, whichever is used, the key steps in dealing with our particular estimation problem involve defining the estimator and proving its differentiability, which is what we do here. The closed ball sigma field is the one used in Buchmann and Grübel (2003). We now state the main result on weak convergence of the inverse estimator of the distribution function F.

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

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

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

$$\begin{aligned} \operatorname{cov}(A(s), A(t)) = \int\int & G^o\left((s-x)\wedge(t-y)\right) H(dx)H(dy) \\ & - G^o \star H(s) G^o \star (H)(t), \quad s, t \geq 0, \end{aligned}$$

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

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

$$R_n(z) = \mathrm{P}\left(\sqrt{n}\|F_n - F\|_{\infty,\tau} \leq z\right), \quad \text{and} \quad R(z) = \mathrm{P}\left(\|A\|_{\infty,\tau} \leq z\right).$$

From 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 $\|F_n - F\|_{\infty,\tau} \leq q(\alpha)/\sqrt{n}$, where $q(\alpha)$ is a continuity point of R with $R(q(\alpha)) \geq \alpha$. 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 $x = (x_1, \ldots, x_n)$, $\mathbb{F}_n(x)$ is the function given by

$$\mathbb{F}_n(x)(z) = \frac{1}{n}\sum_{i=1}^{n} 1_{[x_i,\infty)}(z), \quad z \in [0,\infty).$$

If $Y_1, \ldots, Y_n$ are iid observations from G, then $G_n = \mathbb{F}_n(Y_1, \ldots, Y_n)$. The distribution function $R_n(z)$ is given by

$$R_n(z) = \int_{\mathbb{R}^n} 1_{[0,z]} \left(\sqrt{n} \| \Lambda \left(\mathbb{F}_n(x)\right) - \Lambda(G) \|_{\infty,\tau}\right) G^{\otimes n}(\mathrm{d}x), \tag{5}$$

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 (5). 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

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

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

Theorem 3 (Bootstrap) *Let G_n, G, F_n, F, $\hat{q}_n(\alpha)$ be as above. Assume that $\tau > 0$ is such that $\widetilde{F}(\tau) < \sigma_g$ and $\widetilde{G}^o(\tau) < \sigma_f$. Then*

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

This result gives a confidence region for F given by the set of all distribution functions F' such that $\|F' - F_n\|_{\infty,\tau} \leq \hat{q}_n(\alpha)/\sqrt{n}$.

In practice, an approximation to the bootstrap estimator $\hat{R}_n$ is obtained via Monte Carlo methods. A large number B of (re)samples $Y_j^* = \{Y_{j,1}, \ldots, Y_{j,n}\}$, $j = 1, \ldots, B$, are simulated with replacement from the original data $Y_1, \ldots, Y_n$. For each Y_j^*, we calculate the corresponding estimator $F_j^* = \Lambda\left(\mathbb{F}_n(Y_{j,1}, \ldots, Y_{j,n})\right)$. 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 $\|F' - F_n\|_{\infty,\tau} \leq q_{\text{approx}}/\sqrt{n}$.

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. 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 Fig. 1 for an illustration of this.

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

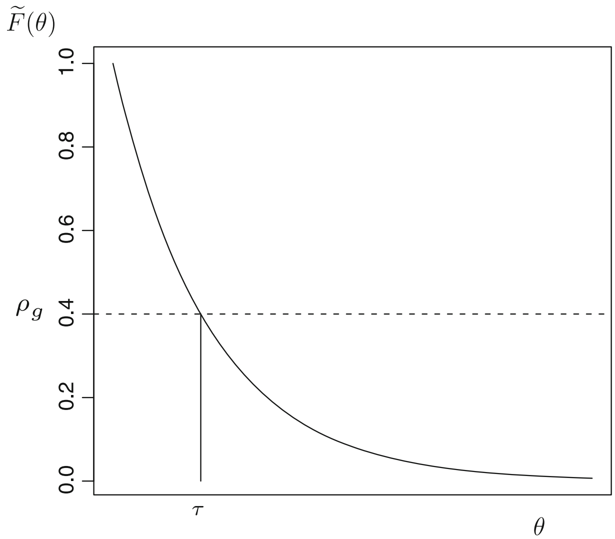

Fig. 1 Choose the smallest τ such that $\widetilde{F}(\tau) \le \rho_g$

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.

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 $0 \le \rho \le 1$ and $p_k(\rho) = (1-\rho)\rho^k$ for all $k = 0, 1, \ldots$. For the Landau bound choose τ such that: $\widetilde{F}(\tau) \le \frac{p_1(\rho)}{4} = \frac{\rho(1-\rho)}{4}$. For the Redheffer bound, recall $A = p_1(\rho)$ and choose τ such that

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

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 Fig. 2, and in the heavy tail case $\lim_{\rho \to 1} B_L/B_R = 1/2$. We notice that in the the heavy tail limit the Redheffer bound is 2 times "better" than the Landau bound.

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

Compound	Landau (Lemma 1)	Redheffer (Remark 1)	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	$\frac{\lambda e^{-\lambda}}{4}$	$\sin(\arctan(\lambda e^{-\lambda}(1-(\lambda e^{-\lambda})^2)^{-1/2}/2))$	$\frac{1}{\lambda}\log(2)$

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 $\widetilde{F}(\tau) \leq (2\rho)^{-1}$. We would expect this bound to be better as it results from a more precise study of the properties of the geometric random sum. This is indeed so and it it most easily seen by an "efficiency" calculation where we divide the bound provided by Redheffer (B_R) with the bound provided by Hansen and Pitts (2006, Theorem 2) (B_D), see Fig. 2 and in the heavy tail case $\lim_{\rho\to 1} B_R/B_D = 0$.

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 $\lambda \geq 0$ 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 Fig. 3. A similar carefully derived bound as existing in the geometric case was derived by Buchmann and Grübel (2003, Theorem 2) and is given by $\widetilde{F}(\tau) \leq (\lambda)^{-1}\log 2$. Henceforth the efficiency comparisons can also be carried over and similar conclusions to the geometric case exists.

4 Applications

At first sight, the decompounding method proposed in this paper appears to be difficult to implement in general. However, there is a wide class of distributions for N,

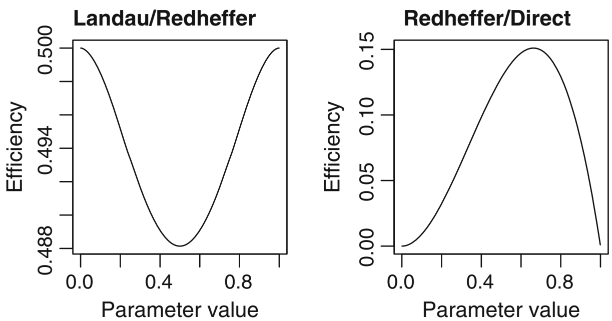

Fig. 2 For the compound geometric case. *Left-hand side* comparison of Laudau bound versus Redheffer bound. *Right-hand side* comparison of Redheffer bound versus the direct bound

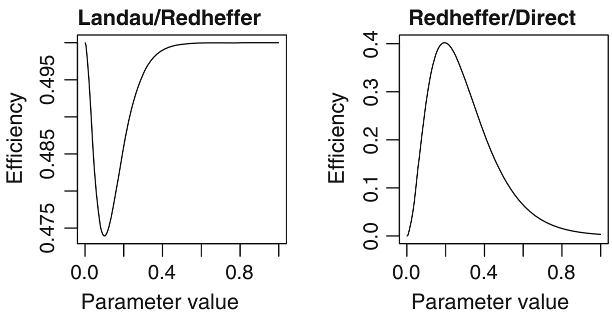

Fig. 3 For the compound Poisson case. *Left-hand side* comparison of Laudau bound versus Redheffer bound. *Right-hand side* comparison of Redheffer bound versus the direct bound

including many of those arising in applications, for which implementation is relatively straightforward and does not involve calculation of the π_k's. This class consists of those compound distributions having a recursion scheme which can be easily inverted to give a recursion scheme for the corresponding decompounding problem. This approach is particularly simple when the p_k's satisfy

$$p_{k+1} = \left(a + \frac{b}{k+1}\right) p_k, \quad k = 0, 1, 2, \ldots, \tag{6}$$

for some constants $a < 1$ and $b \in \mathbb{R}$ (see Dickson 2005, p. 64), when the well known Panjer recursion formula Panjer (1981) may be used for numerical evaluation of the compound distribution. The recursion scheme applies exactly when F is a discrete distribution, and is widely applied in practice for other cases by first discretising F. Suppose now that F is discrete with mass f_k at k, $k = 0, 1, 2, \ldots$, that the p_k's (known) satisfy (6) and have probability generating function $g_N(z)$. Then the compound distribution is also discrete, with mass g_k, say, at k, $k = 0, 1, 2, \ldots$. If the f_k's are known, then Panjer's recursion formula gives

$$g_0 = g_N(f_0), \quad g_k = \frac{1}{1 - af_0} \sum_{j=1}^{k} \left(a + \frac{bj}{k}\right) f_j g_{k-j}, \quad k \geq 1,$$

(see Dickson 2005, p. 69) so that the g_k's may be calculated recursively from the f_k's. It turns out that it is a simple matter to invert this recursion to obtain

$$\begin{aligned} g_0 &= g_N(f_0) \\ f_k &= \frac{1}{(a+b)g_0} \left(((1 - af_0)g_k - \sum_{i=1}^{k-1} \left(a + \frac{ib}{k}\right) f_i g_{k-i} \right), \end{aligned} \tag{7}$$

where it might be necessary to implement a numerical routine to invert (7). This gives a recursion scheme for calculating the f_k's from the g_k's. The class of distributions satisfying (6) consists of exactly the Poisson, the negative binomial and the binomial distributions. The inverse recursions for the Poisson and geometric cases were found in Buchmann and Grübel (2003) and Hansen and Pitts (2006), respectively.

There are other distributions for N which also have known recursions for the associated compound distribution (see Sect. 8.4 of Grandell 1997) and we use one of these as our main example. Let U be a shifted gamma-distribution with probability density function (pdf) given by

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

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

$$g_N(z) = \mathrm{e}^{-\gamma(1-z)}\left(1 + \frac{1-z}{\beta}\right)^{-\alpha}$$

and with

$$p_k = \sum_{n=0}^{k} \frac{\gamma^{k-n}}{(k-n)!}\mathrm{e}^{-\gamma}\binom{\alpha+n-1}{n}\left(\frac{\beta}{\beta+1}\right)^{\alpha}\left(\frac{1}{1+\beta}\right)^{n}.$$

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, and it has been fitted to several data sets in the literature (Ruohonen 1988). It is easy to check that

$$\frac{\mathrm{d}}{\mathrm{d}u}\log f_U(u) = \frac{-\beta u + \alpha + \beta\gamma - 1}{u - \gamma}, \quad \gamma \leq u < \infty$$

This is a ratio of polynomials and therefore belongs to the Willmot class (see Grandell 1997, p. 35).

For the discrete compound Delaporte distribution Schröter (1990) (see also Grandell 1997, Example 8.6) derived the following recursion

$$(1+\beta)g_k = \sum_{j=0}^{k}\left(\left(1 + \frac{(\alpha - 1 + \gamma\beta + \gamma)j}{k}\right)f_j - \frac{\gamma j}{2k}f_j^{\star 2}\right)g_{k-j}, \quad \text{for} \quad k = 1, 2, 3, \ldots, \tag{8}$$

where $f_j^{\star 2} = P(X_1 + X_2 = j)$. This can be inverted to give the inverse recursion

$$g_0 = \mathrm{e}^{-\gamma(1-f_0)}\left(1 + \frac{1-f_0}{\beta}\right)^{-\alpha} \tag{9}$$

$$f_k = \frac{1}{(\alpha + \gamma\beta + \gamma - \gamma f_0) g_0} \Bigg[(1 + \beta - f_0) g_k + \frac{\gamma g_0}{2} \sum_{i=1}^{k-1} f_i f_{k-i} - \sum_{j=1}^{k-1} \left(\left(1 + \frac{\alpha - 1 + \gamma\beta + \gamma}{k} j \right) f_j - \frac{\gamma j}{2k} f_j^{\star 2} \right) g_{k-j} \Bigg] \quad \text{for} \quad k = 1, 2, 3, \ldots. \tag{10}$$

This formula is reasonably simple to implement. We illustrate this by simulating a sample $Y_1, \ldots, Y_n$ of size $n = 500$ from the compound Delaporte distribution with $\alpha = 1, \beta = 1, \gamma = 0.5$ and with F exponentially distributed with mean 1. The resulting empirical distribution function G_n and the 'true' compound Delaporte distribution [calculated using (8)] are shown in the left-hand panel of Fig. 4. The right-hand panel shows the resulting plug-in estimate F_n of F for the resulting decompounding problem, calculated using (9) and (10), together with the true exponential distribution function F.

We now introduce an empirical way of choosing τ. From Theorem 3 we note that τ should be chosen such that $\widetilde{F}(\tau) < \sigma_g$ and $\widetilde{G}^o(\tau) < \sigma_f$. This is the case if ρ_g is chosen as in Remark 1. By the running assumption that the p_k's are known we get $\rho_g \simeq 0.0146$. Now we choose τ on basis of the empirical estimates of $\widetilde{F}$, i.e. choose the smallest $\tau > 0$ such that $\int_{[0,\infty)} \mathrm{e}^{-\tau x} F_n(dx) < \rho_g$. Provided that F_n is not degen-

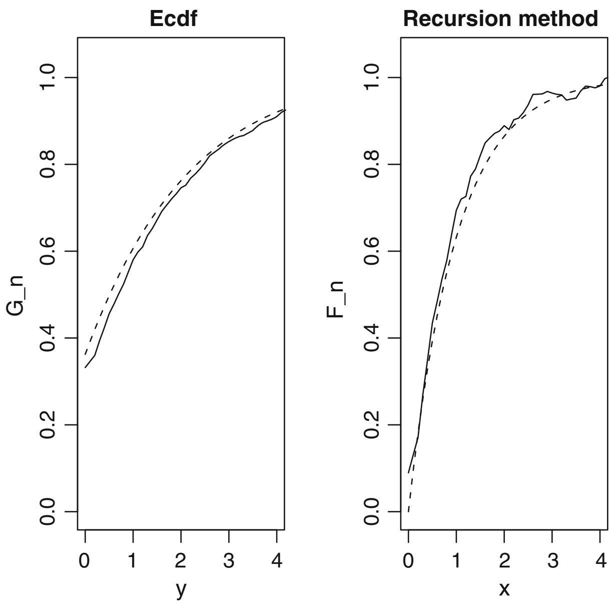

Fig. 4 *Left-hand panel* compound Delaporte distribution with exponentially distributed summands with mean one, $n = 500$, $\alpha = \beta = 1$ and $\gamma = 0.5$: estimate (*solid line*) and true (*dashed line*). *Right-hand panel* summands distribution: estimate via decompounding, calculated by recursion (*solid line*) and true (*dashed line*)

erate at 0, this procedure produces a τ to work with. For the present simulation study, an empirical τ is given by $\tau \simeq 6.6$. Figure 5 shows an approximate 90% confidence band with repsect to $\|\cdot\|_{\infty,\tau}$ with $\tau = 6.6$, and with $B = 100$ bootstrap resamples, as described in Sect. 2.3. The true F is also shown in this figure.

For N-distributions for which there is no recursion formula for the compound distribution, other approaches are possible. It turns out that (numerical approximations to) the first K of the π_k's can easily be found using, for example, Maple. Then it is possible to use a truncated version of the decompounding convolution power series in (2) with K terms, and to calculate the convolution powers either using the fast Fourier transform algorithm (FFT) or using direct convolution. This method gives rise to the additional issue of how to choose K. For illustration, the right-hand panel of Fig. 5 shows the estimate F_n for the above compound Delaporte example calculated using the FFT method with $K = 14$.

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 we have done so in the previous sections by indicating how the problem can be solved in greater generality. We have also shown how Panjer recursion can be inverted, and given an example of such inversion for a Poisson distri-

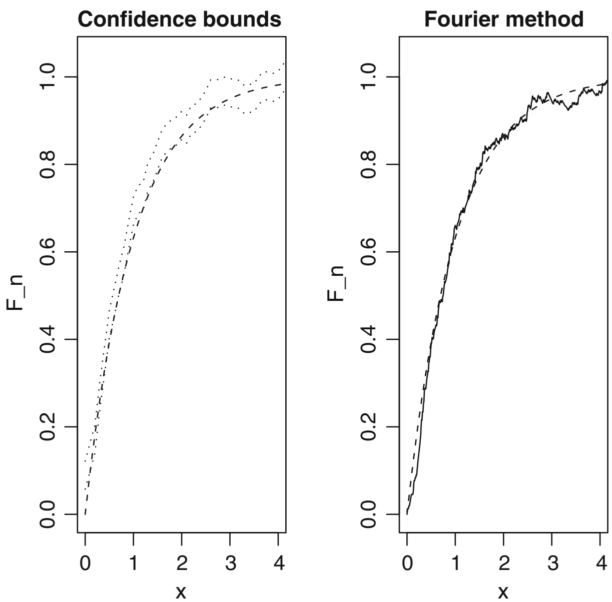

Fig. 5 *Left-hand panel* approximate 90% confidence band for the Delaporte decompounding problem in Fig. 4. *Right-hand panel* summands distribution: estimate via decompounding, calculated via truncated series and FFT (*solid line*) and true (*dashed line*)

bution mixed over a distribution from the Willmot class, thus providing a strong tool to discretize and solve the inverse problem at hand.

There are still a number of unsolved problems for future research. For example, if $p_i = 0$ for $i \leq n-1$ and $p_n \neq 0$, then the equation $w = g(z)$ has a solution

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

where the series of powers of $w^{1/n}$ converges in a neighbourhood of 0. Hence f is an n-valued function of w, having a branch-point at 0 (Copson 1935, p. 123). Using this to define a decompounding convolution power series involves ideas of fractional convolution. For this, and ideas of fractional analysis in general, see 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. Actually, if p_1 is zero, a slower rate of convergence is expected. We have for $p_0, p_1, p_3, \ldots = 0$, so that p_2 is the only nonzero p_k, the classical autoconvolution problem (Gorenflo and Hofmann 1994) which resembles the classical linear devonvolution problem for which it is well-known that there is cube root asymptotics (van Es et al. 1998).

Other possible extensions include the case where the p_k's themselves are unknown, for example we might assume that the distribution of N belongs to an (unspecified) parametric family indexed by a real-valued parameter ρ, and that ρ (as well as F) is to be estimated. This was done in the geometric cases in Hansen and Pitts (2006), but the general case is another area for future research.

Finally, as mentioned in Sect. 2.2, the proposed estimator need not be a distribution function itself. It might take negative values and might be non-monotone. This problem could be handled by truncation as suggested in Buchmann and Grübel (2004).

6 Proofs

The proofs follow the methods of Buchmann and Grübel (2003) and Hansen and Pitts (2006) and so we restrict ourselves to outlines of the proofs here, and refer the reader to the two above papers for further details.

Outline of the proof of Theorem 1. 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 $\mu_H([0, x]) = H(x)$ and hence is a candidate for an integrator in a convolution integral. 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 for H in $D_{\mathrm{m}}^{+}(\infty)$ and g in $D_\tau[0, \infty)$ we have

$$\|g \star H\|_{\infty,\tau} \leq \|g\|_{\infty,\tau} \widetilde{H}(\tau). \tag{11}$$

Turning to the convolution series given by $\Lambda(G)$, we first note that, for any distribution function G on $[0, \infty)$, we have $\sum_{k=1}^{m} \pi_k \,(G^o)^{\star k}$ is in $D_\tau[0, \infty)$ for each finite m. Using (11) we have for $k \geq 1$,

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

Then it follows that

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

Since we assume that $\widetilde{G}^o(\tau) < \sigma_f \ (\leq r(f))$, the series $\sum_{k=1}^{\infty} |\pi_k| \widetilde{G}^o(\tau)^{k-1}$ 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 $\theta > \tau$, $\widetilde{G}^o(\theta) = \sum_{k=1}^{\infty} p_k \widetilde{F}(\theta) = g\left(\widetilde{F}(\theta)\right)$. 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 $z = \sum_{k=1}^{\infty} \pi_k \left(\widetilde{G}^o(\theta)\right)^k$. Since $\widetilde{F}(\theta) < \sigma_g$, this implies that $z = \widetilde{F}(\theta)$ and that $\widetilde{F}(\theta) = \sum_{k=1}^{\infty} \pi_k \left(\widetilde{G}^o(\theta)\right)^k$. The right-hand side is $\widetilde{\Lambda(G)}(\theta)$, from which we conclude that $F = \Lambda(G)$, and Theorem 1 is proved.

Outline of the proof of Theorem 2. This is an application of the infinite-dimensional delta method as explained in Sect. 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*

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

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

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

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 methods similar 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

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

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).

Outline 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.

References

Asmussen, S. (2003). *Applied probability and queues* (2 edn.) New York: Springer.

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

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

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

Buchmann, B., Grübel, R. (2004). Decompounding Poisson random sums: Recursively truncated estimates in the discrete case. *Annals of the Institute of Statistical Mathematics, 56*, 743–756.

Bultheel, A., Martínez-Sulbaran, H. (2006). Recent developments in the theory of the fractional fourier and linear canonical transforms. *Bulletin of the Belgian mathematical society—Simon Stevin, 13*, 971–1005.

Copson, E. T. (1935). *An introduction to the theory of functions of a complex variable*. Oxford: The Clarendon Press.

Delaporte, P. (1959). Quelques problèmes de statistiques mathématiques posés par l'assurance automobile et le bonus pour non sinistre. *Bulletin Trimestriel de l'Institut des Actuaires Français, 227*, 87–102.

Dickson, D. C. M. (2005). *Insurance risk and ruin*. Cambridge: Cambridge University Press.

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

Gorenflo, R., Hofmann, B. (1994). On autoconvolution and regularization. *Inverse Problems, 10*, 353–373.

Grandell, J. (1997). *Mixed poisson processes*. London: Chapman and Hall.

Grübel, R., Pitts, S. M. (1993). Nonparametric estimation in renewal theory I: the empirical renewal function. *Annal of Statistics, 21*, 1431–1451.

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

Hansen, M. B., Pitts, S. M. (2006). Nonparametric inference from the $M/G/1$ workload. *Bernoulli, 12*, 737–759.

Henrici, P. (1974). *Applied and Computational Complex Analysis, Vol. 1: Power series—integration—conformal mapping—location of zeros*. New York: Wiley.

Hohn, N., Veitch, D. (2003). Inverting sampled traffic. In *Proceedings of the 3rd ACM SIGCOMM conference on Internet measurement*, Miami Beach, USA (pp. 222–223). New York: ACM Press.

Johnson, N. L., Kemp, A. W., Kotz, S. (2005). *Univariate discrete distributions* (3 edn.). New York: Wiley.

Neyman, J., Scott, E. L. (1958). A statistical approach to problems of cosmology. *Journal of the Royal Statistical Society, Series B 20*, 1–43.

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.

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 recursions for related compound distributions. *Scandinavian Actuarial Journal, 304*, 161–175.

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

van der Vaart, A. W., Wellner, J. A. (1996). *Weak convergence and empirical processes*. New York: Springer.

van Es, B., Gugushvili, S., Spreij, P. (2007). A kernel type nonparametric density estimator for decompounding. *Bernoulli, 13*, 672–694.

van Es, B., Jongbloed, G., van Zuijlen, M. (1998). Isotonic inverse estimators for nonparametric deconvolution. *Annals of Statistics, 26*, 2395–2406.