The Annals of Statistics
2003, Vol. 31, No. 4, 1054–1074

DECOMPOUNDING: AN ESTIMATION PROBLEM FOR POISSON RANDOM SUMS

BY BORIS BUCHMANN AND RUDOLF GRÜBEL

Technische Universität München and Universität Hannover

Given a sample from a compound Poisson distribution, we consider estimation of the corresponding rate parameter and base distribution. This has applications in insurance mathematics and queueing theory. We propose a plug-in type estimator that is based on a suitable inversion of the compounding operation. Asymptotic results for this estimator are obtained via a local analysis of the decompounding functional.

1. Introduction. The statistical problem to be discussed in this paper is motivated by applications from insurance mathematics and queueing theory. In the standard model of risk theory [see, e.g., Beard, Pentikäinen and Pesonen (1984) or Grandell (1991)], claims of random size $X_1, X_2, X_3, \dots$ arrive at random times $T_1, T_1+T_2, T_1+T_2+T_3, \dots$. The random variables $X_1, X_2, X_3, \dots, T_1, T_2, T_3, \dots$ are assumed to be independent, the X_k, $k \in \mathbb{N}$, have distribution P and the interarrival times T_k, $k \in \mathbb{N}$, are exponentially distributed with parameter λ. In particular, the claim arrival times are given by the points of a Poisson process with constant intensity λ. For all $t \ge 0$,

$$S_t = \sum_{k\,:\,T_1+\cdots+T_k \le t} X_k \tag{1}$$

is the total claim amount up to and including time t. Similarly, in a queueing context as discussed, for example, in Asmussen (1987), if customers arrive at a service point in bulks of size $X_1, X_2, \dots$ at the time points of a Poisson process then (1) gives the total number of customers that arrive in the time interval $(0, t]$.

The assumptions imply that the distribution Q of S_1 can be written as a convolution series,

$$Q = \Psi(\lambda, P) \qquad \text{with } \Psi(\lambda, P) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} P^{\star k}. \tag{2}$$

Q is the compound Poisson distribution with rate λ and base (or claim size or bulk size) distribution P. (Unfortunately, Poisson distributions with a random parameter, i.e., mixed Poisson distributions, are often called compound in the literature.)

Received July 2001; revised April 2002.
AMS 2000 subject classifications. Primary 62G05; secondary 62G20, 62P05.
Key words and phrases. Risk theory, queues with bulk arrival, compound distributions, plug-in principle, asymptotic normality, delta method.

Assume now that we observe the process $S = (S_t)_{t\geq 0}$ at equally spaced time points $h, 2h, 3h, \ldots, nh$. After rescaling if necessary we may take h to be equal to 1. Then the increments

$$Y_k := S_k - S_{k-1}, \qquad k = 1, \ldots, n,$$

of the process are independent and have distribution Q. Is it possible to "recover" P (and λ) from such a sample of Q-observations? This only makes sense if $P(X_i = 0) = 0$ as otherwise the function $(\lambda, P) \mapsto Q$ is not one-to-one and an identifiability problem arises, so we will assume this throughout the paper.

The "direct" problem, from P to Q, has been considered by Pitts (1994a), who used the plug-in estimator derived from (2),

$$Q_n := \Psi(\lambda, P_n) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} P_n^{\star k}, \tag{3}$$

where P_n denotes the empirical distribution function associated with a sample of size n from P. The rate λ was assumed to be known. Regarding $P \mapsto Q = \Psi(\lambda, P)$ as a nonlinear operator (functional) on a suitable function space one can then use the local analytic properties of the functional, such as continuity and differentiability, to deduce statistical properties of Q_n, such as consistency, asymptotic normality and asymptotic validity of bootstrap confidence regions, from the corresponding properties of P_n. A similar approach was used in Grübel and Pitts (1993) and Politis and Pitts (2000) for nonparametric estimation in renewal theory, in Pitts (1994b) for $G/G/1$ queues and in Grübel and Pitts (2000) for nonparametric estimation of perpetuities.

In the context of the "inverse" problem, from Q to P, such a plug-in approach seems not to be feasible, at least on first sight. Compounding transforms a probability distribution into a probability distribution. Compounding can therefore easily be applied to empirical distributions whereas in the other direction, "decompounding" so to speak, we do not have an analogue of (2) in this strict sense. Indeed, as a rule empirical distributions are not in the range of the compounding functional $P \mapsto Q$. Nevertheless, reasonable (in the sense of being algorithmically feasible and accessible to asymptotic analysis) plug-in estimators can be constructed if we are prepared to make some sacrifices. In the discrete situation, by which we mean that $P(\mathbb{N}) = 1$, we can proceed in a relatively straightforward manner as Ψ turns out to be locally invertible if its domain is extended to general summable sequences. The discrete case is of course the one that is of primary interest in queueing applications. In the general case, which is the natural frame for applications in risk theory, we face the difficulty that, roughly speaking, the statistical and the algebraic-topological aspects of the problem do not match as well as in the discrete case where the estimates on the Q-side converge in total variation norm, a norm that relates well to convolution. In the general case the empirical distribution associated with the Q-sample will only converge in a weaker norm, such as the supremum distance of the respective distribution functions, and the corresponding asymptotic normality result will lead to a limit process whose paths are no longer of bounded variation. The concession we make in this situation consists of switching to a relatively weak norm; however, we still have uniform convergence over bounded intervals for our general plug-in estimator.

The paper is organized as follows. Section 2 contains the main results, first for the discrete case and then for the general case. Our results are stronger for the discrete case. We restrict ourselves to asymptotic normality which, as indicated above, follows from a differentiability property of a suitably chosen inverse map. We do not discuss consistency as it is similarly related to the weaker property of continuity. The asymptotic normality results can be used in the discrete case to obtain asymptotically correct confidence intervals for individual claim size probabilities by Studentization, but in order to obtain confidence regions for the whole probability mass function or distribution function we would need the quantiles of the distribution of some functional of an infinite-dimensional Gaussian process. Bootstrap confidence regions are the practical alternative and the differentiability properties that we establish in the course of our proofs of asymptotic normality can also be exploited to prove the asymptotic validity of bootstrap confidence regions. The details of this argument have been carried out in Grübel and Pitts (1993, 2000) and will not be repeated here.

Section 3 discusses algorithmic aspects and gives some illustrative numerical examples. Proofs are collected in Sections 4 and 5. The last section contains some remarks on possible extensions and other aspects of our results. A different approach to decompounding, based on likelihood ideas, will be treated in a separate paper.

2. Main results.

We first consider the discrete case, with P and Q related by (2) and $P(\mathbb{N}) = 1$, which obviously implies $Q(\mathbb{N}_0) = 1$. Let $p = (p_i)_{i\in\mathbb{N}_0}$ and $q = (q_i)_{i\in\mathbb{N}_0}$ with $p_i := P(\{i\})$, $q_i := Q(\{i\})$ be the respective probability mass functions. The compound mass function can be obtained recursively from the rate and the mass function of the base distribution by

$$q_0 = e^{-\lambda}, \quad q_i = \frac{\lambda}{i}\sum_{j=1}^{i} j p_j q_{i-j} \qquad \text{for all } i \in \mathbb{N}. \tag{4}$$

Formulas of this type arise quite generally in the context of discrete infinite divisibility [see, e.g., Johnson, Kotz and Kemp (1992), page 352]. In insurance mathematics, (4) is known as Panjer recursion. The recursion can easily be inverted to give

$$\lambda = -\log q_0, \qquad p_i = -\frac{q_i}{q_0\log q_0} - \frac{1}{iq_0}\sum_{j=1}^{i-1} j p_j q_{i-j} \qquad \text{for all } i \in \mathbb{N}. \tag{5}$$

Now assume that $Y_1, \dots, Y_n$ are independent with common distribution Q. The associated empirical probability mass function $q_n = (q_{n,i})_{i\in\mathbb{N}_0}$ is given by

$$q_{n,i} = \frac{1}{n}\#\{1 \le m \le n : Y_m = i\}.$$

We risk an ambiguity in order to keep the notation compact: q with a single index i or j refers to the components of q, q with index n to the empirical probability mass function. As in the step from (2) to (3) we define the plug-in estimators λ_n and $p_n = (p_{n,i})_{i\in\mathbb{N}_0}$ for λ and p by $\lambda_n = -\log q_{n,0}$,

$$p_{n,i} = -\frac{q_{n,i}}{q_{n,0}\log q_{n,0}} - \frac{1}{i q_{n,0}}\sum_{j=1}^{i-1} j p_{n,j} q_{n,i-j} \qquad \text{for all } i \in \mathbb{N}$$

and $p_{n,0} = 0$. Degenerate cases such as $q_{n,0} = 0$ need separate consideration. We handle this together with a similar aspect relating to p_n: We are interested in statistical properties such as consistency and asymptotic normality, which both refer to a topology on some space for the estimates. Weak convergence for distributions on $\mathbb{N}_0$ is equivalent to convergence in total variation norm by Scheffé's theorem, which leads us to consider the space

$$\ell_1 := \left\{ a \in \mathbb{R}^{\mathbb{N}_0} : \sum_{i=0}^{\infty} |a_i| < \infty \right\}$$

of absolutely summable sequences of real numbers together with the norm

$$\|a\|_1 = \sum_{i=0}^{\infty} |a_i|.$$

We write $\delta_k = (\delta_{ki})_{i\in\mathbb{N}_0}$ for the element of ℓ_1 that has $\delta_{kk} = 1$ and all other entries equal to 0. Obviously, q_n is a random element of ℓ_1 but a priori there is no guarantee that Panjer inversion stays inside this space, that is, we might well have $p_n \notin \ell_1$. In Section 4 we will show that

$$\lim_{n\to\infty} P(q_{n,0} = 0 \text{ or } p_n \notin \ell_1) = 0.$$

Hence, if we simply put $\lambda_n = 1$ and $p_n = \delta_1$ if $q_{n,0} = 0$ or $p_n \notin \ell_1$ then we can regard our estimates as elements of the space $\mathbb{R} \times \ell_1$. In our first result, weak convergence refers to the product topology on this space that is generated by Euclidean distance on the first and by $\|\cdot\|_1$ on the second factor. The condition on p is discussed in Section 4 below.

THEOREM 1. *Assume that $\sum_{i=1}^{\infty} p_i^{1/2} < \infty$ and let $(r_i)_{i\in\mathbb{N}_0}$ be defined recursively by*

$$r_0 := \frac{1}{q_0}, \quad r_i := -\frac{1}{q_0}\sum_{j=1}^{i} q_j r_{i-j} \qquad \textit{for all } i \in \mathbb{N}. \tag{6}$$

Then $(\sqrt{n}(\lambda_n - \lambda), \sqrt{n}(p_n - p))$ *converges in distribution to a centered Gaussian random element* $(\xi, (Z_i)_{i\in\mathbb{N}_0})$ *of* $\mathbb{R} \times \ell_1$ *as* $n \to \infty$ *with* $Z_0 \equiv 0$ *and covariance structure*

$$
\begin{aligned}
E\xi^2 &= r_0 - 1,\\
E\xi Z_i &= \frac{1}{\lambda}(p_i - r_i - p_i r_0) \qquad \textit{for all } i \in \mathbb{N},\\
EZ_i Z_j &= \frac{1}{\lambda^2}\left(p_i r_j + p_j r_i + p_i p_j r_0 - p_i p_j + \sum_{l=0}^{i} r_l r_{l+j-i} q_{i-l}\right)\\
&\qquad\qquad\qquad\qquad \textit{for all } i, j \in \mathbb{N} \textit{ with } j \ge i \ge 1.
\end{aligned}
\tag{7}
$$

We now turn to the general case. For simplicity we assume that λ is known. As in the discrete case we have P and Q related by (2) and $Y_1, \dots, Y_n$ independent with distribution Q. Let F and G be the distribution functions of P and Q respectively; G_n with

$$
G_n(x) = \frac{1}{n}\sum_{m=1}^{n} \mathbb{1}_{[0,x]}(Y_m) \qquad \text{for all } x \ge 0
$$

is the empirical distribution function associated with $Y_1, \dots, Y_n$. (Here and in the following $\mathbb{1}_A$ denotes the indicator function of the set A.) These functions are elements of the space $D = D([0,\infty))$ of functions $h : [0,\infty) \to \mathbb{R}$ that are right-continuous and have left-sided limits; we also require that $\lim_{x\to\infty} f(x)$ exists for elements of this space. For any such function h we write h° for the function $x \mapsto h(x) - h(0)$. If h is the distribution function of some probability measure then the transition from h to h° corresponds to the removal of the atom at zero of this measure. For example, $G_n^\circ(x)$ is the fraction of strictly positive Y-values that are less than or equal to x. We now define an estimator F_n for F by

$$
F_n(x) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1} e^{\lambda k}}{\lambda k} (G_n^\circ)^{\star k}(x) \qquad \text{for all } x \ge 0.
\tag{8}
$$

Of course, "$\star$" continues to denote convolution which, however, is now defined only on a subset of $D \times D$ (details are given in Section 5). Note that the absolute values of the coefficients in this series increase at an exponential rate, so it is not clear a priori that this definition makes sense—indeed, this will be part of our next result. It follows from Lemma 7 below and from the arguments given at the beginning of Section 4 that this new estimator is "backward compatible" to the earlier estimator for the discrete case.

We need one more definition. For any $\tau \in \mathbb{R}$ let $D(\tau)$ be the space of all functions f with the property that $x \mapsto e^{-\tau x} f(x)$ is an element of D. On D we consider the supremum norm

$$
\|f\|_\infty = \sup_{x\ge 0} |f(x)| \qquad \text{for all } f \in D,
$$

which makes D a Banach space. Similarly, when equipped with

$$\|f\|_{\infty,\tau} = \sup_{x\geq 0} e^{-\tau x}|f(x)| \qquad \text{for all } f \in D(\tau),$$

$D(\tau)$ becomes a Banach space. In our second main result weak convergence refers to these spaces, where the σ-field is the one generated by the open balls in the respective norm.

THEOREM 2. *Let $\tau > 0$ be such that $\int e^{-\tau x} F(dx) < (\log 2)/\lambda$. Then $\sqrt{n}(F_n - F)$ converges in distribution as $n \to \infty$ with respect to $(D(\tau), \|\cdot\|_{\infty,\tau})$ to a centered Gaussian process Z with covariance structure*

$$EZ_sZ_t = \iint G^{\circ}\big((s-u)\wedge(t-v)\big)H(du)H(dv) - e^{-2\lambda}H^{\circ}(s)H^{\circ}(t)$$

for all $s, t \geq 0$, with H given by

$$H(x) = \frac{1}{\lambda}\sum_{k=1}^{\infty}(-1)^{k+1}e^{\lambda k}(G^{\circ})^{\star(k-1)}(x).$$

3. Algorithmic aspects and numerical examples. The (in)famous von Bortkewitsch data [see, e.g., Quine and Seneta (1987)] give the number of deaths caused by horse kicks in the Prussian army, for various corps and years. The values 0–4 were observed 109, 65, 22, 3 and 1 time(s), respectively. The interpretation of a possibly compound rather than simple Poisson distribution as horses killing more than one soldier in one go is somewhat far fetched, but it seems interesting to see our procedures at work with a real data set.

Plugging the q-values into the inverse Panjer recursion we obtain the estimates

$$\lambda_n = 0.6069, \quad p_{n,1} = 0.9825, \quad p_{n,2} = 0.0396,$$
$$p_{n,3} = -0.0365, \quad p_{n,4} = 0.0207,$$

all rounded to four decimal places. Note the occurrence of a negative value. Theorem 1 can be used to obtain asymptotically correct confidence intervals for the individual estimators, using plug-in estimates (again) for the unknown asymptotic covariances. Estimates for r can be obtained from the q-estimates via (6), and (7) leads to the estimate

	ξ	Z_1	Z_2	Z_3	Z_4
ξ	0.8349				
Z_1	0.4531	1.0926			
Z_2	−0.5193	−1.5048	2.2433		
Z_3	0.0468	0.4860	−0.9019	0.5674	
Z_4	0.0456	−0.0591	0.1467	−0.1707	0.1171

for the asymptotic covariance matrix. With $n = 200$ we obtain the values 0.0739, 0.1060, 0.0533 and 0.0242 for the standard errors of the individual estimates, again rounded to four decimal places. The estimates for the mass function p are therefore all within one standard deviation of $p = \delta_1$, which corresponds to an ordinary Poisson distribution. Hence, on the basis of these calculations there is no reason to assume that horses run amok.

We now consider a nondiscrete example with simulated data. The right-hand plot in Figure 1 displays the estimates obtained for two samples of size 1000 from a compound Poisson distribution with rate 2, the left-hand plot shows the empirical distribution functions for the compound data. The claim size distribution is a mixture of the exponential distribution with parameter 1 and the distribution concentrated at the single value 1, with mixing coefficients $2/3$ and $1/3$ respectively; the corresponding distribution function is displayed as a dotted line. To obtain the estimates numerically we discretized the data and then applied the inverse Panjer recursion given in (5). The arguments given in Section 5 for the differentiability of the decompounding functional can easily be adapted to obtain a version of continuity that justifies this approximation, hence the choice of the discretization parameter is not a major issue here. We mention in passing that using Panjer recursion instead of transform methods avoids problems that may arise with the latter if the Fourier transform of the q-sequence winds about 0; see Embrechts, Grübel and Pitts (1993), Grübel and Hermesmeier (1999) and the references given there for FFT based calculation of compound distributions, and Buchmann (2001) for the homotopy problem. Using recursion rather than transform methods also makes it possible to calculate a finite initial segment of the distribution functions of interest.

While the two compound empirical distribution functions are relatively close to each other, this is not the case for the two estimates of the base distribution function. Figure 2 shows that increasing the sample size improves the estimate, but that increasing the rate leads to a deterioration.

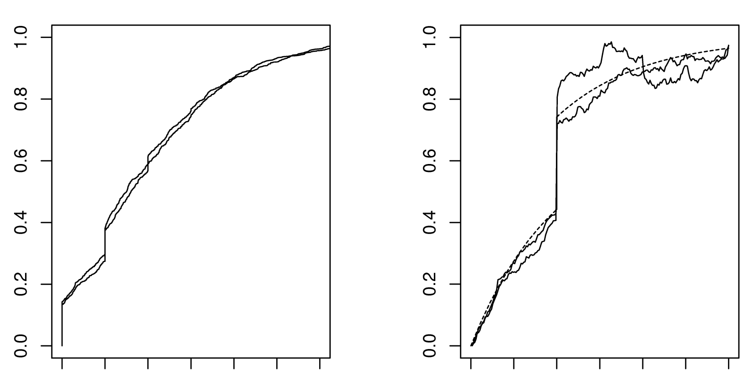

FIG. 1. *Two estimates for the total and the individual claim size distribution* ($n = 1000, \lambda = 2$).

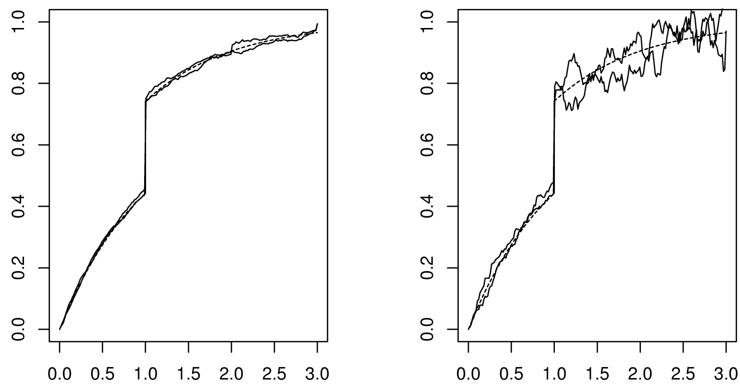

FIG. 2. *Estimates for the individual claim size distribution* (*left*: $n = 10{,}000$, $\lambda = 2$, *right*: $n = 10{,}000$, $\lambda = 5$).

We notice that the estimates for the base distribution are not distribution functions as they are, as a rule, not increasing; see Section 6.3 for possible modifications. The estimates capture the jump at 1. Also, the precision seems to decrease for increasing x-values, in accordance with our results.

4. Proof of Theorem 1. In the discrete case the basic convolution inequality

$$\|a \star b\|_1 \le \|a\|_1 \|b\|_1 \qquad \text{for all } a, b \in \ell_1$$

can be used to transfer the familiar power series calculus to ℓ_1. In particular,

$$\exp(a) = \sum_{k=0}^{\infty} \frac{1}{k!} a^{\star k}$$

is well defined on the whole of ℓ_1 and writing

$$\hat{a}(z) = \sum_{i=0}^{\infty} a_i z^i, \qquad -1 < z < 1,$$

for the generating function associated with $a = (a_i)_{i \in \mathbb{N}_0} \in \ell_1$ we have

$$\hat{b}(z) = \exp\big(\hat{a}(z)\big) \qquad \text{for } b := \exp(a)$$

(it should always be clear from the context which space the exponential function refers to). This implies

$$\frac{d}{dz}\hat{b}(z) = \left(\frac{d}{dz}\hat{a}(z)\right)\hat{b}(z),$$

which upon comparing coefficients leads to

$$b_0 = \exp(a_0), \quad i b_i = \sum_{j=1}^{i} j a_j b_{i-j} \qquad \text{for all } i \in \mathbb{N}.$$

This shows that Panjer recursion can be regarded as an algorithm that implements the exponential function on ℓ_1 (in fact, on even larger spaces).

We require two more properties of the exponential function on ℓ_1, and both are easily verified with the help of generating functions. First,

$$\exp(a+b) = \exp(a) \star \exp(b) \qquad \text{for all } a, b \in \ell_1;$$

second, using the fact that we deal with real vector spaces throughout,

$$\exp(a) = \exp(b) \quad \Longrightarrow \quad a = b \qquad \text{for all } a, b \in \ell_1.$$

As a first application of these rules we obtain that $q = \exp(\lambda(p - \delta_0))$ has a convolution inverse given by $q^{\star(-1)} = \exp(-\lambda(p - \delta_0))$. Comparing coefficients in $q^{\star(-1)} \star q = \delta_0$ shows that $q^{\star(-1)} = r$ with r as in the statement of Theorem 1; in particular, $r \in \ell_1$. Let λ, p, q and λ_n, p_n, q_n be as in Section 2.

LEMMA 3. *If* $\|q_n - q\|_1 < \|r\|_1^{-1}$ *then*

$$(\lambda_n - \lambda)\delta_0 + \lambda p - \lambda_n p_n = \sum_{k=1}^{\infty} \frac{1}{k} \bigl(r \star (q - q_n)\bigr)^{\star k}.$$

PROOF. The series

$$a_n := \sum_{k=1}^{\infty} \frac{1}{k} \bigl(r \star (q - q_n)\bigr)^{\star k}$$

converges in ℓ_1 because of

$$\|r \star (q - q_n)\|_1 \le \|r\|_1 \|q_n - q\|_1 < 1.$$

We know that

$$-\log(1-z) = \sum_{k=1}^{\infty} \frac{1}{k} z^k, \qquad -1 < z < 1,$$

which results in

$$\exp\bigl(-\hat{a}_n(z)\bigr) = \hat{r}(z)\hat{q}_n(z).$$

This means that we have found an element $b_n := \lambda(p - \delta_0) - a_n$ of ℓ_1 such that $q_n = \exp(b_n)$. As explained above, the components of q_n can be obtained recursively from those of b_n. Inverting the recursion, using the fact that the exponential function is one-to-one on ℓ_1 and using the definition of λ_n and p_n in Section 2 we finally see that

$$a_n = \lambda(p - \delta_0) - \lambda_n(p_n - \delta_0)$$

which implies the statement of the lemma. $\square$

Our next auxiliary result can be regarded as a differentiability property of a function closely related to discrete decompounding. Convergence refers to $\|\cdot\|_1$.

PROPOSITION 4. *If* $\sqrt{n}(q_n - q) \to a$ *as* $n \to \infty$ *for some* $a \in \ell_1$ *then*

$$\sqrt{n}(\lambda - \lambda_n)\delta_0 + \sqrt{n}(\lambda_n p_n - \lambda p) \;\to\; r \star a.$$

PROOF. As the condition implies $q_n \to q$ we may assume because of Lemma 3 that

$$(\lambda_n - \lambda)\delta_0 + \lambda p - \lambda_n p_n = \sum_{k=1}^{\infty} \frac{1}{k}\big(r \star (q - q_n)\big)^{\star k}.$$

This in turn implies

$$\sqrt{n}(\lambda - \lambda_n)\delta_0 + \sqrt{n}(\lambda_n p_n - \lambda p) = r \star \big(\sqrt{n}(q_n - q)\big) + b_n$$

with

$$b_n := -\sqrt{n} \sum_{k=2}^{\infty} \frac{1}{k}\big(r \star (q - q_n)\big)^{\star k}.$$

As convolution is continuous we obtain the limit $r \star a$ for the first term in the decomposition, hence it remains to show that b_n tends to 0 in ℓ_1. This however is obvious from

$$\|b_n\| \le \|\sqrt{n}(q - q_n)\|_1 \|r\|_1 \sum_{k=1}^{\infty} \frac{1}{k+1}\big(\|r\|_1 \|q - q_n\|_1\big)^k$$

and $\|q_n - q\|_1 \to 0$. □

On first sight it seems that this proposition is of little use as we do not have pointwise convergence of the random quantities $\sqrt{n}(q_n - q)$, where q_n denotes the empirical mass function associated with a sample of size n from q. What we do have is the following consequence of the Borisov–Durst theorem [see, e.g., Dudley (1999), Theorem 7.3.1].

PROPOSITION 5. *If* $\sum_{i=0}^{\infty} q_i^{1/2} < \infty$ *then* $\sqrt{n}(q_n - q)$ *converges in distribution to a centered Gaussian process* $V = (V_i)_{i \in \mathbb{N}_0}$ *with covariance*

$$\operatorname{cov}(V_i, V_j) = \delta_{ij} q_i - q_i q_j \qquad \textit{for all } i, j \in \mathbb{N}_0.$$

Further, if $\sqrt{n}(q_n - q)$ *converges in distribution then* $\sum_{i=0}^{\infty} q_i^{1/2} < \infty$.

To see that $\sum_{i=0}^{\infty} q_i^{1/2} < \infty$ follows from the condition $\sum_{i=1}^{\infty} p_i^{1/2} < \infty$ in Theorem 1 we note that the function ϕ,

$$\phi(a) = \sum_{i=0}^{\infty} |a_i|^{1/2},$$

has the properties

$$\phi(a+b) \le \phi(a) + \phi(b), \qquad \phi(\alpha a) \le |\alpha|^{1/2}\phi(a), \qquad \phi(a \star b) \le \phi(a)\phi(b).$$

Using these and monotone convergence we obtain

$$\sum_{i=0}^{\infty} q_i^{1/2} \le \sum_{k=0}^{\infty} \left(e^{-\lambda}\frac{\lambda^k}{k!}\right)^{1/2} \left(\sum_{i=1}^{\infty} p_i^{1/2}\right)^k,$$

which gives the desired implication. In fact, the two conditions are equivalent, the other direction being immediate from $q_i \ge \lambda e^{-\lambda} p_i$, hence $\sum_{i=1}^{\infty} p_i^{1/2} < \infty$ is a necessary condition in Theorem 1.

The Skorohod representation theorem provides the connection between the distributional result in Proposition 5 and the pointwise statement in Proposition 4: We can construct a probability space $(\Omega^\square, \mathcal{A}^\square, P^\square)$ carrying random sequences $V^\square$, $q_n^\square$, $n \in \mathbb{N}$, such that $\mathcal{L}(V^\square) = \mathcal{L}(V)$, $\mathcal{L}(q_n^\square) = \mathcal{L}(q_n)$ for all $n \in \mathbb{N}$, and

$$\lim_{n\to\infty} \sqrt{n}(q_n^\square - q)(\omega^\square) = V^\square(\omega^\square) \qquad \text{for all } \omega^\square \in \Omega^\square$$

[we write $\mathcal{L}(X)$ for the distribution of the random quantity X]. Within this construction we can use Proposition 4 to obtain

$$\sqrt{n}\big((\lambda - \lambda_n^\square)\delta_0 + (\lambda_n^\square p_n^\square - \lambda p)\big)(\omega^\square) \to r \star V^\square(\omega^\square) \qquad \text{for all } \omega^\square \in \Omega^\square,$$

where $(\lambda_n^\square, p_n^\square)$ depends on $q_n^\square$ exactly as (λ_n, p_n) depends on q_n, that is, via (5). Switching back to the original quantities and using the distributional equalities built into the construction we obtain

$$\sqrt{n}(\lambda - \lambda_n)\delta_0 + \sqrt{n}(\lambda_n p_n - \lambda p) \to W \qquad \text{in distribution},$$

with $W := r \star V$, V as in Proposition 5. [This is one of the standard methods for proving weak convergence, known as the infinite-dimensional delta method; see Grübel and Pitts (1993, 2000), Pitts (1994a, b), Politis and Pitts (2000) and the references given in these papers for a similar treatment of estimation problems in other areas.] The distributional convergence implies $\lambda_n \to \lambda$ in probability. Using this and

$$\sqrt{n}(p_{ni} - p_i) = \frac{1}{\lambda_n}\big(\sqrt{n}(\lambda_n p_{ni} - \lambda p_i) + p_i\sqrt{n}(\lambda - \lambda_n)\big)$$

together with some standard rules for weak convergence we obtain

$$\big(\sqrt{n}(\lambda_n - \lambda), \sqrt{n}(p_{ni} - p_i)\big) \to (\xi, Z) \qquad \text{in distribution},$$

with $\xi := -W_0$, $Z_0 \equiv 0$ and

$$Z_i = \frac{1}{\lambda}(W_i + p_i W_0) \qquad \text{for all } i \in \mathbb{N}.$$

The steps transforming V into Z are bounded linear operators on ℓ_1, hence Z is a centered Gaussian process. It remains to calculate the covariance structure. We do this for the intermediate process W; the formulas for ξ and Z then follow easily from the above definitions of these quantities in terms of W.

Using $r \star q = \delta_0$ we obtain

$$EW_0^2 = E(r_0 V_0)^2 = r_0^2 E V_0^2 = r_0^2 q_0 - r_0^2 q_0^2 = r_0 - 1,$$

and similarly, for $i \geq 1$,

$$\begin{aligned} EW_0 W_i &= E r_0 V_0 \sum_{j=0}^{i} r_j V_{i-j} \\ &= r_0 \sum_{j=0}^{i} r_j (q_0 \delta_{0,i-j} - q_0 q_{i-j}) \\ &= r_0 r_i q_0 - r_0 q_0 (r \star q)_i = r_i. \end{aligned}$$

The same arguments lead to

$$EW_i W_j = \sum_{l=0}^{i} r_l r_{l+j-i} q_{i-l} \qquad \text{for } j \geq i \geq 1.$$

5. Proof of Theorem 2. We put $D(\infty) := \bigcup_{\tau > 0} D(\tau)$. Let $D_{\mathrm{m}}(\infty) \subset D(\infty)$ be the subset of those functions that have finite variation on all intervals $[0, x]$, $x > 0$. We will use capital letters F, G, H for elements of $D_{\mathrm{m}}(\infty)$. Equivalently, $D_{\mathrm{m}}(\infty)$ can be considered as the space of signed measures μ (hence the index "m") with the bound

$$|\mu|([0, x]) = O(e^{\tau x}) \qquad \text{as } x \to \infty$$

for some $\tau = \tau(\mu) < \infty$ on the increase of the total variation, the connection being provided by $H(x) = \mu([0, x])$. These measures in turn can be characterized by the condition that the measure μ_τ with μ-density $x \mapsto e^{-\tau x}$ is a finite signed measure on the Borel subsets of the nonnegative half line for some $\tau < \infty$. (Our arguments here and below use exponential tilting in a somewhat implicit manner.) The measure associated with a function $H \in D_{\mathrm{m}}(\infty)$ is nonnegative if and only if H is (weakly) increasing. Let $D_{\mathrm{m}}^+(\infty)$ denote the corresponding subset of $D_{\mathrm{m}}(\infty)$. We write D_{m}, $D_{\mathrm{m}}(\tau)$ and D_{m}^+, $D_{\mathrm{m}}^+(\tau)$ for the intersection of D, $D(\tau)$ with $D_{\mathrm{m}}(\infty)$ and $D_{\mathrm{m}}^+(\infty)$, respectively.

Elements of $D_{\mathrm{m}}(\infty)$ are characterized by their Laplace transform,

$$\tilde{H}(\theta) = \int e^{-\theta x} H(dx) \qquad \text{for all } \theta > \tau(\mu).$$

If $H \in D(\tau)$ then the integral is finite for all $\theta > \tau$. Convergence with respect to $\|\cdot\|_{\infty,\tau}$ of a sequence $(H_n)_{n \in \mathbb{N}}$ in $D_{\mathrm{m}}^+(\tau)$ to some $H \in D_{\mathrm{m}}^+(\tau)$ implies vague convergence of the corresponding tilted measures, which in turn implies $\tilde{H}_n(\theta) \to \tilde{H}(\theta)$ as $n \to \infty$ for $\theta > \tau$. An alternative and more direct argument for this fact can be based on $\tilde{H}(\theta) = \theta \int_0^\infty e^{-\theta y} H(y)\, dy$, $\theta > 0$, which follows from an integration by parts.

Lower case letters f, g, h denote generic elements of $D(\infty)$ that might have infinite total variation on finite intervals. For the convolution product to be defined we need some variation condition for at least one of the factors; this is discussed in some detail in Grübel and Pitts (1993). The situation here is simpler as we deal with the "one-sided" case only; that is, all measures are concentrated on $[0, \infty)$, so we can simply write

$$g \star H(x) = \int g(x-y) H(dy) \qquad \text{for all } x \geq 0,$$

which should be selfexplanatory in view of the notational conventions introduced above. Two useful properties of convolution are collected in the following lemma.

LEMMA 6. (a) *If $H, H_n \in D_{\mathrm{m}}^+$ are such that* $\lim_{n\to\infty} \|H_n - H\|_\infty = 0$, *then*

$$\lim_{n\to\infty} \|g \star (H_n - H)\|_\infty = 0 \qquad \textit{for all } g \in D.$$

(b) *If $H \in D_{\mathrm{m}}^+(\infty)$ then $\|g \star H\|_{\infty,\tau} \leq \|g\|_{\infty,\tau} \tilde{H}(\tau)$ for all $\tau > 0$.*

PROOF. (a) The statement is easily checked for $g = \mathbb{1}_{[0,a)}$, $0 < a \leq \infty$, and then immediately generalizes to functions g_0 that can be written as finite linear combinations of such indicator functions. It is not difficult to show that the latter class is dense in D [see also Billingsley (1968), page 110], hence the assertion follows from

$$\|g \star (H_n - H)\|_\infty \leq \|g_0 \star (H_n - H)\|_\infty + \|g - g_0\|_\infty \big(H_n(\infty) + H(\infty)\big),$$

together with a standard ε–δ argument.

(b)

$$\begin{aligned}
\|g \star H\|_{\infty,\tau} &= \sup_{x\geq 0} e^{-\tau x} \left| \int_{[0,x]} g(x-y) H(dy) \right| \\
&\leq \sup_{x \geq 0} \int_{[0,x]} |e^{-\tau(x-y)} g(x-y)| e^{-\tau y} H(dy) \\
&\leq \|g\|_{\infty,\tau} \int e^{-\tau y} H(dy).
\end{aligned}$$

□

The following auxiliary result takes over the role of Lemma 3 in the discrete case. Note that we again use the fact that our transforms are real-valued.

LEMMA 7. *If G is a distribution function with $\widetilde{G^\circ}(\tau) < e^{-\lambda}$ then the series*

$$\Lambda(G) := \sum_{k=1}^{\infty} \frac{(-1)^{k+1} e^{\lambda k}}{\lambda k} (G^\circ)^{\star k}$$

converges in $D(\tau)$. Further, $\Lambda(G) = F$ if $G = \Psi(\lambda, F)$.

PROOF. Using Lemma 6(b) together with the obvious inequality

$$\|f\|_{\infty,\tau} \le \|f\|_\infty \qquad \text{for all } f \in D,$$

we obtain

$$\|H^{\star k}\|_{\infty,\tau} \le \tilde{H}(\tau)^{k-1} \qquad \text{for all } k \in \mathbb{N},$$

if H is the distribution function for some (sub)probability. This implies the convergence of the series.

The series can obviously be written as the difference of two increasing functions and is therefore an element of $D_{\mathrm{m}}(\infty)$. The associated Laplace transform is

$$\widetilde{\Lambda(G)}(\theta) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1} e^{\lambda k}}{\lambda k} \widetilde{G^\circ}(\theta)^k = \frac{1}{\lambda} \log\bigl(1 + e^{\lambda} \widetilde{G^\circ}(\theta)\bigr), \qquad \theta > \tau,$$

hence the final statement of the lemma follows on using

$$\exp\bigl(\lambda \tilde{F}(\theta) - \lambda\bigr) = \tilde{G}(\theta), \qquad \widetilde{G^\circ}(\theta) = \tilde{G}(\theta) - e^{-\lambda} \qquad \text{for all } \theta > 0,$$

and the identifiability of elements of $D_{\mathrm{m}}(\infty)$ by their Laplace transforms. □

For all $H \in D_{\mathrm{m}}^{+}(\infty)$ we have $\widetilde{H^\circ}(\theta) \to 0$ as $\theta \to \infty$ by dominated convergence, hence the condition on G in Lemma 7 and the following proposition, which serves as the analogue of Proposition 4 in the discrete case, is satisfied if τ is chosen large enough.

PROPOSITION 8. *Let G, G_n $(n \in \mathbb{N})$ be elements of D_{m}^{+} with $G(0) = 0$ and $G_n(0) = 0$ for all $n \in \mathbb{N}$. If*

$$\sqrt{n}(G_n - G) \to h \qquad \text{as } n \to \infty$$

with respect to $\|\cdot\|_\infty$ for some $h \in D$ and if τ is such that $\tilde{G}(\tau) < e^{-\lambda}$ then

$$\sqrt{n}\bigl(\Lambda(G_n) - \Lambda(G)\bigr) \to h \star H \qquad \text{as } n \to \infty$$

with respect to $\|\cdot\|_{\infty,\tau}$, with $H := \frac{1}{\lambda} \sum_{k=1}^{\infty} (-1)^{k+1} e^{\lambda k} G^{\star(k-1)}$.

PROOF. We have

$$G_n^{\star k} - G^{\star k} = (G_n - G) \star H_{n,k} \qquad \text{with } H_{n,k} = \sum_{j=0}^{k-1} G_n^{\star j} \star G^{\star(k-1-j)},$$

which leads to the basic decomposition

$$\sqrt{n}\big(\Lambda(G_n) - \Lambda(G)\big) - h \star H = A(N,n) + B(N,n) + C(N,n) - D(N)$$

with

$$A(N,n) := \frac{1}{\lambda} \sum_{k=N+1}^{\infty} \frac{(-1)^{k+1} e^{\lambda k}}{k} \sqrt{n}(G_n - G) \star H_{n,k},$$

$$B(N,n) := \frac{1}{\lambda} \sum_{k=1}^{N} \frac{(-1)^{k+1} e^{\lambda k}}{k} \big(\sqrt{n}(G_n - G) - h\big) \star H_{n,k},$$

$$C(N,n) := \frac{1}{\lambda} \sum_{k=2}^{N} \frac{(-1)^{k+1} e^{\lambda k}}{k} \big(h \star H_{n,k} - kh \star G^{\star(k-1)}\big),$$

$$D(N) := \frac{1}{\lambda} \sum_{k=N+1}^{\infty} (-1)^{k+1} e^{\lambda k} h \star G^{\star(k-1)},$$

valid for all $n, N \in \mathbb{N}$ (because of $H_{n,1} = G^{\star 0} = \mathbb{1}_{[0,\infty)}$ it is enough to start with $k = 2$ in the third term). For a given $\varepsilon > 0$ we need an $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ the sum of the norms of the four terms is less than ε, where we may choose an appropriate N.

Using Lemma 6(b) as in the proof of Lemma 7 we obtain

$$\|A(N,n)\|_{\infty,\tau} \leq \frac{1}{\lambda} \|\sqrt{n}(G_n - G)\|_{\infty} \sum_{k=N+1}^{\infty} \frac{e^{\lambda k}}{k} \tilde{H}_{n,k}(\tau).$$

Let $\eta < 1$ be such that $\tilde{G}(\tau) < \eta e^{-\lambda}$. Since $\tilde{G}_n(\tau) \to \tilde{G}(\tau)$ as $n \to \infty$ we can find an $n_1 < \infty$ such that $\tilde{G}_n(\tau) \leq \eta e^{-\lambda}$ for all $n \geq n_1$. But then

$$\tilde{H}_{n,k}(\tau) = \sum_{j=0}^{k-1} \tilde{G}_n(\tau)^j \tilde{G}(\tau)^{k-1-j} \leq k\eta^{k-1} e^{-\lambda(k-1)}$$

so that

$$\lim_{N\to\infty} \sup_{n \geq n_1} \sum_{k=N+1}^{\infty} k^{-1} e^{\lambda k} \tilde{H}_{n,k}(\tau) = 0.$$

Convergence of $\sqrt{n}(G_n - G)$ implies that the sequence is bounded, hence we obtain

$$\lim_{N\to\infty} \sup_{n \geq n_1} \|A(N,n)\|_{\infty,\tau} = 0.$$

The same arguments work with the fourth term, resulting in

$$\lim_{N\to\infty} \|D(N)\|_{\infty,\tau} = 0.$$

The number of terms in $B(N,n)$ and $C(N,n)$ is finite for any given N, hence it is enough to show that these converge to 0 individually as $n \to \infty$. For those in $B(N,n)$ this follows from the assumptions of the theorem, the boundedness of $\tilde{H}_{n,k}(\tau)$, $n \in \mathbb{N}$ and Lemma 6(b). In order to deal with the terms in $C(N,n)$ we first note that, for $k \geq 2$,

$$\begin{aligned} H_{n,k} - kG^{\star(k-1)} &= \sum_{j=0}^{k-1} \bigl(G_n^{\star j} \star G^{\star(k-1-j)} - G^{\star(k-1)}\bigr) \\ &= \sum_{j=1}^{k-1} \bigl(G_n^{\star j} \star G^{\star(k-1-j)} - G^{\star(k-1)}\bigr) \\ &= \sum_{j=1}^{k-1} (G_n^{\star j} - G^{\star j}) \star G^{\star(k-1-j)} \\ &= (G_n - G) \star \sum_{j=1}^{k-1} H_{n,j} \star G^{\star(k-1-j)}. \end{aligned}$$

Together with Lemma 6(b) this yields

$$\begin{aligned} &\bigl\| h \star \bigl(H_{n,k} - kG^{\star(k-1)}\bigr)\bigr\|_{\infty,\tau} \\ &\qquad \leq \|h \star (G_n - G)\|_\infty \sum_{j=1}^{k-1} \tilde{H}_{n,j}(\tau) \tilde{G}(\tau)^{k-1-j} \\ &\qquad \leq \|h \star (G_n - G)\|_\infty \frac{k(k-1)}{2} \eta^{k-2} e^{-\lambda(k-2)} \end{aligned}$$

with η and n as in the bounds for $A(N,n)$. Lemma 6(a) yields $\|h \star (G_n - G)\|_\infty \to 0$ as $n \to \infty$. Hence we have

$$\lim_{n\to\infty} \|B(N,n)\|_{\infty,\tau} = \lim_{n\to\infty} \|C(N,n)\|_{\infty,\tau} = 0$$

for any fixed finite $N \in \mathbb{N}$.

A routine argument now completes the proof: Given $\varepsilon > 0$, we can find integers n_1 and N such that $\|A(N,n)\|_{\infty,\tau} + \|D(N)\|_{\infty,\tau} < \varepsilon/2$ for all $n \geq n_1$. For this N we can find an n_2 such that $\|B(N,n)\|_{\infty,\tau} + \|C(N,n)\|_{\infty,\tau} < \varepsilon/2$ for all $n \geq n_2$. This shows that the sum of the norms of the four terms in the decomposition is less than ε for all $n \geq n_0 := \max\{n_1, n_2\}$. $\square$

For the proof of Theorem 2 we now proceed as in the discrete case, with the function spaces D, $D(\tau)$ replacing the sequence spaces that we used in Section 4. Convergence in distribution in these function spaces is a technically much more complicated issue: for details we refer the reader to one of the excellent research monographs and textbooks on this subject, such as Dudley (1999), Pollard (1984), Shorack and Wellner (1986) and van der Vaart and Wellner (1996); Billingsley (1968) is the classic in this area.

The analogue of Proposition 5 is the empirical central limit theorem [see, e.g., Pollard (1984), page 97]. In the present setting this theorem states that $\sqrt{n}(G_n - G)$ converges in distribution to a rescaled Brownian bridge $B \circ G$, which is a centered Gaussian process with covariance function given by

$$\operatorname{cov}\big((B \circ G)(s), (B \circ G)(t)\big) = EB(G(s))B(G(t)) = G(s \wedge t) - G(s)G(t)$$

for all $s, t \geq 0$. As $h \mapsto h^\circ$ is a measurable and continuous linear operator on D this implies that $\sqrt{n}(G_n^\circ - G^\circ)$ converges in distribution to $V := (B \circ G)^\circ$, which is again a centered Gaussian process. A straightforward calculation shows that the covariance structure of V is given by

$$\operatorname{cov}\big(V(s), V(t)\big) = G^\circ(s \wedge t) - G^\circ(s)G^\circ(t)$$

for all $s, t \geq 0$. The distribution of the limit process is concentrated on the set of those functions in D that have their discontinuities at the jumps of G°, a separable subspace of D. We may therefore apply the Skorohod representation in the form given in [Pollard (1984), page 71]. In complete analogy with the discrete case discussed in Section 4, now with Proposition 8 instead of Proposition 4, this leads to the convergence in distribution in $D(\tau)$ of

$$\sqrt{n}(F_n - F) = \sqrt{n}\big(\Lambda(G_n^\circ) - \Lambda(G^\circ)\big)$$

to $Z := V \star H$, with H as in the statement of the theorem. Again, Z arises from V by a deterministic linear transformation and therefore is again a centered Gaussian process. Its covariance structure is given by

$$\begin{aligned}
&\operatorname{cov}\big(Z(s), Z(t)\big)\\
&\quad = EV \star H(s) V \star H(t)\\
&\quad = \int_{[0,s]} \int_{[0,t]} EV(s-u)V(t-r)H(du)H(dr)\\
&\quad = \int_{[0,s]} \int_{[0,t]} \big(G^\circ\big((s-u) \wedge (t-r)\big) - G^\circ(s-u)G^\circ(t-r)\big)H(du)H(dr)\\
&\quad = \int_{[0,s]} \int_{[0,t]} G^\circ\big((s-u) \wedge (t-r)\big)H(du)H(dr) - G^\circ \star H(s) G^\circ \star H(t).
\end{aligned}$$

Together with

$$H^\circ = \frac{1}{\lambda} \sum_{k=2}^{\infty} (-1)^{k+1} e^{\lambda k} (G^\circ)^{\star(k-1)} = -\frac{e^\lambda}{\lambda} \sum_{k=1}^{\infty} (-1)^{k+1} e^{\lambda k} (G^\circ)^{\star k} = -e^\lambda G^\circ \star H,$$

this completes the proof. $\square$

6. Comments. In Section 6.1 we discuss the connection to other stochastic processes. Section 6.2 explains a testing application. Two variants of the plug-in estimator are briefly considered in Section 6.3. The final subsection contains some concluding remarks.

6.1. *Other stochastic processes.* In the previous sections we have seen our basic problem as an inference problem of the classical type, where the observations are a sample from a fixed distribution with a specific structure. We briefly point out that the above can also be seen as an inference problem for stochastic processes; in fact, we have already mentioned in the Introduction that the sample typically arises from observing some compound Poisson process $S = (S_t)_{t \ge 0}$ at equally spaced time intervals. The process S can also be regarded as a marked point process, and these and their statistics are treated in Karr (1986). This embedding of the decompounding problem also points toward several generalizations of our basic setup. Some of these are of theoretical interest and have considerable potential for applications, for example, processes with nonconstant rate such as doubly stochastic or Cox processes. Among these, Poisson processes with a Markov modulated intensity have received considerable interest over the years; see, for example, Asmussen (1989). On overview of the literature on the statistical analysis of queueing systems is given in Bhat, Miller and Rao (1997).

6.2. *A testing application.* For Poisson distributions, that is, with the base distribution concentrated at the single value 1, the asymptotic covariance structure of the plug-in estimator given in Theorem 1 can be further evaluated, resulting in

$$E\xi^2 = e^\lambda - 1, \qquad E\xi Z_1 = \lambda^{-1}(1 + \lambda e^\lambda - e^\lambda),$$

$$EZ_1^2 = \lambda^{-2}(\lambda^2 e^\lambda - \lambda e^\lambda + e^\lambda - 1),$$

$$E\xi Z_i = -\lambda^{-1} r_i, \qquad EZ_1 Z_i = \lambda^{-2} r_i(1 - i - \lambda) \qquad \text{for } i \ge 2,$$

$$EZ_i Z_j = \lambda^{-2}(-1)^j r_i m_{i,j} \qquad \text{for } 1 \le i \le j,$$

with

$$r_i = e^\lambda \frac{(-\lambda)^i}{i!}, \qquad m_{i,j} = \sum_{l=0}^{i} \binom{i}{l} \frac{\lambda^{l+j-i}}{(l+j-i)!} \qquad \text{for } 1 \le i \le j.$$

This displays the limit distribution as a function of the rate parameter λ and can be used to obtain asymptotically correct critical regions of tests for Poissonity, if λ is estimated by, for example, the mean of the data. It would be interesting to compare the power of the resulting test with the power of other tests of Poissonity proposed in the literature; see, for example, Klar (1999) and the references given there, especially for compound Poisson alternatives.

6.3. *Two variants of the plug-in estimator.* It is immediate from (3) (or, more probabilistically, from the interpretation of compound distributions as random sums) that $q_i > 0$ for all i in the additive semigroup generated by the support of p. As a consequence probability mass functions associated with compound distribution cannot have bounded support, which means that the empirical mass function q_n cannot be the mass function of a compound Poisson distribution. In particular, even if the Panjer inversion applied to q_n yields an element of ℓ_1 with high probability if n is large, this sequence will always have negative entries. Therefore, the plug-in estimator needs some modification in order for the estimates to be probability distributions. In the discrete case a straightforward remedy is to simply replace the negative entries by 0 and then to normalize to keep the sum of the entries to be equal to 1. This changes the plug-in estimate p_n into $\Phi(p_n)$, say. It is easy to see that Φ is continuous at p, which means that consistency is not lost by applying Φ. However, a similarly straightforward transfer of asymptotic normality by a delta method argument is not possible as Φ is not differentiable at p. A closer analysis, carried out in Buchmann (2001), shows that we still have convergence in distribution of $\sqrt{n}(\Phi(p_n) - p)$, but that the limit is no longer Gaussian.

A second modification of the plug-in estimator is motivated by the observation that the sample $y_1, \dots, y_n$ from the compound distribution cannot possibly contain any information about the base probabilities p_k with $k \ge z_n := \max\{y_1, \dots, y_n\}$. It therefore seems natural to stop the Panjer inversion at z_n, as we have done in the horse kick example in Section 3. It is shown in Buchmann (2001) that this modification does not change the distributional asymptotics.

In the continuous case we have the similar phenomenon that the plug-in estimator for the distribution function of the individual claims is itself not a distribution function (see the right-hand plots in Figures 1 and 2). We could associate with any F that defines a signed measure distribution functions $F^{(1)}$, $F^{(2)}$ via

$$F^{(1)}(x) = \inf\{F(y) \vee 0 : y \ge x\}, \qquad F^{(2)}(x) = \sup\{F(y) \wedge 1 : y \le x\},$$

but at present we do not know the effect of these modifications on the asymptotic distribution of the estimators.

6.4. *Concluding remarks.* The theorems in Section 2 show that the decompounding problem can be solved on the usual $n^{-1/2}$-level, a fact that we continue to find slightly surprising. At least in the general case we were initially regarding the problem as "ill-posed," with the corresponding consequences such as a rate lower than $n^{-1/2}$ for the estimates. Of course, the classification of a problem as ill-posed or inverse, etc. depends on the choice of topologies, so our results may be rephrased as saying that there are statistically meaningful choices for the latter where decompounding can be considered to be a perfectly regular problem. However, numerical experiments such as given in Section 3 remind us of the fact that a good rate is not a guarantee for high precision: comparing the left-hand and the right-hand plots in Figure 1 shows that the "constant in front of the rate" may be rather high. This effect becomes more pronounced with increasing rate λ. Indeed, we know from the central limit theorem for random sums that the precise form of the individual claim size distribution becomes irrelevant as $\lambda \to \infty$ and that only the first two moments survive.

Acknowledgments. We would like to thank both referees and Editor Jon A. Wellner for their comments on the first version of this paper, in particular for drawing our attention to several references that we had overlooked.

REFERENCES

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

ASMUSSEN, S. (1989). Risk theory in a Markovian environment. *Scand. Actuarial J.* 69–100.

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

BHAT, U. N., MILLER, G. K. and SUBBA RAO, S. (1997). Statistical analysis of queueing systems. In *Frontiers in Queueing* (J. H. Dshalalow, ed.) 351–394. CRC Press, Boca Raton, FL.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

BUCHMANN, B. (2001). Decompounding: An estimation problem for the compound Poisson distribution. Ph.D. thesis, Univ. Hannover.

DUDLEY, R. M. (1999). *Uniform Central Limit Theorems*. Cambridge Univ. Press.

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

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

GRÜBEL, R. and HERMESMEIER, R. (1999). Computation of compound distributions. I. Aliasing errors and exponential tilting. *ASTIN Bulletin* **29** 197–214.

GRÜBEL, R. and PITTS, S. M. (1993). Nonparametric estimation in renewal theory. I. The empirical renewal function. *Ann. Statist.* **21** 1431–1451.

GRÜBEL, R. and PITTS, S. M. (2000). Statistical aspects of perpetuities. *J. Multivariate Anal.* **75** 143–162.

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

KARR, A. F. (1986). *Point Processes and Their Statistical Inference*. Dekker, New York.

KLAR, B. (1999). Goodness-of-fit tests for discrete models based on the integrated distribution function. *Metrika* **49** 53–69.

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

PITTS, S. M. (1994b). Nonparametric estimation of the stationary waiting time distribution function for the $GI/G/1$ queue. *Ann. Statist.* **22** 1428–1446.

POLITIS, K. and PITTS, S. M. (2000). Nonparametric estimation in renewal theory. II. Solutions of renewal-type equations. *Ann. Statist.* **28** 88–115.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

QUINE, M. P. and SENETA, E. (1987). Bortkiewicz's data and the law of small numbers. *Int. Statist. Rev.* **55** 173–181.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer, Berlin.

MATHEMATISCHE STATISTIK
TECHNISCHE UNIVERSITÄT MÜNCHEN
D-80290 MÜNCHEN
GERMANY
E-MAIL: bbuch@mathematik.tu-muenchen.de

INSTITUT FÜR MATHEMATISCHE STOCHASTIK
UNIVERSITÄT HANNOVER
POSTFACH 60 09
D-30060 HANNOVER
GERMANY
E-MAIL: rgrubel@stochastik.uni-hannover.de