Ann. Inst. Statist. Math.
Vol. 56, No. 4, 743–756 (2004)

DECOMPOUNDING POISSON RANDOM SUMS: RECURSIVELY TRUNCATED ESTIMATES IN THE DISCRETE CASE

BORIS BUCHMANN[1] AND RUDOLF GRÜBEL[2]

[1]*Mathematische Statistik, Zentrum Mathematik, Technische Universität München, Boltzmannstr. 3, D-85747 Garching, Germany,* e-mail: bbuch@mathematik.tu-muenchen.de
[2]*Institut für Mathematische Stochastik, Universität Hannover, Postfach 60 09, D-30060 Hannover, Germany,* e-mail: rgrubel@stochastik.uni-hannover.de

(Received April 28, 2003; revised January 30, 2004)

Abstract. Given a sample from a discrete compound Poisson distribution, we consider variants of plug-in and likelihood estimators for the corresponding base distribution. These proceed recursively with an intermediate truncation step. We discuss the asymptotic behaviour of the estimators and give some numerical examples. Both procedures compare favourably with the straightforward and the naively projected plug-in estimator that we introduced in Buchmann and Grübel (2003, *The Annals of Statistics*, **31**, 1054–1074).

Key words and phrases: Compound distributions, consistency, discrete distributions, likelihood, limit distributions, plug-in principle.

1. Introduction

Poisson counting processes with bulk arrivals appear in various application areas such as queueing theory. They are one of the standard tools in stochastic modelling. If a process of this type is observed at evenly spaced time intervals then we obtain a sample from a discrete compound Poisson distribution. Formally, let $p = (p_k)_{k\in\mathbb{N}}$ be a probability distribution on the positive integers and let $\lambda > 0$. With '$\star$' denoting convolution we call the distribution $q = (q_k)_{k\in\mathbb{N}_0}$ on the non-negative integers given by

$$q = e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} p^{\star m}$$

the discrete compound Poisson distribution with rate parameter λ and base distribution p. Distributions of this type arise quite generally as random sums: If $N, X_1, X_2, X_3, \ldots$ are independent, N Poisson with parameter λ and p the probability mass function of the X-variables, then q is the probability mass function for $\sum_{m=1}^{N} X_m$. The q-values can be obtained from λ and p by an algorithm known in insurance mathematics as Panjer recursion,

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

Continuing the investigations in Buchmann and Grübel (2003), to which paper we also refer for a more detailed discussion of the problem and its applications, we consider two new estimators for the base distribution associated with a discrete compound Poisson distribution. These are introduced in the next section, which contains four theorems on their asymptotic behaviour. In Section 3 we give some examples with real data, one the canonical horse kick data, the other taken from the ecological literature. Section 4 investigates the finite sample behaviour of our estimators for some specific distributions by simulation.

2. Results

We first recall the definition of the plug-in estimator. In Subsections 2.2 and 2.3 respectively we explain and discuss the new proposals. In the final subsection we describe a connection between the three estimators.

2.1 *The plug-in estimator*

In Buchmann and Grübel (2003) we introduced an estimator which is based on the following inversion of the Panjer recursion,

$$\lambda = -\log q_0, \qquad p_k = \frac{q_k}{\lambda q_0} - \frac{1}{kq_0}\sum_{j=1}^{k-1} jp_j q_{k-j} \qquad \text{for all} \quad k \in \mathbb{N}.$$

Given a sample $Y_1, \ldots, Y_n$ of size n from such a distribution let $\hat{q}_n = (q_{n,k})_{k\in\mathbb{N}_0}$,

$$\hat{q}_{n,k} := \frac{1}{n}\#\{1 \le m \le n : Y_m = k\}$$

be the associated empirical probability mass function. The plug-in estimators $\hat{\lambda}_n$ and $\hat{p}_n^{\mathrm{PI}} = (\hat{p}_{n,k}^{\mathrm{PI}})_{k\in\mathbb{N}}$ for λ and p are then constructed by replacing the q-entries in the above inversion formula by the corresponding relative frequencies $\hat{q}_{n,k}$; in particular,

$$\hat{\lambda}_n := -\log \hat{q}_{n,0}.$$

Here and in the sequel we assume that $\hat{q}_{n,0} > 0$ which in view of our general assumption $\lambda > 0$ will be satisfied if n is large enough. In Buchmann and Grübel (2003) we obtained consistency and asymptotic normality for these estimators, but we also pointed out that the estimate for the base distribution is in general not a probability mass function as it may contain negative entries. One popular if crude remedy consists in replacing such negative entries by 0 and then renormalizing to sum 1, we will refer to this as the projected plug-in estimator $\hat{p}_n^{\mathrm{PPI}} = (\hat{p}_{n,k}^{\mathrm{PPI}})_{k\in\mathbb{N}}$ (in all estimators considered in this paper $\hat{\lambda}_n$ will be the same, so we do not need a distinguishing superscript for the rate parameter). Note that we risk an ambiguity in order to keep the notation compact: q with a single subscript refers to the components of q and $\hat{q}$ with the further subscript n refers to the empirical probability mass function. Below we will also use λ_0, $p_0 = (p_{0,k})_{k\in\mathbb{N}}$ and $q_0 = (q_{0,k})_{k\in\mathbb{N}_0}$ for the true parameters.

2.2 *The truncated plug-in estimator*

The first of our new proposals uses the above recursion but inserts a truncation step in order to insure that the entries are nonnegative and that their sum does not exceed the value 1. Formally, we define the truncated plug-in estimator $\hat{p}_n^{\mathrm{TP}} = (\hat{p}_{n,k}^{\mathrm{TP}})_{k\in\mathbb{N}_0}$ by

$$\hat{p}_{n,k}^{\mathrm{TP}} := \max\left\{0, \min\left\{x_{n,k}, 1 - \sum_{j=1}^{k-1} \hat{p}_{n,j}^{\mathrm{TP}}\right\}\right\},$$

with

$$x_{n,k} := \frac{\hat{q}_{n,k}}{\hat{\lambda}_n \hat{q}_{n,0}} - \frac{1}{k\hat{q}_{n,0}} \sum_{j=1}^{k-1} j\hat{p}_{n,j}^{\mathrm{TP}} \hat{q}_{n,k-j}.$$

By definition, $\hat{p}_n^{\mathrm{TP}}$ is a (sub)probability mass function. Also, $x_{n,k} \le 0$ whenever $\hat{q}_{n,k} = 0$ which shows that the support of the truncated plug-in estimator is contained in the support of $\hat{q}_n$. In particular, $\hat{p}_{n,k}^{\mathrm{TP}} = 0$ for $k > M_n := \max\{Y_1, \dots, Y_n\}$, so that the recursion can always be stopped after a finite number of steps.

The following two theorems deal with the asymptotic behaviour of $\hat{p}_n^{\mathrm{TP}}$. The first of these shows that the truncated plug-in estimators are strongly consistent.

THEOREM 2.1. *Let λ_0 be the true rate parameter and let $p_0 = (p_{0,k})_{k\in\mathbb{N}}$ be the true base distribution. Then $\hat{\lambda}_n \to \lambda_0$ and $\hat{p}_{n,k}^{\mathrm{TP}} \to p_{0,k}$ for all $k \in \mathbb{N}$ almost surely as $n \to \infty$.*

PROOF. We proceed by induction. Since $\hat{q}_{n,0} \to q_{0,0} = e^{-\lambda_0}$ and

$$x_{n,1} = -\frac{\hat{q}_{n,1}}{\hat{q}_{n,0}\log \hat{q}_{n,0}} \to -\frac{q_{0,1}}{q_{0,0}\log q_{0,0}} = -\frac{\lambda_0 p_{0,1} q_{0,0}}{q_{0,0}(-\lambda_0)} = p_{0,1}$$

almost surely, $\hat{\lambda}_n$ and the first component of $\hat{p}_n^{\mathrm{TP}}$ are consistent. Generally,

$$x_{n,k} = \Psi_k(\hat{q}_{n,0}, \dots, \hat{q}_{n,k}; \hat{p}_{n,1}^{\mathrm{TP}}, \dots, \hat{p}_{n,k-1}^{\mathrm{TP}})$$

with

$$\Psi_k(y_0, \dots, y_k; z_1, \dots, z_{k-1}) := -\frac{y_k}{y_0 \log y_0} - \frac{1}{ky_0} \sum_{j=1}^{k-1} j z_j y_{k-j}.$$

For $y_0 > 0$ (an assumption that is satisfied in our setup since y_0 corresponds to $e^{-\lambda_0}$) this is a continuous function, hence consistency of $\hat{q}_{n,j}$ for $j = 0, \dots, k$ and $\hat{p}_{n,j}^{\mathrm{TP}}$ for $j = 1, \dots, k-1$ implies that $x_{n,k}$ converges almost surely to

$$p_{0,k} = \Psi_k(q_{0,0}, \dots, q_{0,k}; p_{0,1}, \dots, p_{0,k-1}),$$

with the equality a consequence of Panjer inversion. The truncation step is continuous and leaves the limit invariant, hence $\hat{p}_{n,k}^{\mathrm{TP}} \to p_{0,k}$ as desired. □

We next consider the distributional asymptotics of the truncated plug-in estimator. In contrast to the situation in Buchmann and Grübel (2003) this new estimator is not a differentiable function of the empirical mass function $\hat{q}_n$ as the truncation introduces a continuous but non-differentiable step. As a consequence we still have the desirable 'parametric' rate $n^{-1/2}$ but the limit will in general not be a Gaussian process. Further, we only obtain (weak) convergence of the finite-dimensional distributions, which we abbreviate as '$\to_{\mathrm{fidi}}$'. This is a consequence of our method of proof, which relies on the recursive structure of the estimators. The truncation step in these recursions prevents the use of the canonical approach of transferring tightness by local linearization.

To define the limit process let $V = (V_k)_{k\in\mathbb{N}_0}$ be a sequence of centred Gaussian random variables with $\operatorname{cov}(V_k, V_j) = \delta_{kj}q_{0,k} - q_{0,k}q_{0,j}$. We define $Z^{\mathrm{TP}} = (Z_k^{\mathrm{TP}})_{k\in\mathbb{N}_0}$ recursively in terms of V, using an intermediate process $W = (W_k)_{k\in\mathbb{N}}$. For this, put $Z_0^{\mathrm{TP}} = -V_0/q_{0,0}$ and, for $k \in \mathbb{N}$,

$$W_k := \frac{q_{0,k}}{\lambda_0^2 q_{0,0}^2} V_0 - \frac{1}{kq_{0,0}} \sum_{j=0}^{k-1} (k-j) p_{0,k-j} V_j + \frac{1}{\lambda_0 q_{0,0}} V_k - \frac{1}{kq_{0,0}} \sum_{j=1}^{k-1} j q_{0,k-j} Z_j^{\mathrm{TP}}.$$

Then

$$Z_k^{\mathrm{TP}} := \begin{cases} W_k, & \text{if } p_{0,k} > 0 \text{ and } \sum_{j=1}^k p_{0,k} < 1, \\ \min\{W_k, -\sum_{j=1}^{k-1} Z_j^{\mathrm{TP}}\}, & \text{if } p_{0,k} > 0 \text{ and } \sum_{j=1}^k p_{0,k} = 1, \\ \max\{0, W_k\}, & \text{if } p_{0,k} = 0 \text{ and } \sum_{j=1}^k p_{0,k} < 1, \\ \max\{0, \min\{W_k, -\sum_{j=1}^{k-1} Z_j^{\mathrm{TP}}\}\}, & \text{if } p_{0,k} = 0 \text{ and } \sum_{j=1}^k p_{0,k} = 1. \end{cases}$$

The truncation step in the definition of the estimator leads to a truncation step in the construction of the limit process that depends on the support of the true base distribution. In particular, if $p_{0,k} > 0$ for all $k \in \mathbb{N}$ then $(Z_k^{\mathrm{TP}})_{k\in\mathbb{N}} = (W_k)_{k\in\mathbb{N}}$ and Z^{TP} is a Gaussian process. For the next theorem, we combine the rate parameter λ and the base distribution p into a single sequence (λ, p).

THEOREM 2.2. *Let λ_0 be the true rate parameter and let $p_0 = (p_{0,k})_{k\in\mathbb{N}}$ be the true base distribution. Then, with $Z^{\mathrm{TP}} = (Z_k^{\mathrm{TP}})_{k\in\mathbb{N}_0}$ as defined above,*

$$\sqrt{n}((\hat{\lambda}_n, \hat{p}_n^{\mathrm{TP}}) - (\lambda_0, p_0)) \to_{\mathrm{fidi}} Z^{\mathrm{TP}} \quad \text{as} \quad n \to \infty.$$

PROOF. The central limit theorem for multinomial distributions implies that $\sqrt{n}(\hat{q}_n - q_0) \to_{\mathrm{fidi}} V$, with $V = (V_k)_{k\in\mathbb{N}_0}$ as given above. Using a suitable construction we may even assume that the convergence holds pointwise for the respective random variables. (This step together with the subsequent local linearizations appears in many proofs of distributional convergence, see, e.g., Section 4 in Buchmann and Grübel (2003).) Since $\hat{\lambda}_n = -\log \hat{q}_{n,0}$ is a differentiable function of $\hat{q}_{n,0}$ and $\lambda_0 = -\log q_{0,0}$ we then obtain

$$\sqrt{n}(\hat{\lambda}_n - \lambda_0) \to Z_0^{\mathrm{TP}} = -\frac{1}{q_{0,0}} V_0.$$

Assume now that we have already shown that

$$\sqrt{n}((\hat{q}_{n,0}, \ldots, \hat{q}_{n,k}; \hat{p}_{n,1}^{\mathrm{TP}}, \ldots, \hat{p}_{n,k-1}^{\mathrm{TP}}) - (q_{0,0}, \ldots, q_{0,k}; p_{0,1}, \ldots, p_{0,k-1}))$$

converges pointwise to the random vector $(V_0, \ldots, V_k; Z_1^{\mathrm{TP}}, \ldots, Z_{k-1}^{\mathrm{TP}})$. Let Ψ_k and $x_{n,k}$ be as in the proof of Theorem 2.1. Then a standard calculus argument yields the pointwise convergence of

$$\begin{aligned} &\sqrt{n}(x_{n,k} - p_{0,k}) \\ &\quad = \sqrt{n}(\Psi_k(\hat{q}_{n,0}, \ldots, \hat{q}_{n,k}; \hat{p}_{n,1}^{\mathrm{TP}}, \ldots, \hat{p}_{n,k-1}^{\mathrm{TP}}) - \Psi_k(q_{0,0}, \ldots, q_{0,k}; p_{0,1}, \ldots, p_{0,k-1})) \end{aligned}$$

to

$$\sum_{j=0}^{k} \frac{\partial \Psi_k}{\partial y_j}(q_{0,0},\dots,q_{0,k};p_{0,1},\dots,p_{0,k-1})V_j$$
$$+\sum_{j=1}^{k-1} \frac{\partial \Psi_k}{\partial z_j}(q_{0,0},\dots,q_{0,k};p_{0,1},\dots,p_{0,k-1})Z_j^{\mathrm{TP}}.$$

A straightforward computation shows that this is equal to

$$\frac{q_{0,k}V_0}{q_{0,0}^2(\log q_{0,0})^2} - \frac{1}{kq_{0,0}}\sum_{j=0}^{k-1}(k-j)p_{0,k-j}V_j - \frac{V_k}{q_{0,0}\log q_{0,0}} - \frac{1}{kq_{0,0}}\sum_{j=1}^{k-1} jq_{0,k-j}Z_j^{\mathrm{TP}},$$

hence $\sqrt{n}(x_{n,k}-p_{0,k}) \to W_k$ with W_k as given above. The definition of $\hat{p}_n^{\mathrm{TP}}$ implies

$$\sqrt{n}(\hat{p}_{n,k}^{\mathrm{TP}} - p_{0,k}) = \max\left\{-\sqrt{n}p_{0,k}, \min\left\{\sqrt{n}(x_{n,k}-p_{0,k}), \sqrt{n}\left(1-\sum_{j=1}^{k}p_{0,j}\right) - y_{n,k}\right\}\right\}$$

with $y_{n,k} := \sum_{j=1}^{k-1}\sqrt{n}(\hat{p}_{n,j}^{\mathrm{TP}} - p_{0,j})$. This representation can be used for a proof by induction that the sequences $(y_{n,k})_{k\in\mathbb{N}}$ are bounded for all $k \in \mathbb{N}$. Hence, if $p_{0,k} > 0$ and $\sum_{j=1}^{k} p_{0,k} < 1$ then the right hand side will be equal to $\sqrt{n}(x_{n,k}-p_{0,k})$ for n large enough and therefore converge to W_k. A similar check of the other three cases shows that

$$\sqrt{n}(\hat{p}_{n,k}^{\mathrm{TP}} - p_{0,k}) \to Z_k^{\mathrm{TP}}.$$

Putting pieces together we obtain that

$$\sqrt{n}((\hat{q}_{n,0},\dots,\hat{q}_{n,k+1};\hat{p}_{n,1}^{\mathrm{TP}},\dots,\hat{p}_{n,k}^{\mathrm{TP}}) - (q_{0,0},\dots,q_{0,k+1};p_{0,1},\dots,p_{0,k}))$$

converges to $(V_0,\dots,V_{k+1};Z_1^{\mathrm{TP}},\dots,Z_k^{\mathrm{TP}})$. Switching back to the original variables we see that this completes the proof of the induction step for the convergence of the finite-dimensional distributions. □

Theorem 2.2 shows that we get a complicated limit process, but the result has some statistical significance. This rests upon two observations: First, the finite-dimensional distributions $\mathcal{L}((Z_j^{\mathrm{TP}})_{j=0,\dots,k} \mid \lambda, p)$, $k \in \mathbb{N}_0$, of the limit process depend on the unknown parameter (λ, p) in a continuous manner, as is obvious from its construction. Hence we can 'studentize', i.e., use $\mathcal{L}((Z_j^{\mathrm{TP}})_{j=0,\dots,k} \mid \hat{\lambda}_n, \hat{p}_n^{\mathrm{TP}})$ to estimate $\mathcal{L}((Z_j^{\mathrm{TP}})_{j=0,\dots,k} \mid \lambda, p)$. In view of Theorem 2.1 this will lead to asymptotically correct confidence regions for finite sets of parameter components if the construction of these regions allows the application of the continuous mapping theorem. Still, it remains to find, e.g., the quantiles of $\mathcal{L}(Z_k^{\mathrm{TP}} \mid \hat{\lambda}_n, \hat{p}_n^{\mathrm{TP}})$. For this, the second observation is useful: A centred Gaussian process $(V_k)_{k\in\mathbb{N}_0}$ with covariance structure

$$\mathrm{var}(V_k) = q_k(1-q_k), \qquad \mathrm{cov}(V_l, V_k) = -q_l q_k \quad \text{for all} \quad k,l \in \mathbb{N}_0 \quad \text{with} \quad k \neq l$$

can be obtained recursively from a sequence $(\xi_k)_{k\in\mathbb{N}_0}$ of independent centred normal random variables with

$$\mathrm{var}(\xi_k) = q_k t_k t_{k+1}, \qquad t_k := \sum_{j=k}^{\infty} q_j \quad \text{for all} \quad k \in \mathbb{N}_0$$

as follows: $V_0 := \xi_0$,

$$V_k := \frac{1}{t_k}\left(\xi_k - q_k \sum_{j=0}^{k-1} V_j\right)$$

for all $k \in \mathbb{N}$ with $t_k > 0$, $V_k = 0$ if $t_k = 0$. Together with the above constructive description of Z^{TP} this makes it easy to generate values from some initial segment of the limit process so that numerical approximations for quantiles etc. can be obtained by simulation. Similarly, we can construct critical regions for tests of simple hypotheses if these involve a finite set of parameters only; see also Subsection 3.3 below.

2.3 *The truncated maximum likelihood estimator*

Our second estimator uses likelihood ideas, but otherwise the approach is very similar. Suppose that we base the estimation of λ_0 on $\hat{q}_{n,0}$ and that of $p_{0,k}$ on $\hat{q}_{n,0}, \dots, \hat{q}_{n,k}$ as we have done in the various forms of plug-in estimation. We obtain the recursive step by assuming when estimating $p_{0,k}$ that the estimates for λ_0 and $p_{0,1}, \dots, p_{0,k-1}$ are exact. Note that this is only a heuristic motivation for the following formal definition. Again, $\hat{\lambda}_n = -\log \hat{q}_{n,0}$. If the original data are truncated at $k+1$ in the sense that we replace Y_l by $\min\{Y_l, k+1\}$ for $l = 1, \dots, n$ then the likelihood associated with $(\lambda, p_1, \dots, p_k)$ is given by

$$L_{n,k}(\lambda, p_1, \dots, p_k) = \sum_{j=0}^{k} \hat{q}_{n,j} \log q_j + \left(1 - \sum_{j=0}^{k} \hat{q}_{n,j}\right) \log \left(1 - \sum_{j=0}^{k} q_j\right),$$

where $q_0, \dots, q_k$ are the corresponding compound probabilities, related to the arguments of $L_{n,k}$ by Panjer recursion. We now define the truncated maximum likelihood estimator $\hat{p}_n^{\mathrm{TL}} = (\hat{p}_{n,k}^{\mathrm{TL}})_{k\in\mathbb{N}}$ recursively: Given $\hat{\lambda}_n$ and $\hat{p}_{n,j}^{\mathrm{TL}}$ for $j = 1, \dots, k-1$ let $\hat{p}_{n,k}^{\mathrm{TL}}$ be the value that maximizes the function

$$x \mapsto L_{n,k}(\hat{\lambda}_n, \hat{p}_{n,1}^{\mathrm{TL}}, \dots, \hat{p}_{n,k-1}^{\mathrm{TL}}, x)$$

on the interval $[0, 1 - \sum_{j=1}^{k-1} \hat{p}_{n,j}^{\mathrm{TL}}]$. This argmax exists, is unique and can be given explicitly. To see this, we first consider the case $k = M_n = \max\{Y_1, \dots, Y_n\}$. Then the second part of $L_{n,k}$ vanishes. In the remaining sum only q_k depends on p_k, q_k is a strictly increasing function of p_k and $\hat{q}_{n,k} > 0$, hence p_k has to be chosen as large as possible. The unique maximizer is therefore given by $1 - \sum_{j=1}^{k-1} \hat{p}_{n,j}^{\mathrm{TL}}$. This also implies that $\hat{p}_{n,j}^{\mathrm{TL}} = 0$ for $j > M_n$; in particular, the truncated maximum likelihood estimator is a (proper) probability mass function and the recursion can be stopped after a finite number of steps. For $k < M_n$ we rewrite the function that has to be maximized as follows,

$$g(x) = C_1 + C_2 \log(C_3 + C_4 x) + C_5 \log(1 - C_6 - C_3 - C_4 x)$$

with

$$C_2 = \hat{q}_{n,k}, \qquad C_3 = \frac{\hat{\lambda}_n}{k} \sum_{j=1}^{k-1} j \hat{p}_{n,j}^{\mathrm{TL}} \hat{q}_{n,k-j}^{\mathrm{TL}}, \qquad C_4 = \hat{\lambda}_n \hat{q}_{n,0}, \qquad C_5 = 1 - \sum_{j=0}^{k} \hat{q}_{n,j}$$

and $C_6 = \sum_{j=0}^{k-1} \hat{q}_{n,j}^{\mathrm{TL}}$. Here $\hat{q}_n^{\mathrm{TL}}$ denotes the compound distribution with rate parameter $\hat{\lambda}_n$ and base distribution $\hat{p}_n^{\mathrm{TL}}$. None of the constants $C_1, \dots, C_6$ depend on x and we may assume that $C_4 > 0$. If $C_2 = 0$ then g is strictly decreasing, which leads to $x = 0$. If $C_2 > 0$ then standard calculations show that the pre-truncation argmax of the strictly concave function g is uniquely given by

$$\begin{aligned} x_{n,k} &= \frac{C_2 - C_2C_3 - C_2C_6 - C_3C_5}{C_2C_4 + C_4C_5} \\ &= \frac{\hat{q}_{n,k}}{\hat{\lambda}_n \hat{q}_{n,0}} \cdot \frac{1 - \sum_{j=0}^{k-1} \hat{q}_{n,j}^{\mathrm{TL}}}{1 - \sum_{j=0}^{k-1} \hat{q}_{n,j}} - \frac{1}{k\hat{q}_{n,0}} \sum_{j=1}^{k-1} j \hat{p}_{n,j}^{\mathrm{TL}} \hat{q}_{n,k-j}^{\mathrm{TL}}, \end{aligned}$$

so that finally

$$\hat{p}_{n,k}^{\mathrm{TL}} = \max\left\{0, \min\left\{x_{n,k}, 1 - \sum_{j=1}^{k-1} \hat{p}_{n,j}^{\mathrm{TL}}\right\}\right\}.$$

It may be interesting to note that the auxiliary quantity $x_{n,k}$ reduces to the one that we introduced in connection with the truncated plug-in estimator if we replace $\hat{q}_{n,k}^{\mathrm{TL}}$ by $\hat{q}_{n,k}$. As in the plug-in case we have that the support of $\hat{p}_n^{\mathrm{TL}}$ is a subset of the support of $\hat{q}_n$.

THEOREM 2.3. *Let λ_0 be the true rate parameter and let $p_0 = (p_{0,k})_{k\in\mathbb{N}}$ be the true base distribution. Then $\hat{\lambda}_n \to \lambda_0$ and $\hat{p}_{n,k}^{\mathrm{TL}} \to p_{0,k}$ for all $k \in \mathbb{N}$ almost surely as $n \to \infty$.*

PROOF. We proceed as in the proof of Theorem 2.1; indeed, the induction start remains unchanged as $\hat{\lambda}_n$, $x_{n,1}$ and therefore the estimator for $p_{0,1}$ are the same for truncated plug-in and truncated maximum likelihood. For the induction step we use $x_{n,k} = \Phi_k(\hat{q}_{n,0}, \dots, \hat{q}_{n,k}; \hat{p}_{n,1}^{\mathrm{TL}}, \dots, \hat{p}_{n,k-1}^{\mathrm{TL}})$ with

$$\Phi_k(y_0, \dots, y_k; z_1, \dots, z_{k-1}) := -\frac{y_k}{y_0 \log y_0} \frac{1 - \sum_{j=0}^{k-1} q_j}{1 - \sum_{j=0}^{k-1} y_j} - \frac{1}{ky_0} \sum_{j=1}^{k-1} j z_j q_{k-j}$$

where the functions q_k are given recursively by $q_0(y_0) = y_0$,

$$q_k = q_k(y_0; z_1, \dots, z_k) = -\frac{\log y_0}{k} \sum_{j=1}^{k} j z_j q_{k-j}(y_0; z_1, \dots, z_{k-j}).$$

Again, Φ_k is continuous at the true parameter value, which provides the basis for the induction step. □

For the corresponding distributional limit result we again give the construction of the limit process first. We need the auxiliary sequences $a_0 = (a_{0,k})_{k\in\mathbb{N}}$ and $b_0 = (b_{0,k})_{k\in\mathbb{N}}$ defined by

$$a_{0,k} := \sum_{j=1}^{k} p_{0,j} q_{0,k-j}, \qquad b_{0,k} := \sum_{j=1}^{k-1} j p_{0,j} a_{0,k-j} \qquad \text{for all} \quad k \in \mathbb{N}.$$

Note that $a_0 = p_0 \star q_0$ is a probability mass function. Further let $(t_{0,k})_{k\in\mathbb{N}_0}$ and $(c_{0,k})_{k\in\mathbb{N}_0}$ denote the tail sequences associated with q_0 and a_0 respectively, i.e.

$$t_{0,k} := \sum_{j=k}^{\infty} q_{0,j}, \qquad c_{0,k} := \sum_{j=k}^{\infty} a_{0,j} \qquad \text{for all} \quad k \in \mathbb{N}_0.$$

As for Theorem 2.2, let $V = (V_k)_{k\in\mathbb{N}_0}$ be a sequence of centred Gaussian random variables with $\operatorname{cov}(V_k, V_j) = \delta_{kj} q_{0,k} - q_{0,k}q_{0,j}$. Again we define $Z^{\mathrm{TL}} = (Z_k^{\mathrm{TL}})_{k\in\mathbb{N}_0}$ recursively, using auxiliary variables W_k, $k \in \mathbb{N}$: Let $Z_0^{\mathrm{TL}} = -V_0/q_{0,0}$ and, for $k \in \mathbb{N}$,

$$\begin{aligned} W_k := \left(\frac{q_{0,k}}{\lambda_0^2 q_{0,0}^2} - \frac{q_{0,k}c_{0,k}}{\lambda_0 q_{0,0}^2 t_{0,k}} + \frac{b_{0,k}}{k q_{0,0}^2} \right) V_0 + \frac{q_{0,k}}{\lambda_0 q_{0,0} t_{0,k}} \sum_{j=0}^{k-1} V_j \\ + \frac{1}{\lambda_0 q_{0,0}} V_k - \sum_{j=1}^{k-1} \left(\frac{q_{0,k}(1-t_{0,k-j})}{q_{0,0}t_{0,k}} + \frac{q_{0,k-j}}{q_{0,0}} \right) Z_j^{\mathrm{TL}}. \end{aligned}$$

Then

$$Z_k^{\mathrm{TL}} := \begin{cases} W_k, & \text{if } p_{0,k} > 0 \text{ and } \sum_{j=1}^k p_{0,k} < 1, \\ \min\{W_k, -\sum_{j=1}^{k-1} Z_j^{\mathrm{TL}}\}, & \text{if } p_{0,k} > 0 \text{ and } \sum_{j=1}^k p_{0,k} = 1, \\ \max\{W_k, 0\}, & \text{if } p_{0,k} = 0 \text{ and } \sum_{j=1}^k p_{0,k} < 1, \\ \max\{0, \min\{W_k, -\sum_{j=1}^{k-1} Z_j^{\mathrm{TL}}\}\}, & \text{if } p_{0,k} = 0 \text{ and } \sum_{j=1}^k p_{0,k} = 1. \end{cases}$$

Note that the truncation step is identical to the one that we used in connection with the limit process for the truncated plug-in estimator.

THEOREM 2.4. *Let λ_0 be the true rate parameter and let $p_0 = (p_{0,k})_{k\in\mathbb{N}}$ be the true base distribution. Then, with $Z^{\mathrm{TL}} = (Z_k^{\mathrm{TL}})_{k\in\mathbb{N}_0}$ as defined above,*

$$\sqrt{n}((\hat\lambda_n, \hat p_n^{\mathrm{TL}}) - (\lambda_0, p_0)) \to_{\mathrm{fidi}} Z^{\mathrm{TL}} \quad \text{as} \quad n \to \infty.$$

PROOF. Let Φ_k and q_k be as in the proof of Theorem 2.3. If we regard the exponential function as a non-linear operator on the space of summable sequences, a view that has been used extensively in Buchmann and Grübel (2003), then the convolution series that gives q in terms of λ and p can be written as $q = \exp(\lambda(p - \delta_0))$. This leads to

$$\frac{\partial q_k}{\partial z_l}(q_{0,0}; p_{0,1}, \dots, p_{0,k}) = \begin{cases} -\log(q_{0,0}) q_{0,k-l}, & \text{if } k \ge l, \\ 0, & \text{if } k < l. \end{cases}$$

Alternatively, this can be verified by induction on using the recursive definition of q_k. The convolution series representation of q also gives

$$\frac{\partial q_k}{\partial y_0}(q_{0,0}; p_{0,1}, \dots, p_{0,k}) = \frac{1}{q_{0,0}}(q_{0,k} - (p_0 \star q_0)_k) = \frac{1}{q_{0,0}}(q_{0,k} - a_{0,k}).$$

Note that $q_j(q_{0,0}; p_{0,1}, \dots, p_{0,j}) = q_{0,j}$. From these we obtain, with $\lambda_0 = -\log q_{0,0}$ and $(\cdots)$ abbreviating $(q_{0,0}, \dots, q_{0,k}; p_{0,1}, \dots, p_{0,k-1})$,

$$\begin{aligned} \frac{\partial \Phi_k}{\partial y_0}(\cdots) &= \frac{q_{0,k}}{\lambda_0^2 q_{0,0}^2} + \frac{q_{0,k}}{\lambda_0 q_{0,0} t_{0,k}} - \frac{q_{0,k}c_{0,k}}{\lambda_0 q_{0,0}^2 t_{0,k}} + \frac{b_{0,k}}{k q_{0,0}^2}, \\ \frac{\partial \Phi_k}{\partial y_j}(\cdots) &= \frac{q_{0,k}}{\lambda_0 q_{0,0} t_{0,k}} \quad \text{for} \quad j = 1, \dots, k-1, \\ \frac{\partial \Phi_k}{\partial y_k}(\cdots) &= \frac{1}{\lambda_0 q_{0,0}}, \\ \frac{\partial \Phi_k}{\partial z_j}(\cdots) &= -\frac{q_{0,k}(1-t_{0,k-j})}{q_{0,0}t_{0,k}} - \frac{q_{0,k-j}}{q_{0,0}} \quad \text{for} \quad j = 1, \dots, k-1. \end{aligned}$$

Using these we see that

$$W_k = \sum_{j=0}^{k} \frac{\partial \Phi_k}{\partial y_j}(\cdots)V_j + \sum_{j=1}^{k-1} \frac{\partial \Phi_k}{\partial z_j}(\cdots)Z_j^{\mathrm{TL}}$$

and we can now continue as in the proof of Theorem 2.2. □

The same remarks as given after Theorem 2.2 apply in this situation too. In fact, a slight simplification occurs as $\partial\Phi_k/\partial y_j$ for $j = 1, \dots, k-1$ does not depend on j: With

$$\xi_k := q_{0,k} \sum_{j=0}^{k-1} V_j + t_{0,k}V_k,$$

which produces a sequence of independent centred normal random variables with $\mathrm{var}(\xi_k) = q_{0,k}t_{0,k}t_{0,k+1}$, we obtain

$$\begin{aligned} W_k = &\left(\frac{q_{0,k}}{\lambda_0^2 q_{0,0}^2} - \frac{q_{0,k}c_{0,k}}{\lambda_0 q_{0,0}^2 t_{0,k}} + \frac{b_{0,k}}{k q_{0,0}^2} \right) \xi_0 \\ &+ \frac{1}{\lambda_0 q_{0,0} t_{0,k}} \xi_k - \sum_{j=1}^{k-1} \left(\frac{q_{0,k}(1 - t_{0,k-j})}{q_{0,0} t_{0,k}} + \frac{q_{0,k-j}}{q_{0,0}} \right) Z_j^{\mathrm{TL}}. \end{aligned}$$

Written in this form the recursion is driven by independent random variables, which is convenient in connection with simulations.

2.4 *Backwards compatibility*

In Buchmann and Grübel (2003) we regarded the plug-in estimator $\hat{p}_n^{\mathrm{PI}} = (\hat{p}_{n,k}^{\mathrm{PI}})_{k\in\mathbb{N}}$ as a point in a suitable sequence space and we directly analyzed its dependence on the sequence (point) $\hat{q}_n = (\hat{q}_{n,k})_{k\in\mathbb{N}_0}$. Alternatively, and in the style of the present paper, we can write

$$\hat{p}_{n,k}^{\mathrm{PI}} = \Psi_k(\hat{q}_{n,0}, \dots, \hat{q}_{n,k}; \hat{p}_{n,1}^{\mathrm{PI}}, \dots, \hat{p}_{n,k-1}^{\mathrm{PI}})$$

and prove consistency and convergence of the finite-dimensional distributions of $\sqrt{n}(\hat{p}_n^{\mathrm{PI}} - p_0)$ as $n \to \infty$, using arguments from the proofs of Theorems 2.1 and 2.2. Apart from providing an alternative method of proof (leading to a weaker distributional result) the recursive structure of the unadorned plug-in estimator, as displayed above, also leads to the following two observations: First, if

$$p_{0,1} > 0, \dots, p_{0,k} > 0 \quad \text{and} \quad \sum_{i=1}^{k} p_{0,i} < 1$$

then the (strong) consistency of the plug-in estimator implies that there exist an $n_0 \in \mathbb{N}$ and a set of probability zero such that, outside this set and for all $n \geq n_0$,

$$\hat{p}_{n,1}^{\mathrm{PI}} > 0, \dots, \hat{p}_{n,k}^{\mathrm{PI}} > 0 \quad \text{and} \quad \sum_{i=1}^{k} \hat{p}_{n,i}^{\mathrm{PI}} < 1.$$

A truncation then does not occur in the first k steps and therefore $\hat{p}_{n,i}^{\mathrm{PI}} = \hat{p}_{n,i}^{\mathrm{TP}}$ for $i = 1, \dots, k$. Essentially the same arguments apply to the truncated maximum likelihood estimator. Indeed, as we will see in the numerical examples in the next section, the three estimates will typically coincide for the first k components and then bifurcate. If truncation occurs at that stage because of $\sum_{i=1}^{k} \hat{p}_{n,i}^{\mathrm{PI}} \geq 1$, then the truncated plug-in and the truncated maximum likelihood estimates will be identical.

Secondly, in the special case with $p_{0,k} > 0$ for all $k \in \mathbb{N}$ (an assumption that holds for some popular parametric families, see Subsection 3.2 below) we can use this argument, together with the familiar fact that the finite-dimensional distributions determine the distribution of a stochastic process with countable index set, to show that $\mathcal{L}(Z^{\mathrm{TP}}) = \mathcal{L}(Z^{\mathrm{TL}}) = \mathcal{L}(W)$, with W the Gaussian limit process obtained in Buchmann and Grübel (2003) for the plug-in estimator.

3. Examples

We consider two real data sets in the first two subsections. In Subsection 3.3 we discuss the applicability of our results to tests of two hypotheses that arise in these examples.

3.1 *The horse kick data*

As in Buchmann and Grübel (2003) we first apply our procedures to the time-honoured Prussian horse kick data; see, e.g., Quine and Seneta (1987). Of the 200 observations 109, 65, 22, 3 and 1 respectively are equal to k = 0, 1, 2, 3 and 4. Table 1 displays the various estimates, for reference we also give the plug-in and projected plug-in estimates in the second and third line. In contrast to our new proposals these have unbounded support. We use the heuristic argument that Y-values smaller than some k cannot possibly contain any information about $p_{0,l}$ for $l \geq k$ and stop the recursion underlying the plug-in estimator at the largest observed value; this is also used as the basis for the projection in the third line. Next are the truncated plug-in and the truncated likelihood estimates; we see that both are closer to the traditional interpretation of these data as being from an ordinary Poisson distribution. Also, both are identical, as announced in Subsection 2.4.

The next three lines give the respective q-values, beginning with the relative frequencies. By construction, these are equal to the q-values for the straight plug-in estimate. The final line contains the result of the usual Poisson approximation, with λ estimated by the mean 0.61 of the data (all decompounding estimators considered in this paper use $\hat{\lambda}_n = -\log \hat{q}_{n,0} = -\log 0.545 = 0.606969\ldots$). We see that the truncation estimators

Table 1. The horse kick data.

k	0	1	2	3	4
$\hat{p}_{n,k}^{\mathrm{PI}}$	—	0.9825	0.0396	−0.0365	0.0207
$\hat{p}_{n,k}^{\mathrm{PPI}}$	—	0.9422	0.0380	0	0.0198
$\hat{p}_{n,k}^{\mathrm{TP}}$, $\hat{p}_{n,k}^{\mathrm{TL}}$	—	0.9825	0.0175	0	0
$\hat{q}_{n,k}$, $\hat{q}_{n,k}^{\mathrm{PI}}$	0.5450	0.3250	0.1100	0.0150	0.0050
$\hat{q}_{n,k}^{\mathrm{PPI}}$	0.5450	0.3117	0.1017	0.0242	0.0112
$\hat{q}_{n,k}^{\mathrm{TP}}$, $\hat{q}_{n,k}^{\mathrm{TL}}$	0.5450	0.3250	0.1027	0.0227	0.0039
Poisson	0.5434	0.3314	0.1011	0.0206	0.0031

give a notably better fit on the q-side than the naively projected plug-in estimator:

$$\sum_{k=0}^{4} |\hat{q}_{n,k}^{\mathrm{PPI}} - \hat{q}_{n,k}| = 0.037, \qquad \sum_{k=0}^{4} |\hat{q}_{n,k}^{\mathrm{TP/TL}} - \hat{q}_{n,k}| = 0.016.$$

3.2 *The plant data*

Compound Poisson distributions ('contagious distributions') also appear in the ecological literature where they are used to model plant and insect populations. In the basic model, apparently due to Neyman (1939), it is assumed that ancestor plants or insects are distributed in a given area according to a two-dimensional Poisson process with constant intensity. These have random numbers of offspring, independent and identically distributed, which stay close to their respective ancestors. Dividing a given (sufficiently homogeneous) area into subareas of equal size and ignoring edge effects one then regards the counts for the subareas as a sample from a compound Poisson distribution. This may be seen as a two-dimensional variant of our motivating example of queues with bulk arrivals. Neyman (1939) advocated the use of a Poisson base distribution, the resulting family of compound distributions is also known as the Neyman Type A family. In the case of a geometric base we similarly arrive at the Pólya-Aeppli distributions; see Chapter 9 in Johnson *et al.* (1992). (Atoms at zero of the base distribution can be incorporated into the rate parameter.) A third popular parametric family in this area is the family of negative binomial distributions. These are also of the compound Poisson type, the special case of geometric distributions is used below in one of the simulation examples.

In an effort to find out which of these three families is appropriate for plant or insect populations Evans (1953) collected and analyzed a variety of data sets. For plant populations he generally regards the Neyman Type A distributions as appropriate, but for one of his data sets (14c in the paper) the Pólya-Aeppli distribution results in a better fit. In Table 2 we give this data set together with our estimates for the base distribution. Again, the plug-in estimate has a negative entry and the truncated plug-in and truncated likelihood estimates are identical. The data here are such that the truncation step in the definition of $\hat{p}_{n,k}^{\mathrm{TP}}$, $\hat{p}_{n,k}^{\mathrm{TL}}$ first takes effect with $k = 8$, hence both are equal to $\hat{p}_{n,k}^{\mathrm{PI}}$ for $k = 1, \ldots, 7$. As a consequence these estimates give a perfect fit of observed and expected frequencies in this k-range, which cannot be obtained with any of the parametric models mentioned above. On the other hand a parametric model, if correct, could be used to extrapolate beyond the range of the observations, for example by providing an estimate for high quantiles of the offspring distribution.

For data such as these our procedures provide a partly nonparametric alternative to the classical approach. In effect, we estimate the offspring distribution directly, without any parametric assumptions, but the assumptions on the spatial distribution of the ancestors remain in force.

Table 2. The plant data.

k	0	1	2	3	4	5	6	7	8	9	10	11	12
Counts	274	71	58	36	20	12	10	7	6	3	0	2	1
$\hat{p}_{n,k}^{\mathrm{PI}}$	—	.431	.296	.137	.049	.023	.029	.018	.018	.002	−.011	.009	.003
$\hat{p}_{n,k}^{\mathrm{TP}}, \hat{p}_{n,k}^{\mathrm{TL}}$	—	.431	.296	.137	.049	.023	.029	.018	.016	0	0	0	0

3.3 *Significance tests*

The numerical examples in the previous two subsections are mainly meant to illustrate the estimators that we introduced in Section 2 and to compare them with the plug-in estimators in Buchmann and Grübel (2003). Of course, the question arises as to what extent our asymptotic results can be used in connection with formal significance tests, for example of the hypothesis that we do have a straight Poisson distribution in the first example or whether the deviation from a geometric distribution is significant in the second.

In the first case the hypothesis can be written as $p_{0,1} = 1$ and the procedure mentioned at the end of Subsection 2.2 can be applied. For the test statistic $T_n := \sqrt{n}(1 - \hat{p}_{n,1}^{\mathrm{PI}})$, for example, Theorem 2.2 leads to the distributional approximation $-Z_1^{\mathrm{TP}} = \max\{-W_1, 0\}$ with $W_1 \sim N(0, \sigma^2(\lambda))$ and

$$\sigma^2(\lambda) = \frac{1 - \lambda + \lambda^2 - e^{-\lambda}}{\lambda^2 e^{-\lambda}}.$$

Inserting $\hat{\lambda}_n = 0.606969$ we arrive at the approximate p-value 0.4058 for the observation $\sqrt{200}(1 - 0.9825) = 0.2474\ldots$ of T_n.

The hypothesis of a geometric base distribution in the second example does not have this simple form and the familiar problems with goodness-of-fit tests using estimated parameters arise; see Pollard ((1984), pp. 99 and 159) for a classical case. A modern approach to problems of this type circumvents the explicit distributional approximation by estimating the distribution of the test statistic directly, using a combination of the plug-in principle and Monte Carlo approximation (bootstrap tests). In this context an extension of our results to the case of a converging sequence of rate parameters and base distributions would be of interest.

4. Some simulation experiments

In our next two examples we use simulated data, with $\lambda_0 = 2$ and p_0 the uniform distribution on the set $\{1, 4, 6\}$ in the first case. Figure 1(a) shows the result of 50 simulations with sample size $n = 500$. Displayed are the corresponding absolute error sums, with $\circ$ and $+$ for the vectors with coordinates

$$\left(\sum_{k=0}^{M_n} |\hat{p}_{n,k}^{\mathrm{TP}} - p_{0,k}|, \sum_{k=0}^{M_n} |\hat{p}_{n,k}^{\mathrm{PPI}} - p_{0,k}|\right) \quad \text{and} \quad \left(\sum_{k=0}^{M_n} |\hat{p}_{n,k}^{\mathrm{TP}} - p_{0,k}|, \sum_{k=0}^{M_n} |\hat{p}_{n,k}^{\mathrm{TL}} - p_{0,k}|\right)$$

respectively. To make the comparisons easier the plots include the line $x \mapsto (x, x)$. The figure shows that, at least in this particular example, the new estimators both considerably improve upon the projected plug-in estimate, and that the two new estimators show a very similar performance.

In the second example with artificial data we take q_0 to be the geometric distribution with parameter $\alpha = 0.25$. It is known that this is a compound Poisson distribution with rate parameter $\lambda = -\log(\alpha)$ and with the logarithmic distribution

$$p_{0,k} = -\frac{(1-\alpha)^k}{k \log \alpha}, \qquad k \in \mathbb{N},$$

as base distribution; see, e.g., Chapter 7 in Johnson *et al.* (1992). As explained in Subsection 2.4 the limit processes are then the same for all three estimators, which

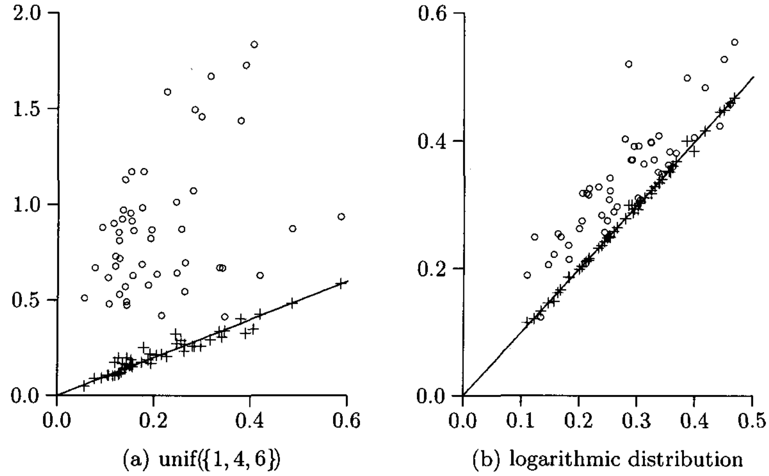

Fig. 1. Error comparisons for simulated data. o: PPI vs. TP, +: TL vs. TP.

Table 3. Support results with unif($\{1,4,6\}$).

	$n=500$, $\lambda=2$						$n=1000$, $\lambda=4$					
k	2	3	5	7	8	≥ 9	2	3	5	7	8	≥ 9
TP	50.2	62.9	48.0	71.0	86.3	88.7	51.1	68.1	42.7	62.5	85.4	93.9
TL	50.2	75.4	47.7	64.5	81.9	76.3	51.1	77.2	35.2	52.0	76.0	74.9
PPI	50.2	47.4	47.8	52.0	48.1	0.0	51.1	43.6	33.3	43.8	44.7	0.0

leads us to suspect that the projected plug-in estimate can compete with the truncation estimates. To some extent this is confirmed by Fig. 1(b).

In our last experiment we consider the performance of the truncated plug-in, the truncated likelihood and the projected plug-in estimators with respect to a structural property of the base distribution such as its support. Again, the base distribution is uniform on the set $\{1,4,6\}$. Table 3 gives the percentages of the correct results in 1000 simulations for two different sample sizes and rate parameters. For example, the last value 93.9 in the first line means that the truncated plug-in estimator gave the correct value $\sum_{k=9}^{\infty} p_{0,k} = 0$ in 939 of the 1000 runs with $n = 1000$, $\lambda = 4$. (The values in the table remain essentially unchanged if we replace the condition $x = 0$ by $|x| < 0.001$.) It appears that the truncated plug-in procedure is slightly superior to the truncated maximum likelihood variant for large k-values, with the order reversed for $k = 3$. Again, both outperform the projected plug-in estimator.

In the degenerate case, with data from an ordinary Poisson distribution, getting the support right means that the base distribution is estimated with zero error. Interestingly, the limiting probability that this occurs is equal to $1/2$, irrespective of the rate parameter:

$$\begin{aligned} P(\hat{p}_{n,1}^{\mathrm{TP}} = 1) &= P(\hat{p}_{n,1}^{\mathrm{TL}} = 1) \\ &= P(\hat{q}_{n,1} \geq \hat{\lambda}_n \hat{q}_{n,0}) \\ &= P(\sqrt{n}(\hat{q}_{n,1} - q_{0,1}) - \hat{q}_{n,0}\sqrt{n}(\hat{\lambda}_n - \lambda_0) - \lambda_0\sqrt{n}(\hat{q}_{n,0} - q_{0,0}) \geq 0) \end{aligned}$$

$$\begin{aligned} &\to P(V_1 + (1-\lambda_0)V_0 \geq 0) \quad \text{as} \quad n \to \infty \\ &= 1/2. \end{aligned}$$

The support results may seem to be a bit disappointing but they can be regarded as another instance of a boundary effect familiar in order-restricted statistical inference. A canonical example is provided by a sample $X_1, \ldots, X_n$ from the normal distribution where we know that the true mean μ is nonnegative. If we estimate μ by $\hat{\mu}_n := \max\{\bar{X}_n, 0\}$, $\bar{X}_n := n^{-1}\sum_{i=1}^n X_i$, then we have $P(\hat{\mu}_n = \mu) = 1/2$ at the boundary $\mu = 0$ of the parameter space; see also the distributional approximation for the test statistic T_n in Subsection 3.3.

As a final comment we mention a drawback of the estimators considered so far: They do not provide a sensible result if no zero values are observed. Ordinary maximum likelihood estimators do not have this drawback but have some others instead, as will be discussed in a separate paper.

Acknowledgements

We would like to thank both referees for their detailed and inspiring comments.

References

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

Evans, D. A. (1953). Experimental evidence concerning contagious distributions in ecology, *Biometrika*, **40**, 186–211.

Johnson, N. L., Kotz, S. and Kemp, A. W. (1992). *Univariate Discrete Distributions*, 2nd ed., Wiley, New York.

Neyman, J. (1939). On a new class of "contagious" distributions applicable in entomology and bacteriology, *The Annals of Mathematical Statistics*, **10**, 35–57.

Quine, M. P. and Seneta, E. (1987) Bortkiewicz's data and the law of small numbers, *International Statististical Review*, **55**, 173–181.

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