



HAL
open science

Semiparametric estimation for count data through weighted distributions

Célestin C. Kokonendji, Tristan Senga Kiessé, Narayanaswamy Balakrishnan

► **To cite this version:**

Célestin C. Kokonendji, Tristan Senga Kiessé, Narayanaswamy Balakrishnan. Semiparametric estimation for count data through weighted distributions. *Journal of Statistical Planning and Inference*, 2009, 139 (10), pp.3625-3638. 10.1016/j.jspi.2009.04.013 . hal-00947788

HAL Id: hal-00947788

<https://hal.science/hal-00947788>

Submitted on 19 Feb 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Semiparametric estimation for count data through weighted distributions

C.C. Kokonendji ^{a,*}, T. Senga Kiessé ^a, N. Balakrishnan ^b

^a*Université de Pau et des Pays de l'Adour, LMA-UMR 5142 CNRS, Pau, France*

^b*McMaster University, Hamilton, Ontario, Canada L8S 4K1*

Abstract

This paper is concerned with semiparametric discrete kernel estimators when the unknown count distribution can be considered to have a general weighted Poisson form. The estimator is constructed by multiplying the Poisson estimate with a nonparametric discrete kernel-type estimate of the Poisson weight function. Comparisons are then carried out with the ordinary discrete kernel probability mass function estimators. The Poisson weight function is thus a local multiplicative correction factor, and is considered as the uniform measure to detect departures from the equidispersed Poisson distribution. In this way, the effects of dispersion and zero-proportion with respect to the standard Poisson distribution are also minimized. This method of estimation is also applied to the weighted binomial form for the count distribution having a finite support. The proposed estimators, in addition to being simple, easy-to-implement and effective, also outperform the competing nonparametric and parametric estimators in finite-sample situations. Two examples illustrate this new semiparametric estimation.

Key words: Bandwidth selection; Discrete kernel method; Discrete weighted distribution; Dispersion; Multiplicative correction factor; Zero-proportion
MSC: Primary 62G07, 62F10; secondary 62G20, 62G99

* *Address for correspondence:* Célestin C. Kokonendji. Université de Pau et des Pays de l'Adour. Laboratoire de Mathématiques Appliquées. UMR 5142 CNRS. Département STID. Avenue de l'Université. 64000 Pau, France. Fax +33559407140.

Email addresses: celestin.kokonendji@univ-pau.fr (C.C. Kokonendji),
tristan.sengakiessé@univ-pau.fr (T. Senga Kiessé),
bala@univmail.cis.mcmaster.ca (N. Balakrishnan).

1 Introduction

Let X_1, \dots, X_n be independent observations from a count distribution with unknown *probability mass function* (pmf) $f(x) = \Pr(X_i = x)$ on the nonnegative integers set $\mathbb{N} = \{0, 1, 2, \dots\}$. A discrete analogue of the (asymmetric) continuous kernel estimator for f can be expressed as

$$\tilde{f}_n(x) = \frac{1}{n} \sum_{i=1}^n K_{x,h}(X_i), \quad x \in \mathbb{N}, \quad (1)$$

where $h = h(n) > 0$ is an arbitrary sequence of smoothing parameters (or bandwidths) that fulfills $\lim_{n \rightarrow \infty} h_n = 0$, while $K_{x,h}(\cdot)$ is a suitably chosen discrete associated-kernel function (to be made precise later, in Section 2). This estimator has been proposed by Kokonendji et al. (2007b). It is connected to the classical kernel estimator of a probability density function in which the kernel function takes the form $K_{x,h}(\cdot) = (1/h)K\{(x - \cdot)/h\}$, where $K(\cdot)$ is typically a bona-fide probability density function with zero mean and unit variance; see, for example, Rosenblatt (1956).

Except for the specification of the count framework, the nonparametric estimator in (1) is known to be completely impartial to special features of the underlying pmf; see, for example, Izenman (1991) for continuous cases. However, this robustness feature of the estimator comes at a price. In general, the choice of kernel function is not very important asymptotically, like ‘frequency estimator’. But, in small samples, the kernel structure may play a more crucial role in approximating the distribution especially for count random variables; see Senga Kiessé (2008) for more details. Hence, the choice of the discrete associated-kernel is also important along with the smoothing parameter for both small and moderate sample sizes. Moreover, the convergence rate of the continuous kernel estimators is in general slower than that of the parametric estimators, and the bias induced by the discrete smoothing procedure can be substantial even for moderate sample sizes. Since count distribution has support on \mathbb{N} bounded on the left, it also becomes necessary to solve the possible problem due to boundary bias or the so-called ‘edge effect’ which depends on the kernel-type or particular count phenomenon. However, in this paper, our aim is not to solve this problem of boundary bias. Note here that, setting apart the frequency (or ‘naive’) estimator and that of Kokonendji et al. (2007b) in (1), there is some literature on nonparametric smoothing of discrete variables or functions dating back to the pioneering work of Aitchison and Aitken (1976); however, the corresponding discrete kernel has a unique form and is ideal for categorical data or finite discrete distributions. One may also refer to Böhning (2000) for a nonparametric approach but in terms of the mixing distribution.

Until recently in the literature, the traditional approach of estimation for count data distribution is totally parametric. In fact, the classical approach starts with a specific structure of count distribution, such as the Poisson model that is equidispersed. It may then become necessary to modify the initial distribution to another related family of count distributions having the same support in order to account for the special features of the counting phenomenon; see, for example, Johnson et al. (2005), Kokonendji et al. (2007a), Shmueli et al. (2005), Kokonendji et al. (2008), and the references therein. The Poisson distribution provides a standard framework for such an analysis of count data, but since it has a single parameter (having no dispersion parameter), it becomes necessary in many instances to construct suitable count distributions by using some indices as measures to detect departures from the Poisson distribution. For example, the well-known and well-studied such departures are the overdispersion and zero-inflation. The opposite phenomena such as underdispersion and zero-deflation are also possible, but are somewhat uncommon. For some properties and applications, one may refer to Puig and Valero (2006) and Nikoloulopoulos and Karlis (2008).

The most general form of modifying Poisson distribution in this way is to multiply it by a weight function; see Kokonendji et al. (2008) and the references therein and also Balakrishnan and Kozubowski (2008). Such *weighted Poisson distributions* (WPDs) provide a unified approach to handle, among others, both dispersion and zero-proportion phenomena. WPDs, used widely as a tool in the selection of appropriate models for observed data drawn without a proper frame, was originally introduced as follows. Let X be a standard Poisson random variable with pmf $p(x; \theta) = \Pr(X = x)$, where $\theta \in \mathbb{R}$ is the canonical parameter. Suppose that when the event $X = x$ occurs, the probability of ascertaining it is $w(x)$. The recorded x is thus a realization of the random variable X^w , which is said to be the *weighted version of X* . Its pmf $p_w(x; \theta) = \Pr(X^w = x)$ is given by

$$p_w(x; \theta) = \frac{w(x) p(x; \theta)}{\sum_{x \in \mathbb{N}} w(x) p(x; \theta)}, \quad x \in \mathbb{N}, \quad (2)$$

where the denominator is the normalizing constant depending on θ . The weight (or recording) function $w(x)$ is a nonnegative function on \mathbb{N} and, from (2), is clearly such that $0 < \sum_{x \in \mathbb{N}} w(x) p(x; \theta) < \infty$. The discrete weight function $w(x) \equiv w(x; \phi)$ can depend on a parameter ϕ representing the recording mechanism, and it may also be connected to the underlying Poisson parameter θ . We easily observe that any count distribution can be formulated as WPD. Of course, this fact does not mean that their generating process is necessarily imperfect recording of the classical Poisson distribution. Clearly, the standard Poisson distribution is a WPD with unit weight function $w(x) = 1 \forall x \in \mathbb{N}$. Also, several other count distributions which are more complex are WPDs. Finally, the Poisson weight function of a count distribution can be considered

as a uniform measure to detect departures from the basic Poisson distribution; see Kokonendji et al. (2008) for more theoretical properties on w in (2).

In this paper, we investigate a semiparametric discrete kernel estimator of a count distribution of the WPD form in (2), where $p(x; \theta)$ is the parametric part and $\frac{w(x)}{\sum_{x \in \mathbb{N}} w(x)p(x; \theta)} = w(x; \theta)$ is the discrete nonparametric part. This semiparametric estimation method was mentioned as a possibility by Kokonendji et al. (2008) at the end of their article. This will be a natural competitor to the nonparametric estimator in (1) as well as the parametric estimator in (2). If the unknown count distribution is concentrated on a finite set of \mathbb{N} , we may consider a weighted binomial distribution; see, for example, Johnson et al. (2005) [pp. 149–150], Chakraborty and Das (2006), and the references therein. This semiparametric estimation for a count distribution is a compromise between pure nonparametric estimation in (1) and the usual parametric modified Poisson estimation. When the discrete Poisson weight function w does not represent theoretically the real recording mechanism or is not well-specified, it is better to allow the count data to yield an estimate of the weight function w by the nonparametric method. The proposed estimator will be the discrete analogue and a particular version of estimator proposed by Hjort and Glad (1995) for continuous data. In the present case, the weight function at each point can be considered as the local multiplicative correction factor aimed to accommodate any pointwise departure from Poisson/binomial distribution. This method of estimation, in addition to being simple and effective for estimating any unknown count distribution, is intended to work well even if the unknown pmf can not be well approximated by either Poisson or binomial distributions. We discuss some basic statistical properties of this estimation procedure and then compare it to other available estimators for the count distribution.

The rest of the paper is organized as follows. In Section 2, we briefly recall the discrete associated-kernel method for the sake of completeness. Section 3 defines the semiparametric estimator of count distribution under the assumption of WPD and then presents some of its properties. We also compare the performance of this estimator with the traditional discrete associated-kernel estimator in (1). In Section 4, we discuss the effects of dispersion and zero-proportion on the proposed estimation procedure through the bandwidth selection. Section 5 extends this idea to the case of the weighted binomial distribution for estimating a count data distribution with finite support. In Section 6, we provide two examples from the literature in order to illustrate the model fitting and the model diagnostics: the first is the data set of goals in championship football games (see Kokonendji et al., 2007b); the second data set is from a sociological experiment about the number of days per week in which alcohol is consumed (see Alanko and Lemmens, 1996). Finally, some concluding remarks are made in Section 7.

2 Discrete associated-kernel method

2.1 Definition

In order to accommodate the discreteness of f , for any target x in \mathbb{N} and any bandwidth $h > 0$, we will select an associated-kernel $K_{x,h}(\cdot)$ of (1) which is itself a discrete pmf with support \aleph_x (not depending on h). To avoid having a zero function, we will require that \aleph_x contains at least x implying $\cup_{x \in \mathbb{N}} \aleph_x \supseteq \mathbb{N}$. Besides, for reasons that will become apparent later (see below), we also impose the following two conditions:

$$\lim_{h \rightarrow 0} \mathbb{E}(\mathcal{K}_{x,h}) = x, \quad (3)$$

$$\lim_{h \rightarrow 0} \text{Var}(\mathcal{K}_{x,h}) = 0, \quad (4)$$

where $\mathcal{K}_{x,h}$ is the discrete random variable whose pmf is $K_{x,h}(\cdot)$.

That definition unifies the notion of associated-kernel that might be either continuous or discrete (Senga Kiessé, 2008). Then, the conditions (3) and (4) are important because they allow us to obtain pointwise convergence of the discrete associated-kernel estimator (1). In fact, the basic condition in (3) clearly points out that \tilde{f}_n defined by (1) is a kind of variable kernel estimate by giving a general discrete kernel form K . It also allows for more flexibility to construct different discrete associated-kernels from any discrete distribution K : for example, $\mathbb{E}(\mathcal{K}_{x,h}) = x + h$ and $\mathbb{E}(\mathcal{K}_{x,h}) = x$. It is implicitly used in asymmetric continuous cases by Chen (1999, 2000) and then by Scaillet (2004). It should be noted that all associated-kernels satisfying (3) share the property that the shape of kernel changes according to the value of the target x . This discrete associated-kernel (or varying kernel shape) changes the amount of smoothing applied to the asymmetric kernel since its variance $\text{Var}(\mathcal{K}_{x,h})$ may or may not depend on the target x as we move away from the boundary. The last condition (4) insures an asymptotic behaviour equivalent to the frequency estimator for the discrete associated-kernel estimator \tilde{f}_n of f defined by (1).

2.2 Properties

Now, we can deduce several properties of the discrete associated-kernel estimator \tilde{f}_n of the unknown count distribution f as follows. Up to the normalizing constant $\tilde{C} = \sum_{x \in \mathbb{N}} \tilde{f}_n(x)$, we assume that $x \mapsto \tilde{f}_n(x)$ is a pmf. Then:

$$\mathbb{E}\{\tilde{f}_n(x)\} = \sum_{y \in \aleph_x \cap \mathbb{N}} K_{x,h}(y) f(y) = \sum_{y \in \aleph_x \cap \mathbb{N}} f(y) \Pr(\mathcal{K}_{x,h} = y) = \mathbb{E}\{f(\mathcal{K}_{x,h})\}. \quad (5)$$

This enables us to approximate the pointwise bias, using the discrete Taylor expansion (see, for example, Schumaker, 1981, p. 343), as

$$\begin{aligned} \text{bias}\{\tilde{f}_n(x)\} &= \mathbb{E}\{f(\mathcal{K}_{x,h})\} - f(x) \\ &= f\{\mathbb{E}(\mathcal{K}_{x,h})\} - f(x) + \frac{1}{2}\text{Var}(\mathcal{K}_{x,h})f^{(2)}(x) + o(h), \end{aligned} \quad (6)$$

where $f^{(2)}$ is the finite difference of second order

$$f^{(2)}(x) = \begin{cases} \{f(x+2) - 2f(x) + f(x-2)\}/4 & \text{if } x \in \mathbb{N} \setminus \{0, 1\} \\ \{f(3) - 3f(1) + 2f(0)\}/4 & \text{if } x = 1 \\ \{f(2) - 2f(1) + f(0)\}/2 & \text{if } x = 0 \end{cases} \quad (7)$$

which is recursively obtained through the finite difference of order $k \in \mathbb{N} \setminus \{0\}$:

$$f^{(k)}(x) = \{f^{(k-1)}(x)\}^{(1)} \text{ and } f^{(1)}(x) = \begin{cases} \{f(x+1) - f(x-1)\}/2 & \text{if } x \in \mathbb{N} \setminus \{0\} \\ f(1) - f(0) & \text{if } x = 0. \end{cases} \quad (8)$$

Here, it is not necessary to suppose certain regularity or differentiability on f because it is a pmf and the finite difference replaces the derivative in the continuous case. The pointwise variance can be expressed as

$$\begin{aligned} \text{Var}\{\tilde{f}_n(x)\} &= \frac{1}{n} \sum_{y \in \mathbb{N}_x} f(y) \{\text{Pr}(\mathcal{K}_{x,h} = y)\}^2 - \frac{1}{n} \left\{ \sum_{y \in \mathbb{N}_x} f(y) \text{Pr}(\mathcal{K}_{x,h} = y) \right\}^2 \\ &= \frac{1}{n} f(x) \{\text{Pr}(\mathcal{K}_{x,h} = x)\}^2 - \frac{1}{n} f^2(x) + R_n(x; h), \end{aligned} \quad (9)$$

with

$$\begin{aligned} R_n(x; h) &= \frac{1}{n} \sum_{y \in \mathbb{N}_x \setminus \{x\}} f(y) \{\text{Pr}(\mathcal{K}_{x,h} = y)\}^2 + \frac{1}{n} f^2(x) \\ &\quad - \frac{1}{n} \left[f(x) + \sum_{y \in \mathbb{N}_x} \{f(y) - f(x)\} \text{Pr}(\mathcal{K}_{x,h} = y) \right]^2. \end{aligned} \quad (10)$$

Under the condition (4) of the discrete associated-kernel, we can show that $R_n(x; h) \rightarrow 0$ when $n \rightarrow \infty$ and $h = h(n) \rightarrow 0$; see, for example, Senga Kiessé (2008).

Since the mean integrated squared error

$$MISE(n, h, K, f) = \sum_{x \in \mathbb{N}} Var\{\tilde{f}_n(x)\} + \sum_{x \in \mathbb{N}} bias^2\{\tilde{f}_n(x)\}$$

of the estimator \tilde{f}_n of f defined in equation (1) is the common measure of accuracy for an estimator, we can establish the following result of convergence (see Senga Kiessé, 2008):

Theorem 1 *Let f be a pmf on \mathbb{N} with $\lim_{x \rightarrow \infty} f(x) = 0$. Then, the discrete associated-kernel estimator in (1) satisfies*

$$MISE \leq \frac{C_1}{n} \sum_{x \in \mathbb{N}} \{\Pr(\mathcal{K}_{x,h} = x)\}^2 + \sum_{x \in \mathbb{N}} \left[f\{\mathbb{E}(\mathcal{K}_{x,h})\} - f(x) + \frac{1}{2}Var(\mathcal{K}_{x,h})f^{(2)}(x) \right]^2,$$

with $C_1 = f_{\max} \leq 1$. Furthermore, for $n \rightarrow \infty$ and $h = h(n) \rightarrow 0$, we have $MISE(n, h, K, f) \rightarrow 0$ if

$$\frac{1}{n} \sum_{x \in \mathbb{N}} \{\Pr(\mathcal{K}_{x,h} = x)\}^2 \rightarrow 0 \quad \text{and}$$

$$\sum_{x \in \mathbb{N}} \left[f\{\mathbb{E}(\mathcal{K}_{x,h})\} - f(x) + \frac{1}{2}Var(\mathcal{K}_{x,h})f^{(2)}(x) \right]^2 \rightarrow 0.$$

Remark 1. For nonparametric estimator in (1), a relative efficiency between two (associated) discrete kernels $\mathcal{K}_{x,h}^{(1)}$ and $\mathcal{K}_{x,h}^{(2)}$ with $\mathbb{E}(\mathcal{K}_{x,h}^{(1)}) = \mathbb{E}(\mathcal{K}_{x,h}^{(2)})$ can be measured via (6) in terms of the difference between their variances $Var(\mathcal{K}_{x,h}^{(1)}) - Var(\mathcal{K}_{x,h}^{(2)})$ for discrete kernels. Thus, to choose an appropriate discrete associated-kernel, we retain in general the one which has a small variance.

Remark 2. From (1), we can write the well-known frequency estimator \tilde{f}_n^F of f as

$$\tilde{f}_n^F(x) = \frac{1}{n} \sum_{i=1}^n D_{x,0}(X_i), \quad x \in \mathbb{N},$$

where $D_{x,0}(\cdot)$ is the discrete associated-kernel with $h = 0$, and which is connected to the Dirac random variable \mathcal{D}_x at x with $\mathbb{E}(\mathcal{D}_x) = x$ and $Var(\mathcal{D}_x) = 0$. Its exact $MISE$ given by

$$MISE(n, 0, D, f) = \frac{1}{n} \sum_{x \in \mathbb{N}} f(x)\{1 - f(x)\} = \frac{1}{n} \left\{ 1 - \sum_{x \in \mathbb{N}} f^2(x) \right\}$$

may be considered as the reference for the convergence of discrete associated-kernel estimators.

2.3 Examples

Presented below are two examples of the usual and competitive families of discrete kernels taken from Senga Kiessé (2008). Figure 1 presents a quick glance at the set and we summarize the main properties in Table 1.

Example 2.1 (Binomial). Consider the binomial distribution $\mathcal{B}(N, p)$, $N \in \mathbb{N}$, $p \in [0, 1]$. The binomial kernel $B_{x,h}$ follows the binomial distribution $\mathcal{B}\{x+1, (x+h)/(x+1)\} =: \mathcal{B}_{x,h}$ with $h \in (0, 1]$ and $\aleph_x = \{0, 1, \dots, x+1\}$. From Remark 1, it is better in the class of the so-called *standard asymmetric discrete kernels* $\mathcal{K}_{x,h}$, such as Poisson and negative binomial, having exactly $\mathbb{E}(\mathcal{K}_{x,h}) = x+h$ and $\cup_{x \in \mathbb{N}} \aleph_x = \mathbb{N}$ but they do not satisfy (4). This is because it is underdispersed: $Var(\mathcal{B}_{x,h}) = (x+h)(1-h)/(x+1) < x+h$. Hence, from (1), the corresponding binomial kernel estimator of f is

$$\tilde{f}_n^B(x) = \frac{1}{n} \sum_{i=1}^n \frac{(x+1)!}{X_i!(x+1-X_i)!} \left(\frac{x+h}{x+1}\right)^{X_i} \left(\frac{1-h}{x+1}\right)^{x+1-X_i}, \quad x \in \mathbb{N},$$

with $X_i \leq x+1$. Its pointwise variance can be deduced from (9) as

$$Var\{\tilde{f}_n^B(x)\} = \frac{(1-h)^2}{n} f(x) \left(\frac{x+h}{x+1}\right)^{2x} - \frac{1}{n} f^2(x) + R_n^B(x; h),$$

with $R_n^B(x; h) \rightarrow 0$ when $n \rightarrow \infty$ and $h = h(n) \rightarrow 0$. For the pointwise bias, a direct calculus gives

$$\text{bias}\{\tilde{f}_n^B(x)\} = f(x) \left\{ (1-h) \left(\frac{x+h}{x+1}\right)^x - 1 \right\} + \sum_{y \in \aleph_x \setminus \{x\}} f(y) B_{x,h}(y)$$

which does not tend to 0 when $n \rightarrow \infty$ and $h = h(n) \rightarrow 0$. Note that, from (6) we can express this pointwise bias of \tilde{f}_n^B as

$$\text{bias}\{\tilde{f}_n^B(x)\} = hf^{(1)}(x) + \frac{1}{2} \left(\frac{x+h-xh}{x+1}\right) f^{(2)}(x) + o(h)$$

because $f\{\mathbb{E}(\mathcal{B}_{x,h})\} = f(x+h) = f(x) + hf^{(1)}(x) + o(h)$ and $Var(\mathcal{B}_{x,h}) = (x+h)/(x+1) - xh/(x+1) + o(h)$. Thus, it follows that \tilde{f}_n^B does not converge in the sense of *MISE*. However, Senga Kiessé (2008) showed that the estimator \tilde{f}_n^B of f can be better (in the sense of *MISE*) than the frequency estimator \tilde{f}_n^F for some finite sample sizes. This estimator and all others of the class are not subject to boundary bias, but they do involve the first finite difference of the unknown pmf. This is because x is not the mean of the corresponding asymmetric kernel, but rather its mode. This is different for the family of discrete triangular kernel estimators below, whose bias involves $f^{(2)}$

only, the finite difference of second order of the unknown pmf f . Finally, the bandwidth selection for the standard asymmetric discrete kernel estimators can be obtained by the well-known cross-validation method, or through the particular situation of excess of zeros (*zero-proportion method*) by solving the equation

$$\sum_{i=1}^n \Pr(\mathcal{K}_{X_i, h_0} = 0) = n_0, \quad (11)$$

where n_0 represents the number of observations equal to zero. In particular, the adapted bandwidth h_0 for the binomial kernel estimator is such that $\sum_{i=1}^n \{(1 - h_0)(X_i + 1)^{-1}\}^{X_i+1} = n_0$, which is determined numerically.

(Table 1 and Figure 1 about here)

Example 2.2 (Triangular). The discrete triangular distributions, introduced by Kokonendji et al. (2007b), are useful to construct a family of *symmetric discrete kernel* estimators for a pmf. For given $(a, x, h) \in \mathbb{N} \times \mathbb{N} \times (0, \infty)$, the discrete triangular associated-kernel $T_{a;x,h}$ is defined through the pmf of its corresponding random variable $\mathcal{T}_{a;x,h}$ on $\mathfrak{N}_{a;x} = \{x, x \pm 1, \dots, x \pm a\}$ as

$$\Pr(\mathcal{T}_{a;x,h} = y) = \frac{(a+1)^h - |y-x|^h}{P(a,h)}, \quad y \in \mathfrak{N}_{a;x},$$

where $P(a, h) = (2a+1)(a+1)^h - 2\sum_{k=0}^a k^h$ is the normalizing constant. The three parameters are such that a denotes the *arm* and is fixed, $x = \mathbb{E}(\mathcal{T}_{a;x,h})$ is the *center* and represents the target, and h is the *order* which corresponds to the bandwidth. The particular case $\mathcal{T}_{0;x,h}$ provides the Dirac random variable at x . From (1), the class of discrete triangular kernel estimators is given, for fixed $a \neq 0$, as

$$\tilde{f}_n^{T_a}(x) = \frac{1}{n} \sum_{i=1}^n \frac{(a+1)^h - |X_i - x|^h}{P(a,h)}, \quad x \in \mathbb{N}.$$

Its pointwise variance can be written from (9) as

$$\text{Var}\{\tilde{f}_n^{T_a}(x)\} = \frac{1}{n} f(x) \left\{ \frac{(a+1)^h}{P(a,h)} \right\}^2 - \frac{1}{n} f^2(x) + R_n^{T_a}(x; h),$$

with $\lim_{h \rightarrow 0} (a+1)^h / P(a, h) = 1$ and $R_n^{T_a}(x; h) \rightarrow 0$ when $n \rightarrow \infty$ and $h = h(n) \rightarrow 0$. While its pointwise bias can be directly obtained by using (6) as

$$\text{bias}\{\tilde{f}_n^{T_a}(x)\} = \frac{1}{2} V(a, h) f^{(2)}(x) + o(h),$$

where $f^{(2)}$ is as given in (7), we have

$$V(a, h) = \text{Var}(\mathcal{T}_{a;x,h}) = \frac{1}{P(a,h)} \left\{ \frac{a(2a+1)(a+1)^{h+1}}{3} - 2 \sum_{k=0}^a k^{h+2} \right\} \quad (12)$$

tends to 0 when $h \rightarrow 0$. Both conditions (3) and (4) hold for $\mathcal{T}_{a;x,h}$ and we can therefore apply Theorem 1 to get the convergence of $\tilde{f}_n^{T_a}$ in the sense of *MISE*. From Remark 1, $\mathcal{T}_{a_1;x,h}$ is more efficient than $\mathcal{T}_{a_2;x,h}$ when $a_1 < a_2$. The bias of $\tilde{f}_n^{T_a}$ do not depend on the first finite difference $f^{(1)}$ as in the case of symmetric continuous kernels. However, for fixed $a \neq 0$, these discrete triangular kernel estimators induce a boundary bias on the left of \mathbb{N} because the set $\cup_{x \in \mathbb{N}} \mathcal{N}_{a;x} = \{-a, \dots, -1\} \cup \mathbb{N}$ contains strictly the support \mathbb{N} of the unknown pmf f . One of the remedies for this is, for significant observations to the boundary $\{0, 1, \dots, m\}$ (m too small, like 0, 1 or 2), to consider the *modified arm* a_0 of a such that, for given $k \in \mathbb{N} \setminus \{0\}$ and $x \in \mathbb{N}$,

$$a_0 = k \iff a_0 = \begin{cases} j & \text{if } x = j, j \in \{0, 1, \dots, k-1\} \\ k & \text{if } x \in \{k, k+1, \dots\}. \end{cases} \quad (13)$$

This procedure preserves the structure of the local symmetry of the discrete kernel around every target, and could resolve the zero-inflation phenomenon as well. The bandwidth selection is made essentially by cross-validation method, but not by using the zero-proportion equation (11) which has no solution under the consideration of discrete triangular kernels (Kokonendji et al., 2007b).

3 Semiparametric estimator under WPD

From (2), any count distribution or random variable X with unknown pmf $f(x) = \Pr(X = x)$ can be written as a WPD:

$$f(x) = w(x; \mu) p(x; \mu) =: f_w(x; \mu), \quad \forall x \in \mathbb{N}, \quad (14)$$

where $p(x; \mu) = \mu^x e^{-\mu} / x! > 0$ is the pmf of the Poisson distribution with mean parameter $\mu > 0$, and $x \mapsto w(x; \mu) = w(x) \{\sum_{x \in \mathbb{N}} w(x) p(x; \mu)\}^{-1}$ is the non-negative (normalized) Poisson weight function. To estimate $f(\cdot) \equiv f_w(\cdot; \mu)$ in (14), a natural way is to consider the estimation problem in a semiparametric setup with $p(x; \mu)$ as the parametric part relative to μ and $x \mapsto w(x; \mu)$ as the nonparametric part on \mathbb{N} . The process is to start out with a parametric pmf estimate $p(x; \hat{\mu}_n)$ where $\hat{\mu}_n$ can be obtained by maximum likelihood method, and then multiply with a nonparametric discrete associated-kernel estimate of the Poisson weight function $w(x; \hat{\mu}_n) = f(x) / p(x; \hat{\mu}_n)$.

More precisely, let X_1, \dots, X_n be independent observations from the unknown pmf f on \mathbb{N} presented in (14). Following Hjort and Glad (1995), our semiparametric estimator of f is expressed as

$$\hat{f}_n(x) = p(x; \hat{\mu}_n) \frac{1}{n} \sum_{i=1}^n \frac{K_{x,h}(X_i)}{p(X_i; \hat{\mu}_n)} \quad (15)$$

$$= \frac{1}{n} \sum_{i=1}^n K_{x,h}(X_i) \frac{p(x; \hat{\mu}_n)}{p(X_i; \hat{\mu}_n)}, \quad x \in \mathbb{N}, \quad (16)$$

where $\hat{\mu}_n = \bar{X}_n = n^{-1}(X_1 + \dots + X_n)$ is the sample mean which is the maximum likelihood estimator of the Poisson mean μ , $h > 0$ is the bandwidth and $K_{x,h}(\cdot)$ is a given discrete associated-kernel as detailed in the previous section. Like the estimator \tilde{f}_n in (1), the estimator \hat{f}_n of the pmf f is defined up to the normalizing constant $\hat{C} = \sum_{x \in \mathbb{N}} \hat{f}_n(x)$ and it is certainly simple enough to implement.

Let us now look at these different expressions of the estimator \hat{f}_n of the pmf f in (14). Firstly, from (15), we obviously deduce the nonparametric discrete associated-kernel estimate

$$\tilde{w}_n(x; \hat{\mu}_n) = \frac{1}{n} \sum_{i=1}^n \frac{K_{x,h}(X_i)}{p(X_i; \hat{\mu}_n)}, \quad x \in \mathbb{N}, \quad (17)$$

of the Poisson weight function $x \mapsto w(x; \hat{\mu}_n)$, which is the local multiplicative correction factor depending on the parameter estimate $\hat{\mu}_n$. Since $p(x; \hat{\mu}_n)$ is already estimated in the expression of $w(x; \hat{\mu}_n) = f(x)/p(x; \hat{\mu}_n)$, an alternative estimator to (17) would be the so-called ‘external’ estimator $\tilde{w}_n^E(x; \hat{\mu}_n) = n^{-1} \sum_{i=1}^n K_{x,h}(X_i)/p(x; \hat{\mu}_n) = \tilde{f}_n(x)/p(x; \hat{\mu}_n)$ which would imply $\hat{f}_n = \tilde{f}_n$; see, for example, Patil et al. (1994) in other context. Hence, we can indicate (17) as the ‘internal’ estimator of $w(x; \hat{\mu}_n)$ and it is considered as the empirical mean of the ratio $K_{x,h}(\cdot)/p(\cdot; \hat{\mu}_n)$. Of course, this internal estimator (17) is more appropriate than the external estimator $\tilde{w}_n^E(x; \hat{\mu}_n)$ because the semiparametric estimate \hat{f}_n is really different to the nonparametric estimator \tilde{f}_n of f defined in (1). Also, we estimate from (17) the pointwise measures of departures from the Poisson distribution. The correction or weight function estimate $\tilde{w}_n(x; \hat{\mu}_n)$ is uniformly equal to one if the Poisson start is well-specified. Secondly, from (16), we note the new (parametric) part

$$\frac{p(x; \hat{\mu}_n)}{p(X_i; \hat{\mu}_n)} = \frac{X_i! (\hat{\mu}_n)^{x-X_i}}{x!}, \quad x \in \mathbb{N} \quad (18)$$

is brought into the ordinary discrete associated-kernel estimator in (1) for improving it. Note also that a constant (18) gives back the purely nonparametric estimator in (1) with an implicit initial parametric start which is given by an improper discrete uniform distribution.

We now examine the bias and variance of the proposed estimator (16) and compare them to those of the traditional estimator (1). We carry this out here under two assumptions, viz., known and unknown Poisson distribution.

3.1 Known Poisson start

Let $p_0(x) = p(x; \mu_0)$ be a fixed pmf of Poisson start in (14), an enlightened supposition of f which is chosen, for example, from a goodness-of-fit test. We therefore write $f = p_0 w$ and then estimate the nonparametric weight function w by $\tilde{w}_n(x) = n^{-1} \sum_{i=1}^n K_{x,h}(X_i) / \tilde{p}_0(X_i)$ with a discrete associated-kernel method, resulting in the estimator

$$\hat{f}_n(x) = p_0(x) \tilde{w}_n(x) = \frac{1}{n} \sum_{i=1}^n K_{x,h}(X_i) \frac{p_0(x)}{p_0(X_i)}, \quad x \in \mathbb{N}. \quad (19)$$

Theorem 2 For $n \rightarrow \infty$ and $h = h(n) \rightarrow 0$, the semiparametric estimator in (19) with fixed Poisson start p_0 possesses the following bias and variance:

$$\text{bias}\{\hat{f}_n(x)\} = p_0(x) \left[w\{\mathbb{E}(\mathcal{K}_{x,h})\} - \frac{f(x)}{p_0(x)} + \frac{1}{2} \text{Var}(\mathcal{K}_{x,h}) w^{(2)}(x) \right] \{1 + o(1)\},$$

$$\text{Var}\{\hat{f}_n(x)\} = \frac{1}{n} f(x) \{\text{Pr}(\mathcal{K}_{x,h} = x)\}^2 - \frac{1}{n} f^2(x) + R_n(x; h),$$

where $w^{(2)}$ is the finite difference of second order as in (7), $o(1)$ does not depend on n and tends to 0 when $h \rightarrow 0$, and $R_n(x; h)$ is as in (10).

Consequently, the new estimator \hat{f}_n in (19) can be better than the ordinary one \tilde{f}_n in (1), in the classical sense of *MISE*. While it is easy to verify that this variance $\text{Var}\{\hat{f}_n(x)\}$ is same as $\text{Var}\{\tilde{f}_n(x)\}$ given in (9), the difference in the two estimators arises from their bias as follows. According to the discrete kernel-type (Examples 2.1 and 2.2), the influential terms in the comparison ensue from

$$f^{(1)} = (p_0 w)^{(1)} = p_0 w^{(1)} + p_0^{(1)} w \leq p_0 w^{(1)}, \quad \text{and/or} \quad (20)$$

$$f^{(2)} = (p_0 w)^{(2)} = p_0 w^{(2)} + 2p_0^{(1)} w^{(1)} + p_0^{(2)} w \leq p_0 w^{(2)}, \quad (21)$$

where \leq denotes \leq or \geq . The inequalities (20) and (21) would depend on the form and, thus, on the variations of the pmf start. For this approximation of second order, the standard asymmetric discrete kernels (binomial) use both inequalities (20) and (21), whereas the symmetric discrete kernels (triangular) need only (21) as symmetric continuous kernels.

As mentioned by Hjort and Glad (1995), this describes a certain neighbourhood of count distribution around the fixed Poisson p_0 where the proposed method of estimation is better or not than the traditional one.

Proof of Theorem 2. From (19), it is enough to calculate $\mathbb{E}\{\tilde{w}_n(x)\}$ and $\text{Var}\{\tilde{w}_n(x)\}$ due to the facts that $\text{bias}\{\hat{f}_n(x)\} = p_0 \mathbb{E}\{\tilde{w}_n(x)\} - f(x)$ and $\text{Var}\{\hat{f}_n(x)\} = p_0^2(x) \text{Var}\{\tilde{w}_n(x)\}$. Now, from (5)–(7), we obtain

$$\begin{aligned}
\mathbb{E}\{\tilde{w}_n(x)\} &= \sum_{y \in \mathbb{N}_x} K_{x,h}(y) p_0^{-1}(y) f(y) = \sum_{y \in \mathbb{N}_x} f(y) p_0^{-1}(y) \Pr(\mathcal{K}_{x,h} = y) \\
&= \mathbb{E}\{f(\mathcal{K}_{x,h}) p_0^{-1}(\mathcal{K}_{x,h})\} = \mathbb{E}\{w(\mathcal{K}_{x,h})\} \\
&= \left[w\{\mathbb{E}(\mathcal{K}_{x,h})\} + \frac{1}{2} \text{Var}(\mathcal{K}_{x,h}) w^{(2)}(x) \right] \{1 + o(1)\}.
\end{aligned}$$

Proceeding in the same manner as in (9), we get

$$\begin{aligned}
\text{Var}\{\tilde{w}_n(x)\} &= \frac{1}{n} \text{Var}\{K_{x,h}(X_1) p_0^{-1}(X_1)\} \\
&= \frac{1}{n} \sum_{y \in \mathbb{N}_x} \{\Pr(\mathcal{K}_{x,h} = y)\}^2 p_0^{-2}(y) f(y) \\
&\quad - \frac{1}{n} \left\{ \sum_{y \in \mathbb{N}_x} \Pr(\mathcal{K}_{x,h} = y) p_0^{-1}(y) f(y) \right\}^2 \\
&= p_0^{-2}(x) \left[\frac{1}{n} f(x) \{\Pr(\mathcal{K}_{x,h} = x)\}^2 - \frac{1}{n} f^2(x) + R_n(x; h) \right].
\end{aligned}$$

Hence, the result. \square

3.2 Unknown Poisson start

Consider our general semiparametric estimator \hat{f}_n presented in (16) of f in (14) such that the estimator $\hat{\mu}_n$ of μ is obtained by the maximum likelihood method; see Hjort and Glad (1995) for quite a general estimator of μ . It is well-known that when the parametric model $p(x; \mu)$ is misspecified, this $\hat{\mu}_n$ converges in probability to the pseudo true value μ_0 which minimizes the Kullback-Leibler distance

$$\sum_{x \in \mathbb{N}} f(x) \log \frac{f(x)}{p(x; \mu)} =: d\{f(\cdot), p(\cdot; \mu)\}$$

of $p(x; \mu)$ from the true pmf $f(x)$; see, for example, White (1982).

Let us write $p_0(x) = p(x; \mu_0)$ for this best parametric approximant, but this p_0 is not explicitly expressible as the one in (19). Denoting $u_0(x) = \partial \log p(x; \mu_0) / \partial \mu = x / \mu_0 - 1$ for the corresponding score function and $v_0(x) = \partial^2 \log p(x; \mu_0) / \partial \mu^2 = -x / \mu_0^2$, a second-order Taylor expansion of (18) provides

$$\begin{aligned}
\frac{p(x; \hat{\mu}_n)}{p(X_i; \hat{\mu}_n)} &= \exp\{\log p(x; \hat{\mu}_n) - \log p(X_i; \hat{\mu}_n)\} \\
&\doteq \frac{p_0(x)}{p_0(X_i)} \left[1 - \{u_0(X_i) - u_0(x)\}(\hat{\mu}_n - \mu_0) + \frac{1}{2}\tau(x, X_i)(\hat{\mu}_n - \mu_0)^2 \right] \\
&= \frac{p_0(x)}{p_0(X_i)} \left[1 - \frac{X_i - x}{\mu_0}(\hat{\mu}_n - \mu_0) + \frac{(x - X_i)(x - X_i - 1)}{2\mu_0^2}(\hat{\mu}_n - \mu_0)^2 \right],
\end{aligned}$$

with $\tau(x, X_i) = v_0(x) - v_0(X_i) + \{u_0(x) - u_0(X_i)\}^2$.

Hence, we can represent the proposed estimator \hat{f}_n in (16) as

$$\begin{aligned}
\hat{f}_n(x) &\doteq \frac{1}{n} \sum_{i=1}^n K_{x,h}(X_i) \frac{p_0(x)}{p_0(X_i)} \\
&\times \left\{ 1 + \frac{x - X_i}{\mu_0}(\hat{\mu}_n - \mu_0) + \frac{(x - X_i)(x - X_i - 1)}{2\mu_0^2}(\hat{\mu}_n - \mu_0)^2 \right\},
\end{aligned}$$

for which the following result provides approximate bias and variance. We omit the proof of this theorem since it is analogous to the result of Hjort and Glad (1995) [Proposition 1] in continuous case and our Theorem 2.

Theorem 3 *Let $p_0(x) = p(x; \mu_0)$ be the best Poisson approximant of the unknown pmf f under the Kullback-Leibler criterion, and let $w = f/p_0$ be the corresponding Poisson weight function. As $n \rightarrow \infty$ and $h = h(n) \rightarrow 0$, the semiparametric estimator in (16) possesses the following bias and variance:*

$$\text{bias}\{\hat{f}_n(x)\} = p_0(x) \left[w\{\mathbb{E}(\mathcal{K}_{x,h})\} - \frac{f(x)}{p_0(x)} + \frac{1}{2}\text{Var}(\mathcal{K}_{x,h})w^{(2)}(x) \right] \{1 + o(1) + n^{-2}\},$$

$$\text{Var}\{\hat{f}_n(x)\} = \frac{1}{n} f(x) \{\text{Pr}(\mathcal{K}_{x,h} = x)\}^2 - \frac{1}{n} f^2(x) + R_n(x; h),$$

where $w^{(2)}$ is the finite difference of second order as in (7), $o(1)$ does not depend on n and tends to 0 when $h \rightarrow 0$, and $R_n(x; h)$ is as in (10).

As in the case of p_0 known (Section 3.1), the proposed estimator \hat{f}_n in (16) of f in (14) can once again be shown to be better or not than the traditional one \tilde{f}_n in (1). Hence, a simulation study is not necessary here to compare the performance of the semiparametric and nonparametric estimates since the performance depends on count data-type and discrete kernel-type; see also Section 6.3 for some comments in this direction and, more generally, Senga Kiessé (2008).

Finally, an important fact for count data or count distribution is that the effect of $\text{Var}(\mathcal{K}_{x,h})$ in the expression of $\text{bias}\{\hat{f}_n(x)\}$ is minimized within a class of discrete kernels, such as the class of standard asymmetric discrete kernels con-

taining over/equi/under-dispersed kernels. This is because the Poisson start p_0 is a multiplicative coefficient of the bias expression in Theorems 2 and 3. This is realized when the Poisson approximation is well-specified for the count data; see the first illustrative example presented in Section 6.1. Because of this, the proposed estimator can be made even better by choosing an appropriate bandwidth.

4 Bandwidth selection

In this section, we examine two methods of selecting the bandwidth for the traditional discrete kernel estimator in (1) used in the semiparametric estimator in (16), viz., cross-validation and zero-proportion methods. It is to be noted that the first one is applicable in the present context, but the second one is not (see below for the proof).

For the proposed estimator in (16), the optimal bandwidth $h > 0$ is obtained, by the popular technique of cross-validation, as $h_{cv} = \arg \min_{h>0} CV(h)$, where

$$CV(h) = \frac{1}{n^2} \sum_{i,j} \frac{1}{p(X_i; \hat{\mu}_n) p(X_j; \hat{\mu}_n)} \sum_{x \in \mathbb{N}} p^2(x; \hat{\mu}_n) K_{x,h}(X_i) K_{x,h}(X_j) - \frac{2}{n(n-1)} \sum_{i,j} K_{X_i,h}(X_j) \frac{p(X_i; \hat{\mu}_{n,-i})}{p(X_j; \hat{\mu}_{n,-i})} \quad (22)$$

with $\hat{\mu}_{n,-i}$ being computed as $\hat{\mu}_n$ by excluding X_i . Note that another alternative for the bandwidth selection is the minimization of the empirical Integrated Squared Error that we define in Section 6.

In the particular situation of zero-inflation (or a certain proportion of excess of zeros) for count data and in the case of the traditional discrete kernel estimator in (1), Eq. (11) of nonparametric zero-proportion can be used to find an adapted bandwidth for given standard asymmetric discrete kernels but not for symmetric discrete triangular ones; see Kokonendji et al. (2007b). For the semiparametric estimator in (16), all effects of excess zeros are taken into account by the Poisson part.

Indeed, to show that the zero-proportion method is not applicable in the semiparametric context, we point out here that the only solution of the corresponding equation of semiparametric zero-proportion is $h_0 = 0$ for the binomial kernel. Similar results hold for Poisson and negative binomial kernels as well, and that there is no solution for all discrete triangular associated-kernels.

From (16) and (18), we have

$$\mathbb{E}\{\widehat{f}_n(x)\} = \int_{z \geq 0} \sum_{y \in \mathbb{N}} \Pr(\mathcal{K}_{x,h} = y) f(y) \frac{y!(z)^{x-y}}{x!} \varphi(z) dz, \quad (23)$$

where φ is the probability density function of $\widehat{\mu}_n$ on $[0, \infty)$. For excess zeros, the semiparametric zero-proportion equation can be obtained by identifying the number of theoretical zeros with $y = z = 0$ in (23) and $f(0) = \varphi(0) = 1$ into the empirical number $n_0 = \#(X_i = 0)$ of observations equal to zero in the sample, yielding

$$\begin{aligned} n_0 &= \sum_{i=1|X_i=0}^n \Pr(\mathcal{K}_{X_i, h_0} = 0) 0^0 + \sum_{i=1|X_i \neq 0}^n \Pr(\mathcal{K}_{X_i, h_0} = 0) \frac{0^{X_i}}{X_i!} \\ &= \sum_{i=1|X_i=0}^n \Pr(\mathcal{K}_{X_i, h_0} = 0) \end{aligned} \quad (24)$$

with $0^0 = 1 = 0!$ and $0^{X_i} = 0$ for $X_i \neq 0$. Consequently, it is easy to verify Eq. (24) for the binomial kernel $\mathcal{K}_{x,h} = \mathcal{B}_{x,h}$ (see Example 2.1) as

$$\sum_{i=1|X_i=0}^n \frac{(X_i + 1)!}{0!(X_i + 1)!} \left(\frac{X_i + h_0}{X_i + 1} \right)^0 \left(\frac{1 - h_0}{X_i + 1} \right)^{X_i + 1} = n_0,$$

which simply reduces to $n_0(1 - h_0) = n_0$ and, therefore, $h_0 = 0$.

5 Restriction to the weighted binomial case

In this section, we assume that the count distribution or random variable X with unknown pmf $f(x) = \Pr(X = x)$ has, without loss of generality, the finite support $\{0, 1, \dots, N\}$ with fixed and known $N \in \mathbb{N} \setminus \{0\}$. Hence, in the same spirit of the WPD discussed earlier in (14), we write f as the *weighted binomial distribution* (see, for example, Johnson et al., 2005 [pp. 149–150]):

$$f(x) = w(x; q) b(x; q) =: f_w(x; q), \quad \forall x \in \{0, 1, \dots, N\}, \quad (25)$$

where $b(x; q) = \frac{N!}{x!(N-x)!} q^x (1-q)^{N-x}$ is the pmf of binomial distribution with probability of success $q \in (0, 1)$, and $x \mapsto w(x; q) = \frac{w(x)}{\sum_{x \in \{0, 1, \dots, N\}} w(x) b(x; q)}$ is the nonnegative (normalized) binomial weight function.

Now, let X_1, \dots, X_n be independent observations from the unknown pmf f on $\{0, 1, \dots, N\}$ presented in (25). The semiparametric estimator of f is then obtained, for $x \in \{0, 1, \dots, N\}$, as

$$\begin{aligned}
\hat{f}_n(x) &= b(x; \hat{q}_n) \frac{1}{n} \sum_{i=1}^n \frac{K_{x,h}(X_i)}{b(X_i; \hat{q}_n)} \\
&= \frac{1}{n} \sum_{i=1}^n K_{x,h}(X_i) \frac{X_i!(N-X_i)!}{x!(N-x)!} \hat{q}_n^{x-X_i} (1-\hat{q}_n)^{X_i-x},
\end{aligned} \tag{26}$$

where $\hat{q}_n = N^{-1}\bar{X}_n$ is the sample proportion of successes, $h > 0$ is the bandwidth, and $K_{x,h}(\cdot)$ is a given discrete kernel as discussed in Section 2.

As in the case of earlier estimators, this estimator \hat{f}_n of the pmf f is defined up to the normalizing constant $\hat{C} = \sum_{x \in \{0,1,\dots,N\}} \hat{f}_n(x)$. All properties and discussions presented earlier in Sections 3 and 4 hold for this \hat{f}_n on the finite support $\{0, 1, \dots, N\}$ as well. However, in this situation, we must also take into account the boundary bias on the right of $\{0, 1, \dots, N\}$. For this situation, we can only use two types of existing discrete kernels: binomial and triangular, which also have compact supports. The modified binomial kernel is the same as in Example 2.1, except at the last target $x = N$ where we can consider the binomial distribution $\mathcal{B}\{N, (N+h)/(N+1)\}$ with $h \in (0, 1]$. However, this (modified) binomial kernel is unstable for smoothing a pmf on a compact support since it considers all information on the left of each target value. For the modified discrete triangular kernels, we also apply the modification of arm (13) on the right of $\{0, 1, \dots, N\}$ (that is, at the neighbourhood of the point $x = N$) as follows: for given $k \in \{1, 2, \dots, N\}$ and $x \in \{0, 1, \dots, N\}$,

$$a_N = k \iff a_N = \begin{cases} j & \text{if } x = N - j, \ j \in \{0, 1, \dots, k-1\} \\ k & \text{if } x \in \{k, k+1, \dots, N-k\}. \end{cases} \tag{27}$$

These modified discrete triangular kernels with (13) and (27) are more appropriate for compact pmf and also possibly for ordered categorical distribution.

6 Illustrative examples and model diagnostics

For evaluating the performance of any (semiparametric or nonparametric) estimator f_n^* of f , we simply use the empirical *Integrated Squared Error* defined by

$$ISE^0 = \sum_{x \in \mathbb{N}} \{f_n^*(x) - f_0(x)\}^2,$$

where $f_0 \equiv \tilde{f}_n^F$ is the frequency estimator. This can be directly observed from graphical presentations; see, for example, Marron and Padgett (1987).

In the special case of count data, we can also measure this performance through

the following χ^2 distance:

$$\chi_0^2 = \sum_{x' \in \{0, 1, \dots, N_0\}} \frac{n \{f_n^*(x') - f_0(x')\}^2}{f_n^*(x')},$$

where $N_0 + 1$ represents the number of valid classes in the sense of χ^2 -test; see, for example, Greenwood and Nikulin (1996). Thus, the statistic χ_0^2 can be suitably approximated by the χ^2 -distribution with $N_0 - r$ degrees of freedom (df), where r is the number of estimated parameters (h , μ or q) in f_n^* . This is a common measure of accuracy for fitting data and it is appropriate to compare with a parametric model using the χ^2 goodness-of-fit test; see below.

6.1 Goals in championship of football

Let us consider the data set of Table 2, used earlier by Kokonendji et al. (2007b). Summary statistics reveal that these count data are slightly overdispersed ($D = 0.113$) and zero-inflated ($Z = 0.059$) with respect to the Poisson distribution with mean $2.13421 = \hat{\mu}_n$. Although the Poisson distribution can fit these data adequately ($\chi^2 = 8.13326$ with 6 df corresponding to the p -value of 0.2285), we show here that the proposed semiparametric method yields better results than the parametric and nonparametric methods upon using certain discrete kernels. Note that, for a given number of goals $g \in \mathbb{N}$, the nearest integer to $n \times f_n^*(g)$ is the semiparametric/nonparametric estimation of the corresponding number of matches.

(Tables 2 and 3 about here)

Table 3 presents the numerical results of a comparison between the traditional nonparametric estimate in (1) and the semiparametric estimate in (16) based on Poisson start with different discrete kernels without boundary bias. While the results for the semiparametric method look somewhat similar in the binomial model, the difference becomes quite noticeable between semiparametric and nonparametric methods in the cases of Poisson, negative binomial and discrete triangular $a_0 \in \{1, 2\}$. For these count data, the semiparametric estimator in (16) provides a good improvement over the nonparametric estimator. Further, we observe no difference between semiparametric estimators with standard asymmetric discrete kernels that are over/equi/under-dispersed. The dispersion effect of associated discrete kernels are being taken into account by the parametric part of this semiparametric method. Thus, for semiparametric smoothings, the choice between (modified) discrete triangular kernels and asymmetric discrete kernels is clearly in favour of symmetric (modified) discrete triangular kernels. All these conclusions are confirmed through both ISE^0 and χ_0^2 measures calculated by using the optimal bandwidth in the sense of cross-validation.

6.2 Daily alcohol consumption

A sociological experiment was carried out concerning the number of days per week in which alcohol was consumed; see Alanko and Lemmens (1996). A randomly selected sample of $n = 399$ Dutch respondents were asked to keep a diary for two consecutive weeks in which they recorded their daily alcohol consumption. To fit these data only for week 1 presented in Table 4, Alanko and Lemmens (1996) used a beta-binomial distribution which is a mixture of binomial distributions, with the probability of success in an individual trial having a beta distribution; see, for example, Johnson et al. (2005) [pp. 253–256]. The results obtained by these authors were not very satisfactory as the p -values were obtained as 0.086 for the χ^2 goodness-of-fit test. Thus, the beta-binomial distribution and *a fortiori* binomial distribution do not seem to be suitable models for these data. Furthermore, we observe that the data have two modal values: the first for 1 day and the second for 7 days. Based on these facts, we feel that it may be better to fit these data by the proposed semiparametric method with a binomial start in (26).

(Table 4 about here)

Table 4 presents the numerical results of fits obtained by means of the semi-parametric estimator with a binomial start in (26). For these count data, only the modified discrete triangular kernels ($a_0 = a_N \in \{1, 2\}$) were used with their optimal bandwidth ($h_{cv} = 0.001$). In the sense of the χ^2 goodness-of-fit test, the fits are quite clearly the best as compared to the results obtained from the beta-binomial distribution. For fixed modified arms $a_0 = a_N \in \{1, 2, \dots, 7\}$ of the discrete triangular kernel, we could choose any bandwidth h ($\neq h_{cv}$) for getting new discrete smoothings (or fits) having a better p -value than the beta-binomial distribution. For fixed bandwidth $h > 0$ in the modified discrete triangular kernel, it is easy to observe (see also Kokonendji et al. (2007b) [Remark 2.3 (ii)]) that if $a_0 = a_N$ increases then both ISE^0 and χ_0^2 increase too; thus, the associated p -values decrease.

6.3 Model diagnostics

The estimated weight function [see Eq. (17)] provides useful information for model diagnostics. The Poisson/binomial weight function should equal one if the Poisson/binomial start pmf is indeed the true pmf. Hjort and Glad (1995) [Section 8.2] proposed to check the adequacy of the model by examining a plot of the weight function w for various potential models with pointwise confidence bands to see if $w(x) = 1$ is reasonable or not. This plot allows to spot easily where misspecification is locally the largest. For the estimated Poisson weight function in (17), the bias and variance can be readily deduced from Theorem

3 as:

$$\mathbb{E}\{\tilde{w}_n(x)\} \doteq w\{\mathbb{E}(\mathcal{K}_{x,h})\} + \frac{1}{2}Var(\mathcal{K}_{x,h})w^{(2)}(x) - \frac{1}{n}w(x)u_0(x)\{1 + u_0(x)/J\},$$

$$Var\{\tilde{w}_n(x)\} \doteq \frac{1}{n} \frac{\omega(x)}{p_0(x)} \{\Pr(\mathcal{K}_{x,h} = x)\}^2 - \frac{1}{n}\omega^2(x)\{1 + u_0^2(x)/J\},$$

with $J = -1/\mathbb{E}\{v_0(X_i)\} = \mu/\mu_0^2$ and $u_0(x) = x/\mu_0 - 1$.

(Figures 2 and 3 about here)

Another possibility, also discussed by Hjort and Glad (1995) [Section 8.2], is to plot the log weight function $\log \tilde{w}_n(x; \hat{\mu}_n) = \log\{\hat{f}_n(x)/p(x; \hat{\mu}_n)\}$ to see how far away it is from zero. Then, a simple graphical goodness-of-fit emerges as follows: plot x against

$$Z(x) = \frac{\log \tilde{w}_n(x; \hat{\mu}_n) + (2n)^{-1}\{p(x; \hat{\mu}_n)\}^{-1} \Pr(\mathcal{K}_{x,h} = x)}{\left[n^{-1}\{p(x; \hat{\mu}_n)\}^{-1} \Pr(\mathcal{K}_{x,h} = x)\right]^{1/2}}.$$

When the Poisson start is indeed the true pmf, $Z(x)$ is approximately distributed as standard normal for each target x , meaning that the $Z(x)$ -values should lie within ± 1.96 about 95% of the time. Notice that the appreciated behaviours of $x \mapsto Z(x)$ are reduced to a suitable subset $\{x \in \mathbb{N}; p(x; \hat{\mu}_n) \geq \alpha > 0\}$ of the support \mathbb{N} because of very small values of individual Poisson probabilities $p(x; \hat{\mu}_n)$. With regard to the binomial start in (26), we omit presenting a similar expression of $Z(x)$; it would be necessary in this case to examine the corresponding $x \mapsto Z(x)$ over its entire support $\{0, 1, \dots, N\}$. Note, in addition, that an investigation of the convexity and concavity of $x \mapsto \log \tilde{w}_n(x; \hat{\mu}_n)$ provides a connection to the theoretical results presented in Kokonendji et al. (2008) with respect to overdispersion and underdispersion.

Thus, applying first to different discrete kernels in Table 3 on the subset $\{0, 1, \dots, 9\}$ of \mathbb{N} with $0.9998 = \sum_{x \in \{0, 1, \dots, 9\}} p(x; \hat{\mu}_n)$, at least 80% of $Z(x)$ -values stay within ± 1.96 (see Figure 2). This suggests that it would be also of interest to consider the pure parametric Poisson distribution for modeling these data than the semiparametric estimation with Poisson start. Hence, we can validate here the choice of a parametric Poisson model for the count data in Table 2. However, for the second example (see Table 4), only 25% of $Z(x)$ -values associated with the semiparametric estimation with binomial start belong to the confidence band ± 1.96 (see Figure 3). Hence, it is necessary in this case to use the semiparametric estimation method for the count data instead of the parametric binomial and beta-binomial models. Note however that if at most 10% of $Z(x)$ -values stay within ± 1.96 (see, for example, Figure 4), an appropriate choice of method would be either the pure nonparametric estimate in (1) or the one given by Böhning (2000).

7 Concluding remarks

In this work, we have introduced a simple and efficient semiparametric estimation for count data through weighted distributions. According to the original idea of Hjort and Glad (1995) for continuous data, we could also start with a different parametric discrete distribution than the Poisson or binomial distributions considered here. The proposed semiparametric estimators in (16) and (26) outperform the nonparametric and parametric estimators in case of finite samples when the pmf to estimate belongs to a neighbourhood of the pmf start. We have more flexibility through choice of discrete kernels as well as bandwidth choices which results in more improved fits for count data by the semiparametric method as compared to the classical parametric method. Moreover, the model diagnostics discussed in Section 6.3 will enable us to make an appropriate choice of the method. Contrary to the pure nonparametric estimate in (1), the semiparametric estimate proposed here has a natural interpretation for count distribution through the popular weighted Poisson and weighted binomial distributions discussed in the literature. Further improvements are still possible with the discrete kernel choice in case of small sample sizes. Two extensions of the results developed here that will be of interest are to consider the multidimensional case and the nonparametric weighted Poisson regression problems. Works in these directions are in progress with respect to the discrete associated-kernel approach which has been recently introduced by Kokonendji et al. (2007b) only for the univariate case.

Acknowledgments. We sincerely thank an associate editor and, in particular, one anonymous referee for their valuable comments.

References

- Aitchison, J., Aitken, C.G.G., 1976. Multivariate binary discrimination by the kernel method. *Biometrika* 63, 413–420.
- Alanko, T., Lemmens, P.H., 1996. Response effects in consumption surveys: an application of the beta-binomial model to self-reported drinking frequencies. *J. Official Statist.* 12, 253–273.
- Balakrishnan, N., Kozubowski, T.J., 2008. A class of weighted Poisson processes. *Statist. Probab. Lett.* 78, 2346–2352.
- Böhning, D., 2000. Computer-assisted analysis of mixtures and applications. Meta-analysis, disease mapping and others. Chapman and Hall, London.
- Chakraborty, S., Das, K.K., 2006. On some properties of a class of weighted quasi-binomial distributions. *J. Statist. Plann. Inference* 136, 156–182.
- Chen, S.X., 1999. Beta kernel estimators for density functions. *Comput. Statist. Data Anal.* 31, 131–145.

- Chen, S.X., 2000. Gamma kernel estimators for density functions. *Ann. Inst. Statist. Math.* 52, 471–480.
- Greenwood, P.E., Nikulin, M.S., 1996. *A Guide to Chi-Squared Testing*. John Wiley and Sons, New York.
- Hjort, N.L., Glad, I.K., 1995. Nonparametric density estimation with a parametric start. *Ann. Statist.* 23, 882–904.
- Izenman, A.J., 1991. Recent developments in nonparametric density estimation. *J. Amer. Statist. Assoc.* 86, 205–224.
- Johnson, N.L., Kemp, A.W., Kotz, S., 2005. *Univariate Discrete Distributions*. 3rd edition. John Wiley and Sons, Hoboken, New Jersey.
- Kokonendji, C.C., Demétrio, C.G.B., Zocchi, S.S., 2007a. On Hinde-Demétrio regression models for overdispersed count data. *Statist. Methodol.* 277–291.
- Kokonendji, C.C., Senga Kiessé, T., Zocchi, S.S., 2007b. Discrete triangular distributions and non-parametric estimation for probability mass function. *J. Nonparam. Statist.* 19, 241–254.
- Kokonendji, C.C., Mizère, D., Balakrishnan, N., 2008. Connections of the Poisson weight function to overdispersion and underdispersion. *J. Statist. Plann. Inference* 138, 1287–1296.
- Marron, J.S., Padgett, W.J., 1987. Asymptotically optimal bandwidth selection for kernel density estimators from randomly right censored samples. *Ann. Statist.* 15, 1520–1535.
- Nikoloulopoulos, A.K., Karlis, D., 2008. On modeling count data: a comparison of some well-known discrete distributions. *J. Statist. Comput. Simul.* 78, 437–457.
- Patil, P.N., Wells, M.T., Marron, J.S., 1994. Some heuristics of kernel-based estimators of ratio functions. *J. Nonparam. Statist.* 4, 203–209.
- Puig, P., Valero, J., 2006. Count data distributions: some characterizations with applications. *J. Amer. Statist. Assoc.* 101, 332–340.
- Rosenblatt, M., 1956. Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* 27, 642–669.
- Scaillet, O., 2004. Density estimation using inverse and reciprocal inverse Gaussian kernels. *J. Nonparam. Statist.* 16, 217–226.
- Schumaker, L.L., 1981. *Spline Functions: Basic Theory*. Wiley, New York.
- Senga Kiessé, T., 2008. *Approche non-paramétrique par noyaux associés discrets des données de dénombrement*. Ph.D Thesis, University of Pau, Pau, France.
- Shmueli, G., Minka, T.P., Kadane, J.P., Borle, S., Boatwright, P., 2005. A useful distribution for fitting discrete data: revival of the Conway–Maxwell–Poisson distribution. *J. Roy. Statist. Soc. Ser. C* 54, 127–142.
- White, H., 1982. Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1–26.

Table 1
Summary of properties of some discrete kernel estimators (Senga Kiessé, 2008)

Type of discrete kernel	$\mathbb{E}(\mathcal{K}_{x,h})$	$\text{Var}(\mathcal{K}_{x,h})$	$\lim_{h \rightarrow 0} \text{Var}(\mathcal{K}_{x,h})$	Convergence of the <i>MISE</i>	Cross-validation	Excess of zero	Symmetry of $\mathcal{K}_{x,h}$	Remarks
Dirac	x	0	0	YES ($n \nearrow \infty$)	--	--	YES	No bandwidth
Poisson	$x + h$	$x + h$	$x \in \mathbb{N}$	NO	YES	YES	NO	Equi-dispersion
Binomial	$x + h$	$(x + h) \left(\frac{1-h}{x+1} \right)$	$0 \leq \frac{x}{x+1} < 1$	NO	YES	YES	NO	Under-dispersion
Negative binomial	$x + h$	$(x + h) \left(1 + \frac{x+h}{x+1} \right)$	$\frac{x(2x+1)}{x+1} \geq 0$	NO	YES	YES	NO	Over-dispersion
Triangular $a \in \mathbb{N} \setminus \{0\}$	x	$V(a, h) : \text{see (12)}$	0	YES ($n \nearrow \infty$ and $h \searrow 0$)	YES	NO	YES	Boundary bias

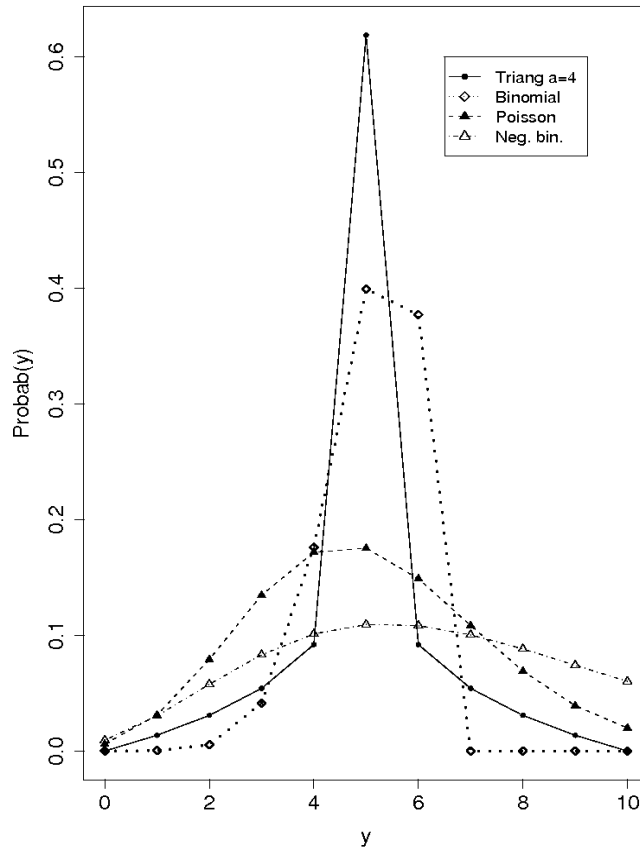


Fig. 1. Behavior of some discrete kernels for count distributions of Senga Kiessé (2008) at the target $y = x = 5$ with the same bandwidth $h = 0.1$.

Table 2

Data of matches having a certain number of goals in the French League 1 of football with $n = 380$ and season 2005-2006 taken from Kokonendji et al. (2007b).

Goals (g)	0	1	2	3	4	5	6	7	8	9	Total
Matches	51	90	109	61	44	12	9	3	0	1	380

Table 3

Some comparative results of semiparametric and nonparametric estimates based on the data in Table 2.

Discrete kernel:	Triang $a_0 = 1$	Triang $a_0 = 2$	Binomial	Poisson	Neg. bin.
Semiparametric					
h_{cv}	0.001	0.001	0.001	0.540	0.820
\hat{C}	1.00004	1.00007	1.01650	1.07392	1.12218
ISE^0	$1.65 \cdot 10^{-8}$	$3.50 \cdot 10^{-8}$	0.00250	0.00301	0.00305
χ_0^2 with 5 df	0.0036	0.0062	10.29	13.99	15.81
p -value	1.0	1.0	0.0674	0.0157	0.0074
Nonparametric					
h_{cv}	0.028	0.019	0.177	0.054	0.039
\tilde{C}	0.99810	0.99780	0.95872	1.05082	1.12028
ISE^0	0.00001	0.00004	0.00395	0.01580	0.02904
χ_0^2 with 6 df	0.1396	0.2816	11.34	63.33	116.80
p -value	0.999946	0.999581	0.0784	$9.45 \cdot 10^{-12}$	0.0

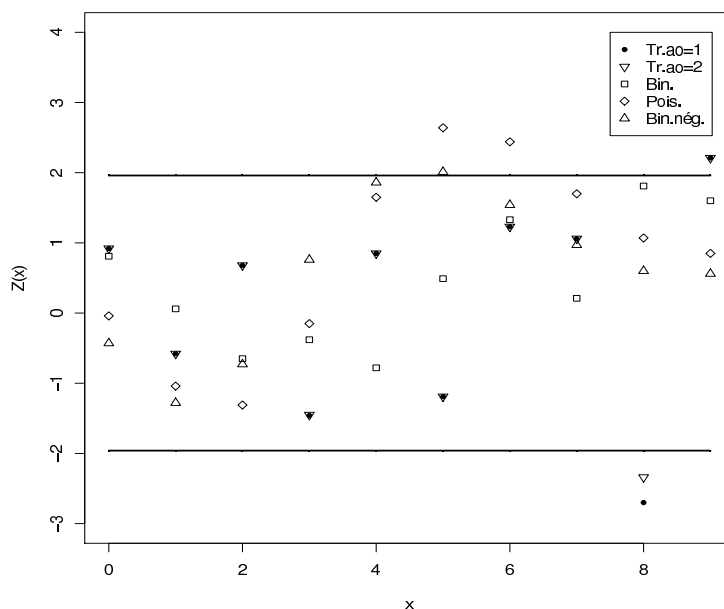


Fig. 2. The $Z(x)$ -values associated with the results in Table 3.

Table 4

Number of alcohol drinking days for week 1 from Alanko and Lemmens (1996) and semiparametric estimates in (26) using modified discrete triangular kernels with $h_{cv} = 0.001$

Number of days per week	Observed frequencies	Expected beta-binomial frequencies	Expected SP Triang-1 frequencies	Expected SP Triang-2 frequencies
0	47	54.6	46.87	46.63
1	54	42.0	54.13	54.02
2	43	38.9	43.03	43.47
3	40	38.5	39.97	39.99
4	40	40.1	39.98	39.89
5	41	44.0	40.98	41.39
6	39	53.1	39.29	39.35
7	95	87.8	94.75	94.26
Total	399	399.00	399.00	399.00
\hat{C}			1.00131	1.00487
ISE^0		$3.02 \cdot 10^{-3}$	$1.15 \cdot 10^{-6}$	$7.29 \cdot 10^{-6}$
χ_0^2		9.6	0.0035	0.0204
df		5	5	5
p -value		0.086	1.0	0.999997

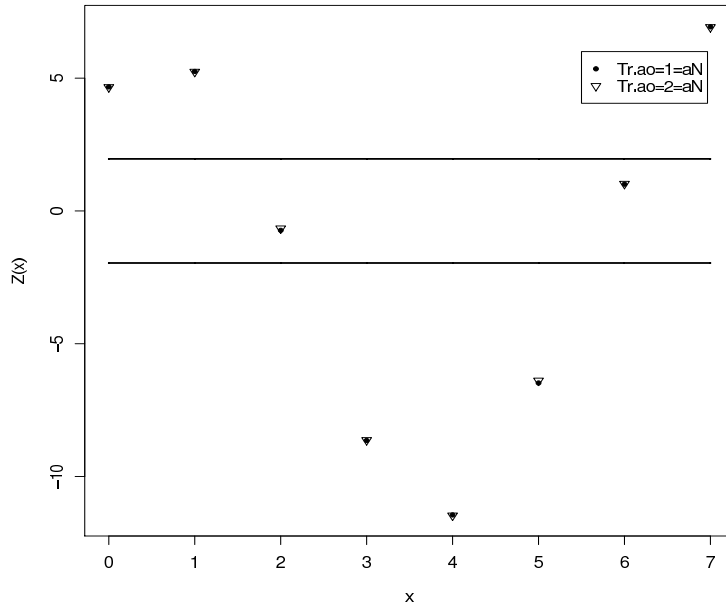


Fig. 3. The $Z(x)$ -values associated with the results in Table 4.

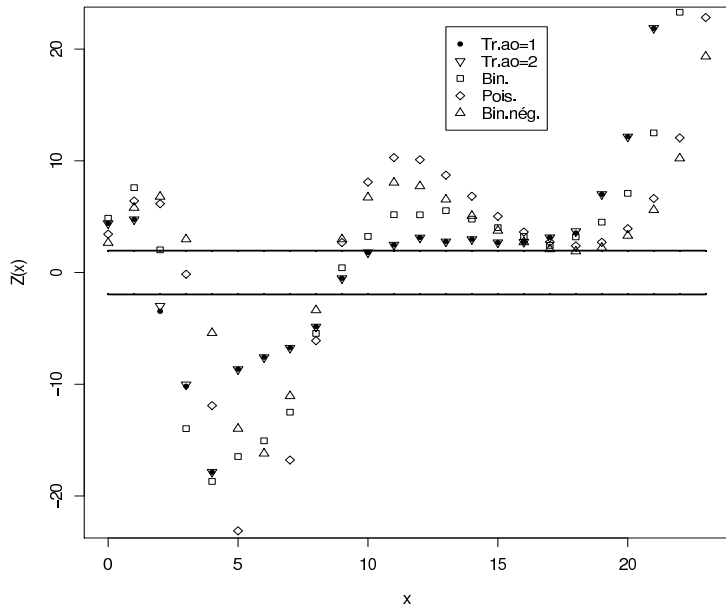


Fig. 4. The $Z(x)$ -values associated with the results of simulated count data from the Poisson mixture distribution $f = 0.4\mathcal{P}(0.5) + 0.6\mathcal{P}(10)$ for $n = 300$