Appropriate kernel regression on a count explanatory variable and applications

Célestin C. Kokonendji, Tristan Senga Kiessé, Clarice G.B. Demétrio

Appropriate kernel regression on a count explanatory variable and applications

CÉLESTIN C. KOKONENDJI[(a)]*, TRISTAN SENGA KIESSÉ[(a)]
[(a)] *University of Pau - LMA-UMR 5142 CNRS, Pau, France*

CLARICE G.B. DEMÉTRIO
University of São Paulo - ESALQ, Piracicaba - SP, Brazil

October 31, 2008

Abstract

Abstract: We propose an appropriate nonparametric regression on a single count regressor using a recent discrete kernel approach. We adapt the Nadaraya-Watson estimator to this discrete kernel for smoothing the regression function on count data. Some properties are studied; in particular, the bandwidth selection is investigated through the cross-validation method. The proposed regression, in addition to being simple, easy to implement and effective, outperforms the competing usual regressions for small and moderate sample sizes. Using simulations and two examples from real life, the importance and the performance of discrete kernels are pointed out and compared with the optimal continuous kernel.

Key words and phrases: Count data, cross-validation, discrete kernel estimator, discrete triangular distribution, nonparametric regression.

AMS Subject Classification: Primary : 62G08; Secondary : 62G99

1 Introduction

Let (x_i, y_i), $i = 1, \ldots, n$, be a sample of observations in $\mathbb{N} \times \mathbb{R}$, where $\mathbb{N}$ denotes the non-negative integers set and $\mathbb{R}$ is the real line. Without any specification of

Address for correspondence: C.C. KOKONENDJI, Université de Pau et des Pays de l'Adour, Laboratoire de Mathématiques et de leurs Applications - UMR 5142 CNRS, Département STID, Avenue de l'Université, 64000 Pau, France. Tel: +33 559 407 145; Fax: +33 559 407 140.
Email: celestin.kokonendji@univ-pau.fr

distributions and in the absence of an evident relation between these observations, the most natural nonparametric regression model of y_i on x_i is

$$y_i = m(x_i) + e_i, \tag{1}$$

where y_i denotes the observation of a real random variable Y_i, x_i is the observation of the count explanatory variable (or regressor) X_i, e_i is the disturbance term from the real random variable ϵ_i satisfying commonly $\mathbb{E}(\epsilon_i) = 0$ and var$(\epsilon_i) = \sigma^2$, and $m : \mathbb{N} \to \mathbb{R}$ is the unknown regression function on count data that we will call *regression count function* for brevity. That function can be represented as $m(x_i) = \mathbb{E}(Y_i|X_i = x_i)$ for a random effect. We are interested in estimating or (discrete) smoothing this regression count function m by taking into account especially its counting structure.

Setting apart the 'frequency (or naif) estimator', 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 kernels are essentially for categorial data or finite discrete distributions. Bierens and Hartog (1988) proposed regression models which take the form of a polynomial of a linear function of the regressors with discrete explanatory variables. A non-appropriate but always used way of smoothing the regression count function m in (1) is simply to consider the count regressor as continuous and then apply one of the numerous techniques of nonparametric estimation of the regression continuous function (see, for example, Chen, 2000b; Collomb, 1981; Gasser and Müller, 1979; Michels, 1992). Regrettably, the particular structure of counting of the regressor is not taken into account. Recently, in the particular situation of the so-called nonparametric binomial regression where the response variable Y_i follows the binomial distribution $\mathcal{B}\{N_i, m(x_i)\}$ at each covariate x_i, Okumura and Naito (2004, 2006a) needed to transform the discrete variable x before using the well-known symmetric kernel estimator of Nadaraya (1964) and Watson (1964) for regression function on continuous data. An extension of this method for multinomial data is also given by Okumura and Naito (2006b).

In this paper we use the discrete analog of the continuous kernel estimator, introduced by Kokonendji et al. (2007b) and Senga Kiessé (2008), for estimating the regression count function m in (1) without transforming the count (or discrete) explanatory variable. That is done by adapting the (continuous) estimator of Nadaraya-Watson to the discrete case, which is one of the oldest and simplest weighted estimators for nonparametric regression function. We attempt to illustrate the necessity of this procedure and its capacity for producing better explanations of real data by means of simulations and also two examples.

For the first example, a sales data set (Table 1), we analyse the number of a new product sold per day during the 25 first days. The 160 observations (x_i, y_i), $i = 1, \ldots, 25$, represent the day x_i of the sales and the corresponding mean numbers $y_i \in \{y_{Ai}, y_{Bi}, \ldots, y_{Hi}\}$ of sales. The number of sale centres for each state

$(A, B, \dots, H)$ is not available except for the state H which is just one. Note that $x_1 = 0$ represents the first day of the product in the market and its sale results only from a previous advertising campaign; $x_2, x_3, \dots$ represent the other days when the sales of the product are results also from informal advertisings of the customers. Because of the novelty of the product, we note that at the beginning there is a fast increase in the mean number of sales up to x_5 in general. After that, the mean number of sales starts to decrease, which can be a result of a limited number of customers or the lack of success of the product. We also observe that after the date x_{14} there are some missing data from some centres. The aim of this type of study is to have a policy for the advertising campaign which in general is very expensive.

(Tables 1, 2 and Figure 1 about here)

The second example (Table 2) concerns the study of average daily fat (kg/day) yields from milk of a single cow for each of the 35 first weeks (McCulloch, 2001, p. 40-45). The quantity of fat in the milk increases during the first fourteen weeks and decreases after. The fitted curves for two generalized linear models (see also McCullagh and Nelder, 1989) are presented in Figure 1. In fact, the first represents a normal model for the log-transformed response variable Y_i with an identity link and the second represents a normal model for the response variable Y_i with a logarithmic link. Both have the same linear predictor $\beta_0 + \beta_1 x_i + \beta_2 \log x_i$, where x_i denotes the week. We can see that neither model fits well to the data although there is an improvement for the log link model. In particular, they do not detect the plateau associated with observations $x = 19, 20, \dots, 27$. We will compare these results with those obtained by our new nonparametric regression.

The rest of the paper is arranged as follows. Section 2 reviews the recent discrete kernel methods to make the paper self-contained as possible. In Section 3, we define the associated discrete kernel estimator for regression count function from the Nadaraya-Watson estimator and give some properties. In particular, we establish a result on the pointwise squared error and the bandwidth selection is made by adapting the least-squares cross-validation method. The new procedure is illustrated through simulations and the two motivating examples in Section 4. Furthermore, some comparisons are pointed out within discrete kernels and also with the optimal continuous kernel (see Epanechnikov, 1969). Section 5 concludes the paper and also suggests some extensions and future research topics.

2 Discrete kernel methods

We briefly recall the more recent discrete kernel methods for estimating (or smoothing) a *probability mass function* (pmf) on $\mathbb{N}$. Such methods are basic for regression function on count data. See, for example, Izenman (1991) for continuous cases.

Let $X_1, \ldots, X_n$ be independent observations from a count distribution with an unknown pmf $f(x) := \Pr(X_i = x)$, the discrete kernel estimator of f is expressed as

$$\widetilde{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_{x,h}(X_i) = \frac{1}{n} \sum_{i=1}^{n} K(X_i; x, h), \quad x \in \mathbb{N}. \tag{2}$$

Given $x \in \mathbb{N}$ and $h > 0$, the *associated discrete kernel* $K_{x,h}(\cdot) = K(\cdot; x, h)$ of $\widetilde{f}_n(x) \equiv \widetilde{f}_n(x; h)$ connecting to a discrete random variable $\mathcal{K}_{x,h}$ on its support $\aleph_x$ (not depending on h) is such that $\aleph_x \cap \mathbb{N} \neq \emptyset$, $\cup_x \aleph_x \supseteq \mathbb{N}$, var$(\mathcal{K}_{x,h}) < \infty$ and we also have:

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

$$\lim_{h \to 0} \text{var}(\mathcal{K}_{x,h}) = 0. \tag{4}$$

That definition unifies the notion of associated kernel that might be either continuous or discrete (Senga Kiessé, 2008). Furthemore, the first two conditions on the supports $\mathbb{N}$ of the unknown pmf and $\aleph_x$ can be changed by $\emptyset \neq \aleph_x \subseteq \mathbb{N}$. Then, the conditions (3) and (4) are important because they allow to obtain the pointwise convergence of the discrete kernel estimator (2). In fact, the basic relation in (3) reflects one of the main differences from the continuous symmetric case where the kernel functions are centred in 0 by using the ratio $(x - X_i)/h$, and such that $K_{x,h}(\cdot) = (1/h)K\{(x - \cdot)/h\}$. This condition (3) clearly points out that the discrete kernel estimator $\widetilde{f}_n$ of f defined by (2) is a kind of variable kernel estimate by giving a general form K. It also allows for more flexibility to construct different (associated) discrete kernels from any discrete ditribution K. So, we shall distinguish two families of discrete kernels satisfying $E(\mathcal{K}_{x,h}) = x + h$ or $E(\mathcal{K}_{x,h}) = x$. It is implicitly used in asymmetric continuous cases by Chen (1999, 2000a) and then by Scaillet (2004). It should be noted that all (associated) kernels satisfying (3) share the property that the shape of the kernels changes according to the value of the target x. This associated discrete kernel (or varying kernel shape) changes the amount of smoothing applied to the asymmetric kernel since its variance var$(\mathcal{K}_{x,h})$ may or may not depend on the target x as we move away from the boundary. According to the behaviour of the variance with respect to the expectation at each target, the (associated) discrete kernels can be underdispersed (*i.e.* var$(\mathcal{K}_{x,h}) < \mathbb{E}(\mathcal{K}_{x,h})$), equidispersed (*i.e.* var$(\mathcal{K}_{x,h}) = \mathbb{E}(\mathcal{K}_{x,h})$) or overdispersed (*i.e.* var$(\mathcal{K}_{x,h}) > \mathbb{E}(\mathcal{K}_{x,h})$). The last condition (4) insures an asymptotic behaviour equivalent to the frequency estimator for the discrete kernel estimator $\widetilde{f}_n$ of f defined by (2).

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

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

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

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

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

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

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

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

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

$$\begin{aligned}\text{var}\left\{\widetilde{f}_n(x)\right\} &= \frac{1}{n}\sum_{y \in \aleph_x} f(y)\left\{\Pr(\mathcal{K}_{x,h} = y)\right\}^2 - \frac{1}{n}\left\{\sum_{z \in \aleph_x} f(z)\Pr(\mathcal{K}_{x,h} = z)\right\}^2 \\ &= \frac{1}{n} f(x)\left\{1 - f(x)\right\}\left\{\Pr(\mathcal{K}_{x,h} = x)\right\}^2 + R_n(x;h),\end{aligned} \tag{9}$$

with

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

Under the condition (4) of the associated discrete kernel, we can verify that $R_n(x;h) = o(1/n)$.

Remark 1. For nonparametric estimator in (2), a relative efficiency between two (associated) discrete kernels $\mathcal{K}^1_{x,h}$ and $\mathcal{K}^2_{x,h}$ with $\mathbb{E}(\mathcal{K}^1_{x,h}) = \mathbb{E}(\mathcal{K}^2_{x,h})$ can be measured via (6) in terms of the difference between their variances $\text{var}(\mathcal{K}^1_{x,h}) - \text{var}(\mathcal{K}^2_{x,h})$ for discrete kernels.

Since the mean integrated squared error

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

of the estimator $\widetilde{f}_n$ of f defined in equation (2) is the common measure of accuracy for an estimator, we can establish the following result of convergence (see Senga Kiessé, 2008):

Theorem 2.1 *Let f be a pmf on $\mathbb{N}$ with $\lim_{x\to\infty} f(x) = 0$. Then, the nonparametric estimator in (2) with an associated discrete kernel satisfies*

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

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

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

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

Remark 2. The $MISE$ of the 'frequency estimator' (with the Dirac kernel type) given by $(1/n)\sum_{x\in\mathbb{N}} f(x)\{1-f(x)\} = (1/n)\left\{1-\sum_{x\in\mathbb{N}} f^2(x)\right\}$ may be considered as the reference for the convergence of discrete kernel estimators.

Presented below are two examples of usual and competitive families of discrete kernels taken from Senga Kiessé (2008). Figure 2 gives a quick glance at the set and we summarize the main properties in Table 3. Note here that, in general, the choice of (discrete) kernel function is not asymptotically very important, such as frequency estimator; but in small and moderate samples, the kernel structure may play a more crucial role in approximating the sample distribution, especially for count random variables.

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 $\aleph_x = \{0,1,\ldots,x+1\}$. 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 $\cup_x\aleph_x = \mathbb{N}$ but they do not satisfy (4). This is because it is underdispersed:

$\text{var}(\mathcal{B}_{x,h}) = (x+h)(1-h)/(x+1) < x+h$. Hence, from (2), the corresponding binomial kernel estimator of f is

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

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

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

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

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

which does not tend to 0 when $n \to \infty$ and $h = h(n) \to 0$. Thus, it follows that $\widetilde{f}_n^B$ does not converge in the sense of $MISE$. However, Senga Kiessé (2008) showed that the estimator $\widetilde{f}_n^B$ of f can be better (in the sense of $MISE$) than the frequency estimator for some finite sample sizes.
This estimator and all others of the class are not subject to boundary bias. Furthemore, the target x is not the mean of the corresponding asymmetric kernel, but rather its mode. For the choice of the optimal bandwidth we use the well-known procedure of cross-validation.

(Figure 2 and Table 3 about here)

Example 2.2 (Discrete 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 associated discrete triangular kernel $T_{a;x,h}$ is defined through the pmf of its corresponding random variable $\mathcal{T}_{a;x,h}$ on $\aleph_{a;x} = \{x, x\pm 1, \ldots, x\pm a\}$ as

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

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 (2), the class of discrete triangular kernel estimators is given, for fixed $a \neq 0$, as

$$\widetilde{f}_n^{T_a}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{(a+1)^h - |X_i - x|^h}{(2a+1)(a+1)^h - 2\sum_{k=0}^{a} k^h}, \quad x \in \mathbb{N}.$$

Its pointwise variance can be written from (9) as

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

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

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

where $f^{(2)}$ is as given in (7) and

$$V(a,h) = \mathrm{var}(\mathcal{T}_{a;x,h}) = \{a(2a+1)(a+1)^{h+1}/3 - 2\sum_{k=0}^{a} k^{h+2}\}/P(a,h) \quad (10)$$

tends to 0 when $h \to 0$. The condition (4) holds for $\mathcal{T}_{a;x,h}$ and we can therefore apply Theorem 2.1 to get the convergence of $\widetilde{f}_n^{T_a}$ 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 $a_1 < a_2$. However, for fixed $a \neq 0$, these discrete triangular kernel estimators induce a boundary bias on the left of $\mathbb{N}$ because the set $\cup_x \aleph_{a;x} = \{-a, \ldots, -1\} \cup \mathbb{N}$ contains strictly the support $\mathbb{N}$ of the unknown pmf f. An original solution is proposed by Kokonendji, Senga Kiessé and Zocchi (2007) to solve this situation while preserving the structure of the local symmetry of the associated discrete kernel around every target. The bandwidth selection is here made essentially by cross-validation method.

3 Nadaraya-Watson discrete estimator

Following the (continuous) weighted estimator of Nadaraya (1964) and Watson (1964), the discrete analog for m in (1) is here defined by

$$\widehat{m}_n(x) = \sum_{i=1}^{n} \omega_x(X_i) Y_i, \quad x \in \mathbb{N},$$

where

$$\omega_x(X_i) = \frac{K_{x,h}(X_i)}{\sum_{i=1}^{n} K_{x,h}(X_i)} = \omega_{x,h}(X_i) \quad (11)$$

represents the weight such that $\sum_{i=1}^{n} \omega_{x,h}(X_i) = 1$ with the convention $0/0 = 0$, and $K_{x,h}(\cdot)$ denotes an associated discrete kernel as given in (2). The bandwidth $h \equiv h(n, K)$ determines the (discrete) smoothness of the estimate. For an appropriate discrete kernel, very small bandwidths almost reproduce the data while extremely large bandwidths yield a constant estimate for the regression count function.

For investigating the bias and variance of the discrete kernel estimator $\widehat{m}_n(x) \equiv \widehat{m}_n(x;h)$ of m, it is common to write $\widehat{m}_n(x)$ as the ratio:

$$\widehat{m}_n(x) = N_n(x;h)/D_n(x;h),$$

with

$$D_n(x;h) = \frac{1}{n}\sum_{i=1}^{n} K_{x,h}(X_i) = \widetilde{f}_n(x) \text{ and } N_n(x;h) = \frac{1}{n}\sum_{i=1}^{n} Y_i K_{x,h}(X_i).$$

The expressions of the expectation and the variance of $D_n(x;h)$ are given in (6) and (9), respectively, where f is assumed to be the unknown pmf of any X_i. Then, from $m(x) = E(Y_i|X_i = x)$, we may write successively:

$$\begin{aligned}\mathbb{E}\{N_n(x;h)\} &= \mathbb{E}\{Y_1 K_{x,h}(X_1)\} = \sum_{z\in\aleph_x} m(z)f(z)\Pr(\mathcal{K}_{x,h} = z)\\ &= \mathbb{E}\{(mf)(\mathcal{K}_{x,h})\}\\ &= (mf)\{\mathbb{E}(\mathcal{K}_{x,h})\} + \frac{\text{var}(\mathcal{K}_{x,h})}{2}(mf)^{(2)}(x) + o(h),\end{aligned}$$

with $(mf)^{(2)} = m^{(2)}f + 2m^{(1)}f^{(1)} + mf^{(2)}$. By using the conditional expectations of Y_i and of Y_i^2 on X_i, we calculate the variance of $N_n(x;h)$ as follows:

$$\begin{aligned}\text{var}\{N_n(x;h)\} &= \frac{1}{n}\sum_{y\in\aleph_x} \mathbb{E}(Y_1^2|X_1 = y)f(y)\{\Pr(\mathcal{K}_{x,h} = y)\}^2\\ &\quad - \frac{1}{n}\left\{\sum_{z\in\aleph_x} \mathbb{E}(Y_1|X_1 = z)f(z)\Pr(\mathcal{K}_{x,h} = z)\right\}^2\\ &= \frac{1}{n}\left\{\mathbb{E}(Y_1^2|X_1 = x) - f(x)\mathbb{E}^2(Y_1|X_1 = x)\right\} f(x)\left\{\Pr(\mathcal{K}_{x,h} = x)\right\}^2\\ &\quad + r_n(x;h), \end{aligned} \tag{12}$$

with

$$\begin{aligned}r_n(x;h) &= \frac{1}{n}\sum_{y\in\aleph_x\setminus\{x\}} \mathbb{E}(Y_1^2|X_1 = y)f(y)\{\Pr(\mathcal{K}_{x,h} = y)\}^2\\ &\quad + \frac{1}{n}\{\mathbb{E}(Y_1|X_1 = x)f(x)\Pr(\mathcal{K}_{x,h} = x)\}^2\\ &\quad - \frac{1}{n}\left\{\sum_{z\in\aleph_x\setminus\{x\}} \mathbb{E}(Y_1|X_1 = z)f(z)\Pr(\mathcal{K}_{x,h} = z)\right\}^2.\end{aligned}$$

3.1 Mean squared error

We here formulate a result on the mean squared error

$$MSE(x) = \text{var}\{\widehat{m}_n(x)\} + \text{bias}^2\{\widehat{m}_n(x)\}, \quad x \in \mathbb{N},$$

of the estimator $\widehat{m}_n$ of the regression count function m connecting to the pmf f of the regressor.

Proposition 3.1 *For $x \in \mathbb{N}$, let $m(x) = \mathbb{E}(Y|X = x)$ and $f(x) = \Pr(X = x)$ defined on $\mathbb{N} \to \mathbb{R}$. As $n \to \infty$ and $h = h(n) \to 0$, for all x such that $f(x) \neq 0$, the nonparametric regression estimator $\widehat{m}_n(x)$ of $m(x)$ with an associated discrete kernel possesses the following bias and variance:*

$$\textit{bias}\{\widehat{m}_n(x)\} = \left\{m^{(2)}(x) + 2m^{(1)}(x)\left(\frac{f^{(1)}}{f}\right)(x)\right\}\frac{\textit{var}(\mathcal{K}_{x,h})}{2} + O(1/n)^2 + o(h)$$

and

$$\textit{var}\{\widehat{m}_n(x)\} = \frac{\mathbb{E}(Y_1^2|X_1 = x) - f(x)\mathbb{E}^2(Y_1|X_1 = x)}{nf(x)}\left\{\Pr(\mathcal{K}_{x,h} = x)\right\}^2 + o\left(\frac{1}{n}\right).$$

Then, we have

$$\begin{aligned} MSE(x) = &\left\{m^{(2)}(x) + 2m^{(1)}(x)\left(\frac{f^{(1)}}{f}\right)(x)\right\}^2\frac{\textit{var}^2(\mathcal{K}_{x,h})}{4} \\ &+ \frac{\mathbb{E}(Y_1^2|X_1 = x) - f(x)\mathbb{E}^2(Y_1|X_1 = x)}{nf(x)}\left\{\Pr(\mathcal{K}_{x,h} = x)\right\}^2 + o\left(h^2 + \frac{1}{n}\right). \end{aligned}$$

Proof. From Bosq and Lecoutre (1987; p. 119-121) for continuous case with here $g = mf$, we can write

$$\begin{aligned} \widehat{m}_n(x) = m(x) &+ \frac{N_n(x;h) - g(x)}{f(x)} - \frac{g(x)\{D_n(x;h) - f(x)\}}{f^2(x)} \\ &- \frac{\{N_n(x;h) - g(x)\}\{D_n(x;h) - f(x)\}}{f^2(x)} \\ &+ \frac{N_n(x;h)}{f^3(x)}\{D_n(x;h) - f(x)\}^2\{1 + o(1)\} \text{ p.s.} \end{aligned} \tag{13}$$

with $D_n(x;h) = n^{-1}\sum_{i=1}^n K_{x,h}(X_i)$ and $N_n(x;h) = n^{-1}\sum_{i=1}^n Y_i K_{x,h}(X_i)$. Under the condition (3) of the associated discrete kernel, we get

$$\mathbb{E}\{D_n(x;h)\} - f(x) = \frac{\text{var}(\mathcal{K}_{x,h})}{2}f^{(2)}(x) + o(h)$$

and

$$\mathbb{E}\{N_n(x;h)\} - (mf)(x) = \frac{\text{var}(\mathcal{K}_{x,h})}{2}(mf)^{(2)}(x) + o(h),$$

with $(mf)^{(2)} = m^{(2)}f + 2m^{(1)}f^{(1)} + mf^{(2)}$. Furthermore, we can show that

$$\begin{aligned}\mathbb{E}[N_n(x;h)\{D_n(x;h) - f(x)\}^2] &= O(1/n)^2 + O(1/n)\\ &\quad + \mathbb{E}\{D_n(x;h) - f(x)\}^2\mathbb{E}\{N_n(x;h)\}.\end{aligned}$$

Thus, we obtain

$$\begin{aligned}\mathbb{E}\{\widehat{m}_n(x)\} - m(x) = &\left\{\frac{(mf)^{(2)}(x)}{f(x)} - \frac{mf^{(2)}(x)}{f(x)}\right\}\frac{\text{var}(\mathcal{K}_{x,h})}{2}\\ &+ \frac{\mathbb{E}\{D_n(x;h) - f(x)\}^2\mathbb{E}\{N_n(x;h)\}}{f^3(x)}\\ &- \frac{f(x)\mathbb{E}[\{N_n(x;h) - g(x)\}\{D_n(x;h) - f(x)\}]}{f^3(x)} + O(1/n)^2 + o(h).\end{aligned}$$

Then, in the previous expression we express

$$\begin{aligned}&\mathbb{E}\{D_n(x;h) - f(x)\}^2\mathbb{E}\{N_n(x;h)\} - f(x)\mathbb{E}[\{N_n(x;h) - g(x)\}\{D_n(x;h) - f(x)\}]\\ &= O(1/n)^2 + o(h),\end{aligned}$$

which finally leads to

$$\text{bias}\{\widehat{m}_n(x)\} = \left\{m^{(2)}(x) + 2m^{(1)}(x)\left(\frac{f^{(1)}}{f}\right)(x)\right\}\frac{\text{var}(\mathcal{K}_{x,h})}{2} + O(1/n)^2 + o(h). \tag{14}$$

For the variance of $\widehat{m}_n$, from (13), we obtain:

$$\begin{aligned}\text{var}\{\widehat{m}_n(x)\} = &\frac{\text{var}\{N_n(x;h)\}}{f^2(x)} + \frac{g^2(x)}{f^4(x)}\text{var}\{D_n(x;h)\}\\ &- 2\frac{g(x)}{f^3(x)}\text{cov}\{N_n(x;h), D_n(x;h)\} + o\left(\frac{1}{n}\right).\end{aligned}$$

By using the variance of N_n in (12), $\text{var}\{D_n(x;h)\} = O(1/n)$ and $\text{cov}\{N_n(x;h), D_n(x;h)\} = O(1/n)$, we get

$$\text{var}\{\widehat{m}_n(x)\} = \frac{\mathbb{E}(Y_1^2|X_1 = x) - f(x)\mathbb{E}^2(Y_1|X_1 = x)}{nf(x)}\{\Pr(\mathcal{K}_{x,h} = x)\}^2 + o\left(\frac{1}{n}\right). \tag{15}$$

The expressions of the expectation (14) and the variance (15) of $\widehat{m}_n(x)$ allow to deduce the mean squared error at $x \in \mathbb{N}$. ■

We can apply the previous result to binomial and associated discrete triangular kernels as in Examples 2.1 and 2.2.

Hence, the global error can be investigated through the $MISE$. For a practical comparison of the amount of nonparametric regression obtained by some discrete or continuous kernels (with a given bandwidth h), we will use the well-known coefficient of determination, R^2, which quantifies the proportion of variation of the response variables y_i explained by the non-intercept regressor x_i:

$$R^2 = \frac{\sum_{i=1}^n (\widehat{y}_i - \overline{y})^2}{\sum_{i=1}^n (y_i - \overline{y})^2}, \tag{16}$$

with $\widehat{y}_i = \widehat{m}_n(x_i; h)$, $\overline{y} = n^{-1}(y_1 + \ldots + y_n)$ and

$$\frac{1}{n}\sum_{i=1}^n (y_i - \overline{y})^2 = \frac{1}{n}\sum_{i=1}^n (\widehat{y}_i - \overline{y})^2 + \frac{1}{n}\sum_{i=1}^n (y_i - \widehat{y}_i)^2.$$

3.2 Bandwidth selection

In this context of discrete kernel regression on count data, the bandwidth selection is also obtained by the well-known least-squares cross-validation method (see, for example, Hardle and Marron, 1985). In fact, for a given associated discrete kernel, the optimal bandwidth is $h_{cv} = \arg\min_{h>0} CV(h)$ with

$$CV(h) = \frac{1}{n}\sum_{i=1}^n \{Y_i - \widehat{m}_{-i}(X_i; h)\}^2 M(X_i), \tag{17}$$

where $\widehat{m}_{-i}(X_i; h) = \sum_{j\neq i}^n Y_j K_{X_i,h}(X_j) / \sum_{j\neq i}^n K_{X_i,h}(X_j)$ is the leave-one-out kernel estimator of $\widehat{m}_n(X_i; h)$ and $0 \leq M(\cdot) \leq 1$ is a weight to avoid difficulties caused by dividing by zero or by the slow convergence rate caused by boundary effects.

Note that, for practical use of the weight $M(\cdot)$ in (17), we can consider $M(X_i) = \omega_{X_i,h}(X_i) \equiv M(X_i; h)$ defined by (11) and depending on h.

4 Illustrations with discussion

In this section we present a simulation study and we apply the discrete kernel estimator $\widehat{m}$ to both data sets of Tables 1 and 2. In order to analyse and then compare the effects of discrete and continuous kernels, we complete Example 2.1 with two other discrete kernels, Poisson and negative binomial, and we add the optimal continuous kernel of Epanechnikov (1969). Indeed, the Poisson and negative binomial kernels are standard asymmetric discrete kernels on the same support $\aleph_x = \mathbb{N}$ with $h > 0$ such that

$$P_{x,h}(z) = \frac{(x+h)^z e^{-(x+h)}}{z!}, \quad z \in \mathbb{N},$$

and

$$NB_{x,h}(z) = \frac{(x+z)!}{x!z!} \left(\frac{x+h}{2x+1+h} \right)^z \left(\frac{x+1}{2x+1+h} \right)^{x+1}, \quad z \in \mathbb{N},$$

respectively. As for the Epanechnikov kernel, it is defined by

$$K^E(z) = \frac{3}{4}(1-z^2), \ \ z \in [-1,1].$$

For the following illustrations, and given (discrete or continuous) kernel, we first determine a bandwidth using generally the cross-validation procedure with the weight $M \equiv 1$. Then, we calculate the corresponding (weigthed) coefficient of determination R^2. Finally, we give some plots using linear interpolation between the regression points. We do not examine here the model diagnostics in terms of the residual study. However, we will use sometimes the term of better fit in the sense of $MISE$ if the bandwidth is selected by the cross-validation method.

4.1 Simulation study

The regression count function considered is

$$m(x) = \frac{2^x}{x!}, \ x \in \mathbb{N}.$$

Table 4 contains the optimal average integrated squared errors and their standard errors for the estimators based on 1000 replications. For each simulation, the optimal discrete smoothing bandwidths are given by the cross-validation method. The optimal integrated squared errors are determined by using the optimal bandwidths. The results in Table 4 show that the associated discrete triangular kernels with small arms perform much better than the binomial and the Epanechnikov ones, even if the sample size is so large. We do not recommend the use of Poisson and negative binomial kernels because they are neither underdispersed nor satisfied (4). Finally, we note that the simulated average integrated squared errors of the best discrete and continuous kernel estimators are closed as the sample size n increased.

(Table 4 about here)

4.2 Average daily fat

Table 5 and Figure 3 present the corresponding results for the nonparametric regressions for the data in Table 2. The associated discrete triangular kernel with $a = 2$, $h_{cv} = 0.1$ and $R^2 = 99.140\%$ represents the most interesting results (in both senses of MISE and R^2) among all these nonparametric regressions. Then, we have the binomial ($R^2 = 97.179\%$) and the Epanechnikov

($R^2 = 96.967\%$). These three kernels point out the plateau associated with observations $x_i = 19, 20, \ldots, 27$. Note that the Poisson and negative binomial kernels do not detect this behaviour, similarly to the generalized linear models discussed in the Introduction; they underestimate or overestimate most of the y-values.

For these data with both single regressor and endogenous observations, we note that the nonparametric regression by the discrete triangular kernel with $a = 1$ gives a better fit in both senses of $MISE$ and R^2, but not in the sense of better representation (or regression curve) because it almost reproduces the data. So, we improve the (discrete) smoothing by taking other values of the arm a which is a free parameter and depends on the user. However, for the associated discrete triangular kernel with $a = 2$, we also obtained a smoother regression curve (or representation) by changing the value of the discrete smoothing parameter ($h = 0.5$ with $R^2 = 97.321\%$). For $a = 4$, the associated discrete triangular kernel provides $R^2 \in \{96.445, 93.101, 90.126, 88.843\}$ according to the values of the bandwidths $h \in \{0.1, 0.3, 0.7, 1\}$, respectively. This shows that the obtained regression curves are smoother while the R^2 does not change too much. (We omit here to present their corresponding tables and figures.) The main question here is to find an optimal arm when we use discrete triangular kernels for getting a smoother regression curve with a better R^2 associated to the optimal bandwidth, which is an open problem.

(Table 5 and Figure 3 about here)

In general, if the optimal bandwidth h_{cv} gives an optimal nonparametric regression in the sense of R^2 (and therefore $MISE$). Use of other values of the discrete smoothing parameter h may also give a good fit in the sense of R^2 and a smoother curve or representation, providing broader choice for h according to the purpose of the user.

4.3 Sales data

Table 6 and Figure 4 show the results for the nonparametric regressions for the data in Table 1, obtained by discrete kernels (triangular with $a \in \{1, 2\}$, binomial, Poisson and negative binomial) and the Epanechnikov kernel.

(Table 6 and Figure 4 about here)

Among discrete kernels, the discrete triangular with $a \in \{1, 2\}$ and binomial kernels give the best results for the nonparametric regressions in both senses of $MISE$ and R^2 which are around 95%, 92% and 80%, respectively. Again the Poisson and negative binomial kernel regressions either underestimate ($x < 10$) or overestimate ($x > 10$) the y-values. Omitting the six strange values (with *), the R^2 values increase for all the kernels but the order of performance remains the same as the previous results.

The R^2 value for the (continuous) Epanechnikov kernel is smaller than the R^2 value for the discrete triangular. So, the associated discrete triangular kernels are more appropriate for this data set.

5 Concluding remarks

In this work, we have introduced an appropriate and efficient nonparametric regression on a single count regressor (1). The discrete kernels used are easy to implement and directly applicable to the count variable without any transformation. According to the discrete kernels, in particular binomial and associated discrete triangular with small arms, simulation results and both data sets demonstrate the usefulness of the proposed method which is the best one or competitive with respect to the optimal continuous kernel regression. Since there is not for instance an optimal associated discrete kernel, one of the reasons of these good fits comes from the small variance and also the finite support of these interesting discrete kernels which are binomial and discrete triangular with small arms. Another reason is suggested by the very good approximation between these discrete kernels and the Gaussian kernel, which has the advantage of being one of the best in the (symmetric) continuous kernel nonparametric regression.

Some evident and practical extensions of the model (1) are for several regressors. The first one is to consider that all the regressors are counts. The multivariate (associated) discrete kernel can be a product of some univariate ones from the same or different discrete distributions. The second and more useful extension is to investigate a nonparametric regression which admits a mixture of discrete and continuous explanatory variables using also the method of appropriate kernels. For the part of the discrete (or discretized) regressors, we must distinguish count regressors from categorial ones. At the same time, we can improve by using semiparametric regression, where the model of the endogenous variable y_i on the $p \times 1$ vector of regressors $x_i = (x_{1i}, \ldots, x_{pi})'$ contains a parametric function $g(x_i; \beta)$ and a nonparametric factor $m(x_i)$ such that $y_i = g(x_i; \beta)m(x_i) + e_i$ for all $i = 1, 2, \ldots, n$.

Acknowledgment

This work was partially supported by grants from FAPESP and CNPq.

References

[1] J. Aitchison and C.G.G. Aitken, Multivariate binary discrimination by the kernel method, Biometrika 63 (1976), 413-420.

[2] H.J. Bierens and J. Hartog, Non-linear regression with discrete explanatory variable with an application to the earnings function, J. Econometr. 38 (1988), 269-299.

[3] D. Bosq and J.P. Lecoutre, Théorie de l'Estimation Fonctionnelle, Economica, Paris, 1987.

[4] G. Collomb, Estimation non-paramétrique de la régression: revue bibliographique, Intern. Statist. Rev. 49 (1981), 75-93.

[5] S.X. Chen, Beta kernel estimators for density functions, Comput. Statist. Data Anal. 31 (1999), 131-145.

[6] S.X. Chen, Gamma kernel estimators for density functions, Ann. Instit. Statist. Math. 52 (2000a), 471-480.

[7] S.X. Chen, Beta kernel smoothers for regression curves, Statist. Sinica 10 (2000b), 73-91.

[8] V. Epanechnikov, Nonparametric estimates of a multivariate probability density, Theory Probab. Appl. 14 (1969), 153-158.

[9] Th. Gasser and H.-G. Müller, Kernel estimation of regression functions, In *Smoothing Techniques for Curve Estimation* (Edited by Th. Gasser and M. Rosenblatt), 23-68, Springer, Heideberg, 1979.

[10] W. Härdle and J. S. Marron, Optimal bandwidth selection in nonparametric regression function estimation, Ann. Statist. 13 (1985), 1465-1481.

[11] A.J. Izenman, Recent developments in nonparametric density estimation, J. Amer. Statist. Assoc. 86 (1991), 205-224.

[12] C.C. Kokonendji, T. Senga Kiessé and S.S. Zocchi, Discrete triangular distributions and non-parametric estimation for probability mass function, J. Nonparam. Statist. 19 (2007), 241-254.

[13] O.L. Mangasarian and L.L. Schumaker, Best summation formulae and discrete splines, SIAM J. Numer. Anal. 10 (1973), 448-459.

[14] P. McCullagh and J.A. Nelder, Generalized Linear Models, 2nd edn., Chapman and Hall, London, 1989.

[15] C.E. McCulloch, An Introduction to Generalized Linear Mixed Models, 46a Reunião Anual da RBRAS - 9o SEAGRO, University of São Paulo - ESALQ, Piracicaba, 2001.

[16] P. Michels, Asymmetric kernel functions in non-parametric regression analysis and prediction, The Statistician 41 (1992), 439-454.

[17] E.A. Nadaraya, On estimating regression, Theory Probab. Appl., 9 (1964), 141-142.

[18] H. Okumara and K. Naito, Weighted kernel estimators in nonparametric binomial regression, J. Nonparam. Statist. 16 (2004), 39-62.

[19] H. Okumara and K. Naito, Bandwidth selection for kernel binomial regression, J. Nonparam. Statist. 18 (2006a), 343-356.

[20] H. Okumara and K. Naito, Non-parametric kernel regression for multinomial data, J. Multivariate Anal. 97 (2006b), 2009-2022.

[21] O. Scaillet, Density estimation using inverse and reciprocal inverse Gaussian kernels, J. Nonparam. Statist. 16 (2004), 217-226.

[22] T. Senga Kiessé, Approche non-paramétrique par noyaux associés discrets des données de dénombrement, Ph.D Thesis of Statistics at University of Pau, 2008.

[23] G. S. Watson, Smooth regression analysis, Sankhyā Ser. A 26 (1964), 359-372.

Table 1: Sales data (with − denotes a missing observation and * can be considered as a strange value)

x_i	y_{Ai}	y_{Bi}	y_{Ci}	y_{Di}	y_{Ei}	y_{Fi}	y_{Gi}	y_{Hi}
0	7.2	9.7	5.0	9.0	12.0	7.0	9.2	7
1	16.4	15.9	*5.0	16.1	16.2	15.4	16.2	16
2	21.4	18.9	22.2	19.7	17.2	20.7	19.9	23
3	22.0	19.7	24.2	20.5	18.1	22.4	21.1	20
4	20.4	19.2	23.0	19.8	17.6	21.7	20.5	23
5	18.2	18.1	20.4	18.4	16.8	19.8	*8.8	24
6	16.1	16.6	17.7	16.6	15.7	17.6	*6.5	12
7	14.2	15.0	15.2	14.8	14.5	15.3	*4.0	13
8	12.6	13.5	13.0	13.0	13.3	13.3	*1.8	9
9	11.1	12.0	11.1	11.5	12.1	11.5	9.9	9
10	9.9	10.6	9.5	10.0	10.9	9.9	8.5	8
11	8.7	9.4	8.2	8.8	9.8	8.6	7.5	10
12	7.8	8.3	7.1	7.7	8.7	7.4	6.8	8
13	6.9	7.3	6.2	6.7	7.8	6.5	6.3	7
14	−	6.4	5.4	5.9	6.9	5.6	5.9	2
15	5.4	5.6	4.7	−	6.1	4.9	5.6	*12
16	4.8	−	4.1	4.6	5.4	−	−	3
17	4.3	−	3.6	−	−	3.8	−	5
18	−	3.8	−	−	−	3.3	−	4
19	−	−	−	3.2	3.7	−	4.3	2
20	−	2.9	−	2.8	3.2	−	−	2
21	−	−	−	−	2.9	2.3	3.5	5
22	−	−	−	−	2.5	−	3.2	5
23	−	−	2.1	1.8	2.2	−	−	2
24	−	−	1.9	−	−	−	2.5	−

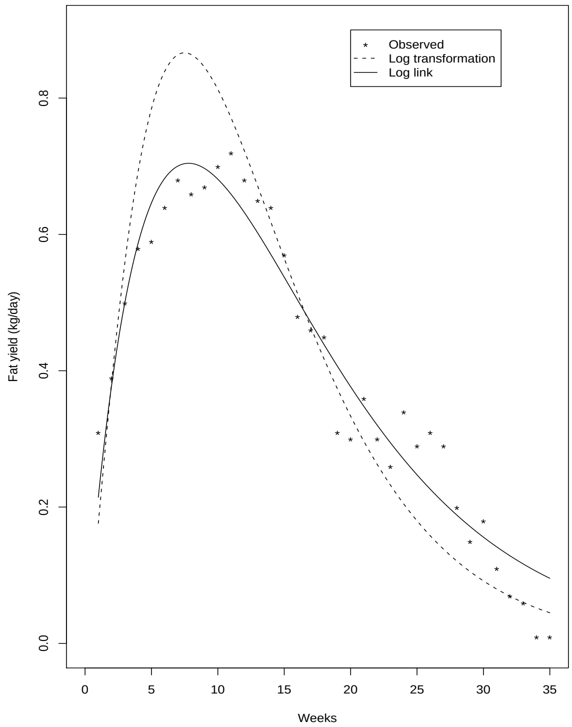

Figure 1: Two generalized linear models from data of Table 2, with the best $R^2 = 0.76790$ (McCulloch, 2001)

Table 2: Average daily fat (kg/day) yields from milk of a single cow for each of the 35 first weeks (McCulloch, 2001)

x_i	1	2	3	4	5	6	7	8	9	10	11	12
y_i	0.31	0.39	0.50	0.58	0.59	0.64	0.68	0.66	0.67	0.70	0.72	0.68
x_i	13	14	15	16	17	18	19	20	21	22	23	24
y_i	0.65	0.64	0.57	0.48	0.46	0.45	0.31	0.33	0.36	0.30	0.26	0.34
x_i	25	26	27	28	29	30	31	32	33	34	35	
y_i	0.29	0.31	0.29	0.20	0.15	0.18	0.11	0.07	0.06	0.01	0.01	

Table 3: Summary of properties of some discrete kernel estimators (Senga Kiessé, 2008)

Type of discrete kernel	$\mathbb{E}(\mathcal{K}_{x,h})$	$\text{var}(\mathcal{K}_{x,h})$	$\lim_{h\to 0}\text{var}(\mathcal{K}_{x,h})$	Convergence of the $MISE$	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	$V(a,h)$: see (10)	0	YES ($n \nearrow \infty$ and $h \searrow 0$)	YES	NO	YES	Boundary bias

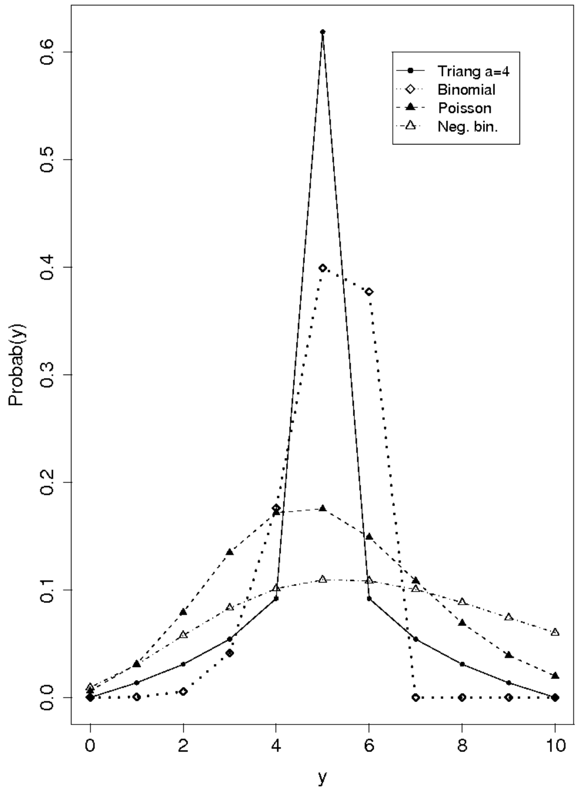

Figure 2: Behaviour of some discrete kernels for count distributions of Senga Kiessé (2008) at the target $y = x = 5$ with the same bandwidth $h = 0.1$

Table 4: Simulated optimal average integrated squared errors and their standard errors (in parentheses) for discrete and Epanechnikov kernel estimators. The results multiplied by 10^3 are given

n	Triang.1	Triang.2	Binomial	*Epanech.*	Poisson	Neg. bin.
20	11.77 (8.2)	24.51 (19.9)	43.87 (49.6)	55.06 (107.2)	65.98 (51.2)	82.73 (122.5)
50	7.01 (8.4)	17.00 (14.1)	27.03 (26.5)	32.02 (31.6)	69.10 (36.1)	87.01 (43.6)
80	4.47 (4.5)	14.13 (8.7)	18.18 (13.6)	23.30 (19.8)	69.42 (27.6)	88.45 (42.9)
100	3.67 (2.9)	12.68 (7.9)	17.03 (11.4)	21.93 (16.5)	70.51 (25.7)	92.12 (37.3)
200	2.43 (1.4)	10.20 (4.0)	12.60 (5.6)	14.81 (15.0)	76.48 (21.2)	102.17 (23.4)
500	1.86 (0.7)	8.46 (2.0)	9.92 (2.5)	9.96 (10.0)	83.45 (17.3)	116.13 (22.4)
1000	1.60 (0.4)	8.09 (1.4)	9.19 (1.5)	8.27 (7.9)	89.69 (13.0)	126.78 (20.1)

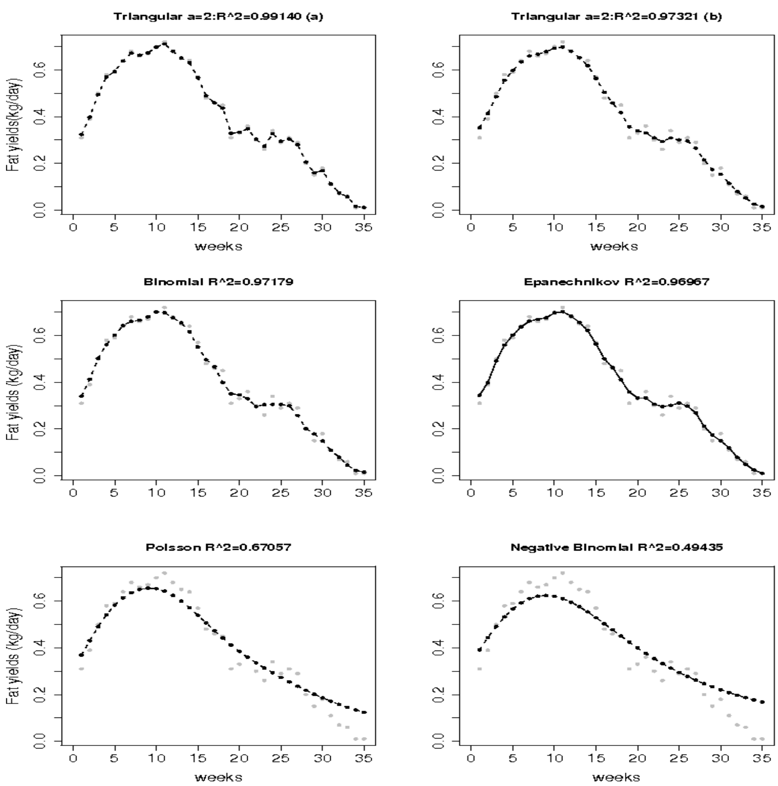

Figure 3: Nonparametric regressions of average daily fat (Table 2) by some discrete kernels and the continuous Epanechnikov kernel

Table 5: R-squared (in %) of nonparametric regressions of average daily fat (Table 2) by some discrete kernels and the continuous Epanechnikov kernel

Kernel	Triang.2 (a)	Triang.2 (b)	Binomial	*Epanech.*	Poisson	Neg. bin.
h_{cv}	0.1	0.5*	0.101	4.0	0.151	0.224
R^2	99.140	97.321	97.179	96.967	67.057	49.435

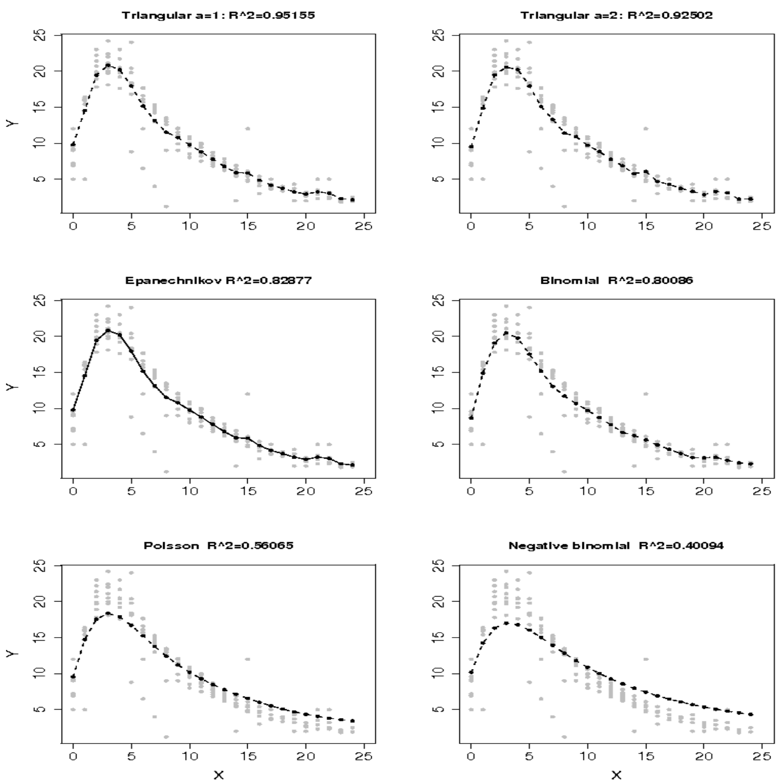

Figure 4: Nonparametric regressions of sales data (Table 1) by some discrete kernels and the continuous Epanechnikov kernel

Table 6: R-squared (in %) of nonparametric regressions of sales data (Table 1) by some discrete kernels and the continuous Epanechnikov kernel

Data	Kernel	Triang.1	Triang.2	*Epanech.*	Binomial	Poisson	Neg. bin.
Complete	h_{cv}	0.558	0.132	2.427	0.064	0.206	0.327
	R^2	95.155	92.502	82.877	80.086	56.065	40.094
Incomplete	h_{cv}	0.170	0.052	2.121	0.078	0.209	0.327
(without *)	R^2	94.075	92.922	93.915	89.151	63.159	45.230