HAL
open science

Discrete triangular distributions and non-parametric estimation for probability mass function

Célestin Kokonendji, Tristan Senga Kiessé, Silvio S. Zocchi

To cite this version:

Célestin Kokonendji, Tristan Senga Kiessé, Silvio S. Zocchi. Discrete triangular distributions and non-parametric estimation for probability mass function. Journal of Nonparametric Statistics, 2007, 19 (6-8), pp.241-254. 10.1080/10485250701733747 . hal-00947709

HAL Id: hal-00947709

https://hal.science/hal-00947709

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.

Nonparametric Statistics
Vol. 00, No. 00, Month-Month 200x, 1–13

Discrete triangular distributions and nonparametric estimation for probability mass function

C.C. KOKONENDJI*†, T. SENGA KIESSE† and S.S. ZOCCHI‡

†University of Pau - LMA UMR 5142 CNRS, Pau, France

‡University of São Paulo - ESALQ, Piracicaba - SP, Brazil

(v3.1 released January 2006)

Discrete triangular distributions are introduced in order to serve as kernels in the nonparametric estimation for probability mass function. They are locally symmetric around every point of estimation. Their variances depend on the smoothing bandwidth, and they establish a bridge between a Dirac distribution and a discrete uniform one. The boundary bias related to the discrete triangular kernel estimator is solved through a modification of the kernel near the boundary. The mean integrated squared errors and then the optimal bandwidth are investigated. We also study the adequate bandwidth for excess zeros. The performance of the discrete triangular kernel estimator is illustrated by using simulated count data. An application to count data from football is described and compared to a binomial kernel estimator.

Keywords: Boundary bias; Count data; Discrete kernel estimator; Excess zeros; Variable kernel estimate.
2000 MSC: Primary: 62E15, 62G07; Secondary: 62G99.

1 Introduction

Let $X_1, \cdots, X_n$ be a random sample from a discrete distribution with an unknown probability mass function f which is defined on the discrete set $\aleph$. From [1], the discrete kernel estimator of f has been defined by [2] as follows:

$$\widehat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{x,h}(X_i) =: \widehat{f}_{n,h,K}(x), \quad x \in \aleph, \tag{1}$$

where $h > 0$ is the bandwidth (or smoothing parameter) and K is a discrete kernel. Given $x \in \aleph$ and $h > 0$, the associated discrete kernel $K_{x,h}(\cdot)$ connecting to a discrete random variable $\mathcal{K}_{x,h}$ on $\aleph_{x,h}$ is such that $\aleph_{x,h} \cap \aleph \neq \varnothing$, $\cup_x \aleph_{x,h} \supseteq \aleph$, $\mathbb{E}(\mathcal{K}_{x,h}) \sim x$ when $h \to 0$, and

$$\begin{aligned}\mathbb{E}[\widehat{f}_{n,h,K}(x)] &= \sum_{y\in\mathbb{N}} K_{x,h}(y) f(y) \\ &= \sum_{y\in\mathbb{N}} \Pr(\mathcal{K}_{x,h} = y) f(y) \\ &= \mathbb{E}[f(\mathcal{K}_{x,h})]. \end{aligned} \tag{2}$$

*Corresponding address: 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. Tel.: +33 559 407 145; fax: +33 559 407 140.
Email addresses: celestin.kokonendji@univ-pau.fr (Célestin C. Kokonendji), tristan.sengakiesse@univ-pau.fr (Tristan Senga Kiessé), sszocchi@esalq.usp.br (Silvio S. Zocchi).

Nonparametric Statistics
ISSN 1048-5252 print/ ISSN 1029-0311 online ©2006 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/10485250xxxxxxxxxxxxx

Thus, the estimator $\widehat{f} = \widehat{f}_{n,h,K}$ of f can be explicitly written and the basic (asymptotical) properties investigated under the following general hypothesis:

$$\mathbb{E}(\mathcal{K}_{x,h}) = x + h + o(h) \quad \text{and} \quad Var(\mathcal{K}_{x,h}) = V_K(x,h) + o(h), \tag{3}$$

with $V_K(x,h) \geq 0$. Naturally, the discrete kernel estimators are more appropriated for estimating any (weighted or regression) discrete function than the use of continuous kernels. Standard discrete distributions such as Poisson and binomial (see also in [3]) can be used as discrete kernels under the conditions (3) according to the intrinsic parameters of the considered discrete distribution. However, it would be necessary to be skillful in the choice of parameters satisfying (3) for obtaining pointwise and global properties of the estimator (1). Since classical discrete distributions are not thought-out to be kernels in (1), the role of standard discrete kernels for small samples is essential than for continuous kernels. 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 sample distribution especially for count random variables. About secondary importance of continuous kernels, the readers can refer to [4] and [5] for generality of (assumed) continuous data, [6] for multivariate cases, [7] for functional data, [8] for ordered categorial data, and [9] for discrete data. Let us note here that there exists another way for smoothing discrete probability distribution; see, for example, Ouyang *et al.* [10] and references therein for categorical data or finite discrete distributions. In its conception, the smoothed estimator (1) of this paper is the discrete analogue of the continuous case which is well known (*e.g.* [11–13]).

The kernel function K is usually symmetric and is regarded as less important than the smoothing bandwidth. While using a symmetric (discrete) kernel is appropriate for estimating probability densities with unbounded supports, it is not adequate for at least one-side bounded supports as it causes boundary bias. Precise here that the cause of boundary bias (or edge effect) is due to the fact that the fixed symmetric kernel which assignes weight outside the density support when smoothing is made near the boundary. An efficacious remedy is to use kernels that never allocate weight outside the support. See, for example, Chen [12, 13] and references therein for continuous cases.

In this paper we introduce a new family of symmetric discrete distributions called *discrete triangular distributions* in order to serve as kernels in (1). From (3), the associated random variables $\mathcal{K}_{x,h}$ are symmetric around the mean value which is the point x of estimation, and its variances do not depend on $x \in \aleph$. Without loss of generality, we consider that the support $\aleph$ of f is the set $\mathbb{N}$ of nonnegative integers. Bias, variance and mean (integrated) squared errors are investigated to deduce the optimal bandwidth of a discrete triangular kernel estimator. In this context, the boundary bias is remedied by modifying the kernel near $x = 0$ such that the associated discrete triangular kernels do not assign weight outside $\mathbb{N}$. We also provide a solution to the phenomena of excess zeros.

The paper is structured as follows. In Section 2 discrete triangular distributions are introduced. Their connections to discrete kernel estimators are studied in Section 3. Mean integrated squared error, optimal bandwidth, adapted bandwidth for excess zeros and boundary bias are therefore derived. Section 4 first illustrates various aspects of the performance of the estimators using simulated count data. Then, there is an illustration for estimating count data from football choosen among many count data which exist in different areas as economy, finance, ecology, medicine, insurance. This is followed by a comparison with a binomial kernel estimator. Section 5 contains concluding remarks.

2 Discrete triangular distributions

Let $a, c \in \mathbb{N}$. A discrete random variable $\mathcal{T}_{a,c}$ is said *triangular* with arm a and center c, if its support is $\aleph_{a,c} = \{c, c \pm 1, \cdots, c \pm a\}$ and the probability mass function is given by

$$\Pr(\mathcal{T}_{a,c} = y) = \frac{a + 1 - |y - c|}{(a+1)^2}, \qquad y \in \aleph_{a,c}.$$

A graphical representation points out the triangular form on the support $\aleph_{a,c}$ and it is symmetric around its mean value $c = \mathbb{E}(\mathcal{T}_{a,c})$. An extension of discrete triangular random variable on the same support $\aleph_{a,c}$ is the following.

Definition 2.1 Let $h > 0$ and $(a, c) \in \mathbb{N}^2$. A discrete random variable $\mathcal{T}_{a,h,c}$ is said triangular of order h, centered in c with arm a, if its support is $\aleph_{a,c} = \{c, c \pm 1, \cdots, c \pm a\}$ and its probability mass function is

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

with the normalizing constant

$$P(a,h) = (2a+1)(a+1)^h - 2\sum_{k=0}^{a} k^h.$$

Note that $h = 1$ corresponds to the discrete triangular random variable $\mathcal{T}_{a,c}$. The cases $h \leq 0$ are not defined to the center $c \in \mathbb{N}$ and, in particular, it gives the null random variable for $h = 0$. For positive integers $h \in \mathbb{N}^*$, the normalizing constant $P(a,h)$ can be written as

$$P(a,h) = (2a+1)(a+1)^h - 2\sum_{k=0}^{a} \frac{(-1)^{h-k+1} h! B_{h-k+1}}{k!(h-k+1)!} a^k, \tag{4}$$

where B_j is the Bernoulli number (see, for example, [14]). We are able to provide some expressions of $P(a,h)$ for $h = 1, 2, \cdots, 8$, by using formula for sum of powers of integers. Figure 1 presents some graphics of discrete triangular distributions which we can denote by $\mathcal{T}(a,h,c)$.

(Figure 1 about here)

The following proposition shows some elementary but basic properties of discrete triangular random variables.

PROPOSITION 2.2 *Let $\mathcal{T}_{a,h,c}$ be a discrete triangular random variable of order $h > 0$, centered in $c \in \mathbb{N}$ with arm $a \in \mathbb{N}$. Then $\mathcal{T}_{a,h,c}$ is symmetric around its mean value $c = \mathbb{E}(\mathcal{T}_{a,h,c})$, and its variance $Var(\mathcal{T}_{a,h,c}) = V(a,h)$ does not depend on c with*

$$V(a,h) = \frac{1}{P(a,h)} \left(\frac{a(2a+1)(a+1)^{h+1}}{3} - 2\sum_{k=0}^{a} k^{h+2} \right).$$

Proof From Definition 2.1, one has successively

$$\begin{aligned}
\mathbb{E}(\mathcal{T}_{a,h,c}) &= \sum_{y \in \aleph_{a,c}} y \Pr(\mathcal{T}_{a,h,c} = y) \\
&= \frac{1}{P(a,h)} \left((a+1)^h \sum_{y \in \aleph_{a,c}} y - \sum_{y \in \aleph_{a,c}} y|y-c| \right) \\
&= c \times \frac{(a+1)^h(2a+1) - 2\sum_{k=0}^{a} k^h}{P(a,h)} \\
&= c,
\end{aligned}$$

and trivially for $y \in \aleph_{a,c}$

$$\begin{aligned}\Pr(\mathcal{T}_{a,h,c} = c - y) &= \frac{(a+1)^h - |y|^h}{P(a,h)} \\ &= \Pr(\mathcal{T}_{a,h,c} = c + y).\end{aligned}$$

Finally, one gets

$$\begin{aligned}Var(\mathcal{T}_{a,h,c}) &= \sum_{y \in \aleph_{a,c}} y^2 \Pr(\mathcal{T}_{a,h,c} = y) - [\mathbb{E}(\mathcal{T}_{a,h,c})]^2 \\ &= \frac{1}{P(a,h)} \left((a+1)^h \sum_{y \in \aleph_{a,c}} y^2 - \sum_{y \in \aleph_{a,c}} y^2 |y-c|^h \right) - c^2 \\ &= \frac{1}{P(a,h)} \left(\frac{2a(a+1)^{h+1}(2a+1)}{6} - 2\sum_{k=0}^{a} k^{h+2} + c^2 P(a,h) \right) - c^2 \\ &= \frac{1}{P(a,h)} \left(\frac{a(a+1)^{h+1}(2a+1)}{3} - 2\sum_{k=0}^{a} k^{h+2} \right).\end{aligned}$$

□

From expression (4) of $P(a,h)$ obtained by the Bernoulli number, one can calculate corresponding forms of $V(a,h) = Var(\mathcal{T}_{a,h,c})$ for $h = 1, 2, \cdots, 8$.

A part of the following remark is easy to check through Definition 2.1, Proposition 2.2, and the fact that the support $\aleph_{a,c} = \{c, c \pm 1, \cdots, c \pm a\}$ of $\mathcal{T}_{a,h,c}$ depends on the arm $a \in \mathbb{N}$ but not on the order $h > 0$. In general, it can be also observed by using a computer algebra program.

Remark 1

(i) For fixed $h > 0$, the first partial variance function $a \mapsto V(a,h) = Var(\mathcal{T}_{a,h,c})$ of $\mathcal{T}_{a,h,c}$ is increasing and unbounded with respect to the arm $a \in \mathbb{N}$.
(ii) For fixed $a \in \mathbb{N}$, the second partial variance function $h \mapsto V(a,h) = Var(\mathcal{T}_{a,h,c})$ of $\mathcal{T}_{a,h,c}$ is increasing from 0 to $a(a+1)/3$ with respect to the order $h > 0$.

In order to see the boundaries of the function $h \mapsto V(a,h) = Var(\mathcal{T}_{a,h,c})$ in Part (ii) of Remark 1, the following proposition shows the limits with respect to h of the random variables $\mathcal{T}_{a,h,c}$.

PROPOSITION 2.3 *Let $\mathcal{T}_{a,h,c}$ be a discrete triangular random variable of order $h > 0$, centered in $c \in \mathbb{N}$ with arm $a \in \mathbb{N}$. Then for all $(a,c) \in \mathbb{N}^2$:*

(i) when $h \to 0$ the limit random variable $\mathcal{T}_{a,0,c}$ follows the Dirac law at c;
(ii) when $h \to \infty$ the limit random variable $\mathcal{T}_{a,\infty,c}$ follows the discrete uniform law on the support $\aleph_{a,c} = \{c, c \pm 1, \cdots, c \pm a\}$.

Proof Since the support of $\mathcal{T}_{a,h,c}$ is the discrete set $\aleph_{a,c}$, it is enough to calculate the limits of individual probabilities of $\mathcal{T}_{a,h,c}$ with respect to h.
(i) On the first hand we have

$$\begin{aligned}\lim_{h \to 0} \Pr(\mathcal{T}_{a,h,c} = c) &= \lim_{h \to 0} \frac{1}{2a + 1 - 2\sum_{k=0}^{a} [k/(a+1)]^h} \\ &= \frac{1}{2a + 1 - 2a} = 1.\end{aligned}$$

On the other hand, for all $y \neq c$, there exists $j \in \{1, 2, \cdots, a\}$ such that $y = c \pm j$; and, therefore, we get

$$\lim_{h\to 0} \Pr(\mathcal{T}_{a,h,c} = y) = \lim_{h\to 0} \frac{(a+1)^h - j^h}{(2a+1)(a+1)^h - 2\sum_{k=0}^{a} k^h}$$
$$= \frac{1-1}{2a+1-2a} = 0.$$

(ii) Let $y \in \aleph_{a,c}$. We can always write y as $y = c \pm j$ with $j \in \{0, 1, \cdots, a\}$. Then, we obtain

$$\Pr(\mathcal{T}_{a,h,c} = y) = \frac{(a+1)^h - j^h}{(2a+1)(a+1)^h - 2\sum_{k=0}^{a} k^h}$$
$$= \frac{1}{2a+1} \times \frac{1 - [j/(a+1)]^h}{1 - [2/(2a+1)] \sum_{k=0}^{a} [k/(a+1)]^h}$$

for $j \in \{0, 1, 2, \cdots, a\}$. It follows that $\lim_{h\to\infty} \Pr(\mathcal{T}_{a,h,c} = y) = 1/(2a+1)$, for any $y \in \aleph_{a,c}$. □

3 Discrete triangular kernel estimators

Now we can define the discrete triangular kernel estimators. Let $X_1, \cdots, X_n$ be a random sample from a count distribution with an unknown probability mass function $x \mapsto f(x) := \Pr(X_i = x)$ on $\mathbb{N}$. Given $a \in \mathbb{N}$, a *discrete triangular kernel estimator of arm* a of f is defined by

$$\widehat{f}_{n,h,a}(x) = \frac{1}{n} \sum_{i=1}^{n} T_{a,h,x}(X_i)$$
$$= \frac{1}{n} \sum_{i=1}^{n} \frac{(a+1)^h - |X_i - x|^h}{(2a+1)(a+1)^h - 2\sum_{k=0}^{a} k^h}, \qquad x \in \mathbb{N}, \tag{5}$$

where $h > 0$ is the smoothing bandwidth, and $T_{a,h,x}$ is the associated discrete kernel to the discrete triangular random variable $\mathcal{T}_{a,h,x}$ on $\aleph_{a,x} = \{x, x \pm 1, \cdots, x \pm a\}$. According to [2], we have $\widehat{f}_{n,h,a}(x) \in [0, 1]$ for all $x \in \mathbb{N}$ and, up to normalizing constant $C = \sum_{x\in\mathbb{N}} \widehat{f}_{n,h,a}(x)$, we assume that the function $x \mapsto \widehat{f}_{n,h,a}(x)$ is a density of probability on $\mathbb{N}$. Note also that the equivalent of (2) is here as

$$\mathbb{E}[\widehat{f}_{n,h,a}(x)] = \mathbb{E}[f(\mathcal{T}_{a,h,x})]. \tag{6}$$

Also, the support $\aleph_{a,x}$ of the random variable $\mathcal{T}_{a,h,x}$ does not depend on h as in general case (2), and we have $\cup_x \aleph_{a,x} \supseteq \mathbb{N}$. If $a = 0$ we get $\cup_x \aleph_{0,x} = \mathbb{N}$. In this case $a = 0$, the estimator $\widehat{f}_{n,h,0}$ of f is only the empirical distribution f_0 of observations. While if $a \neq 0$ we obtain

$$\bigcup_{x\in\mathbb{N}} \aleph_{a,x} = \{-a, \cdots, -1\} \cup \mathbb{N}. \tag{7}$$

The fact that the support $\cup_x \aleph_{a,x}$ (7) of the discrete triangular kernel at fixed $a \neq 0$ contained strictly the support $\mathbb{N}$ of f induces a boundary bias on the left of $\mathbb{N}$. We remedy at the end of subsection 3.2 by modifying lightly a to a_0 such that, for any a_0, one has

$$\bigcup_{x\in\mathbb{N}} \aleph_{a_0,x} = \mathbb{N}. \tag{8}$$

Let us indicate here that if the support of f is $\mathbb{Z}$ (unbounded integers set) then $\cup_{x\in\mathbb{Z}}\aleph_{a,x} = \mathbb{Z}$ and, therefore, there is no problem of boundary bias. Whereas if the support of f is $\{0, 1, \cdots, N\}$ (compact) then $\cup_{x\in\{0,1,\cdots,N\}}\aleph_{a,x} = \{-a, \cdots, -1\}\cup\{0, 1, \cdots, N\}\cup\{N+1, \cdots, N+a\}$; what would lead to the boundaries bias at the same moment to the left and to the right of the support of f.

3.1 *Mean integrated squared errors*

From (6) we investigate the behaviour of the smoothing parameter $h > 0$ and, also, of the arm $a \in \mathbb{N}$ through the common measure of accuracy for an estimator which is the Mean Integrated Squared Error:

$$MISE_a(n,h)(f) = \mathbb{E}\sum_{x\in\mathbb{N}}\left[\widehat{f}_{n,h,a}(x) - f(x)\right]^2 \tag{9}$$

$$= \sum_{x\in\mathbb{N}} MSE_a(n,h,x) \tag{10}$$

$$= \sum_{x\in\mathbb{N}} Var[\widehat{f}_{n,h,a}(x)] + \sum_{x\in\mathbb{N}} Bias^2[\widehat{f}_{n,h,a}(x)] \tag{11}$$

of the estimator $\widehat{f}_{n,h,a}$ of f defined in (5), where $MSE_a(n,h,x)$ is the Mean Squared Error at a point $x \in \mathbb{N}$.

Indeed, given $x \in \mathbb{N}$. Since $\mathbb{E}(\mathcal{T}_{a,h,x}) = x$ and $Var(\mathcal{T}_{a,h,x}) = V(a,h)$ (Proposition 2.2), the approximate expression of the bias of $\widehat{f}_{n,h,a}(x)$ is given classically by the discrete Taylor expansion as

$$\begin{aligned} Bias[\widehat{f}_{n,h,a}(x)] &= \mathbb{E}[\widehat{f}_{n,h,a}(x)] - f(x) \\ &= \mathbb{E}[f(\mathcal{T}_{a,h,x})] - f(x) \\ &= \frac{1}{2}Var(\mathcal{T}_{a,h,x})f''(x) + o(h) \\ &= \frac{1}{2}V(a,h)f''(x) + o(h), \end{aligned} \tag{12}$$

and its variance can be approximated successively by

$$\begin{aligned} Var\left[\hat{f}_{n,h,a}(x)\right] &= \frac{1}{n}Var\left[T_{a,h,x}(X_1)\right] \\ &= \frac{1}{n}\mathbb{E}\left[(T_{a,h,x}(X_1))^2\right] + O(n^{-1}) \\ &= \frac{1}{n}\sum_{y\in\aleph_{a,x}} f(y)\left[\Pr(\mathcal{T}_{a,h,x} = y)\right]^2 + O(n^{-1}) \\ &= \frac{1}{n}f(x)\Pr(\mathcal{T}_{a,h,x} = x) + o(n^{-1}a^{-h}) \\ &= \frac{(a+1)^h}{nP(a,h)}f(x) + o(n^{-1}a^{-h}), \end{aligned} \tag{13}$$

where $P(a,h)$ is given in Definition 2.1, $V(a,h)$ in Proposition 2.2, and the second derivative of f at each

point $x \in \mathbb{N}$ is replaced by the following finite differences

$$f''(x) = \begin{cases} [f(x+2) - 2f(x) + f(x-2)]/4 & \text{if } x \in \mathbb{N} \setminus \{0,1\} \\ [f(3) - f(2) - f(1) + f(0)]/2 & \text{if } x = 1 \\ f(2) - 2f(1) + f(0) & \text{if } x = 0. \end{cases} \tag{14}$$

The fourth equality of (13) can be obtained through properties of $\mathcal{T}_{a,h,c}$ and the global majorations:

$$\sum_{y \in \aleph_{a,x}} f(y) \left[\Pr(\mathcal{T}_{a,h,x} = y)\right]^2 \leq f_{\max} \max_{y \in \aleph_{a,x}} \Pr(\mathcal{T}_{a,h,x} = y) \leq 1.$$

Using (13) the approximate of the integrated variance is given by

$$\sum_{x \in \mathbb{N}} Var[\widehat{f}_{n,h,a}(x)] = \frac{(a+1)^h}{nP(a,h)} + o(n^{-1}a^{-h}), \tag{15}$$

which does not depend on f. Observe that $(a+1)^h/P(a,h) = \Pr(\mathcal{T}_{a,h,x} = x)$ is the modal probability of the random variable $\mathcal{T}_{a,h,x}$ (Definition 2.1 and also Figure 1), it follows that the function $h \mapsto (a+1)^h[nP(a,h)]^{-1}$ is decreasing from n^{-1} to $[n\,(2a+1)]^{-1}$ for all n and a in $\mathbb{N}^*$.

From pointwise bias (12), the approximate of the integrated squared bias is written

$$\sum_{x \in \mathbb{N}} Bias^2[\widehat{f}_{n,h,a}(x)] = \frac{1}{4}\left[V(a,h)\right]^2 \sum_{x \in \mathbb{N}} \left[f''(x)\right]^2 + o(h^2). \tag{16}$$

Since $f''(x)$ is a finite linear combination of $f(x \pm j) \in [0,1]$ for $j \in \{0,1,2\}$ (14), one has $\sum_{x \in \mathbb{N}} \left[f''(x)\right]^2 < \infty$. We deduce from Part (ii) of Remark 1 that the function $h \mapsto \left[V(a,h)\right]^2/4 = \left[Var(\mathcal{T}_{a,h,x})/2\right]^2$ is increasing from 0 to $a^2(a+1)^2/36$ for all $a \in \mathbb{N}$.

Replacing (15) and (16) in (11), we obtain

$$MISE_a(n,h)(f) = \frac{(a+1)^h}{nP(a,h)} + \frac{1}{4}\left[V(a,h)\right]^2 \sum_{x \in \mathbb{N}} \left[f''(x)\right]^2 + o(n^{-1}a^{-h} + h^2), \tag{17}$$

which leads to choosing the optimal parameter of h and an adequate arm a.

3.2 *Optimal parameters*

The optimal value of $h > 0$ and of $a \in \mathbb{N}^*$ are obtained (if there exists) by minimizing the asymptotic $MISE$ of (17):

$$AMISE_a(n,h)(f) = \frac{(a+1)^h}{nP(a,h)} + \frac{1}{4}\left[V(a,h)\right]^2 \sum_{x \in \mathbb{N}} \left[f''(x)\right]^2. \tag{18}$$

Firstly, since $a \in \mathbb{N}^*$ and the function $a \mapsto AMISE_a(n,h)(f)$ is always increasing for fixed $h > 0$, the most interesting positive arm is $a = 1 = a^*$ which we can call "adequate" or "optimal" arm for convenience. Secondly, for given arm $a \in \mathbb{N}^*$, the *optimal bandwidth* for a discrete triangular kernel estimator is defined by

$$h^* = arg \min_{h>0} AMISE_a(n,h)(f) = h^*(n,a,f). \tag{19}$$

Its existence is guaranteed by the diminishing of the approximate of the integrated variance (15) and the growth of the approximate of the integrated squared bias (16) in the global squared error (18). For h small, the bias is small too but the variance is large. On the other hand, if h is large, the variance becomes small and the bias large. To find the optimal value of h, one should balance the approximate of the squared bias and the variance terms. This means that there exists $\varepsilon > 0$ such that the function $h \mapsto AMISE_a(n,h)(f)$ is decreasing for h belongs to $(0,\varepsilon)$ and increasing for h in (ε,∞) for all $a \in \mathbb{N}^*$. Since $h^* = h^*(n,a,f)$ cannot be computed in practice because it depends on the unknown f, one way to "estimate" is to take an empirical distribution f_0 of f and choose the corresponding bandwidth $h_0^* = h^*(n,a,f_0)$. That is equivalent to

$$h_0^* = arg \min_{h>0} AMISE_a(n,h)(f_0) = h_0^*(n,a,f_0). \tag{20}$$

Because of f_0 tends to f when $n \to \infty$, we may have

$$\lim_{n\to\infty} h_0^*(n,a,f_0) = \lim_{n\to\infty} h^*(n,a,f), \tag{21}$$

for fixed $a \in \mathbb{N}$. This procedure allows to choose the optimal bandwidth (20) to estimate a probability mass function from an empirical distribution.

An alternative to the choice of optimal bandwidth is the use of the well-known cross validation method (*e.g.*, [15], [16] and also [17]). Even in the discrete case, this procedure has the advantage that it does not approximate $MISE_a(n,h)(f)$ by using finite differences of f. Hence, we can differently develop the $MISE$ (9) as

$$MISE_a(n,h)(f) = \mathbb{E}\left[\sum_{x\in\mathbb{N}} \hat{f}^2_{n,h,a}(x)\right] - 2\mathbb{E}\left[\sum_{x\in\mathbb{N}} \widehat{f}_{n,h,a}(x) f(x)\right] + \sum_{x\in\mathbb{N}} f^2(x);$$

and, therefore, the optimal bandwidth is obtained by

$$h_{cv} = arg \min_{h>0} CV_a(h) = h_{cv}(a), \tag{22}$$

where

$$CV_a(h) = \sum_{x\in\mathbb{N}} \left[\frac{1}{n}\sum_{i=1}^{n} T_{a,h,x}\left(X_i\right)\right]^2 - \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} T_{a,h,X_i}\left(X_j\right)$$

is the unbiased estimator of $\mathbb{E}\left[\sum_{x\in\mathbb{N}} \hat{f}^2_{n,h,a}(x)\right] - 2\mathbb{E}\left[\sum_{x\in\mathbb{N}} \widehat{f}_{n,h,a}(x) f(x)\right]$.

Following [2], another way to find an *adapted bandwidth* h_0 in the particular situation of excess zeros for count data would be to solve the equation $\sum_{i=1}^{n} \Pr\left(\mathcal{T}_{a,h_0,X_i} = 0\right) = n_0$, equivalent to

$$\left[n - (2a+1)\, n_0\right](a+1)^h + 2n_0 \sum_{k=0}^{a} k^h - \sum_{i=1}^{n} X_i^h = 0, \tag{23}$$

where $n_0 = \natural(X_i = 0)$ represents the number of observations equal to zero. But this equation has no solution. Indeed, because $k \leq a$ in the second term of the first member in (23), we have

$$\left[n - (2a+1)\, n_0\right](a+1)^h + 2n_0 \sum_{k=0}^{a} k^h < \left[n - (2a+1)\, n_0\right](a+1)^h + 2n_0 a (a+1)^h$$

$$= (n - n_0)(a+1)^h.$$

Since $\sum_{i=1}^n X_i^h = \sum_{j=0}^k n_j X_j^h$ with $\sum_{j=0}^k n_j = n$ and $X_j \geq a+1$, $j = 1, 2, \cdots, k$, we can see that

$$\sum_{i=1}^n X_i^h \geq (n - n_0)(a+1)^h.$$

This result, combined with the previous inequality, allows to deduce that equality (23) does not hold and we finally have

$$[n - (2a+1)\, n_0]\, (a+1)^h + 2n_0 \sum_{k=0}^a k^h - \sum_{i=1}^n X_i^h < 0.$$

Let us now propose our remedy to the problem of boundary bias. For significant observations to the boundary $\{0, 1, \cdots, m\}$ of the support $\mathbb{N}$ of f (m too small, like 0, 1 or 2), we consider the *modified arm* a_0 of a satisfying (8). Our solution is such that, for given $k \in \mathbb{N}^*$ and $x \in \mathbb{N}$,

$$a_0 = k \Longleftrightarrow a = \begin{cases} j \text{ if } x = j, \quad j \in \{0, 1, \cdots, k-1\} \\ k \text{ if } x \in \{k, k+1, \cdots\}. \end{cases} \tag{24}$$

Thus, the discrete triangular kernel associated to this modified arm a_0 preserves the structure of local symmetry of the discrete kernel around every point of estimation. An alternative to our method would be to consider an asymmetric triangular kernel for the boundary. In practice, this procedure is done before the normalization of the estimator $\widehat{f}_{n,h,a^*}$ by the constant $C^* = \sum_{x\in\mathbb{N}} \widehat{f}_{n,h,a^*}(x)$. So that $\widehat{f}_{n,h,0}(0)$ is generally not equal to the empirical estimation $f_0(0)$ of f at the edge $x = 0$ under circumstances of an excess of zero.

For $a = 1$ and $a_0 = 1$, we respectively have

$$\begin{aligned} AMISE_{a_0=1}(n,h)(f) &= AMSE_{a=0}(n,h,0) + \sum_{x\in\mathbb{N}^*} AMSE_{a=1}(n,h,x) \\ &= \frac{f(0)}{n} + \sum_{x\in\mathbb{N}^*} AMSE_{a=1}(n,h,x) \end{aligned}$$

and

$$\begin{aligned} AMISE_{a=1}(n,h)(f) &= AMSE_{a=1}(n,h,0) + \sum_{x\in\mathbb{N}^*} AMSE_{a=1}(n,h,x) \\ &= \frac{2^h}{P(1,h)} \frac{f(0)}{n} + \frac{1}{4}[V(1,h)]^2 (f''(0))^2 + \sum_{x\in\mathbb{N}^*} AMSE_{a=1}(n,h,x), \end{aligned}$$

where $AMSE_a$ is the asymptotic MSE_a of (10). Consequently, comparison between $AMISE_{a_0}(n,h)(f)$ and $AMISE_a(n,h)(f)$ is reduced to compare $f(0)/n$ to $2^h f(0)/[nP(1,h)] + [V(1,h)]^2 (f''(0))^2/4$, for which we cannot conclude. Indeed, it depends on smoothing parameter h and on both size n and proportion $f(0)$ from sample. So, in the next section we present some graphs of $AMISE_{a_0}(n,h)(f)$ and $AMISE_a(n,h)(f)$ to illustrate their behaviour (see Figure 2).

Finally, the choices of bandwidths h_0^* or h_{cv} and of arms a^* or a_0^* are wide enough to obtain a very good quality for smoothing any probability mass function. The use of a_0^* instead of a^* is recommended for count data having a large proportion of zeros.

4 Illustrations from data

In this section we illustrate some previous results. We firstly use simulated data and then real data from football already presented in [2]. All computations are done by the R software [18].

4.1 *Simulated data*

Consider the following discrete function

$$f(x) = 0.4\ e^{-0.5}\ 0.5^x/x! + 0.6\ e^{-10}\ 10^x/x!, \qquad x \in \mathbb{N},$$

which is a mixture of two Poisson distributions with means $\mu_1 = 0.5$ and $\mu_2 = 10$. This probability mass function f defined on $\mathbb{N}$ has a bimodality with the maximum at $x = 0$ ($f(0) = 0.243$), a local minimum at $x = 3$ ($f(3) = 9.594 \times 10^{-3}$), a local maximum at $x = 9$ ($f(9) = 7.506603 \times 10^{-2}$) and a tail from $x = 22$ ($1 - \sum_{x=0}^{21} f(x) = 4.198 \times 10^{-4}$).

In general, for given random sample n, we consider 1000 independent simulated samples f_0 of f and in most of 98% of cases it is easy to observe the following results. First, each corresponding function $h \mapsto AMISE_a(n,h)(f_0)$ observed is bounded, of the same order (very small) as theoretical $h \mapsto AMISE_a(n,h)(f)$ and is increasing for $a \in \mathbb{N}^*$. One gets the optimal triangular kernel associated to optimal bandwidth h^* of (19) with arm $a = 1 = a^*$. Furthermore, the behaviour of the equality (21) holds easily for a sample size $n = 1000$ and $a = 1 = a^*$. Thus, the approximate h_0^* of (20) is a well indicatory of h^* for large n.

From simulated data that we omit here to present all results, we obtained that the functions $h \mapsto AMISE_{a_0=1}(n,h)(f_0)$ and $h \mapsto AMISE_{a=1}(n,h)(f_0)$ have the same aspect for different samples sizes. For a sample with both large size n and important proportion $f_0(0)$, the values of h which realize the minimum of each of the two previous functions are of the same order.

For a simulated sample f_0 with a significant proportion (*e.g.* $f_0(0) = 0.240$), we observed that we can estimate parameter h^* by h_0^*. The use of the modified arm a_0 or of the arm a is possible for any discrete distribution f_0, but the arm a_0 is suitable with a large number of zeros.

Figure 2 below will show the behaviour of $AMISE_{a_0}(n,h)(f)$ and $AMISE_a(n,h)(f)$ for real data with a moderate sample size $n = 380$.

4.2 *Real data*

We realize some nonparametric estimation $\widehat{f}$ of the distribution of count data from the French football championship of League 1 with $n = 380$ (Table 1). We first use some discrete triangular kernel estimators and we then compare them with a binomial kernel estimator. Note that, for a given goal $g \in \mathbb{N}$, the nearest integer to $n \times \widehat{f}(g)$ is the nonparametric estimation of the corresponding matches number.

(Tables 1 and 2 about here)

In Figure 2 we plot $h \mapsto AMISE_a(n,h)(f_0)$ and $h \mapsto AMISE_{a_0}(n,h)(f_0)$ for the arms $a \in \{1,2\}$ and $a_0 \in \{1,2\}$. We observe that for $a = 1$ and $a_0 = 1$, curve of $AMISE_{a_0}(n,h)(f_0)$ is over then under those of $AMISE_a(n,h)(f_0)$. The behaviour is the same for $a = 2$ and $a_0 = 2$, except that the point of

intersection is near to zero. In what follows we do not use the values of bandwidth h which realize the minimum of $AMISE_a(n,h)(f_0)$ for smoothing but the optimal smoothing parameter $h_{cv}(a)$ of (22) found by cross validation procedure.

In order to measure the performance of the estimators, we simply use the practical Integrated Squared Error

$$ISE^0 = \sum_{x\in\mathbb{N}} \left[\widehat{f}_{n,h_{cv}(a)}(x) - f_0(x)\right]^2$$

(which can be directly observed in graphics) and also the estimate $\widehat{AMISE}(\widehat{f}_{n,h_{cv}(a)})$ of $AMISE_a(n,h)(f)$ for fixed $a \in \mathbb{N}$. In Table 2 we can see the results from smoothing of data of League 1 by discrete triangular kernel for $a \in \{1,2\}$ and $a_0 = 1$ with the corresponding values of $h_{cv}(a)$ and the normalizing constant C. According to the measure $\widehat{AMISE}(\widehat{f}_{n,h_{cv}(a)})$, the best performance among triangular kernels is obtained with $a = 1$. It is due to the proportion $f_0(0) = 0.134$ of zeros in the sample which is smaller than $f(0) = 0.243$ for the previous study. So, the arm $a_0 = 1$ does not give a better result than $a = 1$. But, by using practical measure ISE^0, the modified arm $a_0 = 1$ is better than the arm $a = 1$, even if both are of the same order. The difference comes from $\widehat{AMISE}(\widehat{f}_{n,h_{cv}(a)})$ which is a biased theoretical measure approximated by the use of finite differences of f. Discrete triangular kernels improve discrete standard kernels. One can observe this last fact in the following comparison with the best of them: the binomial kernel corresponding to the estimator

$$\begin{aligned}\widehat{f}_{n,h}(x) &= \frac{1}{n}\sum_{i=1}^{n} B_{x,h}(X_i)\\ &= \frac{1}{n}\sum_{i=1}^{n} \frac{(x+1)!}{X_i!(x+1-X_i)!}\left(\frac{x+h}{x+1}\right)^{X_i}\left(\frac{1-h}{x+1}\right)^{x+1-X_i}, \qquad x \in \mathbb{N},\end{aligned}$$

where $h \in (0,1]$ is the smoothing parameter and $B_{x,h}$ is the discrete kernel associated to the binomial distribution $\mathcal{B}(x+1,(x+h)/(x+1))$ with its support $\aleph_{x,h} = \{0,1,\cdots,x+1\}$; see [2], Exemple 2.2, for more details. The corresponding results of the binomial kernel estimation with $h_{cv} = 0.177$ are compared to those obtained by discrete triangular kernels in Figure 3 and Table 2. By using the estimate error $\widehat{AMISE}(\widehat{f}_{n,h_{cv}(a)})$, discrete triangular kernels are better than the binomial kernel with orders 2.26×10^{-3} (for triangular $a = 1$) and 2.41×10^{-3} (binomial). As for error ISE^0 generally used in practice, one can clearly see that the best results are obtained by discrete triangular kernel ($ISE^0 = 10^{-5}$ for $a_0 = 1$) compared to binomial kernel ($ISE^0 = 3.95 \times 10^{-3}$). Since ISE^0 is more appropriated than $\widehat{AMISE}$ we can conclude that triangular ($a_0 = 1$) kernel estimation is better than binomial one for these count data.

(Figures 2 and 3 about here)

5 Concluding remarks

Univariate discrete triangular distributions $\mathcal{T}_{a,h,c}$ complete the families of so many discrete distributions [3]. We can define other versions of $\mathcal{T}_{a,h,c}$ such multivariate case. Discrete triangular kernel estimators which include frequency estimator (with $a = 0$ and $h = 1$) are now ready to use and may be adapted for discrete regression functions or other discrete functions. However, it remains to investigate performance of discrete triangular kernels with respect to any underdispersed discrete kernel in lieu of the prototype binomial. In practice, a binomial kernel seems to be appropriated for small size count data which are not sparsed. In comparison with binomial estimator, triangular kernel estimators with small arm ($a \in \mathbb{N}^*$) are more consistent than the frequency one.

Acknowledgement

We thank the anonymous referee for her/his valuable comments.

References

[1] Marsh, L.C., Mukhopadhyay, K., 1999, Discrete Poisson kernel density estimation with an application to wildcat coal strikes. *Appl. Economics Lett.,* **6**, 393-396.
[2] Kokonendji, C.C., Senga Kiessé, T., 2006, Estimateurs à noyau discret standard d'une densité de probabilité discrète, Technical report LMA No. 0632, University of Pau.
[3] Johnson, N.L., Kemp, A.W. and Kotz, S., 2005, *Univariate Discrete Distributions (3rd ed.).* New York: Wiley.
[4] Devroye, L., 1987, *A Course in Density Estimation.* Boston: Birkhäuser.
[5] Tsybakov, A.B., 2004, *Introduction à l'Estimation Non-Paramétrique.* Paris: Springer.
[6] Scott, D.W., 1992, *Multivariate Density Estimation - Theory, Practice, and Visualization.* New York: Wiley.
[7] Ferraty, F., Vieu, P., 2006, *Nonparametric Functional Data Analysis: Theory and Practice.* Berlin: Springer.
[8] Simonoff, J.S., 1996, *Smoothing Methods in Statistics.* New York: Springer.
[9] Simonoff, J.S., Tutz, G., 2000, Smoothing methods for discrete data. In *Smoothing and Regression: Approaches, Computation, and Application,* ed. M.G. Schimek, pp. 193-228, New York: Wiley.
[10] Ouyang, D., Li, Q. and Racine, J., 2006, Cross-validation and the estimation of probability distributions with categorical data. *J. of Nonparametric Statistics,* **18**, page 69-100.
[11] Parzen, E., 1962, On estimation of a probability density function and mode, *Ann. Math. Statist.,* **33**, 1065-1076.
[12] Chen, S.X., 1999, Beta kernel estimators for density functions. *Computat. Statist. Data Anal.,* **31**, 131-145.
[13] Chen, S.X., 2000, Gamma kernel estimators for density functions. *Ann. Instit. Statist. Math.,* **52**, 471-480.
[14] Bouvier, A., George, M. and Le Lionnais, F., 2005, *Dictionnaire de Mathématiques.* Paris: Presses Universitaires de France.
[15] Bowman, A., 1984, An alternative method of cross-validation for the smoothing of density estimates. *Biometrika,* **71**, 352-360.
[16] Marron, J.S., 1987, A comparison of cross-validation techniques in density estimation. *Ann. Statist.,* **15**, 152-162.
[17] Yang, Y., 2007, Consistency of cross validation for comparing regression procedures, *Ann. Statist.* (to appear).
[18] R Development Core Team, 2006, *A Language and Environment for Statistical Computing.* Vienna - Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, URL http://www.R-project.org.

Tables and figures

Table 1. Data of matches having a certain number of goals in the French football championship of League 1 (season 2005-2006)

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 2. Results from smoothing by both discrete triangular and binomial kernels of real data from the French League 1 with $n = 380$

	Triang($a = 1$)	Triang($a = 2$)	Triang($a_0 = 1$)	Binomial
h_{cv}	0.204	0.055	0.028	0.177
C	0.98600	0.98618	0.99810	0.95872
ISE^0	0.00042	0.00023	0.00001	0.00395
$\widehat{AMISE}$	0.00226	0.00462	0.00255	0.00241

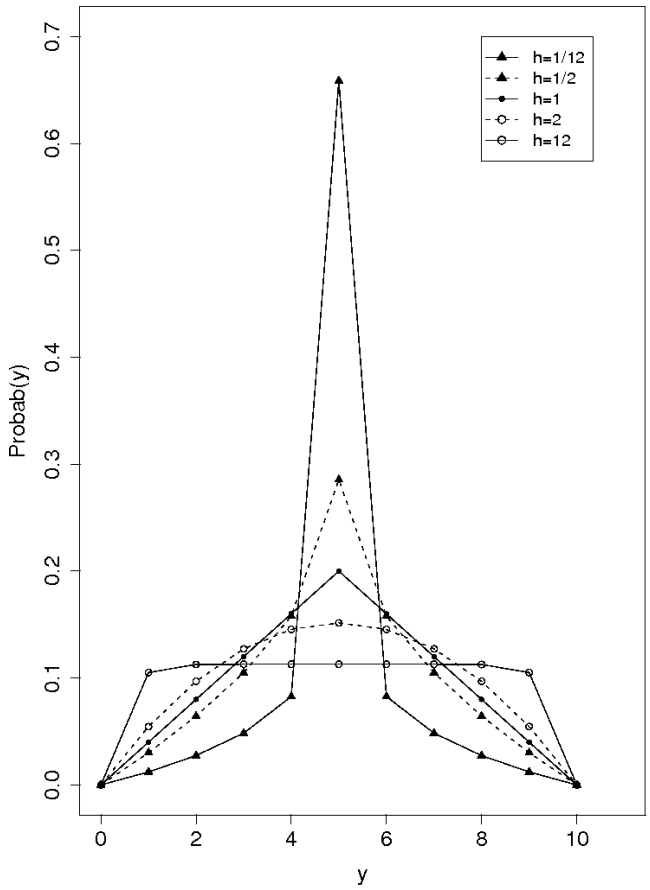

Figure 1. Some discrete triangular distributions of order $h \in \{1/12, 1/2, 1, 2, 12\}$ centered in $c = 5$ with arm $a = 4$

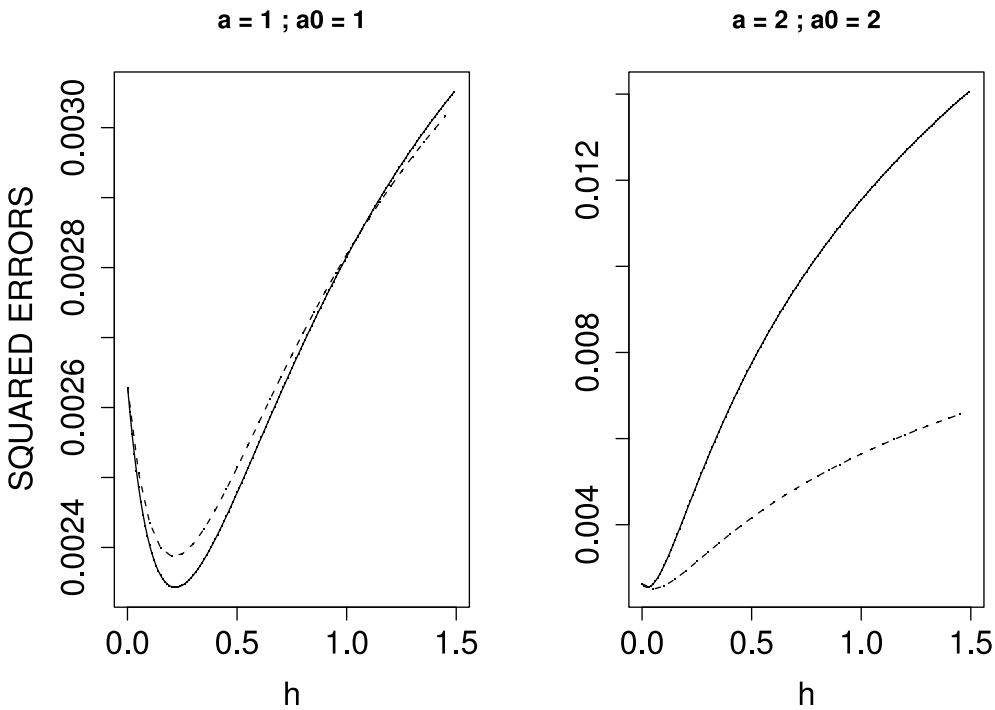

Figure 2. Graphs of functions $AMISE_a(n,h)(f_0)$ [solid line] and $AMISE_{a_0}(n,h)(f_0)$ [dashed line] of the discrete triangular kernel estimator for data from the French League 1 (season 2005-2006) with $n = 380$

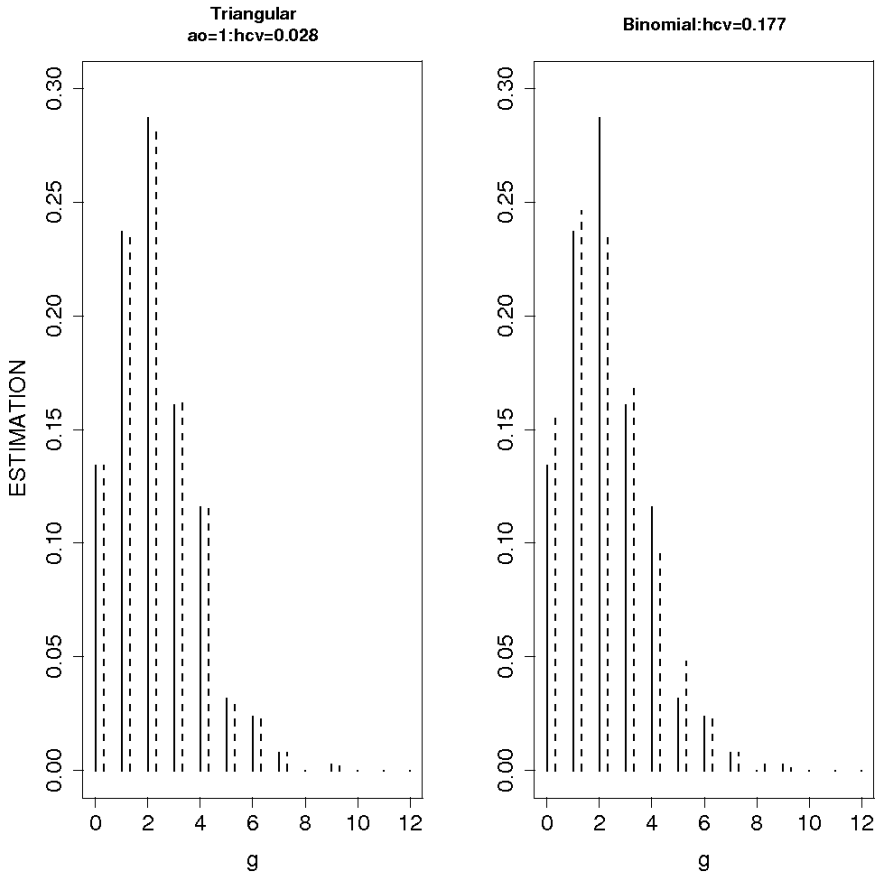

Figure 3. Discrete smoothing [dashed line] by discrete triangular $a_0 = 1$ and binomial kernels for empirical distribution [solid line] of data from the French League 1 with $n = 380$