Discrete triangular associated kernel and bandwidth choices in semiparametric estimation for count data
Tristan Senga Kiessé, Henri H. Cuny

To cite this version:
Tristan Senga Kiessé, Henri H. Cuny. Discrete triangular associated kernel and bandwidth choices in semiparametric estimation for count data. Journal of Statistical Computation and Simulation, 2014, 64 (8), pp.1813-1829. 10.1080/00949655.2013.768995. hal-01097983

HAL Id: hal-01097983
https://hal.science/hal-01097983
Submitted on 28 Oct 2021

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.

Distributed under a Creative Commons Attribution 4.0 International License
Discrete triangular associated kernel and bandwidth choices in semiparametric estimation for count data

T. Senga Kiessé* and H.E. Cuny

INRA, UMR1092, Laboratoire d’Etude des Ressources Forêt Bois (LERFoB), Centre INRA de Nancy, F-54280 Champenoux, France

This work deals with semiparametric kernel estimator of probability mass functions which are assumed to be modified Poisson distributions. This semiparametric approach is based on discrete associated kernel method appropriated for modelling count data; in particular, the famous discrete symmetric triangular kernels are used. Two data-driven bandwidth selection procedures are investigated and an explicit expression of optimal bandwidth not available until now is provided. Moreover, some asymptotic properties of the cross-validation criterion adapted for discrete semiparametric kernel estimation are studied. Finally, to measure the performance of semiparametric estimator according to each type of bandwidth parameter, some applications are realized on three real count data-sets from sociology and biology.

Keywords: bootstrap methods; cross-validation; discrete associated kernel; mean integrated squared error; optimal bandwidth; wood formation

1. Introduction

The traditional approach for estimating count data distribution has been essentially parametric until recently. This approach classically consists in a departure with a structure of count distribution such that the Poisson model; however, the estimation provided in this way is not always sufficient and it becomes necessary to modify the initial distribution. In this work, the estimation approach adopted assumes that any count distribution or random variable $X$ having a probability mass function (p.m.f.) $f(x) = \Pr(X = x) > 0$ on support $\mathbb{N}$ can be written as a modified Poisson distribution:

$$f(x) = p(x; \Theta) \omega(x) \quad \forall x \in \mathbb{N},$$

where $p(x; \Theta) = \Theta^x \exp(-\Theta)/x!$ is the p.m.f. of a Poisson distribution with mean parameter $\Theta > 0$, and $\omega(x) > 0$ is a nonparametric function playing the role of a correction factor. Equation (1) is an assumption related to the works on weighted Poisson distributions (WPDs) which are investigated as some alternatives to parametric Poisson model classically applied for count data; WPDs allow one to take into account the counting phenomenon as over/under-dispersion or zero-inflation/deflation [1]. Thus, the advantage of using Equation (1) is to express

*Corresponding author. Email: tristan.senga@nancy.inra.fr, sengatristan@yahoo.fr
the deviation from classical Poisson models and thus to take intrinsically into account special features of counting phenomenon mentioned above. Let $X_1, X_2, \ldots, X_n$ be a sample of independent observations with an unknown count distribution $f$ given in Equation (1). A discrete semiparametric estimator of $f$ is proposed in [2] as a parametric estimation $\hat{p}(x) = p(x; \hat{\Theta})$ of $p$ followed by a nonparametric kernel estimation $\hat{\omega}(x)$ of $\omega(x) = f(x)\hat{p}(x)$ given by

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

such that we have

$$
\hat{f}_n(x) = \hat{p}(x) \times \hat{\omega}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_{x,h}(X_i) \frac{p(x; \hat{\Theta})}{p(X_i; \hat{\Theta})} =: \hat{f}_{n,K,h}(x), \quad x \in \mathbb{N}.
$$

The estimator $\hat{\Theta} = n^{-1} \sum_{i=1}^{n} X_i$ is the sample mean which is the maximum likelihood estimator of the Poisson mean $\theta$, the bandwidth $h = h(n) > 0$ is an arbitrary sequence of smoothing parameters that fulfills $\lim_{n \to \infty} h(n) = 0$, and the discrete associated kernel $K_{x,h}(\cdot) = Pr(K_{x,h} = \cdot)$ of random variable $K_{x,h}$ is a p.m.f. with support $\mathcal{S}$ satisfying the hypotheses

$$
H_1 : \lim_{h \to 0} E(K_{x,h}) = x \quad \text{and} \quad H_2 : \lim_{h \to 0} \text{Var}(K_{x,h}) = 0.
$$

Note that the continuous version of $\hat{f}_n$ can be found in [3].

This paper pursues the works on estimator $\hat{f}_n$ in Equation (2) using the famous example of discrete symmetric triangular associated kernels introduced by Kokonendji et al. [4]. Under some assumptions, a mathematical result on pointwise consistency of $\hat{f}_n$ is formulated followed by a proposition on global consistency of $\hat{f}_n$ using discrete triangular kernels. About smoothing parameter $h > 0$, some data-driven bandwidth selection procedures are investigated. Indeed, the well-known cross-validation procedure, which consists in the minimization of a score function, has been just applied for $\hat{f}_n$; here, we further study this function by establishing some asymptotic properties such as bias and variance. It results in an other mathematical result on these properties of cross-validation score function. In addition, for bandwidth choice, the minimization of an approximate global-squared error of $\hat{f}_n$ is investigated; thus, an explicit expression of optimal bandwidth not available until now for $\hat{f}_n$ is obtained by using discrete triangular kernels. Note that for choosing bandwidth parameters in continuous nonparametric kernel estimation, one can refer to [5–7]. In addition, let us also remark that a Bayesian local approach is developed by Zougab et al. [8,9] for bandwidth selection in discrete nonparametric associated kernel estimation of p.m.f. Finally, concerning count data, the problem of their semiparametric regression is treated by Abdous et al. [10].

We attempt to illustrate our investigations via three real count data-sets. The first data-set, used earlier by Kokonendji et al. [2, Table 4, p. 12], comes from a sociological experiment concerning the number of days per week in which alcohol was consumed. The second application concerns count data characterizing development of spiralling whitefly, which is an insect pest plant collected in Republic of Congo-Brazzaville [11]. This insect causes some damages such as sucking the sap, decreasing photosynthesis activity and drying up the leaves; and the congolese biologists are searching for a suitable modelling of data related to this insect growth. The third application is realized on wood cell count data from times series: it concerns data relative to the annual wood formation dynamics of two silver fir trees. Indeed, the study of wood formation has become an innovative and fast-growing field in plant sciences over the last decade since wood is a major component of the biosphere and plays a key role in ecosystem functioning, representing for example one of the strongest sink of CO$_2$, which is a major contributor to climate change [12].
Therefore, modelling wood development involves crucial issues and has become an important problem of plant sciences.

The remainder of this paper is organized as follows. Section 2 gives details about discrete triangular associated kernels with some expansions of its modal probability and variance. Some results on consistency of semiparametric estimator \( \hat{f}_n \) in Equation (2) of p.m.f. \( f \) are also given. Moreover, the choice of optimal \( h \)-values is studied for \( \hat{f}_n \) according to the two procedures mentioned previously; a mathematical result on the asymptotic bias and variance of cross-validation criterion is established. Then, Section 3 presents applications on count data-sets from sociology and biology; in particular, the bootstrap method is applied on data of insect pest plant for a robust evaluation of the bandwidth parameter choices proposed. Finally, Section 4 contains some concluding remarks. The proofs of mathematical results are postponed to the appendix.

2. Semiparametric kernel estimator

This section first presents a class of discrete symmetric kernels [4,13]. Then, the two optimal bandwidth choices for estimator \( \hat{f}_n \) in Equation (2) are presented.

2.1. Discrete symmetric triangular associated kernels

Let \( a \) be a fixed integer and \( h > 0 \) be a smoothing parameter. For any fixed \( x \in \mathbb{N} \), consider the random variable (r.v.) \( K_{a,x,h} \) of discrete symmetric triangular associated kernel \( K_{a,x,h} \) defined on support \( S_{a,x} = \{x, x \pm 1, \ldots, x \pm a\} \) and whose p.m.f. is given by

\[
K_{a,x,h}(y) = \frac{(a + 1)^h - |y - x|^h}{P(a, h)} \quad \forall y \in S_{a,x},
\]

where \( P(a, h) = (2a + 1)(a + 1)^h - 2 \sum_{k=1}^{a} k^h \) is the normalizing constant. The parameter \( a \) plays a role on the number of observations falling in the set \( S_{a,x} \), while \( h \) is directly the smoothing parameter. This class of kernels satisfies assumptions \( H_1-H_2 \), which implies \( \Pr(K_{a,x,h} = x) \to 1 \) and \( \Pr(K_{a,x,h} = y) \to 0 \) for \( y \in S_{a,x} \setminus \{x\} \), as \( h \to 0 \). However, these assumptions remain quite general and do not allow some investigations as, for example, providing an explicit expression of optimal bandwidth or specifying the convergence of a kernel \( K_{a,x,h} \) to \( x \). Then, to further study the choice of optimal bandwidth, some expansions of modal probability and variance of kernel \( K_{a,x,h} \) are provided in what follows. For \( h \) sufficiently small, we have

\[
\Pr(K_{a,x,h} = x) \simeq \left[ (2a + 1) - 2[1 - h \log(a + 1)]\{a + h \sum_{k=1}^{a} \log(k)\} \right]^{-1} = 1 - 2h \left\{ a \log(a + 1) - \sum_{k=1}^{a} \log(k) \right\} + O(h^2) = 1 - 2hA(a) + O(h^2) \tag{3}
\]

and

\[
\text{Var}(K_{a,x,h}) = 2h \left( \frac{a(2a^2 + 3a + 1)}{6} \log(a + 1) - \sum_{k=1}^{a} k^2 \log(k) \right) + O(h^2) = 2hV(a) + O(h^2). \tag{4}
\]
The asymptotic behaviours pointed out in Equations (3) and (4) allow the discrete symmetric triangular associated kernel to tend to the Dirac-type kernel $D_{x,h} \equiv D_{x}$ of r.v. $D_{x,h}$ given by

$$D_{x,h}(y) = 1_{y=x} = \begin{cases} 1 & \text{if } y = x \\ 0 & \text{if } y \neq x \end{cases}$$

for any $x \in \mathbb{N}$ and any $h \geq 0$,

such that $S_x = \{x\}$, $\text{Pr}(D_{x,h} = x) = 1$ and $\text{Var}(D_{x,h}) = 0$. In fact, the expansions in Equations (3) and (4) are useful attempts to provide specific behaviours of discrete triangular kernel less general than $H_1$–$H_2$.

The second example concerns the discrete associated kernel deduced from [14] such that

$$K_{c,x,h}(y) = (1-h)1_{y=x} + \frac{h}{c-1}1_{y \neq x} \text{ for all } y \in S_x = \{0, 1, \ldots, c-1\} \text{ and } h \in (0, 1]$$

has also modal probability and variance which can be expressed as in Equations (3) and (4) with $A(c; x, h) = \frac{1}{2}$ and $V(c; x, h) = \{c(6x^2 + 2c^2 - 3c + 1 - 6xe + 6x)\}{12(c - 1)}$.

The third example concerns a discrete kernel proposed by Wang and Van Ryzin [15] such that

$$K_{x,h}(z) = (1-h)1_{z=x} + \frac{1}{2}(1-h)\frac{h}{|z-x|}1_{|z-x| \geq 1} \text{ for all } z \in S = \mathbb{Z}.$$

The modal probability and variance of this kernel can be expressed as in Equations (3) and (4) with $A \equiv \frac{1}{2}$ and $V \equiv (1 + 3h)/2$.

Lastly, the expansions in Equations (3) and (4) are not available for standard asymmetric discrete kernel constructed from usual p.m.f. (binomial, Poisson or negative binomial) which satisfy only assumption $H_1$ and have variance such that $\text{Var}(K_{x,h}) \in \mathcal{V}(0)$ instead of $H_2$, where $\mathcal{V}(0)$ is neighbourhood of 0 [16]. Thus, we are not interested in this work with these kernels since we cannot express them in an explicit optimal bandwidth parameter. However, it stays interesting to use the binomial kernel (which outperforms the other standard kernels) for small or moderate sample in comparison with discrete triangular kernels; but discrete standard kernels do not tend asymptotically to the Dirac kernel, i.e. $\text{Pr}(K_{x,h} = x) \rightarrow 1$ as $h \rightarrow 0$, in contrast with discrete triangular kernels.

**Remark 2.1** It would be of interest to define discrete associated kernel such that

$$H_1' : \text{Pr}(K_{x,h} = x) = 1 - hA(K_{x,h}) + O(h^2) \text{ and } H_2' : \text{Var}(K_{x,h}) = hV(K_{x,h}) + O(h^2)$$

with $\sum_{y \in S_x \setminus \{x\}} \text{Pr}(K_{x,h} = y) = hA(K_{x,h}) + O(h^2) \rightarrow 0$ as $h \rightarrow 0$; but, this is not the subject of the current work.

Next, we study a first data-driven bandwidth selection procedure using expansions in Equations (3) and (4).

### 2.2. Global-squared error

Let us assume $p_0(x) = p(x; \Theta_0)$ be a fixed p.m.f. of Poisson start in Equation (1), i.e. $f = p_0 \omega$, such that $\hat{\Theta}$ converges to $\Theta_0$ in probability. We first formulate the following result on the pointwise consistency of $\hat{f}_n$; the proof is given in the appendix.

**Theorem 2.1** Under assumption $H_1$–$H_2$, for any $x \in \mathbb{N}$, one has

$$\lim_{n \rightarrow \infty} \mathbb{E}(\hat{f}_n(x) - f(x))^2 = 0.$$
To express an explicit expression of optimal bandwidth requires the use of the global error of \( \hat{f}_n \) defined as

\[
\text{MISE}[\hat{f}_n(x)] = \mathbb{E} \sum_{x \in \mathbb{N}} (\hat{f}_n(x) - f(x))^2 = \sum_{x \in \mathbb{N}} \text{Var}[\hat{f}_n(x)] + \sum_{x \in \mathbb{N}} \text{Bias}^2[\hat{f}_n(x)]. \tag{5}
\]

By considering the expressions in Equations (3) and (4), we have the following expansions of pointwise bias and variance of estimator \( \hat{f}_n \) given by

\[
\text{Bias}[\hat{f}_{n,K_{a,x}}(x)] = \frac{h}{2} \mathbb{V}(K_{a,x,h}) p_0(x) \omega(2) + o(h) + O(h^2),
\]

\[
\text{Var}[\hat{f}_{n,K_{a,x}}(x)] = \frac{1}{n} f(x) [1 - h A(K_{a,x,h})]^2 - \frac{1}{n} f^2(x) + R_n(a; x, h) + O(h^2),
\]

where \( \omega(2) \) is the finite difference of second order of \( \omega \) such that

\[
\omega(2)(x) = \begin{cases} 
\frac{[\omega(x+2) - 2\omega(x) + \omega(x-2)]}{4} & \text{if } x \in \mathbb{N} \setminus \{0, 1\}, \\
\frac{[\omega(3) - 3\omega(1) + 2\omega(0)]}{4} & \text{if } x = 1, \\
\frac{[\omega(2) - 2\omega(1) + \omega(0)]}{2} & \text{if } x = 0 
\end{cases}
\]

(see also [16]); and, the term

\[
R_n(a; x, h) = \frac{1}{n} \sum_{y \in \mathbb{S}_x \setminus \{x\}} f(y) K_{a,x,y}^2 + \frac{1}{n} f(x)^2 - \frac{1}{n} \left[ f(x) + \sum_{y \in \mathbb{S}_x} (f(y) - f(x)) K_{a,x,y} \right]^2
\]

\[
(\omega) \in O(1/n) \text{ since } K_{a,x,y}(x) \to 0, \text{ for } y \in \mathbb{S}_x \setminus \{x\}, \text{ as } h \to 0. \text{ Hence, we have}
\]

\[
\text{MISE}[\hat{f}_{n,K_{a,x}}(x)] = \frac{1}{n} \left[ 1 - h A(K_{a,x,h}) \right]^2 - \frac{1}{n} f^2(x) + \frac{h^2}{4} \mathbb{V}^2(K_{a,x,h}) \sum_{x \in \mathbb{N}} \left( (p_0 \omega(2))(x) \right)^2
\]

\[
+ o \left( \frac{1}{n} \right) + O(h^2), \tag{6}
\]

where the leading term is an approximate MISE denoted \( \text{AMISE}(\hat{f}_{n,K_{a,x}}(x)) \equiv \text{AMISE}(a; n, h, f) \), with \( \sum_{x \in \mathbb{N}} (p_0 \omega(2))(x))^2 < \infty \). It ensues the following proposition on global consistency of \( \hat{f}_{n,K_{a,x}} \).

**PROPOSITION 2.2** For \((a, x) \in \mathbb{N} \times \mathbb{N} \text{ and } h > 0\), let us consider the semiparametric estimator \( \hat{f}_n \) using discrete symmetric triangular kernel \( K_{a,x,h} \). One has

\[
\lim_{n \to \infty} \mathbb{E} \sum_{x \in \mathbb{N}} (\hat{f}_{n,K_{a,x}}(x) - f(x))^2 = 0.
\]

Then, the bandwidth \( \hat{h}_{opt} = \arg \min_{h>0} \text{AMISE}(a; n, h, f) \) comes by solving the equation \( d\{\text{AMISE}(a; n, h, f)\}/dh = 0 \) which is equivalent to

\[
\frac{h}{2} \mathbb{V}^2(K_{a,x,h}) \sum_{x \in \mathbb{N}} \{ (p_0 \omega(2))(x) \}^2 - \frac{2}{n} A(K_{a,x,h}) \{ 1 - h A(K_{a,x,h}) \} = 0.
\]

It results in

\[
\hat{h}_{opt}(a; n, f) = \frac{A(a)}{2A^2(a) + (n/2) \mathbb{V}^2(a) \sum_{x \in \mathbb{N}} [p_0(x) \omega(2)(x)]^2} \to 0 \text{ when } n \to \infty. \tag{7}
\]
Moreover, it comes the following asymptotic relationship: \( \hat{h}_{\text{opt}} \sim k_0n^{-1} \) with
\[
k_0 = \frac{4A(a)}{V^2(a) \sum_{x \in \mathbb{N}} \{(p_0\omega^{(2)}(x))\}^2}.
\]

A comparison can be realized with the discrete nonparametric kernel estimator \( \tilde{f}_n \) of p.m.f. \( f \) proposed by Kokonendji and Kiessé [16] such that
\[
\tilde{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K_{x,h}(X_i), \quad x \in \mathbb{N}.
\]

The expression of \( \tilde{f}_n \) can be deduced from that of \( \hat{f}_n \) in Equation (2) by assuming \( p \equiv 1 \) in Equation (1) thus \( \hat{p} \equiv 1 \) in Equation (2). Thus, the bias of \( \tilde{f}_n \) is given by 
\[
\text{Bias}(\tilde{f}_n(x)) = \frac{(h/2)f^{(2)}(x)V(K_{a,x,h}) + o(h) + O(h^2)}{n}, \quad x \in \mathbb{N},
\]
while its variance is identical to that of semi-parametric estimator \( \hat{f}_n \). Hence, by calculating MISE of \( \tilde{f}_n \) as in Equation (5), the optimal bandwidth parameter minimizing the corresponding AMISE is such that
\[
\tilde{h}_{\text{opt}}(a; n, f) = \frac{A(a)}{2A^2(a) + (n/2)V^2(a) \sum_{x \in \mathbb{N}} f^{(2)}(x)^2} \to 0 \quad \text{when} \quad n \to \infty,
\]
with \( 0 < \sum_{x \in \mathbb{N}} f^{(2)}(x)^2 < \infty \).

Finally, the comparison between \( \hat{h}_{\text{opt}} \) in Equation (7) and \( \tilde{h}_{\text{opt}} \) in Equation (9) depends on the finite differences \( f^{(2)} \) and \( p_0\omega^{(2)} \) since \( f^{(2)} - p_0\omega^{(2)} \) can be either positive or negative with \( f^{(2)} = p_0^{(2)}\omega + 2p_0^{(1)}\omega^{(1)}(x) + p_0\omega^{(2)} \). Note that the comparison between estimators \( \hat{f}_n \) and \( \tilde{f}_n \), having the same variance, is related to their respective bias for which the leading terms of their difference given by
\[
\frac{1}{2} [(p_0^{(2)}\omega)(x) + 2(p_0^{(1)}\omega^{(1)}(x))V(K_{a,x,h})]
\]
can also be either positive or negative depending on start parametric function \( p_0 \).

### 2.3. Cross-validation function

A second bandwidth parameter value can be obtained by applying the cross-validation procedure which consists in the minimization of the score function cross-validation (CV), i.e. \( \hat{h}_{cv} = \arg \min_{h > 0} \text{CV}(h) \), such that
\[
\text{CV}(h) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{p(X_i; \Theta_0)p(X_j; \Theta_0)} \sum_{x \in \mathbb{N}} p^2(x; \Theta_0)K_{x,h}(X_i)K_{x,h}(X_j)
\]
\[- \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq 1} K_{x,h}(X_j) \frac{p(X_i; \Theta_{0,-i})}{p(X_i; \Theta_{0,-i})}
\]
\[= A_n - 2B_n
\]
\[= \sum_{x \in \mathbb{N}} \sum_{i=1}^{n} \left( \frac{p(x; \Theta_0)}{p(X_i; \Theta_0)} K_{x,h}(X_i) \right)^2 + \frac{2}{n^2} \sum_{j < i} H_{ij}
\]
with
\[
H_{ij} = \frac{1}{p(X_i; \Theta_0)p(X_j; \Theta_0)} \sum_{x \in \mathbb{N}} p^2(x; \Theta_0)K_{x,h}(X_i)K_{x,h}(X_j) - 2K_{x,h}(X_i) \frac{p(X_i; \Theta_{0,-i})}{p(X_i; \Theta_{0,-i})}.
\]
A similar procedure is presented in [16] for a bandwidth ˜\(\theta\) depending on \(h\) in the following expression of MISE in Equation (5):

\[
\text{MISE}\{\hat{f}_n(x)\} = \mathbb{E} \left( \sum_{x \in \mathcal{N}} \hat{f}_n^2(x) \right) - 2\mathbb{E} \left( \sum_{x \in \mathcal{N}} \hat{f}_n(x)f(x) \right) + \mathbb{E} \left( \sum_{x \in \mathcal{N}} f^2(x) \right).
\]

The expression \(A_n\) is an unbiased estimator of the first term \(\mathbb{E}\{\sum_{x \in \mathcal{N}} \hat{f}_n^2(x)\}\) since

\[
\sum_{x \in \mathcal{N}} \hat{f}_n^2(x) = \sum_{x \in \mathcal{N}} \left\{ p(x; \theta_0) \frac{1}{n} \sum_{i=1}^n \frac{K_{x,h}(X_i)}{p(X_i; \theta_0)} \right\}^2 = \frac{1}{n^2} \sum_{x \in \mathcal{N}} p^2(x; \theta_0) \left\{ \sum_{i=1}^n \frac{K_{x,h}(X_i)}{p(X_i; \theta_0)} \right\}^2 = A_n.
\]

Then, the expression \(B_n\) is an estimator of the second term \(\mathbb{E}\{\sum_{x \in \mathcal{N}} \hat{f}_n(x)f(x)\}\) since we first have

\[
\mathbb{E}(B_n) = \mathbb{E} \left\{ \frac{1}{n-1} \sum_{j \neq 1} K_{X_1,h}(X_j) \frac{p(X_1; \theta_{0,-1})}{p(X_j; \theta_{0,-1})} \right\} = \mathbb{E} \left\{ K_{X_1,h}(X_2) \frac{p(X_1; \theta_0)}{p(X_2; \theta_0)} \right\}
\]

and, then,

\[
\mathbb{E} \left\{ \sum_{x \in \mathcal{N}} \hat{f}_n(x)f(x) \right\} = \mathbb{E} \left\{ \frac{1}{n} \sum_{i=1}^n K_{X_i,h}(X_i) \frac{p(X_1; \theta_0)}{p(X_i; \theta_0)} \right\} = \mathbb{E} \left\{ K_{X_1,h}(X_2) \frac{p(X_1; \theta_0)}{p(X_2; \theta_0)} \right\}.
\]

A similar procedure is presented in [16] for a bandwidth \(\bar{h}_{cv}\) of nonparametric estimator \(\hat{f}_n\) of p.m.f. \(f\) in Equation (8). The CV’s mean and variance for fixed \(h > 0\) are provided in the following theorem under assumptions \(H_1\)–\(H_2\).

**Theorem 2.3** Let \(x\) be a given point in \(\mathbb{N}\) and \(h = h(n) > 0\) be the bandwidth such that \(\lim_{n \to \infty} h = 0\). Consider the cross-validation function CV for semiparametric estimator in Equation (2) of p.m.f. \(f\). Then, the mean and variance of CV admit the following expansions:

\[
\mathbb{E}\{\text{CV}(h)\} = \text{AMISE} - \sum_{x \in \mathcal{N}} f^2(x) + o(h^2) + O\left(\frac{1}{n}\right)
\]

and

\[
\text{Var}\{\text{CV}(h)\} = \frac{1}{n} \sum_{x \in \mathcal{N}} \left\{ \sum_{x \in \mathcal{N}} \sum_{x \in \mathcal{N}} p_0^2(x)K_{x,h}(x) - 2K_{x,h}(x) \right\}^2 - \frac{1}{n} \left\{ \sum_{x \in \mathcal{N}} f^2(x) \right\}^2 + O\left(\frac{1}{n^2}\right),
\]

where \(K_{x,h}\) is a discrete associated kernel satisfying \(H_1\)–\(H_2\).

**Proof** See the appendix. \(\blacksquare\)

**Remark 2.2**

(i) Similar to Theorem 2.3, the bias and variance of cross-validation criterion can be calculated for nonparametric estimator \(\hat{f}_n\) in Equation (8). It results in the same expression for \(\mathbb{E}\{\text{CV}(h)\}\) while the difference comes from their variance since in nonparametric case the score function CV admits the following variance:

\[
\text{Var}\{\text{CV}(h)\} = \frac{1}{n} \sum_{x \in \mathcal{N}} \left\{ \sum_{x \in \mathcal{N}} K_{x,h}^2(x) - 2K_{x,h}(x) \right\}^2 - \frac{1}{n} \left\{ \sum_{x \in \mathcal{N}} f^2(x) \right\}^2 + O\left(\frac{1}{n^2}\right).
\]
(ii) It would be of interest to compare $\text{MISE}(\hat{f}_{n,h_{cv}})$ and $\text{MISE}(\tilde{f}_{n,h_{opt}})$. To derive such a result, for some sufficiently large region $H_n$, some information about $\sup_{H_n} |\text{CV}(h) + \sum_{x \in \mathbb{N}} f^2(x) - \text{MISE}(h)|$ would be necessary. That will be the subject of a forthcoming article.

### 2.4. About choice of parameter ‘a’ for discrete triangular kernel

Looking at expression of $\text{MISE}$ in Equation (6) of semiparametric estimator $\hat{f}_n$ using discrete triangular kernel, we are not able to calculate theoretically the optimal parameter $a \in \mathbb{N}$ minimizing $\text{MISE}(\hat{f}_{n,K_a,h}(x))$. Hence, Figure 1 illustrates the comparative results of function $a \mapsto \text{MISE}(a)$ of $\hat{f}_n$ with discrete triangular kernel for the simulated p.m.f.

$$f(x) = \frac{2}{5} \exp(-0.5) \frac{0.5^x}{x!} + \frac{3}{5} \exp(-10) \frac{10^x}{x!}, \quad x \in \mathbb{N},$$

which is a mixture of two Poisson distributions $\mathcal{P}(\mu)$ with respective means $\mu_1 = 0.5$ and $\mu_2 = 10$. For $h \in \{0.1, 0.5, 1\}$, the optimal value $a_{opt} = \arg \min_{a \in \mathbb{N}} \text{MISE}(a)$ is small and equal to 1, 2 or 3, while for $h = 0.01$ we have $a_{opt} \geq 5$. It appears that the optimal $a_{opt}$ decreases as the sample size $n \in \{50, 100, 150, 200\}$ increases. Note that the case $a = 0$ for discrete triangular kernel results in kernel of Dirac type. In addition, Figure 2 presents the comparative function $a \mapsto \hat{h}_{opt}(a)$ in Equation (7). For fixed sample size $n \in \{50, 100, 150, 200\}$, the optimal $h$-value tends to 0 as $a \in \mathbb{N}$ is increasing. Finally, we propose to consider $a \in \{1, 2, \ldots, 5\}$ for applications in the following section.
Figure 2. Simulated function $a \mapsto \hat{h}_{\text{opt}}(a)$ of semiparametric estimator using discrete triangular kernel for $f = 0.4P(0.5) + 0.6P(10)$ with $\sum_{x \in \mathbb{N}} f^2(x) = 0.0049$.

This procedure is proposed for choosing the parameter $a \in \mathbb{N}$ since we are not able to calculate an explicit expression of this parameter minimizing global-squared error MISE in Equation (6) as $\hat{h}_{\text{opt}}$ in Equation (7) for bandwidth parameter $h > 0$. It would be interesting to propose the general rule for the choice of $a \in \mathbb{N}$ for future works.

3. Applications

In this section, we present the results of the applications of the semiparametric estimator $\hat{f}_n \equiv \hat{f}_{n,K_{a,h}}$ using discrete triangular kernels on three real count data-sets. The performance of $\hat{f}_{n,K_{a,h}}$ with respect to each optimal bandwidth parameter $\hat{h}_{\text{cv}}$ and $\hat{h}_{\text{opt}}$ is evaluated by using the following descriptive measure of degree-of-fit:

$$\text{ISE}(h) = \sum_{x \in \mathbb{N}} (\hat{f}_{n,K_{a,h}}(x) - f_0(x))^2,$$

(integrated squared error) with $f_0$ being the empirical frequency estimate of observations. Concerning discrete triangular associated kernels, we propose the values of the integer parameter $a \in \{1, 2, \ldots, 5\}$.

3.1. Data of alcohol consumption

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. For $a = 1$, the use
Table 1. ISE ($\times 10^3$) calculated according to each type of optimal bandwidth parameters for semiparametric estimates on data of alcohol consumption.

| Parameter $a$ | $h_{cv}$ | ISE($h_{cv}$) | $h_{opt}$ | ISE($h_{opt}$) |
|---------------|----------|---------------|-----------|----------------|
| $a = 1$       | 0.02     | 0.07          | 0.08      | 1.008          |
| $a = 2$       | 0.009    | 0.07          | 0.01      | 0.141          |
| $a = 3$       | 0.005    | 0.06          | 0.003     | 0.028          |
| $a = 4$       | 0.003    | 0.05          | 0.001     | 0.007          |
| $a = 5$       | 0.002    | 0.04          | 0.0004    | 0.002          |

Table 2. Data of preimarginal development time observed on days for several species of fruit trees.

| Fruit Tree | Observations | Frequencies |
|------------|--------------|-------------|
| Safou      | 30 31 32     | 28 21 11    |
| Hura       | 25 26 27 28 29 30 31 32 33 34 35 | 5 5 7 8 11 2 1 4 4 2 2 |
| Citrus     | 25 26 27 28 29 30 31 32 33 34 35 | 10 7 7 8 10 18 4 3 8 6 4 5 |
| Mango      | 31 32 33 34 35 36 37 38       | 11 24 12 9 6 8 1 5 |
| Avocado    | 27 28 29 30 31 32 33 34       | 13 15 3 6 4 12 2 3 |

of bandwidth $\hat{h}_{cv}$ is clearly better (in term of ISE) than that of $\hat{h}_{opt}$; and, for $a \in \{2, 3, 4, 5\}$ the use of $\hat{h}_{opt}$ is similar or better than that of $\hat{h}_{cv}$ (Table 1). Both $h$-values $\hat{h}_{cv}$ and $\hat{h}_{opt}$ tend to 0 as the parameter $a \in \mathbb{N}$ is increasing. Thus, in this example, using the semiparametric estimator $\hat{f}_{n,K,a,h}$ with $h = \hat{h}_{cv}$ is appropriated for a small value of parameter $a \in \mathbb{N}$, while using $\hat{f}_{n,K,a,h}$ with $h = \hat{h}_{opt}$ becomes interesting or more appropriated for large values of this parameter.

3.2. Data of insect growth

Some experimental plantations were realized on several host plants among them some fruit trees well known in Congo safou (Dacryodes edulis), hura (Hura crepitans), mango (Mangifera indica), citrus (Citrus paradisi) and avocado (Persea americana). These plantations were realized with some young trees being 5–6 months old under some conditions of temperature and humidity, and the observations were done using a binocular loupe. The development of the insect parasite studied is described by several count explanatory variables observed on days such as the preimarginal development time from egg to adult stage (Table 2).

Similar to the previous example about alcohol consumption, the semiparametric estimator $\hat{f}_{n,K,a,h}$ with $h = \hat{h}_{cv}$ is appropriated for parameter $a = 1$ while using $\hat{f}_{n,K,a,h}$ with $h_{opt}$ is generally more appropriated for $a \in \{2, 3, 4, 5\}$. There is an exception for safou fruit tree probably due to the small numbers of observed points $x$. Note that for the hura fruit tree, the cross-validation procedure does not converge (a well-known problem for this method) when $a = 1$ while an optimal value $\hat{h}_{opt}$ is available.
Table 3. ISE calculated according to each type of optimal bandwidth parameters for semiparametric estimates on data of preimarginal development time.

| Parameter a | Safou | Hura | Citrus | Mango | Avocado |
|-------------|-------|------|--------|-------|---------|
| h_{opt}     | 0.71  | 0.61 | 0.55   | 0.38  | 0.48    |
| ISE(h_{opt})| 0.0076| 0.0042| 0.0053 | 0.0039| 0.0099  |
| h_{cv}      | 0.16  | –    | 0.38   | 0.33  | 0.18    |
| ISE(h_{cv}) | 0.0010| –    | 0.0035 | 0.0032| 0.0025  |
| a = 2       |       |      |        |       |         |
| h_{opt}     | 0.32  | 0.21 | 0.17   | 0.09  | 0.15    |
| ISE(h_{opt})| 0.0065| 0.0036| 0.0037 | 0.0014| 0.0055  |
| h_{cv}      | 0.05  | 0.68 | 0.15   | 0.13  | 0.08    |
| ISE(h_{cv}) | 0.0004| 0.0105| 0.0032 | 0.0026| 0.0022  |
| a = 3       |       |      |        |       |         |
| h_{opt}     | 0.20  | 0.09 | 0.06   | 0.03  | 0.05    |
| ISE(h_{opt})| 0.0053| 0.0021| 0.0015 | 0.0004| 0.0022  |
| h_{cv}      | 0.02  | 0.37 | 0.10   | 0.07  | 0.05    |
| ISE(h_{cv}) | 0.0001| 0.0099| 0.0032 | 0.0020| 0.0022  |
| a = 4       |       |      |        |       |         |
| h_{opt}     | 0.13  | 0.04 | 0.02   | 0.01  | 0.02    |
| ISE(h_{opt})| 0.0042| 0.0009| 0.0005 | 0.0001| 0.0007  |
| h_{cv}      | 0.013 | 0.26 | 0.08   | 0.05  | 0.04    |
| ISE(h_{cv}) | 0.0007×10^{-1}| 0.0096| 0.0032 | 0.0019| 0.0023  |
| a = 5       |       |      |        |       |         |
| h_{opt}     | 0.09  | 0.02 | 0.01   | 0.004 | 0.009   |
| ISE(h_{opt})| 0.0031| 0.0004| 0.0002 | 0.0003×10^{-1} | 0.0002 |
| h_{cv}      | 0.009 | 0.17 | 0.06   | 0.03  | 0.03    |
| ISE(h_{cv}) | 0.0005×10^{-1}| 0.0083| 0.0029 | 0.0011| 0.0019  |

3.2.1. Bootstrap method

We pursue our study with a more robust evaluation which consists in resampling the observations of citrus tree species used, for example. From the data of this tree species, we draw \( n \in \{25, 50, 75\} \) bootstrap samples on which we apply our estimator and the methods for optimal choice of the bandwidth parameter. It ensues the calculation of the averages \( \text{ISE}, \hat{h}_{\text{opt}} \) and \( \hat{h}_{\text{cv}} \) of ISE and optimal bandwidth parameters, respectively, such that we have \( \text{ISE} = (1/n) \sum_{i=1}^{n} \text{ISE} \) or \( \hat{h} = (1/n) \sum_{i=1}^{n} \hat{h} \). In Table 4, the results confirm the first conclusions formulated previously: semiparametric estimator \( \hat{f}_n \) with optimal \( \hat{h}_{\text{cv}} \)-value provides better estimations for \( a \in \{1, 2\} \) while using \( \hat{f}_n \) with \( \hat{h}_{\text{opt}} \)-value is better for \( a \in \{3, 4, 5\} \). Note that the average \( \text{ISE} \) is an approximate of global-squared error \( \text{MISE} \) in Equation (5) since we have \( \text{MISE} = E(\text{ISE}) \).

3.3. Wood cell count data

3.3.1. Data

We illustrated our investigations using count data from time series relative to the wood formation of silver fir, one of the widest spread European conifer species. Wood derives from the cambium, a tissue consisting of a thin layer of cells able to divide between the wood and the bark in tree
Table 4. Means of ISE and optimal bandwidth parameters for semi-parametric estimates on data of preimarginal development time of citrus tree species using the bootstrap method.

| Bootstrap samples \( n \) | \( h_{cv} \) | ISE\((h_{cv})\) | \( h_{opt} \) | ISE\((h_{opt})\) |
|---------------------------|-------------|----------------|-------------|----------------|
| \( a = 1 \)              |             |                |             |                |
| \( n = 25 \)             | 0.349       | 0.0032         | 0.567       | 0.0069         |
| \( n = 50 \)             | 0.315       | 0.0029         | 0.569       | 0.0065         |
| \( n = 75 \)             | 0.310       | 0.0028         | 0.577       | 0.0073         |
| \( a = 2 \)              |             |                |             |                |
| \( n = 25 \)             | 0.155       | 0.0031         | 0.187       | 0.0046         |
| \( n = 50 \)             | 0.155       | 0.0030         | 0.192       | 0.0044         |
| \( n = 75 \)             | 0.154       | 0.0030         | 0.185       | 0.0041         |
| \( a = 3 \)              |             |                |             |                |
| \( n = 25 \)             | 0.106       | 0.00316        | 0.082       | 0.0026         |
| \( n = 50 \)             | 0.104       | 0.00308        | 0.076       | 0.0023         |
| \( n = 75 \)             | 0.096       | 0.00296        | 0.074       | 0.0022         |
| \( a = 4 \)              |             |                |             |                |
| \( n = 25 \)             | 0.062       | 0.00267        | 0.033       | 0.0010         |
| \( n = 50 \)             | 0.067       | 0.00277        | 0.034       | 0.0010         |
| \( n = 75 \)             | 0.068       | 0.00279        | 0.034       | 0.0011         |
| \( a = 5 \)              |             |                |             |                |
| \( n = 25 \)             | 0.051       | 0.00260        | 0.016       | 0.00045        |
| \( n = 50 \)             | 0.052       | 0.00262        | 0.016       | 0.00043        |
| \( n = 75 \)             | 0.056       | 0.00277        | 0.015       | 0.00040        |

axes. Divisions of the cambial cells give birth to the new wood cells in the centripetal direction. In conifers, a newly produced wood cell (called tracheid) undergoes two differentiation phases: (1) first, its radial diameter increases, it is in the cell-diameter enlargement phase; (2) second, cell-wall thickening and lignification begin, it enters the secondary cell-wall formation phase [17,18]. Once the differentiation process is complete, programmed cell death takes place, giving a mature and functional cell of the wood, i.e. a dead slender tracheid able to transport water (in trees, water is transported from the root to the leaves in the wood) and confers mechanical support to the stem. Under temperate climate conditions, wood formation presents an annual pattern: it is active during the hot season and inactive during the cold season. All the tracheids produced during a year are organized as juxtaposed radial files and form an annual tree-ring which adds to the tree-rings formed in the previous years (Figure 3). Here, we use the number of cells weekly counted, from April to November, in the cell-diameter enlargement phase of the wood formation along the radial files of the forming tree-ring of two silver fir trees: the silver fir tree 1 in 2007 and the silver fir tree 2 in 2008 (Table 5).

3.3.2. Results

The two cell count data-sets presented in Table 5 are parametrically estimated using the two following Poisson distributions. First, the unimodal Poisson p.m.f. \( P(x; \Theta_0) \) with mean \( \Theta_0 \) is used for the number of wood cells \( n_{1,j} \) in the diameter enlargement phase of the silver fir tree 1. Second, the bimodal Poisson p.m.f. \( 0.35P(x; \Theta_0^1) + 0.65P(x; \Theta_0^2) \) is used for the number of wood cells \( n_{2,j} \) in the same phase of the silver fir tree 2. Indeed, the count data distribution \( n_2 \equiv n_{2,j} \) can be considered as a mixture of two data-sets (with respective mean \( \Theta_0^j, j = 1, 2 \)) which consist of 35% and 65% of the total number of size. The unimodal and bimodal Poisson p.m.f. are used as
some starts in the semiparametric estimator $\hat{f}_n$ in Equation (2). Finally, note that we propose to add here a pseudo true optimal bandwidth computed numerically as $h_{opt}^0 = \arg \min_{h > 0} \text{MISE}(h)$ by replacing the unknown distribution $f$ by $f_0$ in Equation (5); the same replacement is also did in the expression $\hat{h}_{opt}$ in Equation (7). All the numeric results are presented in Tables 6 and 7 with $a \in \{1, 2\}$.

In general, the best estimations in term of ISE are provided by using the optimal bandwidth $h_{opt}^0$. Moreover, the value $h_{opt}^0$ decreases when the value of parameter $a \in \{1, 2\}$ increases. For the unimodal data-set related to the silver fir tree 1, the bandwidth $h_{opt}^0$ is better approximated by $\hat{h}_{cv}$ than $\hat{h}_{opt}$ using $a = 1$ (Table 6). For the bimodal data-set related to the silver fir tree 2, the bandwidth value $h_{opt}^0$ is better approximated by $\hat{h}_{opt}$ than $\hat{h}_{cv}$ using $a \in \{1, 2\}$ (Table 7). For the two wood cell count data-sets, the values of the three types of optimal bandwidth become closer when the parameter $a \in \{1, 2\}$ increases. Thus, one can see that the performance of each bandwidth $\hat{h}_{opt}$ and $\hat{h}_{cv}$ s also depends on the start functions in these examples.
Table 6. ISE ($\times 10^3$) calculated according to each type of optimal bandwidth parameters for semiparametric estimates on the unimodal distribution of the number of wood cells ($\times 10$) counted in the diameter enlargement phase of the silver fir 1 during 2007.

| $a = 1$ | $\hat{h}_{opt}$ | $h_{0 opt}$ | $\hat{h}_{cv}$ | ISE($\hat{h}_{opt}$) | ISE($h_{0 opt}$) | ISE($\hat{h}_{cv}$) |
|---------|----------------|-------------|---------------|---------------------|-----------------|-----------------|
| 0.49    | 0.15           | 0.23        | 1.284         |                     | 0.264           | 0.501           |

| $a = 2$ | $\hat{h}_{opt}$ | $h_{0 opt}$ | $\hat{h}_{cv}$ | ISE($\hat{h}_{opt}$) | ISE($h_{0 opt}$) | ISE($\hat{h}_{cv}$) |
|---------|----------------|-------------|---------------|---------------------|-----------------|-----------------|
| 0.13    | 0.09           | 0.13        | 0.591         |                     | 0.348           | 0.591           |

Table 7. ISE ($\times 10^3$) calculated according to each type of optimal bandwidth parameters for semiparametric estimates on the bimodal distribution of the number of wood cells ($\times 10$) counted in the diameter enlargement phase of the silver fir 2 during 2008.

| $a = 1$ | $\hat{h}_{opt}$ | $h_{0 opt}$ | $\hat{h}_{cv}$ | ISE($\hat{h}_{opt}$) | ISE($h_{0 opt}$) | ISE($\hat{h}_{cv}$) |
|---------|----------------|-------------|---------------|---------------------|-----------------|-----------------|
| 0.58    | 0.32           | 1.00        | 0.548         |                     | 0.278           | 0.865           |

| $a = 2$ | $\hat{h}_{opt}$ | $h_{0 opt}$ | $\hat{h}_{cv}$ | ISE($\hat{h}_{opt}$) | ISE($h_{0 opt}$) | ISE($\hat{h}_{cv}$) |
|---------|----------------|-------------|---------------|---------------------|-----------------|-----------------|
| 0.19    | 0.15           | 0.34        | 0.422         |                     | 0.312           | 0.788           |

4. Concluding remarks

In this work, an expression of optimal bandwidth $\hat{h}_{opt}$, unavailable until now in discrete case, is developed for the semiparametric (and nonparametric) kernel estimator with discrete triangular associated kernels. Therefore, an explicit optimal bandwidth is provided like that which is available for continuous kernel density estimation. The expression of $\hat{h}_{opt}$ proposed depends both on parameter $a \in \mathbb{N}$ of discrete triangular kernel and sample size $n$; of course, this new optimal bandwidth goes to 0 when sample size $n$ goes to $\infty$. The performance of this new optimal bandwidth is comparable to optimal bandwidth $\hat{h}_{cv}$ provided by applying the cross-validation procedure when the parameter $a \in \mathbb{N}$ increases; in particular, concerning the examples studied in our work, $\hat{h}_{opt}$
clearly outperforms $\hat{h}_{cv}$ for $a \geq 3$. However, this first attempt for obtaining an explicit optimal bandwidth for estimator using discrete triangular kernel cannot be generalized to other discrete kernels as, in particular, those which have been constructed from usual discrete distributions as binomial or Poisson. Thus, an interesting perspective would be to find a general rule to express of optimal bandwidth for discrete kernel estimation as it is available for continuous kernel one. Finally, following the idea developed in the current paper, some works are in progress to express an optimal bandwidth for discrete nonparametric or semiparametric triangular kernel estimators for the count regression function.

Acknowledgements

The authors are grateful to an anonymous referee whose comments improved this paper.

References

[1] Kokonendji CC, Mizère D, Balakrishnan N. Connection of the Poisson weight function to overdispersion and underdispersion. J. Statist. Plann. Inference. 2008;138:1287–1296.
[2] Kokonendji CC, Senga Kiessé T, Balakrishnan N. Semiparametric estimation for count data through weighted distributions. J. Statist. Plann. Inference. 2009;139:3625–3638.
[3] Hjort NL, Glad IK. Nonparametric density estimation with a parametric start. Ann. Statist. 1995;23:882–904.
[4] Kokonendji CC, Senga Kiessé T, Zocchi SS. Discrete triangular distributions and non-parametric estimation for probability mass function. J. Nonparametr. Stat. 2007;19:241–257.
[5] Bowman A. An alternative method of cross-validation for the smoothing of density estimates. Biometrika. 1984;71:352–360.
[6] Marron JS. A comparison of cross-validation techniques in density estimation. Ann. Statist. 1987;15:152–162.
[7] Simonoff JS, Tutz G. Smoothing methods for discrete data. In: Schimek MG, editor. Smoothing and regression: approaches, computation, and application. New York: Wiley; 2000. pp. 193–228.
[8] Zougab N, Adjabi S, Kokonendji CC. Binomial kernel and Bayes local bandwidth in discrete function estimation. J. Nonparametr. Stat. 2012;24:783–795.
[9] Zougab N, Adjabi S, Kokonendji CC. Adaptive smoothing in associated kernel discrete functions estimation using Bayesian approach. J. Stat. Comput. Simul. 2012. DOI:10.1080/00949655.2012.686615.
[10] Abdous B, Kokonendji CC, Senga Kiessé T. On semiparametric regression for count explanatory variables. J. Statist. Plann. Inference 2012;142:1537–1548.
[11] Mizère D, Kissita G, Kiyindou A. Etude statistique de l’influence de la plante hôte sur le développement de l’aleurode Aleurodicus dispersus Russel via le modèle de Poisson translaté. Annales de l’Université Marien NGOUABI Sciences et Techniques. 2008;9:19–29.
[12] Tans PP, White JWC. In balance, with a little help from the plants. Science. 1988;281:183–184.
[13] Senga Kiessé T, Zocchi SS, Kokonendji CC. The R package for (symmetric) discrete triangular distributions; 2009. Available from: http://cran.r-project.org/web/packages/TRIANG/index.html
[14] Aitchison J, Aitken CGG. Multivariate binary discrimination by the kernel method. Biometrika. 1976;63:413–420.
[15] Wang M-C, Van Ryzin J. A class of smooth estimators for discrete distributions. Biometrika. 1981;68:301–309.
[16] Kokonendji CC, Kiessé TS. Discrete associated kernel method and extensions. Stat. Methodol. 2011;8:497–516.
[17] Plomion C, Leprovos G, Stokes A. Wood formation in trees. Plant Physiol. 2011;127:1513–1523.
[18] Cuny HE, Rathgeber CBK, Fortin M, Fournier M. Life strategies in intra-annual dynamics of wood formation: example of three conifer species in a temperate forest in north-east France. Tree Physiol. 2012;32:612–625.

Appendix

Proof of Theorem 1  Let us consider the following decomposition of mean-squared error given by
\[ E[\hat{f}_n(x) - f(x)]^2 = \text{Var}(\hat{f}_n(x)) + [E[\hat{f}_n(x)] - f(x)]^2. \]
We first simply show that the variance term tends to 0 as $n \to \infty$ and $h \to 0$ since
\[
\text{Var}(\hat{f}_n(x)) \leq \frac{p_0^2(x)}{n} \sum_{y \in S_x} p_0^{-2}(y) K_{x,h}^2(y)f(y) = \frac{1}{n} \sum_{y \in S_x} p_0^{-2}(y) K_{x,h}^2(y)f(y)
\]
with \( K_{x,h}(y) \to 0 \) as \( h \to 0 \) for \( y \in S_x \setminus \{x\} \). Next, the bias term is such that

\[
|E[\hat{f}_n(x)] - f(x)| \leq p_0(x) \sum_{y \in S_x \setminus (S_{x,h} \cup S_{x,\eta})} K_{x,h}(y) |\omega(y) - \omega(x)|
\]

\[
\leq \sum_{y \in S_{x,\eta}} |K_{x,h}(y) (\omega(y) - \omega(x))| + \sum_{y \in S_{x,\eta}} |K_{x,h}(y) (\omega(y) - \omega(x))|,
\]

where, for all \( \eta > 0 \), we denote \( S_{x,\eta} = \{ y \in S_x : |y - x| < \eta \} \) and \( S_{x,\eta} \) its complementary. For the first term in Equation (11), by assuming that there exists a finite constant \( M > 0 \) such that \( 0 < \omega < M \) (since \( 0 < f = p_0 \omega \leq 1 \)) and using the Tchebychev–Markov inequality, we have successively

\[
\sum_{y \in S_{x,\eta}} |K_{x,h}(y) (\omega(y) - \omega(x))| \leq 2M \Pr(|K_{x,h} - x| > \eta) \]

\[
\leq \frac{2M}{\eta^2} E[(K_{x,h} - x)^2]
\]

\[
\leq \frac{2M}{\eta^2} [\text{Var}(K_{x,h}) + (E(K_{x,h}) - x)^2] \to 0 \quad \text{as} \quad h \to 0,
\]

under assumptions \( H_1 - H_2 \). The second term in Equation (11) converges to 0 by using the continuity of discrete function \( \omega \) such that: for all \( \epsilon > 0 \), there exists \( \eta > 0 \) for which \( \sum_{y \in S_{x,\eta}} |K_{x,h}(y) (\omega(y) - \omega(x))| \leq \epsilon \). Hence the desired result. □

**Proof of Theorem 2** Let us consider the cross-validation function in Equation (10). Its expectation may be given by

\[
E(CV(h)) = \frac{1}{n} \sum_{x \in N} \left\{ \frac{p_0(x)}{p_0(X_1)} K_{x,h}(X_1) \right\}^2 + E(H_y) + O\left(\frac{1}{n}\right).
\]

In what follows the developments are essentially realized around target \( x \) under assumptions \( H_1 - H_2 \); in addition, we have just to recall that \( K_{x,h}(x) = \Pr(K_{x,h} = x) \to 1 \) and, for \( y \neq x \), \( K_{x,h}(y) \to 0 \) as \( h \to 0 \). Consequently, for the first term of \( E(CV(h)) \), we express

\[
\sum_{x \in N} \left\{ \frac{p_0(x)}{p_0(X_1)} K_{x,h}(X_1) \right\}^2 = \sum_{x \in N} \left\{ K_{x,h}(x)f(x) + p_0^2(x) \sum_{y \in N \setminus \{x\}} K_{x,h}^2(y) \right\}
\]

\[
= \sum_{x \in N} K_{x,h}^2(x)f(x) + o(h^2)
\]

since \( 0 < \sum_{y \in N \setminus \{x\}} K_{x,h}^2(y)f(y) \leq \sum_{y \in N \setminus \{x\}} K_{x,h}(y) \to 0 \) as \( h \to 0 \). The calculation of the second term of \( E(CV(h)) \) requires to express the three following equalities:

\[
E \left\{ K_{X_1,h}(X_2) \frac{p(X_1; \Theta_0)}{p(X_2; \Theta_0)} \right\} = \sum_{x \in N} E \left\{ K_{x,h}(X_2) \frac{p(x; \Theta_0)}{p(X_2; \Theta_0)} \right\} f(x),
\]

\[
E \left\{ K_{X_1,h}(X_1) \right\} = \sum_{y \in N} p^{-1}(y; \Theta_0) K_{x,h}(y) f(y) = \sum_{y \in N} (p^{-1}(y) K_{x,h}(y) = E[\omega(K_{x,h})],
\]

and this last: \( E[\omega(K_{x,h})] = \omega(x) + (1/2)\omega^2(x) \text{Var}(K_{x,h}) + o(h^2) \), obtained by using discrete Taylor expansion. From here, it ensues

\[
E(H_y) = \sum_{x \in N} p^2(x; \Theta_0) E[p^{-1}(X_1; \Theta_0) K_{x,h}(X_1)] E[p^{-1}(X_2; \Theta_0) K_{x,h}(X_2)]
\]

\[
= -2E \left\{ K_{X_1,h}(X_2) \frac{p(X_1; \Theta_0)}{p(X_2; \Theta_0)} \right\}
\]

\[
= \sum_{x \in N} \left\{ p_0 \omega(x) + \frac{1}{2} \text{Var}(K_{x,h}) p_0 \omega^2 \right\}^2
\]

\[
= \frac{1}{4} \sum_{x \in N} [\text{Var}(K_{x,h}) p_0 \omega^2 \right] - \sum_{x \in N} (p_0 \omega)^2(x) + o(h^2).
\]

The previous calculations result in expression of CV’s expectation in Theorem 2.3.
Second, for the variance of CV, we can assume that the variance of first term is such that \( \text{Var}\{ \sum_{x \in N} \{ n^{-2} \sum_{j=1}^{n} K_{x_i}^2(X_i) \} \} = O(n^{-3}) \). Considering the variance of the second term of CV, we have

\[
\frac{1}{n^2} \text{Var} \left( \sum_{j<i} H_{ij} \right) = \frac{1}{n^2} \text{Var}(H_{ij}) + \frac{1}{n} \left( 1 - \frac{3}{n} \right) \text{Cov}(H_{ij}, H_{ik}) + O \left( \frac{1}{n^2} \right).
\]

First, we have

\[
E(H_{ij}^2) = \sum_{x \in N} \sum_{z \in N} \left\{ \frac{1}{p_0(x)p_0(z)} \sum_{x \in N} p_0^2(x)K_{x_i}(x)K_{x_j}(z) - 2 \frac{p_0(x)}{p_0(z)} K_{x_i}(z) \right\}^2 f(x)f(z)
\]

\[
= \sum_{x \in N} \left\{ \sum_{x \in N} p_0^2(x)K_{x_i}^2(x) - 2K_{x_i}(x) \right\}^2 \omega^2(x) + o(h^2).
\]

Then, by using expression of \( E(H_{ij}) \) calculated previously, we get the variance of \( H_{ij} \) as follows:

\[
\text{Var}(H_{ij}) = \sum_{x \in N} \left\{ \sum_{x \in N} p_0^2(x)K_{x_i}^2(x) - 2K_{x_i}(x) \right\}^2 \omega^2(x) - \left\{ \sum_{x \in N} (p_0 \omega)^2(x) \right\}^2
\]

\[
+ \frac{1}{2} \sum_{x \in N} \left( \text{Var}(K_{x_i})(p_0^2)(x) \right)^2 \sum_{x \in N} (p_0 \omega)^2(x) + o(h^2).
\]

In addition, without give all details here, it can be shown that

\[
E(H_{ij}H_{ik}) = \sum_{x \in N} \left\{ \sum_{x \in N} p_0^2(x)K_{x_i}^2(x) - 2K_{x_i}(x) \right\}^2 \omega^3(x) + o(h^3),
\]

then, by considering \( \text{Cov}(H_{ij}, H_{ik}) = E(H_{ij}H_{ik}) - E(H_{ij})E(H_{ik}) \), we have

\[
\text{Cov}(H_{ij}, H_{ik}) = \sum_{x \in N} \left\{ \sum_{x \in N} p_0^2(x)K_{x_i}^2(x) - 2K_{x_i}(x) \right\}^2 \omega^3(x) - \left\{ \sum_{x \in N} (p_0 \omega)^2(x) \right\}^2
\]

\[
+ \frac{1}{2} \sum_{x \in N} \left( \text{Var}(K_{x_i})(p_0^2)(x) \right)^2 \sum_{x \in N} (p_0 \omega)^2(x) + o(h^3),
\]

with \( \text{Cov}(H_{ij}, H_{ik}) = 0 \). Hence the desired result on variance of CV.