C. R. Acad. Sci. Paris, Ser. I 354 (2016) 735–740

Contents lists available at ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Statistics

On bandwidth parameter choices for discrete nonparametric kernel estimator

Choix de fenêtres de lissage pour un estimateur non paramétrique à noyau discret

Tristan Senga Kiessé

INRA, UMR 1069, Sol Agro et Hydrosystème Spatialisation, 35000 Rennes, France

ARTICLE INFO

Article history:
Received 20 May 2012
Accepted after revision 10 February 2016
Available online 13 May 2016

Presented by the Editorial Board

ABSTRACT

This note concentrates on the nonparametric estimation of a probability mass function (p.m.f.) using discrete associated kernels. An expression of the optimal bandwidth minimizing the asymptotic part of the global squared error is given. Some asymptotic expressions of bias and variance of the cross-validation criterion are also presented. At last, the two bandwidth selection procedures are illustrated through some simulations and an application on a real count data set.

RÉSUMÉ

Cette note se focalise sur l'estimation non paramétrique à noyau associé discret d'une fonction de masse de probabilité. Une expression de la fenêtre optimale minimisant la partie asymptotique de l'erreur quadratique globale est donnée. Des expressions asymptotiques pour le biais et la variance d'un critère de sélection par validation croisée sont également présentées. Enfin, les deux méthodes de choix de fenêtre sont illustrées par des simulations et une application sur des données réelles.

1. Introduction

Let $(X_i)_{i=1,2,\cdots,n}$ be a sample of i.i.d. discrete random variables (r.v.) having a p.m.f. $f(x) = \Pr(X_i = x) > 0$ on support $\mathcal{S}$ (*e.g.*, non-negative integers set $\mathbb{N}$). The discrete kernel estimator $\widetilde{f}_n$ of f can be expressed as

$$\widetilde{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_{x,h}(X_i), \quad x \in \mathcal{S}, \tag{1}$$

E-mail address: tristan.sengakiesse@rennes.inra.fr.

http://dx.doi.org/10.1016/j.crma.2016.02.012
1631-073X/

where $h = h(n) > 0$ is an arbitrary sequence of smoothing parameters that fulfills $\lim_{n\to\infty} h(n) = 0$ and $K_{x,h}(\cdot)$ is a p.m.f., called *discrete associated kernel*, with r.v. $\mathcal{K}_{x,h}$ on support $\mathcal{S}_x$ (containing target x and not depending on h) satisfying

$$\text{H1}: \lim_{h\to 0} \mathrm{E}(\mathcal{K}_{x,h}) = x, \quad \text{and} \quad \text{H2}: \lim_{h\to 0} \mathrm{Var}(\mathcal{K}_{x,h}) = 0;$$

without loss of generality, we assume $\mathcal{S}_x \subseteq \mathcal{S}$ in all this work. The nonparametric kernel estimation has been already largely discussed in the literature for the probability density function (p.d.f.), in particular for choosing smoothing parameters; see Bowman [2] or Marron [6]. In contrast, the discrete kernel estimation for p.m.f. has received more less attention than that for p.d.f. In fact, until now, it commonly results from the discretization of the continuous case by considering the ordinal variables as being continuous; see Aitchison and Aitken [1], Titterington [8] or Simonoff and Tutz [7]. For a specific investigation of count data, Kokonendji et al. [4] and Kokonendji and Senga Kiessé [3] have developed the nonparametric kernel estimation using some discrete associated kernels. This opens way to a specific treatment of the bandwidth selections of discrete kernel estimators by adapting classical methods, which are well known for the continuous case. This note pursues the first works realized on this subject, with a first attempt to provide an expression of the optimal bandwidth and by investigating the asymptotic properties of the cross-validation criterion for $\widetilde{f}_n$ in (1).

2. Discrete associated kernels

Let us first introduce the following hypotheses on modal probability of $\mathrm{K}_{x,h}$:

$$\text{H1}': K_{x,h}(x) = 1 - h\mathrm{A}(\mathcal{K}_{x,h}) + O(h^2),$$

with $\sum_{y\in\mathcal{S}_x\setminus\{x\}} K_{x,h}(y) = h\mathrm{A}(\mathcal{K}_{x,h}) + O(h^2) \to 0$ as $h \to 0$, where $\mathrm{A}(\mathcal{K}_{x,h})$ is bounded away from 0 for $h \to 0$. It results in the following proposition.

Proposition 2.1. *Consider a fixed point $x \in \mathcal{S}$ and the bandwidth $h > 0$. Under assumption* H1′*, the expectation and variance of the discrete associated kernel $K_{x,h}$ fulfill H1 and H2.*

Proof. For the assumption H1, the expectation of $\mathcal{K}_{x,h}$ comes directly from

$$\mathbb{E}(\mathcal{K}_{x,h}) = xK_{x,h}(x) + \sum_{y\in\mathcal{S}_x\setminus\{x\}} yK_{x,h}(y) = x + B(x;h),$$

with $B(x;h) = -xh\mathrm{A}(\mathcal{K}_{x,h}) + \sum_{y\in\mathcal{S}_x\setminus\{x\}} yK_{x,h}(y) + O(h^2) \to 0$ for $h \to 0$. For the assumption H2, the variance of $\mathcal{K}_{x,h}$ can be successively expressed as

$$\mathrm{Var}(\mathcal{K}_{x,h}) = \sum_{y\in\mathcal{S}_x} y^2 K_{x,h}(y) - \left\{\sum_{y\in\mathcal{S}_x} yK_{x,h}(y)\right\}^2 = x^2\{K_{x,h}(x) - 1\} + D(x;h),$$

with $D(x;h) = \sum_{y\in\mathcal{S}_x\setminus\{x\}} y^2 K_{x,h}(y) + x^2 - \left\{x + \sum_{y\in\mathcal{S}_x}(y-x)K_{x,h}(y)\right\}^2 \to 0$ when $h \to 0$ under the hypotheses H1′. Indeed, when $h \to 0$, we have both $K_{x,h}(y) \to 0$ for $y \neq x$ and $K_{x,h}(x) \to 1$ for $y = x$. $\square$

Then, the following assumption can be formulated on the variance of $\mathcal{K}_{x,h}$:

$$\text{H2}': \mathrm{Var}(\mathcal{K}_{x,h}) = h\mathrm{V}(\mathcal{K}_{x,h}) + O(h^2),$$

leading to some hypotheses H1′ and H2′ less general than H1 and H2; $\mathrm{A}(\mathcal{K}_{x,h})$ and $\mathrm{V}(\mathcal{K}_{x,h})$ do not obligatory depend on x and h, as in the next example.

Example. (See Kokonendji and Zocchi [5].) Let a_1, a_2 be fixed integers and h_1, h_2 be smoothing parameters. For any fixed $x \in \mathcal{S} = \mathbb{Z}$, consider the r.v. $\mathcal{K}_{a_1,a_2;x,h_1,h_2}$ of discrete generalized triangular associated kernels defined on supports $\mathcal{S}_{a_1,x} = \{x-1, x-2, \ldots, x-a_1\}$ and $\mathcal{S}_{x,a_2} = \{x, x+1, \ldots, x+a_2\}$ and whose p.m.f. is

$$K_{a_1,a_2;x,h_1,h_2}(y) = \frac{1}{P}\left\{\left[1 - \left(\frac{x-y}{a_1+1}\right)^{h_1}\right]\mathbf{1}_{\mathcal{S}_{a_1,x}}(y) + \left[1 - \left(\frac{y-x}{a_2+1}\right)^{h_2}\right]\mathbf{1}_{\mathcal{S}_{x,a_2}}(y)\right\},$$

where $P \equiv (a_1 + a_2 + 1) - (a_1+1)^{-h_1}\sum_{k=1}^{a_1} k^{h_1} - (a_2+1)^{-h_2}\sum_{k=1}^{a_2} k^{h_2} = P(a_1, a_2, h_1, h_2)$ is the normalizing constant. For h sufficiently small, one has

$$K_{a_i;x,h_i}(x) = 1 - \sum_{i=1}^{2}\left\{h_i A(a_i) + O(h_i^2)\right\} \text{ and } \mathrm{Var}(\mathcal{K}_{a_i;x,h_i}) = \sum_{i=1}^{2}\left\{h_i V(a_i) + O(h_i^2)\right\},$$

with $A(a_i) = a_i\log(a_i+1) - \sum_{k=1}^{a_i}\log(k)$ and $V(a_i) = a_i(2a_i^2 + 3a_i + 1)\log(a_i+1)/6 - \sum_{k=1}^{a_i} k^2\log(k)$. Hence, the assumptions H1′ and H2′ are fulfilled.

3. Nonparametric kernel estimator

This section presents an optimal h-value minimizing an approximate of the mean integrated squared error (MISE) of $\widetilde{f}_n$ in (1), in comparison with the h-value mimimizing cross-validation function (Kokonendji and Senga Kiessé [3]).

3.1. Minimization of MISE

The $\widetilde{f}_n$'s variance and bias have been established by Kokonendji and Senga Kiessé [3] under H1 and H2; by taking into account H1′ and H2′, the global error is

$$\mathrm{MISE}(h)=\sum_{x\in\mathcal{S}}\mathrm{Var}\{\widetilde{f}_n(x)\}+\sum_{x\in\mathcal{S}}\mathrm{Bias}^2\{\widetilde{f}_n(x)\}=\mathrm{AMISE}(h)+o\left(\frac{1}{n}+h^2\right)+O(h^2),$$

where

$$\mathrm{AMISE}(h)=\frac{1}{n}\sum_{x\in\mathcal{S}}f(x)\left[\left\{1-h\mathrm{A}(\mathcal{K}_{x,h})\right\}^2-f(x)\right]+\frac{1}{4}h^2\sum_{x\in\mathcal{S}}\left\{\mathrm{V}(\mathcal{K}_{x,h})f^{(2)}(x)\right\}^2$$

is the approximated MISE for h sufficiently small, with $f^{(2)}$ a finite difference of second order. Hence, an approximate value $\widehat{h}_{\mathrm{opt}}=\arg\min_{h>0}\mathrm{AMISE}(h)$ of the true optimal bandwidth $h_{\mathrm{opt}}=\arg\min_{h>0}\mathrm{MISE}(h)$ can be given by

$$\widehat{h}_{\mathrm{opt}}(n,f)=\frac{\sum_{x\in\mathcal{S}}f(x)\mathrm{A}(\mathcal{K}_{x,h})}{\sum_{x\in\mathcal{S}}f(x)\{\mathrm{A}(\mathcal{K}_{x,h})\}^2+(n/4)\sum_{x\in\mathcal{S}}\left\{\mathrm{V}(\mathcal{K}_{x,h})f^{(2)}(x)\right\}^2},$$

which tends to 0 as $n\to\infty$ with $0<\sum_{x\in\mathcal{S}}\left\{\mathrm{V}(\mathcal{K}_{x,h})f^{(2)}(x)\right\}^2<\infty$. It results in an asymptotic relationship such that $\widehat{h}_{\mathrm{opt}}\sim k_0n^{-1}$ with $k_0=4\sum_{x\in\mathcal{S}}f(x)\mathrm{A}(\mathcal{K}_{x,h})/[\sum_{x\in\mathcal{S}}\left\{\mathrm{V}(\mathcal{K}_{x,h})f^{(2)}(x)\right\}^2]$. Thus, the discrete generalized triangular associated kernel presented as an example has optimal bandwidth

$$\widehat{h}_{\mathrm{opt}}(n,a_i,f)=\frac{\mathrm{A}(a_i)}{2\{\mathrm{A}(a_i)\}^2+(n/2)\{\mathrm{V}(a_i)\}^2\sum_{x\in\mathbb{Z}}\left\{f^{(2)}(x)\right\}^2},\quad i=1,2. \tag{2}$$

3.2. Minimization of cross-validation function

We are interested in a bandwidth h_{cv} minimizing a cross-validation score function CV with respect to h such that $h_{\mathrm{cv}}=arg\min_{h>0}\mathrm{CV}(h)$ with

$$\mathrm{CV}(h)=\sum_{x\in\mathcal{S}}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{x,h}\left(X_i\right)\right\}^2-\frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}K_{X_i,h}\left(X_j\right). \tag{3}$$

The CV's mean and variance for fixed h are provided in the following theorem.

Theorem 3.1. *Consider a fixed point $x\in\mathcal{S}$ and the bandwidth $h=h(n)>0$ such that $\lim_{n\to\infty}h=0$. Assume that CV is the cross-validation function for the nonparametric estimator in (1) of p.m.f. f. Then, we have*

$$\mathbb{E}\{\mathrm{CV}(h)\}=\mathrm{AMISE}(h)-\sum_{x\in\mathcal{S}}f^2(x)+O\left(\frac{1}{n}+h^2\right)$$

and

$$\mathrm{Var}\{CV(h)\}=\frac{1}{n}\sum_{x\in\mathcal{S}}\left\{\sum_{x\in\mathcal{S}}K^2_{x,h}(x)-2K_{x,h}(x)\right\}^2f^3(x)-\frac{1}{n}\left\{\sum_{x\in\mathcal{S}}f^2(x)\right\}^2+O\left(\frac{h^2}{n}+\frac{1}{n^2}\right),$$

where $K_{x,h}$ is a discrete kernel satisfying the assumption H1′.

Proof. Let us present the cross-validation score function in (3) as

$$CV(h)=\frac{1}{n^2}\sum_{i=1}^{n}\sum_{x\in\mathcal{S}}K^2_{x,h}(X_i)+\frac{2}{n^2}\sum\sum_{j<i}H_{ij},$$

with $H_{ij}=\sum_{x\in\mathcal{S}}K_{x,h}(X_i)K_{x,h}(X_j)-2K_{X_i,h}\left(X_j\right)$. One has $\mathbb{E}\{CV(h)\}=(1/n)\sum_{x\in\mathcal{S}}\mathbb{E}\{K^2_{x,h}(X_1)\}+(1-1/n)\mathbb{E}(H_{ij})$. For the first term of $\mathbb{E}\{CV(h)\}$, we get:

$$\sum_{x\in\mathcal{S}}\mathbb{E}\{K_{x,h}^2(X_1)\}=\sum_{x\in\mathcal{S}}K_{x,h}^2(x)f(x)+\sum_{x\in\mathcal{S}}\sum_{y\in\mathcal{S}\setminus\{x\}}K_{x,h}^2(y)f(y);$$

where the second term in the previous equation is finite. Now let us express

$$\mathbb{E}\{K_{X_1,h}(X_2)\}=\sum_{x\in\mathcal{S}}\mathbb{E}\{K_{x,h}(X_2)\}f(x),\quad \mathbb{E}\{K_{x,h}(X_1)\}=\sum_{y\in\mathcal{S}}K_{x,h}(y)f(y)=\mathbb{E}\{f(\mathcal{K}_{x,h})\},$$

and $\mathbb{E}\{f(\mathcal{K}_{x,h})\}=f(x)+(1/2)f^{(2)}(x)\operatorname{Var}(\mathcal{K}_{x,h})+o(h)$ obtained by using a discrete Taylor expansion around x. For the second term of $\mathbb{E}\{CV(h)\}$, it ensues

$$\mathbb{E}(H_{ij})=\frac{h^2}{4}\sum_{x\in\mathcal{S}}\left\{\mathrm{V}(\mathcal{K}_{x,h})f^{(2)}(x)\right\}^2-\sum_{x\in\mathcal{S}}f^2(x)+o(h^2)+O(h^2),$$

where $o(h^2)$ is asymptotically dominated by $O(h^2)$. Hence, it results $\mathbb{E}\{CV(h)\}$.

For CV's variance, we begin by calculating the variance of the first term as

$$\frac{1}{n^3}\operatorname{Var}\left\{\sum_{x\in\mathcal{S}}K_{x,h}^2(X_1)\right\}=\frac{1}{n^3}\left[\left\{\sum_{x\in\mathcal{S}}K_{x,h}^2(x)\right\}^2f(x)-\left\{\sum_{x\in\mathcal{S}}K_{x,h}^2(x)f(x)\right\}^2\right]+Q_n(x;h),$$

with

$$n^3Q_n(x;h)=\sum_{y\in\mathcal{S}\setminus\{x\}}\left\{\sum_{x\in\mathcal{S}}K_{x,h}^2(y)\right\}^2f(y)+\left\{\sum_{x\in\mathcal{S}}K_{x,h}^2(x)f(x)\right\}^2-\left\{\sum_{y\in\mathcal{S}}\sum_{x\in\mathcal{S}}K_{x,h}^2(y)f(y)\right\}^2,$$

where Q_n is $o(1/n^3)$; in fact, we can assume $\operatorname{Var}\left[\sum_{x\in\mathcal{S}}\left\{n^{-2}\sum_{i=1}^nK_{x,h}^2(X_i)\right\}\right]=o(n^{-3})+O(n^{-3})$. Considering the variance of the second term of CV, we have

$$\frac{1}{n^4}\operatorname{Var}\left(\sum_{j<i}\sum H_{ij}\right)=\frac{1}{n^2}\operatorname{Var}(H_{ij})+\frac{1}{n}\left(1-\frac{3}{n}\right)\operatorname{Cov}(H_{ij},H_{ik})+O\left(\frac{1}{n^3}\right).$$

Without giving all details, we get $\mathbb{E}(H_{ij}^2)=\sum_{x\in\mathcal{S}}\left\{\sum_{x\in\mathcal{S}}K_{x,h}^2(x)-2K_{x,h}(x)\right\}^2f^2(x)+o(h^2)$. Then, by using expression of $\mathbb{E}(H_{ij})$ calculated previously, we have

$$\begin{aligned}\operatorname{Var}(H_{ij})=&\sum_{x\in\mathcal{S}}\left\{\sum_{x\in\mathcal{S}}K_{x,h}^2(x)-2K_{x,h}(x)\right\}^2f^2(x)-\left\{\sum_{x\in\mathcal{S}}f^2(x)\right\}^2\\&+\frac{h^2}{2}\sum_{x\in\mathcal{S}}\left\{\mathrm{V}(\mathcal{K}_{x,h})f^{(2)}(x)\right\}^2\sum_{x\in\mathcal{S}}f^2(x)+o(h^2).\end{aligned}$$

In addition, it can be shown that

$$\mathbb{E}(H_{ij}H_{ik})=\sum_{x\in\mathcal{S}}\left\{\sum_{x\in\mathcal{S}}K_{x,h}^2(x)-2K_{x,h}(x)\right\}^2f^3(x)+o(h^3);$$

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

$$\begin{aligned}\operatorname{Cov}(H_{ij},H_{ik})=&\sum_{x\in\mathcal{S}}\left\{\sum_{x\in\mathcal{S}}K_{x,h}^2(x)-2K_{x,h}(x)\right\}^2f^3(x)-\left\{\sum_{x\in\mathcal{S}}f^2(x)\right\}^2\\&+\frac{h^2}{2}\sum_{x\in\mathcal{S}}\left\{\mathrm{V}(\mathcal{K}_{x,h})f^{(2)}(x)\right\}^2\sum_{x\in\mathcal{S}}f^2(x)+o(h^3)+O(h^2),\end{aligned}$$

with $\operatorname{Cov}(H_{ij},H_{kl})=0$. Hence the desired result on $\operatorname{Var}\{CV(h)\}$. □

4. Illustrations

This section illustrates the bandwidth choices; discrete symmetric and generalized triangular kernels are applied for simulated and real data, respectively.

Table 1
Averages ISE, $\widehat{h}_{\text{opt}}$ and h_{cv} for nonparametric estimator using triangular kernels.

	Sample size n	h_{opt}	$ISE(h_{\text{opt}})$	$\overline{\widehat{h}_{\text{opt}}}$	$\overline{ISE}(\widehat{h}_{\text{opt}})$	$\overline{h_{\text{cv}}}$	$\overline{ISE}(h_{\text{cv}})$
$a=2$	25	0.49	0.0028	0.23	0.0116	0.73	0.0130
	100	0.16	0.0011	0.12	0.0043	0.23	0.0053
$a=3$	25	0.22	0.0034	0.10	0.0140	0.43	0.0149
	100	0.06	0.0009	0.04	0.0057	0.07	0.0062
$a=4$	25	0.12	0.0033	0.05	0.0196	0.23	0.0195
	100	0.03	0.0006	0.02	0.0065	0.04	0.0068

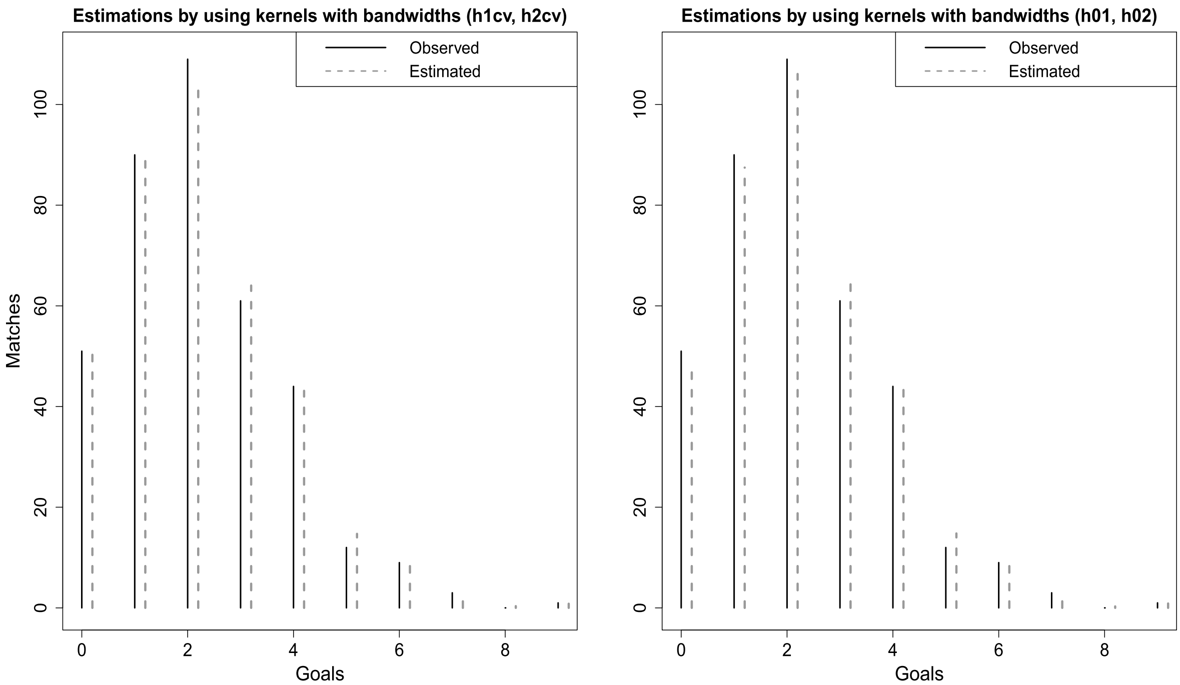

Fig. 1. Estimations of the number of goals per match using a discrete generalized triangular kernel.

4.1. Simulations

Consider simulated data of sample size n from a Poisson distribution f with mean $\mu = 2$. The estimator $\widetilde{f}_n$ using symmetric triangular kernels is applied with $a_1 = a_2 = a \in \{2, 3, 4\}$, as an example. The performance of $\widetilde{f}_n$ is evaluated via the integrated squared error $ISE(h) = \sum_{x\in\mathbb{N}}\{\widetilde{f}_n(x) - f(x)\}^2$ determined using $\widehat{h}_{\text{opt}}$, h_{cv} and true h_{opt} computed numerically since f is known. In Table 1, the averages $\overline{ISE}$, $\overline{\widehat{h}_{\text{opt}}}$ and $\overline{h_{\text{cv}}}$ are calculated based on 100 replications of the simulated data. The value $\widehat{h}_{\text{opt}}$ is competitive since globally we have $\overline{ISE}(\widehat{h}_{\text{opt}}) \leq \overline{ISE}(h_{\text{cv}})$ for chosen values of a; note that h_{opt} is underestimated by $\widehat{h}_{\text{opt}}$ and overestimated by h_{cv}. At last, by estimating f using empirical frequency of data, $\overline{ISE}$ is equal to 0.0308 for $n = 25$ and 0.0088 for $n = 100$.

4.2. Application

The estimator $\widetilde{f}_n$ is applied using a generalized triangular kernel with $(a_1, a_2) = (1, 2)$ on count data, describing a number of goals per match from the French football championship (Kokonendji et al. [4]). By a cross-validation procedure, we have $h_{1\text{cv}} = 0.170$ and $h_{2\text{cv}} = 0.085$, while the expression (2) results in $\widehat{h}_{\text{opt},01} = 0.156$ and $\widehat{h}_{\text{opt},02} = 0.031$, obtained by replacing the unknown p.m.f. f in (2) with its empirical frequency estimate. The quality of the estimate provided using $(h_{1\text{cv}}, h_{2\text{cv}})$ is close to that of the one calculated using $(\widehat{h}_{\text{opt},01}, \widehat{h}_{\text{opt},02})$, since ISE is, respectively, equal to 3.0081×10^{-4} and 3.0082×10^{-4} (Fig. 1).

References

[1] J. Aitchison, C.G.G. Aitken, Multivariate binary discrimination by the kernel method, Biometrika 63 (1976) 413–420.
[2] A. Bowman, An alternative method of cross-validation for the smoothing of density estimates, Biometrika 71 (1984) 352–360.
[3] C.C. Kokonendji, T. Senga Kiessé, Discrete associated kernel method and extensions, Stat. Methodol. 8 (2011) 497–516.

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

[5] C.C. Kokonendji, S.S. Zocchi, Extensions of discrete triangular distributions and boundary bias in kernel estimation for discrete functions, Stat. Probab. Lett. 80 (2010) 1655–1662.

[6] J.S. Marron, A comparison of cross-validation techniques in density estimation, Ann. Stat. 15 (1987) 152–162.

[7] J.S. Simonoff, G. Tutz, Smoothing methods for discrete data, in: M.G. Schimek (Ed.), Smoothing and Regression: Approaches, Computation, and Application, Wiley, New York, 2000, pp. 193–228.

[8] D.M. Titterington, A comparative study of kernel-based density estimates for categorical data, Technometrics 22 (1980) 259–268.