CrossMark

Multiplicative bias correction for discrete kernels

Lynda Harfouche[1] · **Smail Adjabi**[1] ·
Nabil Zougab[1] · **Benedikt Funke**[2]

Abstract In this paper, we prove that two multiplicative bias correction techniques (MBC) can be applied for discrete kernels in the context of probability mass function estimation. First, some properties of the MBC discrete kernel estimators (bias, variance and mean integrated squared error) are investigated. Second, the popular cross-validation technique is adapted for bandwidth selection. Finally, a simulation study and a real data application for discrete data illustrate the performance of the MBC estimators based on dirac discrete uniform and triangular discrete kernels.

Keywords Bandwidth selection · Cross-validation · Discrete kernel · Discrete data · Mean integrated squared error · Multiplicative bias correction

1 Introduction

Given a random sample of observations $X_1, \ldots, X_n$ with unknown probability mass function (pmf) f, which is supported on the discrete set $\mathbb{T}$ ($\mathbb{N}, \mathbb{Z}$ or $\mathbb{T} = \{0, 1, \ldots, N\}$), the discrete kernel estimator $\hat{f}_h(x)$ of $f(x) = \Pr(X_i = x)$ has been defined as follows [see for example Kokonendji and Senga Kiessé (2011)]:

✉ Lynda Harfouche
  lyndaharfouche@hotmail.com

  Smail Adjabi
  adjabi@hotmail.com

  Nabil Zougab
  nabilzougab@yahoo.fr

  Benedikt Funke
  benedikt.funke@mathematik.tu-dortmund.de

[1] Research Unit LaMOS, University of Bejaia, Bejaïa, Algeria

[2] Department of Mathematics, Technical University of Dortmund, Dortmund, Germany

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^{n} K_{x,h}(X_i),$$

where $h = h(n) > 0$ is a bandwidth (or smoothing parameter) and $K_{x,h}$ is a discrete kernel assumed to be a suitable pmf with support $\mathbb{S}_x$ not depending on h; see, e.g., Kokonendji et al. (2007) and Kokonendji and Senga Kiessé (2011). Naturally, the use of the discrete kernels is more appropriate than continuous kernels for estimating any discrete function; see again Kokonendji et al. (2007) and Kokonendji and Senga Kiessé (2011). See also Aitchison and Aitken (1976) for categorical data and finite discrete distributions and Wang and Ryzin (1980) for ordered discrete variables.

In view of the fact that the bias of $\hat{f}_h$ is $O(h)$ as $h \to 0$, this paper considers improvements in discrete kernel estimation that reduce the order of magnitude in bias to $O(h^2)$, while the order of magnitude in variance is maintained. In the case of symmetric kernels, this kind of rate improvements can be typically achieved by employing higher-order kernels; see Jones and Foster (1993) for methods of generating higher-order kernels from a given second-order kernel. To our best knowledge, equivalent techniques are yet to be proposed for discrete kernels. Instead, this paper applies two classes of multiplicative bias correction (MBC) techniques in order to attain the rate improvements. The MBC approaches were proposed and largely studied by several authors in symmetric and asymmetric kernel density estimation (continuous situation), see, e.g., Terrell and Scott (1980), Jones et al. (1995), Hirukawa (2010), Hirukawa and Sakudo (2014), Hirukawa and Sakudo (2015), Zougab and Adjabi (2015) and Funke and Kawka (2015). The first class of our considered MBC method is concerned with the construction of a multiplicative combination of two density estimators using different smoothing parameters. This idea is originally proposed by Terrell and Scott (1980) as an additive bias correction to the logarithm of densities. The second class of MBC in the spirit of Jones et al. (1995) is based on the idea of expressing $f(x) = \hat{f}(x) \left\{ f(x)/\hat{f}(x) \right\}$ and estimating the bias-correction term $f(x)/\hat{f}(x)$ nonparametrically. When applied to discrete kernel estimation, both MBC techniques still yield estimators that are free of boundary bias. In addition, these estimators have a practically appealing property. They always generate nonnegative density estimates everywhere by construction, as $\hat{f}_h$ does.

This paper is organized as follows. Section 2 briefly recalls discrete kernels for pmf estimation. In Sect. 3 we first introduce the MBC discrete kernel estimators. Second, we develop asymptotic properties like bias and variance of the newly proposed estimators. Third, we adapt the unbiased-cross validation (UCV) procedure for choosing the bandwidth. Section 4 conducts Monte Carlo simulations to compare the finite sample performance of standard discrete kernel estimators and the proposed MBC discrete kernel estimators. Section 5 provides an application on real data. Section 6 concludes the paper. All proofs are given in the "Appendix".

2 Discrete kernel estimator

Given a random sample $X_1, \ldots, X_n$ with unknown probability mass function (pmf) f, which is supported on the discrete set $\mathbb{T}$ ($\mathbb{N}$, $\mathbb{Z}$ or $\mathbb{T} = \{0, 1, \ldots, N\}$), the discrete

Table 1 Univariate discrete kernels

Kernel(L)	Explicit form
DT (Kokonendji and Senga Kiessé 2011)	$K_{DT(x,h)}(y) = \frac{(a+1)^h - |y-x|^h}{P(a,h)}$, where $P(a,h) = (2a+1)(a+1)^h - 2\sum_{k=0}^{a} k^h$.
DDU (Aitchison and Aitken 1976)	$K_{DDU(x,h)}(y) = (1-h)\mathbb{I}_{y=x} + \frac{h}{m-1}\mathbb{I}_{y \neq x}$.
WVR (Wang and Ryzin 1980)	$K_{WVR(x,h)}(y) = (1-h)\mathbb{I}_{y=x} + \frac{1}{2}(1-h)h^{|y-x|}\mathbb{I}_{|y-x| \geq 1}$.
LR (Racine and Li 2004)	$K_{LR(x,h)}(y) = \mathbb{I}_{y=x} + h^{|y-x|}\mathbb{I}_{y \neq x}$.

Table 2 Explicit forms of $q(x, f)$

Kernel(L)	$q(x, f)$
DT	$\left(\frac{\log(a+1)}{2}S(a) - 2\sum_{k=1}^{a} k^2 \log(k)\right)\frac{f^{(2)}(x)}{2}$
DDU	$\left(\frac{m}{2} - x - \frac{x}{m-1}\right)f^{(1)}(x) + \left(\frac{m(2m-1)}{6} + x^2 + \frac{x^2}{m-1} - mx\right)\frac{f^{(2)}(x)}{2}$
WVR	$\frac{f^{(2)}(x)}{2}$
LR	$2xf^{(1)}(x) - \left(2x^2 - 2\right)\frac{f^{(2)}(x)}{2}$

where $S(a) = \frac{a(a+1)(2a+1)}{3}$ and $f^{(j)}$ denotes the finite difference of j-th order given by $f^{(j)}(x) = \{f^{(j-1)}(x)\}^{(1)}$ with $f^{(1)}(x) = \begin{cases} [f(x+1) - f(x-1)]/2 & \text{if } x \in \mathbb{N}^*; \\ f(1) - f(0) & \text{if } x = 0. \end{cases}$

kernel estimator of $f(x) = \Pr(X_i = x)$ using kernel L = DT, DDU, WVR, LR can be expressed as

$$\widehat{f_L}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{L(x,h)}(X_i), \tag{1}$$

where $x \in \mathbb{T}$ is the target (point where the pmf is estimated) and $h > 0$ is a bandwidth (or smoothing parameter) and the explicit forms of the kernels are listed in Table 1. The asymptotic properties of the estimator (1) are studied in detail in Kokonendji and Senga Kiessé (2011). The asymptotic bias when $h \to 0$ is given by

$$\text{bias}(\widehat{f_L}(x)) = q(x, f)h + o(h),$$

where the explicit forms of $q(x, f)$ for a specific-kernel L are given in Table 2.

Similarly, when $n \to \infty$ and $h \to 0$ the asymptotic variance is

$$\text{Var}(\widehat{f_L}(x)) = \frac{1}{n}f(x)\{1 - f(x)\}K^2_{L(x,h)}(x) + o\left(\frac{1}{n}\right).$$

The mean integrated squared error (MISE) is given in Kokonendji and Senga Kiessé (2011) and Kokonendji et al. (2007) and is expressed as

$$\text{MISE}(\widehat{f_L}) = \sum_{x \in \mathbb{T}} \text{bias}^2(\widehat{f_L}(x)) + \sum_{x \in \mathbb{T}} \text{Var}(\widehat{f_L}(x))$$
$$= h^2 \sum_{x \in \mathbb{T}} q^2(x, f) + \frac{1}{n} \sum_{x \in \mathbb{T}} f(x)\{1 - f(x)\} K^2_{L(x,h)}(x) + o\left(h^2 + \frac{1}{n}\right).$$

In analogy to kernel density estimation, the choice of a suitable bandwidth is also a crucial issue in the discrete kernel method. For this reason, several approaches have been proposed in the literature. The common methods of continuous kernel estimators that adopt the mean integrated squared error (MISE) as a criterion and the techniques of cross-validation (CV) are also developed for discrete kernel estimation techniques; see, e.g, Kokonendji et al. (2007), Kokonendji and Senga Kiessé (2011) and Chu et al. (2015).

3 MBC for discrete kernel estimators

In this section, we adapt the mentioned MBC methods to the estimation of probability mass functions where the special kernels used for pmf estimation are called discrete kernels, which have support on some discrete set including $\mathbb{N}$, $\mathbb{Z}$ or a finite number of integers. Our proposed approaches have the same intuition as in the continuous case, each of the MBC methods is shown to improve the bias convergence of univariate pmf estimators from $O(h)$ to $O(h^2)$ while their variance convergence remains unchanged at $o(n^{-1})$. Globally the proof strategies of each MBC method in pmf estimation largely follow those of the corresponding method that is originally developed for kernel density estimation of scalar continuous random variables see, e.g., Terrell and Scott (1980), Jones et al. (1995), Hirukawa (2010), Hirukawa and Sakudo (2014), Zougab and Adjabi (2015) and Funke and Kawka (2015).

3.1 Estimators

Following the idea of the geometric estimator of Terrell and Scott (1980) and Hirukawa (2010) abbreviated as "TS". This can be readily extended to the discrete kernel L in the context of probability mass function estimation. For a given kernel L, let $\hat{f}_{L,h}(x)$ and $\hat{f}_{L,h/c}(x)$ are the pmf estimators using smoothing parameters h and h/c, respectively, where $c \in (0, 1)$ is some predetermined constant that does not depend on the design point x. Then, the TS-MBC kernel pmf estimator can be adapted as follows:

$$\tilde{f}_{TS,L}(x) = \left\{ \hat{f}_{L,h}(x) \right\}^{\frac{1}{1-c}} \left\{ \hat{f}_{L,h/c}(x) \right\}^{-\frac{c}{1-c}}. \tag{2}$$

The second approach of MBC techniques for symmetric kernel density estimators is attributed to Jones et al. (1995) [see also Hirukawa (2010), Hirukawa and

Sakudo (2014), Zougab and Adjabi (2015) and Funke and Kawka (2015) by using asymmetric kernels] abbreviated as "JLN" and utilizes a single smoothing parameter h. The JLN technique proposed by Jones et al. (1995) is based on the identity $f(x) = \hat{f}_{L,h}(x) \left\{ f(x)/\hat{f}_{L,h}(x) \right\}$. In analogy to their estimators, using the discrete kernel L, we denote as $\tilde{f}_{JLN,L}(x)$ the following estimator:

$$\tilde{f}_{JLN,L}(x) = \hat{f}_{L,h}(x) \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{K_{L(x,h)}(X_i)}{\hat{f}_{L,h}(X_i)} \right\}, \tag{3}$$

where $K_{L(x,h)}$ is the kernel L. Recognize that the term inside the brackets is a natural estimator of the bias-correction term $f(x)/\hat{f}_{L,h}(x)$. Also, by construction, both $\tilde{f}_{TS,L}(x)$ and $\tilde{f}_{JLN,L}(x)$ always generate nonnegative probability mass function estimates everywhere.

3.2 Asymptotic properties

The asymptotic bias and variance of the MBC estimators are presented in the following theorems. We assume that

A1. The derivatives of f at each point $x \in \mathbb{N}$ are replaced by the finite differences [see Kokonendji and Senga Kiessé (2011)]

$$f^{(j)}(x) = \{f^{(j-1)}(x)\}^{(1)},$$

where

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

A2. The smoothing parameter $h = h(n)$ satisfies $h \to 0$ as $n \to \infty$.

Theorem 1 *Let $\tilde{f}_{TS,L}$ be the TS-MBC estimator using kernel L defined by* (2)*. For a given target x, and assuming A1 and A2, it holds:*

(i) The bias of the TS-MBC discrete kernel estimator takes the following approximation

$$\text{bias}(\tilde{f}_{TS,L}(x)) = \frac{1}{c} \left[\frac{1}{2} \left\{ \frac{l_1^2(x, f)}{f(x)} - l_2(x, f) \right\} \right] h^2 + o(h^2),$$

where the explicit forms of $l_1(x, f)$ and $l_2(x, f)$ are respectively given in Tables 3 *and* 4*. (ii) The variance of the TS-MBC estimator with the kernel L admits the following expansion*

$$\text{Var}(\tilde{f}_{TS,L}(x)) = \frac{f(x)(1 - f(x))}{n(1 - c)^2} \left(K_{L,h}(x) - c K_{L,h/c}(x) \right)^2 + o\left(\frac{1}{n} \right).$$

Table 3 Explicit forms of $l_1(x, f)$

Kernel(L)	$l_1(x, f)$
DT	$\left(\log(a+1)S(a) - 2\sum\limits_{k=1}^{a} k^2 \log(k)\right)\frac{f^{(2)}(x)}{2} + \left(\log(a+1)R(a) - 2\sum\limits_{k=1}^{a} k^4 \log(k)\right)\frac{f^{(4)}(x)}{24}$
DDU	$\left(\frac{m}{2} - x - \frac{x}{m-1}\right)f^{(1)}(x) + \frac{1}{2}\left(\frac{m(2m-1)}{6} + x^2 + \frac{x^2}{m-1} - xm\right)f^{(2)}(x) +$
	$\frac{1}{6}\left(\frac{m^2(m-1)}{4} - x^3 - \frac{x^3}{m-1} + \frac{3x^2 m}{2} - \frac{xm(2m-1)}{2}\right)f^{(3)}(x)$
	$+ \frac{1}{24}\left(\frac{m[6(m-1)^3 + 9(m-1)^2 + m - 2]}{30} + x^4 - \frac{x^4}{m-1} + x^2 m(2m-1) - xm^2(m-1)\right)f^{(4)}(x)$
WVR	$\frac{1}{2}f^{(2)}(x) + \frac{121}{24}f^{(4)}(x)$
LR	$2xf^{(1)}(x) - (x^2 - 1)f^{(2)}(x) + \frac{1}{3}x^3 f^{(3)}(x) - (\frac{1}{12}x^4 - \frac{1}{12})f^{(4)}(x)$

Table 4 Explicit forms of $l_2(x, f)$

Kernel(L)	$l_2(x, f)$
DT	$\left(\frac{\log^2(a+1)}{2}S(a) - \sum\limits_{k=1}^{a} k^2 \log^2(k)\right)\frac{f^{(2)}(x)}{2} + \left(\frac{\log^2(a+1)}{2}R(a) - \sum\limits_{k=1}^{a} k^4 \log^2(k)\right)\frac{f^{(4)}(x)}{24}$
DDU	0
WVR	$-\frac{3}{2}f^{(2)}(x) + \frac{105}{8}f^{(4)}(x)$
LR	$2xf^{(1)}(x) + (x^2 - 4)f^{(2)}(x) + \frac{1}{3}x^3 f^{(3)}(x) + (\frac{1}{12}x^4 - \frac{4}{3})f^{(4)}(x)$

Proof The proof is given in the "Appendix".

Theorem 2 *Let $\tilde{f}_{JLN,L}$ be the JLN-MBC estimator with kernel L defined by* (3). *For a given target x, and assuming A1 and A2, it holds that:*

(i) *The bias of the JLN-MBC discrete kernel estimator takes the following approximation*

$$bias(\tilde{f}_{JLN,L}(x)) = -f(x)l_1(x, g)h^2 + o(h^2),$$

where $l_1(x, g)$ is obtained by replacing f by g in $l_1(x, f)$ with $g = g(x, f) = l_1(x, f)/f(x)$.

(ii) *The variance of the JLN-MBC discrete kernel estimator takes the following expression*

$$Var(\tilde{f}_{JLN,L}(x)) = \frac{f(x)(1 - f(x))}{n}K_{L,h}^2(x) + o\left(\frac{1}{n}\right).$$

Proof The proof is given in the "Appendix". □

3.3 Global property

We use the mean integrated squared error (MISE) as a criterion for the global property defined as

$$\text{MISE}(\tilde{f}_{MBC,L}) = \sum_{x \in \mathbb{T}} \text{bias}^2(\tilde{f}_{MBC,L}(x)) + \sum_{x \in \mathbb{T}} \text{Var}(\tilde{f}_{MBC,L}(x)),$$

where $\tilde{f}_{MBC,L}$ is the TS-L or the JLN-L kernel estimator. The mean integrated squared error (MISE) of the TS-L kernel estimator presented in (2) is given by

$$\text{MISE}(\widehat{f}_{TS,L}) = \frac{h^4}{c^2} \sum_{x \in \mathbb{T}} \left[\frac{1}{2} \left\{ \frac{l_{1,L}^2(x,f)}{f(x)} - l_{2,L}(x,f) \right\} \right]^2$$
$$+ \sum_{x \in \mathbb{T}} \frac{f(x)(1-f(x))}{n(1-c)^2} \left(K_{L,h}(x) - c K_{L,h/c}(x) \right)^2 + o\left(h^4 + \frac{1}{n} \right).$$

Similarly, the mean integrated squared error of the JLN-L kernel estimator presented in (3) is expressed as

$$\text{MISE}(\widehat{f}_{JLN,L}) = h^4 \sum_{x \in \mathbb{T}} f(x)^2 l_1^2(x,g) + \frac{1}{n} \sum_{x \in \mathbb{T}} f(x)\{1 - f(x)\} K_{L(x,h)}^2(x)$$
$$+ o\left(h^4 + \frac{1}{n} \right).$$

Remark 1 We can easily transmit the results when bias and variance have to be established at a point x. As we have seen, the bias remains unchanged and is uniformly of order $O(h^2)$ over the whole support. Moreover, the variance exhibits the following order:

$$Var(\tilde{f}_{TS,L}(x)) = Var(\tilde{f}_{JLN,L}(x)) = Var(\hat{f}_L(x)) = o(n^{-1}).$$

The two theorems demonstrate that both TS and JLN estimators are free of boundary bias. More importantly, these two MBC estimators reduce the order of magnitude in bias from $O(h)$ to $O(h^2)$, while their variances are still $o(n^{-1})$. The variance of JLN estimator is first-order asymptotically equivalent to that of the corresponding classical estimator. Besides, since the variance of TS-MBC estimators depends on $c \in (0, 1)$, the variance of these estimators tends to be larger than that of the classical estimator, but not least importantly, see, e.g., Hirukawa (2010) and Hirukawa and Sakudo (2014) for more details in continuous case.

3.4 Normalization

Neither $\tilde{f}_{TS,L}(x)$ nor $\tilde{f}_{JLN,L}(x)$ sum up to one. In general, MBC leads to lack of normalization, Hirukawa (2010) for example, argues that this issue can be resolved, and propose two renormalized beta MBC kernel density estimators. Taking the structures

of $\tilde{f}_{TS,L}(x)$ and $\tilde{f}_{TS,L}(x)$ into account. Following Hirukawa (2010), we adopt their macro approach to obtain the renormalized of our MBC estimators.

$$\tilde{f}_{TS,L}^{R}(x) = \frac{\tilde{f}_{TS,L}(x)}{\sum_{x\in\mathbb{T}}\tilde{f}_{TS,L}(x)},$$

$$\tilde{f}_{JLN,L}^{R}(x) = \frac{\tilde{f}_{JLN,L}(x)}{\sum_{x\in\mathbb{T}}\tilde{f}_{JLN,L}(x)}.$$

Since

$$\mathbb{E}\left(\sum_{x\in\mathbb{T}}\tilde{f}_{TS,L}(x)\right) = \sum_{x\in\mathbb{T}}\left(\mathbb{E}(\tilde{f}_{TS,L}(x))\right),$$

$$= 1 + \frac{1}{c}\sum_{x\in\mathbb{T}}\left[\frac{1}{2}\left\{\frac{l_1^2(x,f)}{f(x)} - l_2(x,f)\right\}\right]h^2 + o(h^2)$$

and

$$\mathbb{E}\left(\sum_{x\in\mathbb{T}}\tilde{f}_{JLN,L}(x)\right) = \sum_{x\in\mathbb{T}}\left(\mathbb{E}(\tilde{f}_{JLN,L}(x))\right),$$

$$= 1 - \sum_{x\in\mathbb{T}}f(x)l_1(x,g)h^2 + o(h^2),$$

biases of $\tilde{f}_{TS,L}^{R}(x)$ and $\tilde{f}_{JLN,L}^{R}(x)$ can be approximated by

$$bias\left(\tilde{f}_{TS,L}^{R}(x)\right) \sim \frac{1}{c}\left[\frac{1}{2}\left\{\frac{l_1^2(x,f)}{f(x)} - l_2(x,f)\right\} - \sum_{x\in\mathbb{T}}\frac{1}{2}\left\{\frac{l_1^2(x,f)}{f(x)} - l_2(x,f)\right\}\right]h^2,$$

$$bias\left(\tilde{f}_{JLN,L}^{R}(x)\right) \sim \left[-f(x)l_1(x,g) + \sum_{x\in\mathbb{T}}f(x)l_1(x,g)\right]h^2.$$

3.5 Choice of smoothing parameter for discrete MBC kernel estimators

We adapt in this section the popular unbiased cross validation (UCV) method. First, we consider the case of TS-kernel estimators based on the kernel L. The optimal smoothing parameter $h_{TS,L}^{opt}$ is given by

$$h_{TS,L}^{opt} = arg\ \min_{h}\ UCV_{TS,L}(h),$$

where

$$
UCV_{TS,L}(h) = \sum_{x \in \mathbb{T}} \tilde{f}^2_{TS-L}(x) - \frac{2}{(n-1)} \sum_{i=1}^{n} \tilde{f}^{(-i)}_{TS-L}(X_i)
$$

$$
= \sum_{x \in \mathbb{T}} \left\{ \hat{f}_{L,h}(x) \right\}^{\frac{2}{1-c}} \left\{ \hat{f}_{L,h/c}(x) \right\}^{-\frac{2c}{1-c}} - \frac{2}{n(n-1)}
$$

$$
\times \sum_{i=1}^{n} \left[\left\{ \sum_{j \neq i} K_{L(X_i,h)}(X_j) \right\}^{\frac{1}{1-c}} \left\{ \sum_{j \neq i} K_{L(X_i,h/c)}(X_j) \right\}^{-\frac{c}{1-c}} \right].
$$

In the case of JLN-kernel estimators, the expression of UCV takes the following form:

$$
UCV_{JLN,L}(h) = \sum_{x \in \mathbb{T}} \tilde{f}^2_{JLN,L}(x) - \frac{2}{(n-1)} \sum_{i=1}^{n} \tilde{f}^{(-i)}_{JLN,L}(X_i)
$$

$$
= \frac{1}{n^2} \sum_{x \in \mathbb{T}} \hat{f}^2_{L,h}(x) \left\{ \sum_{i=1}^{n} \frac{K_{L(x,h)}(X_i)}{\hat{f}_{L,h}(X_i)} \right\}^2
$$

$$
- \frac{2}{n(n-1)} \times \sum_{i=1}^{n} \sum_{j \neq i} K_{L(X_i,h)}(X_j) \frac{\hat{f}_{L,h}(X_i)}{\hat{f}_{L,h}(X_j)}
$$

and the bandwidth $h^{opt}_{JLN,L}$ is defined as follows

$$
h^{opt}_{JLN,L} = arg \ \min_{h} \ UCV_{JLN,L}(h).
$$

4 Illustrations from simulated data

This section investigates the performances of TS-DDU, TS-DT, JLN-DDU and JLN-DT kernel estimators considered in the previous section and compares their performances with the standard DDU and DT kernel estimators. Note that for the DT kernel, we used the arm $a = 2$, see for example Kokonendji and Senga Kiessé (2011). We consider six probability mass functions defined as follows:

(a) $\mathbf{F}_1$ a Poisson distribution with parameter $\lambda = 8$:

$$
f(x) = e^{-8} \frac{8^x}{x!}, \quad x \in \mathbb{N}.
$$

(b) $\mathbf{F}_2$ a mixture of three Poisson distributions with parameters $\mu_1 = 3$, $\mu_2 = 12$ and $\mu_3 = 24$:

$$
f(x) = \frac{1}{3} e^{-3} \frac{3^x}{x!} + \frac{1}{3} e^{-12} \frac{12^x}{x!} + \frac{1}{3} e^{-24} \frac{24^x}{x!}, \quad x \in \mathbb{N}
$$

(c) $\mathbf{F}_3$ a Geometric distribution with parameter $p = 0.1$:

$$f(x) = 0.1 \cdot (0.9)^{x-1}, \quad x \in \mathbb{N}.$$

(d) $\mathbf{F}_4$ a mixture of Poisson and Geometric distributions with parameters $\mu = 10$ and $p = 0.1$:

$$f(x) = \frac{2}{5} \cdot e^{-10} \frac{10^x}{x!} + \frac{3}{5} \cdot 0.1 \cdot (0.9)^{x-1}, \quad x \in \mathbb{N}.$$

(e) $\mathbf{F}_5$ a negative binomial distribution with parameters $n_1 = 20$ and $p = 2/3$:

$$f(x) = \frac{(19 + x)!}{x!19!} \left(\frac{2}{3}\right)^{20} \left(\frac{1}{3}\right)^x \quad x \in \mathbb{N}.$$

(f) $\mathbf{F}_6$ a binomial distribution with parameters $n_1 = 5$ and $p = 0.1$:

$$f(x) = \frac{5!}{x!(5 - x)!} 0.1^x \cdot 0.9^{5-x}, \quad x \in \{0, 1, \ldots, 5\}.$$

Note that for these considered pmfs, 500 replications of sizes $n = 20, 50, 100$ and 200 are generated. The MBC-DDU (TS-DDU and JLN-DDU) and MBC-DT (TS-DT and JLN-DT) discrete kernel estimators are applied to estimate the pmfs generated from Poisson($\lambda = 8$), a mixture of three Poisson with ($\mu_1 = 3, \mu_2 = 12, \mu_3 = 24$), a Geometric($p = 0.1$), mixture of Poisson($\mu = 10$) and Geometric($p = 0.1$) distributions, a negative binomial distribution BN($n_1 = 20, p = 2/3$) and a binomial distribution B($n_1 = 5, p = 0.1$). Note that, for our simulations the value of c which is obtained in the sense of mean integrated squared error (MISE) see, e.g., Hirukawa (2010) is fixed at $c = 0.5$. We use the standard DDU and DT kernel estimators to compare their performance with the MBC-DDU and MBC-DT kernel estimators. For the choice of the bandwidth, we use the UCV technique proposed in the previous section. Finally, the performances of the different standard estimators and MBC estimators are examined via the integrated squared error (ISE) and the integrated squared bias (ISB) given respectively as follows:

$$\text{ISE} := \sum_{x \in \mathbb{T}} \left[\widehat{f}(x) - f(x)\right]^2$$

and

$$\text{ISB} := \sum_{x \in \mathbb{T}} \left[\mathbb{E}\{\widehat{f}(x)\} - f(x)\right]^2.$$

We also compute the integrated variance IV given by

$$\text{IV} := \sum_{x \in \mathbb{T}} \left[Var\{\widehat{f}(x)\}\right].$$

Through simulation results (Tables 5, 6, 7), we can observe immediately that:

Table 5 Some expected values of ISE based on 500 replications for the previous considered pmfs

f	Kernel	Estimator	$n = 20$	$n = 50$	$n = 100$	$n = 200$
F_1	DDU	$\widehat{f}$	0.042787	0.042609	0.042574	0.042563
		$\tilde{f}_{TS}$	**0.033705**	**0.032229**	**0.032031**	**0.031361**
		$\tilde{f}_{JLN}$	0.034777	0.034401	0.032380	0.031987
	DT	$\widehat{f}$	0.013403	0.012382	0.007470	0.001309
		$\tilde{f}_{TS}$	**0.002905**	**0.001878**	**0.001366**	**0.000366**
		$\tilde{f}_{JLN}$	0.025200	0.004287	0.001050	0.000034
F_2	DDU	$\widehat{f}$	0.053881	0.053495	0.053313	0.053182
		$\tilde{f}_{TS}$	0.022398	0.022391	0.022377	0.020300
		$\tilde{f}_{JLN}$	**0.006122**	**0.005489**	**0.005135**	**0.004648**
	DT	$\widehat{f}$	0.007583	0.006054	0.005454	0.005031
		$\tilde{f}_{TS}$	0.016221	0.014836	0.013071	0.009145
		$\tilde{f}_{JLN}$	**0.005181**	**0.004894**	**0.001957**	**0.001172**
F_3	DDU	$\widehat{f}$	0.026628	0.026608	0.026578	0.026070
		$\tilde{f}_{TS}$	0.014414	0.007664	**0.004681**	**0.004666**
		$\tilde{f}_{JLN}$	**0.010331**	**0.006939**	0.005592	0.005442
	DT	$\widehat{f}$	0.008378	0.007500	0.007187	0.005424
		$\tilde{f}_{TS}$	0.008790	0.007584	0.004624	0.004080
		$\tilde{f}_{JLN}$	**0.006373**	**0.005404**	**0.003235**	**0.003192**
F_4	DDU	$\widehat{f}$	0.035007	0.027178	0.027054	0.027009
		$\tilde{f}_{TS}$	0.038446	0.038436	0.038410	0.026179
		$\tilde{f}_{JLN}$	**0.024550**	**0.024524**	**0.022049**	**0.010206**
	DT	$\widehat{f}$	0.004518	0.002838	0.002777	0.002506
		$\tilde{f}_{TS}$	0.008109	0.006315	0.001436	0.000200
		$\tilde{f}_{JLN}$	**0.002297**	**0.001980**	**0.000230**	**0.000151**
F_5	DDU	$\widehat{f}$	0.061334	0.060788	0.055170	0.050989
		$\tilde{f}_{TS}$	0.031456	0.030866	0.023564	**0.018992**
		$\tilde{f}_{JLN}$	**0.022082**	**0.019926**	**0.019238**	0.019047
	DT	$\widehat{f}$	0.009689	0.000810	0.000145	0.000030
		$\tilde{f}_{TS}$	0.005364	0.004618	0.0035203	0.001700
		$\tilde{f}_{JLN}$	0.000159	0.000256	0.000108	0.000009
F_6	DDU	$\widehat{f}$	0.000014	0.000006	0.000002	0.000001
		$\tilde{f}_{TS}$	0.000008	0.000004	0.0000026	0.0000022
		$\tilde{f}_{JLN}$	**0.000008** $\times 10^{-1}$	**0.000005** $\times 10^{-1}$	**0.000002** $\times 10^{-1}$	**0.000001** $\times 10^{-1}$
	DT	$\widehat{f}$	**0.002889**	0.001826	0.001808	0.001434
		$\tilde{f}_{TS}$	0.003393	**0.001583**	0.001542	0.000857
		$\tilde{f}_{JLN}$	0.004142	0.003615	**0.000836**	**0.000537**

Bold values indicate the best results, which are in general for the MBC approaches

Table 6 Empirical ISB values based on 500 replications for the previous considered pmfs

f	Kernel	Estimator	$n = 20$	$n = 50$	$n = 100$	$n = 200$
F_1	DDU	$\widehat{f}$	0.001044	0.001041	0.001030	0.001028
		$\tilde{f}_{TS}$	0.000416	0.000135	**0.000051**	**0.000023**
		$\tilde{f}_{JLN}$	**0.000204**	**0.000129**	0.000065	0.000042
	DT	$\widehat{f}$	0.006911	0.003771	0.001309	0.000225
		$\tilde{f}_{TS}$	**0.000508**	0.000350	0.000239	**0.000030**
		$\tilde{f}_{JLN}$	0.004287	**0.000073**	**0.000034**	0.000033
F_2	DDU	$\widehat{f}$	0.004498	0.004448	0.004447	0.004463
		$\tilde{f}_{TS}$	0.001642	0.001479	0.001448	0.001521
		$\tilde{f}_{JLN}$	**0.000258**	**0.000145**	**0.000140**	**0.000132**
	DT	$\widehat{f}$	0.006280	0.005772	0.005392	0.004935
		$\tilde{f}_{TS}$	0.007141	0.003604	0.002514	0.001346
		$\tilde{f}_{JLN}$	**0.005146**	**0.002067**	**0.001280**	**0.001003**
F_3	DDU	$\widehat{f}$	0.002007	0.001993	0.001992	0.001941
		$\tilde{f}_{TS}$	0.002461	0.000488	0.000475	0.000020
		$\tilde{f}_{JLN}$	**0.000238**	**0.000137**	**0.000063**	**0.000007**
	DT	$\widehat{f}$	0.007637	0.007351	0.005861	0.003762
		$\tilde{f}_{TS}$	0.007279	0.006882	0.004247	0.003259
		$\tilde{f}_{JLN}$	**0.006215**	**0.003963**	**0.003135**	**0.002800**
F_4	DDU	$\widehat{f}$	0.003082	0.002755	0.002382	0.002365
		$\tilde{f}_{TS}$	0.003401	0.003200	0.002375	0.002001
		$\tilde{f}_{JLN}$	**0.002176**	**0.002152**	**0.001894**	**0.000968**
	DT	$\widehat{f}$	0.003027	0.002814	0.002709	0.001951
		$\tilde{f}_{TS}$	0.005236	0.004054	0.000979	**0.000011**
		$\tilde{f}_{JLN}$	**0.000503**	**0.000354**	**0.000216**	0.000093
F_5	DDU	$\widehat{f}$	0.003155	0.003933	0.003800	0.002954
		$\tilde{f}_{TS}$	0.001244	0.000203	0.000180	**0.000043**
		$\tilde{f}_{JLN}$	**0.000308**	**0.000146**	**0.000077**	0.000054
	DT	$\widehat{f}$	0.002525	0.000336	0.000115	0.000083
		$\tilde{f}_{TS}$	0.004808	0.004231	0.001527	0.000898
		$\tilde{f}_{JLN}$	**0.000130**	**0.000219**	**0.000016**	**0.000003**
F_6	DDU	$\widehat{f}$	0.000013	0.000006	0.000002	0.000001
		$\tilde{f}_{TS}$	0.000007	0.000004	0.0000026	0.0000022
		$\tilde{f}_{JLN}$	**0.000007** $\times 10^{-1}$	**0.000004** $\times 10^{-1}$	**0.000002** $\times 10^{-1}$	**0.000001** $\times 10^{-1}$
	DT	$\widehat{f}$	0.000972	0.000703	0.000692	0.000212
		$\tilde{f}_{TS}$	**0.000527**	0.000363	0.000334	0.000444
		$\tilde{f}_{JLN}$	0.004053	**0.000287**	**0.000155**	**0.000021**

Bold values indicate the best results, which are in general for the MBC approaches

Table 7 Empirical IV values based on 500 replications for the previous considered pmfs

f	Kernel	Estimator	$n = 20$	$n = 50$	$n = 100$	$n = 200$
F_1	DDU	$\widehat{f}$	0.041743	0.041568	0.041544	0.041535
		$\tilde{f}_{TS}$	**0.033289**	**0.032094**	**0.031980**	**0.031338**
		$\tilde{f}_{JLN}$	0.034573	0.034272	0.032315	0.031945
	DT	$\widehat{f}$	0.006492	0.008611	0.006161	0.001084
		$\tilde{f}_{TS}$	**0.002397**	**0.001528**	0.001127	0.000309
		$\tilde{f}_{JLN}$	0.020913	0.004213	**0.001016**	**0.000001**
F_2	DDU	$\widehat{f}$	0.049383	0.049047	0.048866	0.048719
		$\tilde{f}_{TS}$	0.020756	0.020912	0.020929	0.018779
		$\tilde{f}_{JLN}$	**0.005864**	**0.005344**	**0.004995**	**0.004516**
	DT	$\widehat{f}$	0.001303	**0.000282**	**0.000062**	**0.000096**
		$\tilde{f}_{TS}$	0.009080	0.011232	0.010557	0.007799
		$\tilde{f}_{JLN}$	**0.000035**	0.002827	0.000677	0.000169
F_3	DDU	$\widehat{f}$	0.024621	0.024615	0.024586	0.024129
		$\tilde{f}_{TS}$	0.011953	0.007176	**0.004206**	**0.004646**
		$\tilde{f}_{JLN}$	**0.010093**	**0.006802**	0.005529	0.005435
	DT	$\widehat{f}$	0.000741	0.000149	0.001326	0.001662
		$\tilde{f}_{TS}$	0.001511	**0.000702**	0.000377	0.000821
		$\tilde{f}_{JLN}$	**0.000158**	0.001441	**0.000099**	**0.000392**
F_4	DDU	$\widehat{f}$	0.031925	0.024423	0.024672	0.024644
		$\tilde{f}_{TS}$	0.035045	0.035236	0.036035	0.024178
		$\tilde{f}_{JLN}$	**0.022374**	**0.022372**	**0.010035**	**0.009238**
	DT	$\widehat{f}$	**0.001491**	**0.000024**	0.000068	0.000555
		$\tilde{f}_{TS}$	0.002873	0.002261	0.000457	0.000189
		$\tilde{f}_{JLN}$	0.001794	0.001626	**0.000014**	**0.000058**
F_5	DDU	$\widehat{f}$	0.058179	0.056855	0.051370	0.048035
		$\tilde{f}_{TS}$	0.030212	0.030663	0.023384	**0.018949**
		$\tilde{f}_{JLN}$	**0.021774**	**0.019780**	**0.019161**	0.018993
	DT	$\widehat{f}$	0.007164	0.000474	0.000029	0.000224
		$\tilde{f}_{TS}$	0.000556	0.000387	0.001993	0.000802
		$\tilde{f}_{JLN}$	**0.000029**	**0.000037**	**0.000091**	**0.000005**
F_6	DDU	$\widehat{f}$	0.000007×10^{-1}	0.000001×10^{-1}	0.000003×10^{-2}	0.000005×10^{-3}
		$\tilde{f}_{TS}$	0.000008×10^{-1}	0.000002×10^{-1}	0.000004×10^{-2}	0.000001×10^{-2}
		$\tilde{f}_{JLN}$	$\mathbf{0.000001 \times 10^{-1}}$	$\mathbf{0.000005 \times 10^{-2}}$	$\mathbf{0.000001 \times 10^{-2}}$	$\mathbf{0.000002 \times 10^{-3}}$
	DT	$\widehat{f}$	**0.001917**	0.001123	0.001115	0.001221
		$\tilde{f}_{TS}$	0.002866	0.001220	0.001208	**0.000413**
		$\tilde{f}_{JLN}$	0.003328	**0.000657**	**0.000681**	0.000516

Bold values indicate the best results, which are in general for the MBC approaches
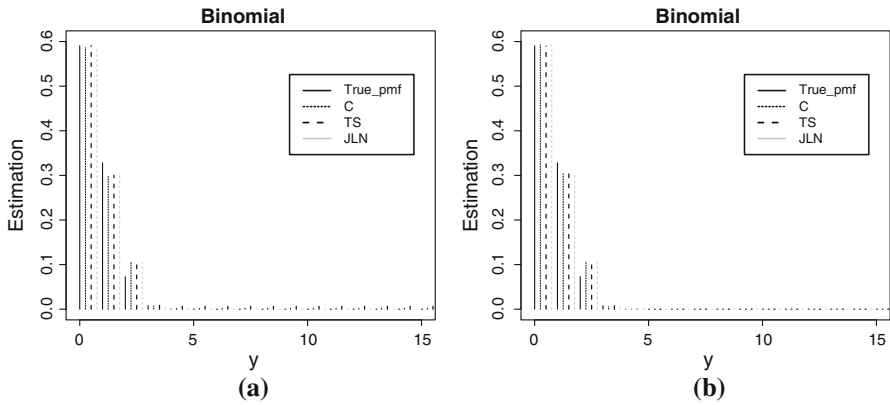
Fig. 1 The pmf estimation of Binomial data with $n = 200$ using the standard and MBC discrete associated kernel estimators. **a** DDU kernel. **b** DT($a = 2$) kernel

1. for all estimators, the means of ISE and ISB based on 500 simulations decrease as sample size n increases, which indicates that our estimators are consistent;
2. in terms of ISE and ISB, the performances of JLN-DDU and TS-DDU kernel estimators are mixed depending on the distribution. For example, in case of the binomial distribution, the JLN-DDU kernel estimator in general works better than the other competitors in the sense of ISB;
3. for all sample sizes the TS-DDU, JLN-DDU, TS-DT, JLN-DT kernel estimators outperform the standard DDU and DT kernel estimators in the senses of ISE and ISB.

Note that the performances of the WVR and LR kernels are similar to those obtained by the DDU kernel, for this reason and to avoid making the manuscript more cumbersome, we have considered the DDU kernel rather than the WVR nor the LR in simulations and empirical illustrations.

The comparison is also illustrated in Figs. 1 and 2. We have plotted the estimates for sample size $n = 200$ with DDU and DT($a = 2$) kernel. The solid lines represent the true pmf, the dotted lines represent the classical (C) estimator with DDU or DT kernel, the TS-DDU and TS-DT estimator are represented by the dashed lines, the JLN-DDU and JLN-DT estimators by the solid lines in gray. The plot shows that in general the MBC-DDU and MBC-DT estimators improves the standard DDU and DT kernel estimator for all pmfs. The smoothing quality is considered satisfactory. We have obtained the best smoothing quality by using the MBC-DDU (TS-DDU and JLN-DDU) or the MBC-DT (TS-DT and JLN-DT) kernel estimators.

The TS-MBC estimator depends on two smoothing parameters h and h/c, these two smoothing parameters also play a role in determining the boundary region. Controlling both h and h/c is a cumbersome task. Because $0 < c < 1$, the pmf estimator using h/c tends to be oversmoothed, which is potentially a source of a large bias in every TS-MBC estimator. On the other hand, we make c too long in order to have a reasonable short of h/c.
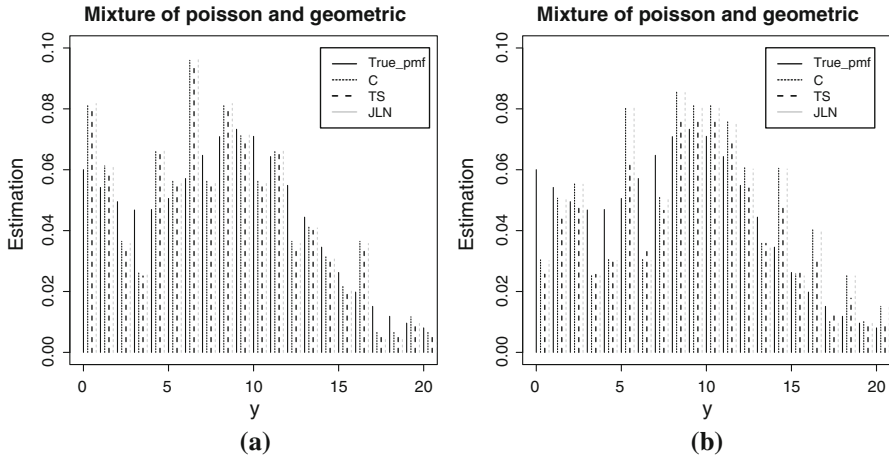
Fig. 2 The pmf estimation of mixture of Poisson and Geometric distributions data with $n = 200$ using the standard and MBC discrete associated kernel estimators. **a** DDU kernel. **b** DT($a = 2$) kernel

Table 8 Summary Statistics for Travel Mode Choice Data

Categories	Air	Train	Bus	Car	Total
Number choosing	58	63	30	59	210
p	0.28	0.30	0.14	0.28	1

Table 9 Data of longevity of adult insects observed in days

Days	1	2	3	4	5	6	7	8	9	Total
Observed frequencies	29	16	22	8	2	4	0	0	1	82

5 Illustrations from real data

To complete our Monte Carlo simulations, we consider in this section two real data applications. First we illustrate the performances of the MBC techniques for discrete kernel estimators based on DDU and DT($a = 2$) kernels for the travel mode choice (between Sydney and Melbourne, Australia) data from Greene (2011). This data consists of $n = 210$ observations and $m = 4$ categories (air, train, bus and car). Note that the relative proportions of air, train, bus and car are 0.28, 0.30, 0.14 and 0.28 respectively. Table 8 provides the summary statistics of these real data observations. The second real application is related to the development of an insect parasite called the spiraling whitefly and observed in Republic of Congo, see Senga Kiessé and Mizère (2012). This insect pest plant causes some damages as sucking the sap, decreasing photosynthesis activity and drying up the leaves. The congolese biologists are searching for a suitable modeling by studying some count data characterizing the growth of spiraling whitefly such as the longevity of the adult insect (see Table 9).

Table 10 Results from bandwidth and ISE^0 by discrete and MBC kernels estimators of real data from the travel mode choice (between Sydney and Melbourne, Australia) of $n = 210$

Bold values indicate the best results, which are for the MBC approaches

Estimators	h_{ucv}	ISE^0
$\widehat{f}_{DDU}$	0.015000	0.082230
$\tilde{f}_{TS-DDU}$	0.999361	**0.000708**
$\tilde{f}_{JLN-DDU}$	0.206504	0.017897
$\widehat{f}_{DT}$	0.043106	0.020105
$\tilde{f}_{TS-DT}$	0.995202	**0.015453**
$\tilde{f}_{JLN-DT}$	0.952180	0.016453

Table 11 Results from bandwidth and ISE^0 by discrete and MBC kernels estimators of real data from longevity of adult insects observed in days of $n = 82$

Bold values indicate the best results, which are for the MBC approaches

Estimators	h_{ucv}	ISE^0
$\widehat{f}_{DDU}$	0.047749	0.097224
$\tilde{f}_{TS-DDU}$	0.999513	**0.005144**
$\tilde{f}_{JLN-DDU}$	0.556757	0.042317
$\widehat{f}_{DT}$	0.084456	0.049919
$\tilde{f}_{TS-DT}$	0.920273	**0.020639**
$\tilde{f}_{JLN-DT}$	0.902025	0.043699

Now we apply the MBC-DDU and MBC-DT kernel estimators to estimate the pmfs for the considered real data. The value of c is fixed at 0.5 for TS-DDU and TS-DT kernel estimators. The standard DDU and DT kernel estimators are also used for comparison.

In order to measure the performance of all estimators, we simply use the practical integrated squared error given by [see Kokonendji and Senga Kiessé (2011)]:

$$\text{ISE}^0 := \sum_{x \in \mathbb{N}} \left[\widehat{f}(x) - f_0(x) \right]^2,$$

where $f_0(x)$ is the empirical (naive) estimator. Categorical independent variables can be used in a nonparametric pmf estimation, but they need to be coded. In our study we use the following code: 1="air"; 2="train"; 3="bus"; 4="car". The bandwidths for the estimators are chosen by using the popular UCV technique. The obtained values of h_{ucv} and ISE^0 for both applications are given in Tables 10 and 11.

We can see that in terms of the ISE^0, the MBC discrete kernel estimators with UCV bandwidths perform better than the standard kernel estimators for both applications. We have also plotted the estimations obtained by the classical (C) estimator and the MBC (TS and JLN) estimators with the Dirac Discrete Uniform and the triangular kernels for the second real data of longevity of adults insects observed in days. From Fig. 3, we observe that the smoothing quality is satisfactory. But the smoothing contributed by the JLN estimator is more suitable.
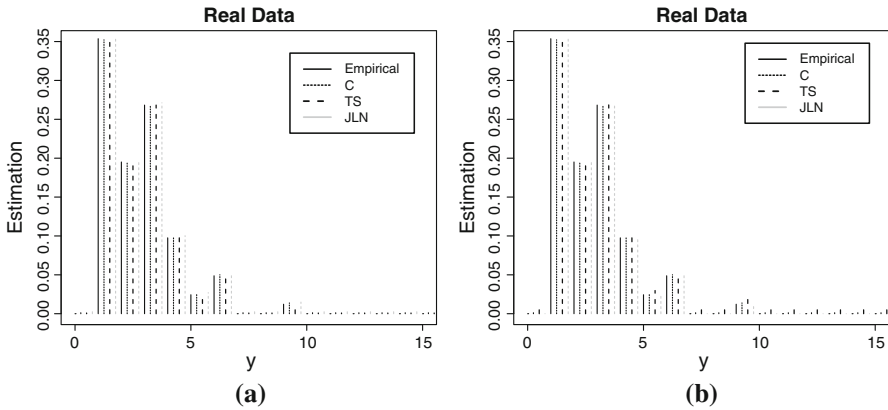
Fig. 3 The pmf estimation of real data of longevity of adults insects observed in days with $n = 82$ using the standard and MBC discrete associated kernel estimators. **a** DDU kernel. **b** DT($a = 2$) kernel

6 Conclusion

This paper has proposed two multiplicative bias correction (MBC) techniques for discrete kernels in the context of probability mass function (pmf) estimation. We have shown that these two classes of MBC techniques improve the order of magnitude in bias from $O(h)$ to $O(h^2)$. The performances of the MBC techniques for discrete kernel estimators (TS-DDU, TS-DT, JLN-DT and JLN-DDU kernel estimators) with unbiased cross-validation (UCV) bandwidth selectors are investigated through a simulation study and a real data application for count and categorical data. In general, the MBC discrete kernel estimators perform better than the standard discrete kernel estimators in the sense of integrated squared error (ISE) and integrated squared bias (ISB).

This paper deals only with the univariate case. An extension is obviously given by the estimation of multivariate pmfs. We are aware of two recent publications, which deal with multivariate (discrete) kernels. Kokonendji and Somé (2015) investigated multivariate kernels for the estimation of the density of continuously distributed random vectors. Moreover, discrete multivariate kernels have been studied by Belaid et al. (2016). In the latter one, a Bayesian bandwidth selection method for those kernels has been proposed.

Hence, let $X = \{(X_{i1}, \ldots, X_{id}), i = 1, \ldots, n\}$ be a sample of i.i.d. random vectors of dimension $d \geq 1$. When following the approach by Belaid et al. (2016), we define the multivariate version of a discrete kernel with diagonal bandwidth matrix $H = diag(h_1, \ldots, h_d)$ and kernel L according to

$$\hat{f}_L(\mathbf{x}) := \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{d} K_{L(x_j, h_j)}^{[j]}(X_{ij}),$$

where $K_{L(x_j, h_j)}^{[j]}(X_{ij})$ denotes the univariate discrete kernel studied in this paper and

$$\mathbf{x} := (x_1, \ldots, x_d) \in \mathbb{T}^d := \times_{j=1}^{d} \mathbb{T}_1^{[j]} \subseteq \mathbb{Z}^d$$

denotes the target vector. Moreover, $\mathbb{T}^d$ denotes the support of the underlying pmf f, which has to be estimated at $\mathbf{x}$.

In view of our univariate findings and in analogy of the asymmetric kernel based approach by Funke and Kawka (2015), we define the multivariate version of the TS estimator as

$$\hat{f}_{TS,L}(\mathbf{x}) := \left(\hat{f}_L(\mathbf{x})\right)^{\frac{1}{1-c}} \left(\hat{f}_L(\mathbf{x})\right)^{-\frac{c}{1-c}}.$$

In an analogous way, the multivariate JLN estimator is defined according to

$$\hat{f}_{JLN}(\mathbf{x}) := \hat{f}_L(\mathbf{x}) \frac{1}{n} \sum_{i=1}^{n} \frac{\prod_{j=1}^{d} K_{L(x_j,h_j)}^{[j]}(X_{ij})}{\hat{f}_L(\mathbf{X}_i)}.$$

Under appropriate assumptions, it can be shown that the mean squared errors of both estimators are given by

$$MSE\left(\hat{f}_{TS,L}(\mathbf{x})\right) = O\left(h^4 + \frac{\prod_{j=1}^{d}\left(K_{L(x_j,h)}(x_j) - cK_{L(x_j,h/c)}(x_j)\right)^2}{n}\right) \text{ as } n\to\infty,$$

as well as

$$MSE\left(\hat{f}_{JLN}(\mathbf{x})\right) = O\left(h^4 + \frac{1}{n}\prod_{j=1}^{d} K_{L(x_j,h)}^2(x_j)\right) \text{ as } n\to\infty,$$

where, for the sake of simplicity, the bandwidth vector h is given by $h \equiv h_1 = \cdots = h_d$. Analytical exact expressions of both bias terms are under investigation and will be covered in a following paper.

Appendix

We present a sketch of proofs of Theorems 1 and 2. We provide the proofs when the discrete triangular kernel is used. The proofs using the other kernels can be given similarly.

Sketch of the proof of Theorem 1

Bias

First, note that $\mathbb{E}\left(\widehat{f}_{DT,h}(x)\right) = \mathbb{E}(f(\mathcal{T}))$, where the random variable $\mathcal{T} \sim DT(a; x, h)$. By using a fourth-order discrete Taylor expansion around $\mathcal{T} = x$ for

$$I_h(x) = \mathbb{E}(\widehat{f}_{DT,h}) = \sum K_{DT,h}(y) f(y) = \mathbb{E}(f(\mathcal{T})),$$

we have

$$I_h(x) = f(x) + \sum_{j=1}^{4} \frac{f^{(j)}}{j!} \mathbb{E}(\mathcal{T} - x)^j + o(\mathbb{E}(\mathcal{T} - x)^4).$$

By using the property of the discrete triangular random variable and a Taylor expansion around $h = 0$,

$$\mathbb{E}(\mathcal{T} - x) = 0,$$

$$\mathbb{E}(\mathcal{T} - x)^2 = \left\{ \log(a+1)S(a) - 2\sum_{k=1}^{a} k^2 \log(k) \right\} h$$

$$+ \left\{ \frac{\log^2(a+1)}{2} S(a) - \sum_{k=1}^{a} k^2 \log^2(k) \right\} h^2 + o(h^2),$$

$$\mathbb{E}(\mathcal{T} - x)^3 = 0,$$

$$\mathbb{E}(\mathcal{T} - x)^4 = \left\{ \log(a+1)R(a) - 2\sum_{k=1}^{a} k^4 \log(k) \right\} h$$

$$+ \left\{ \frac{\log^2(a+1)}{2} R(a) - \sum_{k=1}^{a} k^4 \log^2(k) \right\} h^2 + o(h^2),$$

where

$$R(a) = \frac{2}{5}a^5 + a^4 + \frac{2}{3}a^3 - \frac{1}{15}a.$$

The Taylor expansion of $I_h(x)$ around $h = 0$ is then given by

$$I_h(x) = f(x) \left\{ 1 + \frac{l_1(x, f)}{f(x)} h + \frac{l_2(x, f)}{f(x)} h^2 + o(h^2) \right\},$$

where

$$l_1(x, f) = \frac{f^{(2)}(x)}{2} \left(\log(a+1)S(a) - 2\sum_{k=1}^{a} k^2 \log(k) \right)$$

$$+ \frac{f^{(4)}(x)}{24} \left(\log(a+1) R(a) - 2 \sum_{k=1}^{a} k^4 \log(k) \right)$$

and

$$l_2(x, f) = \frac{f^{(2)}(x)}{2} \left(\frac{\log^2(a+1)}{2} S(a) - \sum_{k=1}^{a} k^2 \log^2(k) \right)$$

$$+ \frac{f^{(4)}(x)}{24} \left(\frac{\log^2(a+1)}{2} R(a) - \sum_{k=1}^{a} k^4 \log^2(k) \right).$$

Similarly, $I_{h/c}(x) = \mathbb{E}\left(\widehat{f}_{DT,h/c}(x)\right)$ can be approximated by

$$I_{h/c}(x) = f(x) \left\{ 1 + \frac{1}{c} \frac{l_1(x, f)}{f(x)} h + \frac{1}{c^2} \frac{l_2(x, f)}{f(x)} h^2 + o(h^2) \right\}.$$

Now, we define

$$\widehat{f}_{DT,h}(x) = I_h(x) + Z$$

and

$$\widehat{f}_{DT,h/c}(x) = I_{h/c}(x) + W.$$

The estimator $\tilde{f}_{TS,DT}$ can be written as follows:

$$\tilde{f}_{TS,DT} = \{I_h(x)\}^{\frac{1}{1-c}} \left\{ 1 + \frac{Z}{I_h(x)} \right\}^{\frac{1}{1-c}} \{I_{h/c}(x)\}^{-\frac{c}{1-c}} \left\{ 1 + \frac{W}{I_{h/c}(x)} \right\}^{-\frac{c}{1-c}}.$$

Using the expansion $(1+t)^\alpha = 1 + \alpha t + o(t^2)$, we then have

$$\tilde{f}_{TS,DT}(x) = \{I_h(x)\}^{\frac{1}{1-c}} \{I_{h/c}(x)\}^{-\frac{c}{1-c}} + \frac{1}{1-c} Z \left\{ \frac{I_h(x)}{I_{h/c}(x)} \right\}^{-\frac{c}{1-c}}$$

$$- \frac{c}{1-c} W \left\{ \frac{I_h(x)}{I_{h/c}(x)} \right\}^{\frac{1}{1-c}} + O\left\{ (Z+W)^2 \right\}. \tag{4}$$

Based on Assumption 2 and using the same calculations as in Hirukawa (2010) and Terrell and Scott (1980), we can show easily that

$$\mathbb{E}\left(\tilde{f}_{TS,DT}(x)\right) = f(x) + \frac{1}{c} \left[\frac{1}{2} \left\{ \frac{l_1^2(x, f)}{f(x)} - l_2(x, f) \right\} \right] h^2 + o(h^2).$$

Variance

For the variance, from Eq. (4) we have

$$\text{Var}\left(\tilde{f}_{TS,DT}(x)\right) = \mathbb{E}\left(\frac{1}{1-c}Z - \frac{c}{1-c}W\right)^2 + o(n^{-1})$$

$$= \text{Var}\left(\frac{1}{1-c}\widehat{f}_{DT,h}(x) - \frac{c}{1-c}\widehat{f}_{DT,h/c}(x)\right) + o(n^{-1})$$

$$= \frac{1}{(1-c)^2}\text{Var}\left(\widehat{f}_{DT,h}(x)\right) + \frac{c^2}{(1-c)^2}\text{Var}\left(\widehat{f}_{DT,h/c}(x)\right)$$

$$- \frac{2c}{(1-c)^2}\text{cov}\left(\widehat{f}_{DT,h}(x), \widehat{f}_{DT,h/c}(x)\right).$$

First, note that the terms $\text{Var}\left(\widehat{f}_{DT,h}(x)\right)$ and $\text{Var}\left(\widehat{f}_{DT,h/c}(x)\right)$ are given by [see Kokonendji et al. (2007)]:

$$\text{Var}\left(\widehat{f}_{DT,h}(x)\right) = f(x)(1 - f(x))K^2_{DT,h}(x) + o\left(\frac{1}{n}\right)$$

$$= \frac{f(x)}{n}(1 - f(x))\frac{(1+a)^{2h}}{P^2(a,h)} + o\left(\frac{1}{n}\right),$$

and

$$\text{Var}\left(\widehat{f}_{DT,h/c}(x)\right) = f(x)(1 - f(x))K^2_{DT,h/c}(x) + o\left(\frac{1}{n}\right)$$

$$= \frac{f(x)}{n}(1 - f(x))\frac{(1+a)^{2h/c}}{P^2(a,h/c)} + o\left(\frac{1}{n}\right).$$

Now,

$$\text{cov}\left(\widehat{f}_{DT,h}(x), \widehat{f}_{DT,h/c}(x)\right)$$

$$= \mathbb{E}\left(\widehat{f}_{DT,h}(x)\widehat{f}_{DT,h/c}(x)\right) - \mathbb{E}\left(\widehat{f}_{DT,h}(x)\right)\mathbb{E}\left(\widehat{f}_{DT,h/c}(x)\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left(K_{DT,h}(X_i)K_{DT,h/c}(X_j)\right) - \mathbb{E}\left(K_{DT,h}(X_i)\right)\mathbb{E}\left(K_{DT,h/c}(X_j)\right)$$

$$= \frac{1}{n}\mathbb{E}\left(K_{DT,h}(X_i)K_{DT,h/c}(X_i)\right) + \frac{(n-1)}{n}\mathbb{E}\left(K_{DT,h}(X_i)\right)\mathbb{E}\left(K_{DT,h/c}(X_j)\right)$$

$$- \mathbb{E}\left(K_{DT,h}(X_i)\right)\mathbb{E}\left(K_{DT,h/c}(X_j)\right)$$

$$= \frac{1}{n}\mathbb{E}\left(K_{DT,h}(X_i)K_{DT,h/c}(X_i)\right) - \frac{1}{n}\mathbb{E}\left(K_{DT,h}(X_i)\right)\mathbb{E}\left(K_{DT,h/c}(X_j)\right)$$

$$= \frac{1}{n} K_{DT,h}(x) K_{DT,h/c}(x) f(x) - \frac{1}{n} K_{DT,h}(x) K_{DT,h/c}(x) f^2(x) + o\left(\frac{1}{n}\right)$$

$$= \frac{1}{n} f(x)(1 - f(x)) \frac{(1+a)^h}{P(a,h)} \frac{(1+a)^{h/c}}{P(a,h/c)} + o\left(\frac{1}{n}\right).$$

Therefore, the variance of $\tilde{f}_{TS,DT}(x)$ is given by

$$\text{Var}\left(\tilde{f}_{TS,DT}(x)\right) = \frac{f(x)(1-f(x))}{n(1-c)^2} \left(\frac{(1+a)^h}{P(a,h)} - c\frac{(1+a)^{h/c}}{P(a,h/c)}\right)^2 + o\left(\frac{1}{n}\right),$$

which corresponds to the results in Theorem 1.

Sketch of the proof of Theorem 2

Bias

At first, the estimator $\tilde{f}_{JLN,DT}$ can be written as [see Hirukawa (2010)]

$$\tilde{f}_{JLN,DT}(x) = f(x) \left\{ 1 + \frac{\widehat{f}_{DT}(x) - f(x)}{f(x)} \right\} \{1 + (\psi(x) - 1)\},$$

where $\psi(x) = n^{-1} \sum_{i=1}^n K_{DT,h}(X_i)/\widehat{f}_{DT}(X_i)$. Then, we have

$$\mathbb{E}\left(\tilde{f}_{JLN,DT}(x)\right) = f(x) + f(x)\mathbb{E}\left\{\frac{\widehat{f}_{DT}(x) - f(x)}{f(x)}\right\} + f(x)\mathbb{E}\{\psi(x) - 1\}$$

$$+ f(x)\mathbb{E}\left\{\left(\frac{\widehat{f}_{DT}(x) - f(x)}{f(x)}\right)(\psi(x) - 1)\right\}.$$

By using Assumption 2 and the properties of DT random variables, the terms $\mathbb{E}\left\{\frac{\widehat{f}_{DT}(x) - f(x)}{f(x)}\right\}$, $\mathbb{E}\{\psi(x) - 1\}$ and $\mathbb{E}\left\{\left(\frac{\widehat{f}_{DT}(x) - f(x)}{f(x)}\right)(\psi(x) - 1)\right\}$ can be approximated following the same procedures as in Hirukawa (2010). Thus, $\mathbb{E}(\widehat{f}_{JLN-DT})$ is approximated by

$$\mathbb{E}(\tilde{f}_{JLN-DT}(x)) = f(x) - f(x)\left[\frac{1}{2}\left\{\log(a+1)S(a) - 2\sum_{k=1}^a k^2 \log(k)\right\} q^{(1)}(x)\right]h^2$$

$$- f(x)\left[\frac{1}{24}\left\{\log(a+1)R(a) - 2\sum_{k=1}^a k^4 \log(k)\right\} q^{(2)}(x)\right]h^2 + o(h^2),$$

$$\mathbb{E}(\tilde{f}_{JLN-DT}(x)) = f(x) - f(x)\left[\frac{1}{2}\left\{\log(a+1)S(a) - 2\sum_{k=1}^a k^2 \log(k)\right\} q^{(1)}(x)\right]$$

$$+ \frac{1}{24} \left\{ \log(a+1)R(a) - 2 \sum_{k=1}^{a} k^4 \log(k) \right\} q^{(2)}(x) \right] h^2 + o(h^2),$$

where $q(x) = l_1(x, f)/f(x)$ with $l_1(x, f)$ given in the proof of Theorem 1.

Variance

Note that following Hirukawa (2010) and Jones et al. (1995), we can show that $\tilde{f}_{JLN,DT}(x)$ is equivalent to

$$\tilde{f}_{JLN-DT}(x) = f(x) \frac{1}{n} \sum_{i=1}^{n} \frac{K_{DT,h}(X_i)}{f(X_i)}.$$

It follows that

$$
\begin{aligned}
\text{Var}\left(\tilde{f}_{JLN-DT}(x)\right) &= f^2(x)\frac{1}{n}\text{Var}\left\{\frac{K_{DT,h}(X_i)}{f(X_i)}\right\} \\
&= f^2(x)\frac{1}{n}\left\{\mathbb{E}\left(\frac{K_{DT,h}^2(X_i)}{f^2(X_i)}\right) - \left[\mathbb{E}\left(\frac{K_{DT,h}(X_i)}{f(X_i)}\right)\right]^2\right\} \\
&= \frac{f^2(x)}{n}\left\{\frac{K_{DT,h}^2(x)}{f(x)} - K_{DT,h}^2(x)\right\} + o\left(\frac{1}{n}\right) \\
&= \frac{f(x)}{n}(1 - f(x))\frac{(1+a)^{2h}}{P^2(a, h)} + o\left(\frac{1}{n}\right).
\end{aligned}
$$

Therefore, we obtain the approximation for the variance given in Theorem 2. $\quad\square$

References

Aitchison J, Aitken CGG (1976) Multivariate binary discrimination by the kernel method. Biometrika 63:413–420

Belaid N, Adjabi S, Zougab N, Kokonendji CC (2016) Bayesian bandwidth selection in discrete multivariate associated kernel estimators for probability mass functions. J Korean Stat Soc 45:557–567

Chu CY, Henderson DJ, Parmeter CF (2015) Plug-in bandwidth selection for kernel density estimation with discrete data. Econometrics 3:199–214

Funke B, Kawka R (2015) Nonparametric density estimation for multivariate bounded data using two non-negative multiplicative bias correction methods. Comput Stat Data Anal 92:148–162

Greene W (2011) Econometric analysis. Pearson, Cambridge

Hirukawa M (2010) Nonparametric multiplicative bias correction for kernel-type density estimation on the unit interval. Comput Stat Data Anal 54:473–495

Hirukawa M, Sakudo M (2014) Nonnegative bias reduction methods for density estimation using asymmetric kernels. Comput Stat Data Anal 75:112–123

Hirukawa M, Sakudo M (2015) Family of the generalised gamma kernels: a generator of asymmetric kernels for nonnegative data. J Nonparametric Stat 27:41–63

Jones MC, Foster PJ (1993) Generalized jackknifing and higher order kernels. J Nonparametric Stat 3:81–94

Jones MC, Linton O, Nielsen JP (1995) A simple bias reduction method for density estimation. Biometrika 82:327–338

Kokonendji CC, Senga Kiessé T (2011) Discrete associated kernels method and extensions. Stat Methodol 8:497–516

Kokonendji CC, Senga Kiessé T, Zocchi SS (2007) Discrete triangular distributions and non-parametric estimation for probability mass function. J Nonparametric Stat 19:241–254

Kokonendji CC, Somé SM (2015) On multivariate associated kernels for smoothing general density function. arXiv: 1502.01173

Racine JS, Li Q (2004) Nomparametric estimation of regression functions with both categorical and continuous data. J Econom 119:99–130

Senga Kiessé T, Mizère D (2012) Weighted Poisson and semiparametric kernel models applied for parasite growth. Aust N Z J Stat 55:1–13

Terrell GR, Scott DW (1980) On improving convergence rates for nonnegative kernel density estimators. Ann Stat 8:1160–1163

Wang M, Ryzin J (1980) A class of smooth estimators for discrete distributions. Biometrika 68:301–309

Zougab N, Adjabi S (2015) Multiplicative bias correction for generalized Birnbaum–Saunders kernel density estimators and application to nonnegative heavy tailed data. J Korean Stat Soc 45:51–63