



HAL
open science

Discrete optimal symmetric kernels for estimating count data distributions

Tristan Senga Kiessé, Gilles Durrieu

► **To cite this version:**

Tristan Senga Kiessé, Gilles Durrieu. Discrete optimal symmetric kernels for estimating count data distributions. 2020. hal-02503789

HAL Id: hal-02503789

<https://hal.science/hal-02503789>

Preprint submitted on 10 Mar 2020

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.

Discrete optimal symmetric kernels for estimating count data distributions

Tristan Senga Kiessé · Gilles Durrieu

Received: date / Accepted: date

Abstract The frequency estimator has long been regarded as the non-parametric reference for count data distributions. Meanwhile, the non-parametric estimator using discrete kernels has been developed as one competing alternative to the frequency estimator. Various discrete kernels are now available in the literature, which raises the issue of finding one discrete optimal kernel for non-parametric estimation of count data. To address this issue, we investigated discrete symmetric kernels that minimise the global squared error of non-parametric estimator of count data distributions. Basic asymptotic properties of both discrete optimal symmetric kernels and the corresponding non-parametric estimator were studied, in comparison with other discrete non-parametric kernel estimators. The performance of one discrete optimal symmetric kernel was also illustrated through simulations and applications on real data set.

Keywords Discrete kernel · Modal probability · Non-parametric estimation · Squared error

Mathematics Subject Classification (2010) 62G07 · 62G20 · 62G99

1 Introduction

Modelling the unknown probability density function (pdf) f of a random variable X on the real line \mathbb{R} is the subject of much research to find the best representation of f . In this context, the continuous kernel method was presented

T. Senga Kiessé
UMR SAS, INRAE, Agrocampus Ouest, 35000 Rennes, France
ORCID:0000-0003-2710-5825
E-mail: tristan.senga-kiesse@inrae.fr

G. Durrieu
Laboratoire de Mathématiques de Bretagne Atlantique LMBA, Université Bretagne Sud,
UMR CNRS 6205, 56000 Vannes, France

as one simple and efficient non-parametric technique for density estimation (Parzen, 1962; Rosenblatt, 1956). A continuous kernel estimator $\hat{f}_{K,h}$ of f was constructed as a kernel weighted average given by

$$\hat{f}_{K,h}(x) = \frac{1}{n} \sum_{i=1}^n K_{x,h}(X_i), \quad x \in \mathbb{R},$$

where $(X_i)_{i=1,\dots,n}$ are independent and identically distributed (i.i.d.) observations from a random sampling of f , $K_{x,h}$ is a continuous (symmetric or asymmetric) kernel and $h = h(n) > 0$ is an arbitrary sequence of smoothing parameters (or bandwidths) such that $\lim_{n \rightarrow \infty} h(n) = 0$ (Wand and Jones, 1995). Studies are increasingly interested in how the kernel method can contribute to non-parametric estimation of functions in many various applied research fields. For instance, in hydrology, the kernel method was applied for non-parametric modelling of flood frequencies (Adamowski, 1985). Likewise, in economy, the non-parametric Nadaraya-Watson regression estimator was proposed for modelling the relation between economic growth and environmental degradation (Taskin and Zaim, 2000). The kernel method was also applied in medicine for estimating a relative risk of disease cases (Davies and Hazelton, 2010). Recently, the non-parametric estimation of extreme tail of conditional distribution functions involved a weighted kernel estimator of Pareto distribution parameter (Durrieu et al., 2015).

As an alternative to continuous kernels used for non-parametric estimation of a pdf, discrete kernels have been developed for non-parametric estimation of a probability mass function (pmf) $f(\cdot) = \Pr(X = \cdot)$ of a random variable X on a support \mathcal{S} , such as sets \mathbb{Z} and \mathbb{N} of all integers and non-negative integers, respectively (Kokonendji and Senga Kiessé, 2011). Various discrete kernels are now available in the literature such as geometric and uniform weight kernels (Wang and Ryzin, 1981), discrete symmetric and asymmetric triangular kernels (Kokonendji and Senga Kiessé, 2011; Kokonendji and Zocchi, 2010) and specific kernels to deal with features of categorical variables (Aitchison and Aitken, 1976). These discrete kernels are competing with the Dirac type kernel uses in the frequency estimator, which was considered until recently as the non-parametric reference for estimating pmf.

The large variety of discrete kernels available in the literature raises the issue of finding one discrete optimal kernel. For instance, the Epanechnikov continuous kernel has been presented as being the optimal symmetric one to estimate a pdf (Epanechnikov, 1969; Tsybakov, 2004), although the Gaussian kernel is one of the most often used in the literature. Likewise, optimal asymmetric kernels have been proposed also for continuous density estimation (Abadir and Lawford, 2004). Meanwhile, to our knowledge, no discrete optimal kernel has been proposed yet for non-parametric estimation of pmf. Thus, the main contribution of this work consists of introducing discrete optimal symmetric kernels for non-parametric estimation of pmf.

2 Discrete associated kernels

We first recall fundamental notions of the discrete associated kernel method. Let us consider a fixed point x on a discrete support $\mathcal{S} \subseteq \mathbb{Z}$ and a bandwidth parameter $h = h(n) > 0$, such that $h(n)$ tends to 0 as n goes to ∞ . The discrete kernel $K_{x,h}$ associated with a random variable $\mathcal{K}_{x,h}$ defined on a support \mathcal{S}_x (which contains x) is a pmf such that

$$\sum_{y \in \mathcal{S}_x} K_{x,h}(y) = 1, \quad \text{with } 0 \leq K_{x,h}(y) = \Pr(\mathcal{K}_{x,h} = y) \leq 1.$$

The underlying idea behind the notion of discrete kernels $K_{x,h}$ is to propose a pmf that converges to the Dirac type kernel $D_x(y) = \mathbf{1}_{y=x}$ on $\mathcal{S}_x = \{x\}$, such that its modal probability behaves asymptotically as follows:

$$K_{x,h}(x) = \Pr(\mathcal{K}_{x,h} = x) \rightarrow D_x(x) = 1 \quad \text{as } h \rightarrow 0, \quad (1)$$

where $\mathbf{1}_A$ takes the value 1 when the condition A is verified and 0 otherwise. The discrete associated kernel $K_{x,h}(\cdot)$ was intended to attribute the more important probability mass (i.e. closest to one) at the point $x \in \mathcal{S}$, while having a bandwidth parameter $h > 0$ to take into account the probability mass at points $y \in \mathcal{S}_x \setminus \{x\}$ in the neighbourhood of x . The following general expressions of $\mathcal{K}_{x,h}$'s expectation and variance result from equation (1):

$$(E_1) : \mathbb{E}(\mathcal{K}_{x,h}) = x + a(x, h) \quad \text{and} \quad (E_2) : \text{Var}(\mathcal{K}_{x,h}) = b(x, h),$$

where both $a(x, h)$ and $b(x, h)$ tend to 0 as h goes to 0 (Senga Kiessé, 2017).

2.1 Examples

We now present examples of discrete associated kernels that have been proposed in the literature.

Example 1 For any $x \in \mathcal{S} = \mathbb{Z}$ and $h \in (0, 1]$, Wang and van Ryzin (1981) have proposed a geometric weight kernel given by

$$K_{x,h}(z) = \begin{cases} 1 - h & x = z, \\ 1/2(1 - h)h^{|x-z|} & |x - z| \geq 1, \end{cases} \quad z \in \mathcal{S}_x = \mathbb{Z}.$$

The expectation and variance of this kernel are, respectively,

$$\mathbb{E}(\mathcal{K}_{x,h}) = x \quad \text{and} \quad \text{Var}(\mathcal{K}_{x,h}) = h(1 + h)/(1 - h)^2.$$

Example 2 Likewise, for any $x \in \mathcal{S} = \mathbb{Z}$, $h \in (0, 1]$ and $k \in \mathbb{N} \setminus \{0, 1\}$, Wang and van Ryzin (1981) have proposed a uniform weight kernel expressed by

$$K_{k;x,h}(z) = \begin{cases} 1 - h & x = z, \\ 1/2(h/k) & |x - z| = 1, \dots, k, \\ 0 & |x - z| \geq k + 1, \end{cases} \quad z \in \mathcal{S}_x = \mathbb{Z},$$

with its expectation and variance given by

$$\mathbb{E}(\mathcal{K}_{k;x,h}) = x \quad \text{and} \quad \text{Var}(\mathcal{K}_{k;x,h}) = h(k + 1)(2k + 1)/6.$$

Example 3 Let us consider a fixed point $x \in \{0, 1, \dots, c\}$, a bandwidth parameter $h \in (0, 1]$ and an entire $c \in \mathbb{N} \setminus \{0, 1\}$. Aitchison and Aitken (1976) have also proposed one kernel to deal with features of a discrete r.v. X with c unordered categories whose pmf is given by

$$K_{c;x,h}(z) = \begin{cases} 1-h & x = z, \\ h/(c-1) & |x-z| \geq 1, \end{cases} \quad z \in \mathcal{S}_x = \{0, 1, \dots, c-1\},$$

with

$$\mathbb{E}(\mathcal{K}_{x,h}) = x + h \left(1 - x - \frac{x}{c-1} + \frac{hc}{2} \right)$$

and

$$\text{Var}(\mathcal{K}_{x,h}) = - \left(\frac{c^2(-2x+c-1)^2}{4(c-1)^2} \right) h + \left(\frac{c(6x^2+2c^2-3c+1-6xc+6x)}{6(c-1)} \right) h^2.$$

Example 4 For $x \in \mathcal{S} = \mathbb{Z}$, $h > 0$ and $p \in \mathbb{N}$, the symmetric discrete triangular kernel $K_{p;x,h}$ had a pmf defined by

$$K_{p;x,h}(z) = \begin{cases} (p+1)^h / C(p,h) & x = z, \\ (p+1)^h / C(p,h) \left(1 - \frac{|z-x|^h}{(p+1)^h} \right) & |x-z| \geq 1, \end{cases} \quad z \in \mathcal{S}_x = \{x, \dots, x \pm p\},$$

such that

$$\mathbb{E}(\mathcal{K}_{p;x,h}) = x \text{ and } \text{Var}(\mathcal{K}_{p;x,h}) = \frac{1}{C(p,h)} \left(\frac{p(2p+1)(p+1)^{h+1}}{3} - 2 \sum_{k=0}^p k^{h+2} \right),$$

where $C(p,h) = (2p+1)(p+1)^h - 2 \sum_{k=0}^p k^h$ is a normalising constant (Kokonendji and Senga Kiessé, 2011).

The four discrete associated kernels previously cited have their modal probability $K_{x,h}(x)$ that tends to 1 as h goes to 0 (equation (1)). Their expectation and variance also satisfy equations (E₁) – (E₂). Meanwhile, the Aitchison and Aitken kernel for categorical variables (Example 3) has a mean that is not equal to $x \in \mathcal{S}$ and a variance that depends on both x and $h > 0$.

Other discrete kernels associated to standard discrete probability distributions (e.g. binomial) have been proposed (Kokonendji and Senga Kiessé, 2011). The modal probability of these discrete kernels do not satisfy the asymptotic property in equation (1). However, the non-parametric estimator using the binomial kernel is competing with those using discrete symmetric triangular and Dirac type kernels for modelling pmf of count data with medium and small sizes (Senga Kiessé, 2017).

2.2 Mean integrated squared error

For $x \in \mathcal{S}$ and $h > 0$, without loss of generality, we now consider discrete associated kernels $K_{x,h}(\cdot)$ on the support \mathcal{S}_x that are symmetric pmf, i.e. $K_{x,h}(y) = K_{x,h}(-y)$, such that

$$\mathbb{E}(\mathcal{K}_{x,h}) = x \quad \text{and} \quad \mathbb{E}(\mathcal{K}_{x,h}^2) = q(x, h) < \infty.$$

Like discrete associated kernels in Examples 1, 2 and 4, the variance of $K_{x,h}$ $\text{Var}(\mathcal{K}_{x,h}) = q(x, h) - x^2 = b(h) < \infty$ does not depend on x and $b(h)$ tends to 0 as h goes to 0 (equation (E₂)). The non-parametric estimator $\tilde{f}_{K,h}$ of f with a discrete symmetric kernel $K_{x,h}$ on \mathcal{S}_x is given by

$$\tilde{f}_{K,h}(x) = \frac{1}{n} \sum_{i=1}^n K_{x,h}(X_i) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i), \quad (2)$$

with $(X_i)_{i=1, \dots, n}$ being i.i.d. observations having a pmf $f(\cdot) = \Pr(X_i = \cdot)$ to estimate on \mathcal{S} , with $\lim_{x \rightarrow \infty} f(x) = 0$. General expressions of bias and variance of estimator $\tilde{f}_{K,h}$ are given by

$$\text{Bias}(\tilde{f}_{K,h}(x)) = \mathbb{E}(\tilde{f}_{K,h}(x)) - f(x) = \mathbb{E}(f(\mathcal{K}_{x,h})) - f(x) \quad (3)$$

and

$$\text{Var}(\tilde{f}_{K,h}(x)) = \frac{1}{n} \sum_{y \in \mathcal{S}_x} K_{x,h}^2(y) f(y) - \frac{1}{n} \mathbb{E}^2(\tilde{f}_{K,h}(x)). \quad (4)$$

Concerning the $\tilde{f}_{K,h}$'s bias in equation (3), $f(\mathcal{K}_{x,h})$ can be expanded around x by using the following discrete version of Taylor's formula for a pmf $f : \mathcal{S} \rightarrow [0, 1]$ around $t_0 \in \mathcal{S} \subseteq \mathbb{Z}$:

$$f(t) = f(t_0) + \sum_{j=1}^k \frac{f^{(j)}(t_0)}{j!} (t - t_0)^j + \frac{1}{k!} \sum_{i=t_0}^{t-k-1} (t - i - 1)^k f^{(k+1)}(i), \quad (5)$$

where $f^{(j)}$ are finite differences of order $j = 1, 2, \dots, k$, of f and the last term is $o\{(t - t_0)^k\}$; refer to Argawal and Bonher (1999) and Schumaker (1996)[Theorem 8.61, page 351]. More precisely, the finite differences of f are such that $f^{(0)} = f$ and $f^{(j)} = \{f^{(j-1)}\}^{(1)}$, with

$$f^{(1)} = \frac{1}{2} (f(x+1) - f(x-1)), \quad x \in \mathbb{Z}.$$

By applying the formula in equation (5), we get the following expansion of $f(\mathcal{K}_{x,h})$ around $\mathbb{E}(\mathcal{K}_{x,h}) = x$:

$$\begin{aligned} f(\mathcal{K}_{x,h}) &= f(\mathbb{E}(\mathcal{K}_{x,h})) + (\mathcal{K}_{x,h} - \mathbb{E}(\mathcal{K}_{x,h})) f^{(1)}(\mathbb{E}(\mathcal{K}_{x,h})) + \frac{1}{2} (\mathcal{K}_{x,h} - \mathbb{E}(\mathcal{K}_{x,h}))^2 f^{(2)}(\mathbb{E}(\mathcal{K}_{x,h})) \\ &\quad + \sum_{k \geq 3} \frac{1}{k!} (\mathcal{K}_{x,h} - \mathbb{E}(\mathcal{K}_{x,h}))^k f^{(k)}(\mathbb{E}(\mathcal{K}_{x,h})), \end{aligned}$$

which finally yields the expression:

$$\text{Bias}(\tilde{f}_{K,h}(x)) = \frac{f^{(2)}(x)}{2} \text{Var}(\mathcal{K}_{x,h}) + o(h). \quad (6)$$

The $\tilde{f}_{K,h}$'s bias tends to 0 as h goes to 0, under the condition (E₂) on the variance of $\mathcal{K}_{x,h}$.

Concerning the $\tilde{f}_{K,h}$'s variance in equation (4), by using again the discrete version of Taylor's formula, $f(y)$ can be expanded around x as follows $f(y) = f(x) - (y-x)f^{(1)}(x) + o((y-x))$, such that we get

$$\text{Var}(\tilde{f}_{K,h}(x)) = \frac{1}{n} f(x) \sum_{y \in \mathcal{S}_x} K_{x,h}^2(y) + o\left(\frac{1}{n}\right).$$

The previous expressions of $\tilde{f}_{K,h}$'s bias and variance enable us to calculate the mean integrated squared error of $\tilde{f}_{K,h}$ given by:

$$\text{MISE}(\tilde{f}_{K,h}) = \sum_{x \in \mathcal{S}} \text{Var}(\tilde{f}_{K,h}(x)) + \sum_{x \in \mathcal{S}} \text{Bias}^2(\tilde{f}_{K,h}(x)) = \text{AMISE}(\tilde{f}_{K,h}) + o\left(\frac{1}{n} + h^2\right), \quad (7)$$

with AMISE being the leading term in MISE. Next section presents a discrete optimal symmetric kernel that minimises (A)MISE of $\tilde{f}_{K,h}$.

3 Discrete optimal symmetric kernel

For $x \in \mathcal{S}$ and $h > 0$, to find a discrete optimal associated kernel $K_{x,h}^{opt}(\cdot)$ remains to minimising $\sum_{y \in \mathcal{S}_x} K_{x,h}^2(y)$ from the expression of MISE (7), such that $K_{x,h}(y) \geq 0$ and

$$C_1 : \sum_{y \in \mathcal{S}_x} K_{x,h}(y) = 1, C_2 : \sum_{y \in \mathcal{S}_x} y K_{x,h}(y) = x \text{ and } C_3 : \sum_{y \in \mathcal{S}_x} y^2 K_{x,h}(y) = q.$$

That consists in a minimisation problem under constraints that can be solved using the method of Lagrange multiplier, such that a discrete expression of the Lagrange function is given by

$$\mathcal{L}(K) = \sum_{y \in \mathcal{S}_x} (K_{x,h}^2(y) + \lambda_1 K_{x,h}(y) + \lambda_2 y K_{x,h}(y) + \lambda_3 y^2 K_{x,h}(y)) f(y) + \text{constant},$$

where the *constant* does not depend on $K_{x,h}$ and $\lambda_i, i = 1, 2, 3$, are Lagrange multipliers corresponding to constraints C₁-C₃. A similar problem of finding an optimal kernel by a minimisation under constraints was investigated for symmetric (Epanechnikov, 1969; Tsybakov, 2004) and asymmetric (Abadir and Lawford, 2004) continuous kernels. Likewise, one can refer to Hodges et al. (1956) for the problem of minimising pdf.

The discrete optimal symmetric kernel has the general form

$$K_{x,h}^{opt}(y, \lambda_1, \lambda_2, \lambda_3) = -\frac{1}{2}(\lambda_1 + \lambda_2 y + \lambda_3 y^2), \quad (8)$$

for the minimum value $t^{opt} = \min_{t \geq 0}(t^2 + \lambda_1 t + \lambda_2 t y + \lambda_3 t y^2)$ is achieved at $t^{opt} = -(\lambda_1 + \lambda_2 y + \lambda_3 y^2)/2$. Without loss of generality, we investigate discrete optimal symmetric kernels $K_{x,h}^{opt}$ on the support $\mathcal{S}_x = \{x, x \pm 1, \dots, x \pm k\}$, with $k \geq 1$ a fixed integer. We get the following result based on solving the minimisation problem described above.

Theorem 1 For $x \in \mathcal{S}$ and a fixed integer $k \geq 1$, the discrete optimal symmetric kernel on $\mathcal{S}_x = \{x, x \pm 1, \dots, x \pm k\}$ that minimises the mean integrate squared error of the estimator $\tilde{f}_{K,h}$ in (2) is given by

$$K_{x,h}^{opt}(y) = \frac{\lambda_3}{2} \left(-y^2 + 2xy - x^2 + \frac{3k^2 + 3k - 1}{5} \right) + \frac{h}{2k + 1}, y \in \mathcal{S}_x, \quad (9)$$

where $\lambda_3 = 30(1-h)/((2k+1)(4k^2+4k-3)) > 0$ and $3/5(1-1/k) < h < 1$.

The proof of theorem is postponed to the appendices.

Remark 1 The bandwidth parameter $h > 0$ has a varying interval range with a lower bound $h_{lower} = 3/5(1-1/k)$ that depends on $k \geq 1$. That insures discrete optimal symmetric kernels $K_{x,h}^{opt}$ to be non-negative on the support \mathcal{S}_x when k varies.

For a fixed target $x = 4$ and various values of the parameters $h \in \{0.5, 0.9\}$ and $k \in \{1, 2, 3, 4\}$, the shape of discrete optimal symmetric kernels $K_{x,h}^{opt}$ is illustrated in Figure 1.

Case of discrete optimal symmetric kernel with $k = 1$. For $(x, h) \in \mathcal{S} \times (0, 1)$, we present the case of discrete optimal symmetric kernel with $k = 1$ on $\mathcal{S}_x = \{x - 1, x, x + 1\}$ whose the pmf is given by

$$\begin{aligned} K_{x,h}^{opt}(y) &= (1-h) \left(-y^2 + 2xy - x^2 + \frac{3-2h}{3(1-h)} \right), \quad y \in \mathcal{S}_x, \\ &= \begin{cases} 1-2h/3 & x = y, \\ h/3 & |x-y| = 1, \\ 0 & |x-y| \geq 2. \end{cases} \end{aligned} \quad (10)$$

This kernel has its expectation and variance given by, respectively,

$$\mathbb{E}(K_{x,h}^{opt}) = x \text{ and } \text{Var}(K_{x,h}^{opt}) = 2h/3.$$

We are now able to formulate the following proposition.

Proposition 1 Let $x \in \mathcal{S}$ and $h \in (0, 1)$. Consider the discrete optimal kernel $K_{x,h}^{opt}(\cdot)$ on the support $\mathcal{S}_x = \{x - 1, x, x + 1\}$,

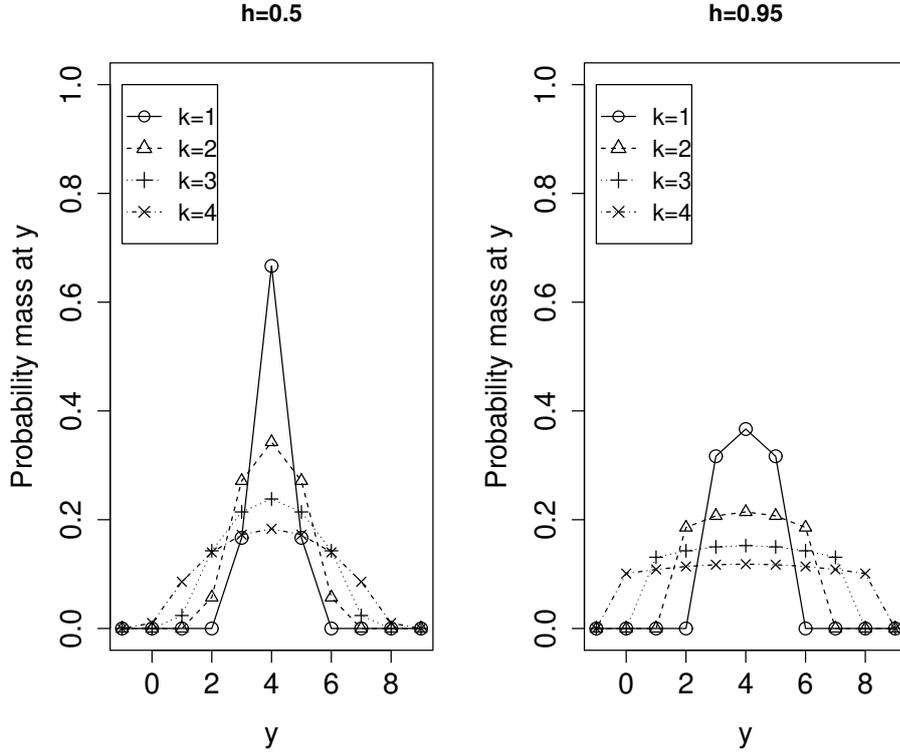


Fig. 1 Shape of discrete optimal symmetric kernels at the target $x = 4$ for h -values $\in \{0.5, 0.9\}$ and k -values $\in \{1, 2, 3, 4\}$.

- (i) when $h \rightarrow 1$, the limit optimal kernel $K_{x,1}^{opt}$ follows the discrete uniform distribution on the support \mathcal{S}_x .
- (ii) when $h \rightarrow 0$, the limit optimal kernel $K_{x,0}^{opt}$ follows the Dirac distribution on the support \mathcal{S}_x .

Proof. The proof of Proposition 1 comes directly from the expression of the discrete optimal symmetric kernel $K_{x,h}^{opt}$ in equation (10). ■

In the following sections, we consider the simple case of discrete optimal kernel $K_{x,h}^{opt}(y)$ with $k = 1$ on $\mathcal{S}_x = \{x-1, x, x+1\}$ both for comparison to other discrete associate kernels and for illustrating properties of the non-parametric estimator $\tilde{f}_{K,h}$ in equation (2).

4 Non-parametric discrete optimal kernel estimator

Consider i.i.d. observations $(X_i)_{i=1,\dots,n}$ having a pmf f on $\mathcal{S} = \mathbb{Z}$, i.e. $\sum_{x \in \mathbb{Z}} f(x) = 1$ and $0 \leq f(x) \leq 1$. Now we focus on asymptotic properties of non-parametric

estimator $\tilde{f}_{K,h}$ of f in equation (2) that uses the discrete optimal symmetric kernel with $k = 1$ on $\mathcal{S}_x = \{x-1, x, x+1\}$ such that

$$\tilde{f}_{K,h}^{opt}(x) = \frac{1}{n} \sum_{i=1}^n K_{x,h}^{opt}(X_i) = \frac{1}{n} \sum_{i=1}^n \left(\left(1 - \frac{2}{3}h\right) \mathbf{1}_{X_i=x} + \frac{1}{3}h \mathbf{1}_{|X_i-x|=1} \right).$$

The estimator $\tilde{f}_{K,h}^{opt}$ is a genuine pmf that does not need to be normalised, i.e. $\sum_{x \in \mathcal{S}} \tilde{f}_{K,h}^{opt}(x) = 1$. Basic properties of the estimator may be presented without using the discrete Taylor's formula, unlike the expression of bias of $\tilde{f}_{K,h}$ in equation (6). Then, other developments of bias and variance of $\tilde{f}_{K,h}^{opt}$ in (3) and (4), respectively, may be given by

$$\text{Bias} \left(\tilde{f}_{K,h}^{opt}(x) \right) = \mathbb{E} \left(\tilde{f}_{K,h}^{opt}(x) \right) - f(x) = \frac{h}{3} \left(-2f(x) + \sum_{y \in \mathcal{S}_x \setminus \{x\}} f(y) \right)$$

and

$$\begin{aligned} \text{Var} \left(\tilde{f}_{K,h}^{opt}(x) \right) &= \frac{1}{n} \mathbb{E} \left(\left(K_{x,h}^{opt}(X_1) \right)^2 \right) - \frac{1}{n} \mathbb{E}^2 \left(K_{x,h}^{opt}(X_1) \right) \\ &= \frac{1}{n} \sum_{y \in \mathcal{S}_x} \left(K_{x,h}^{opt}(y) \right)^2 f(y) - \frac{1}{n} \left(\sum_{y \in \mathcal{S}_x} K_{x,h}^{opt}(y) f(y) \right)^2 \\ &= \frac{1}{n} \left(\left(1 - \frac{2}{3}h\right)^2 f(x) + \frac{h^2}{9} \sum_{y \in \mathcal{S}_x \setminus \{x\}} f(y) \right) \\ &\quad - \frac{1}{n} \left(\left(1 - \frac{2}{3}h\right) f(x) + \frac{h}{3} \sum_{y \in \mathcal{S}_x \setminus \{x\}} f(y) \right)^2 \\ &= \frac{1}{n} \left(1 - \frac{2}{3}h\right)^2 f(x) (1 - f(x)) + R, \end{aligned}$$

with

$$R = \frac{1}{n} \left[\frac{h^2}{9} \sum_{y \in \mathcal{S}_x \setminus \{x\}} f(y) + \left(1 - \frac{2}{3}h\right)^2 f^2(x) - \left(\left(1 - \frac{2}{3}h\right) f(x) + \frac{h}{3} \sum_{y \in \mathcal{S}_x \setminus \{x\}} f(y) \right)^2 \right]$$

such that $R = o(h^2/n)$. That also results in another expression of the $\tilde{f}_{K,h}^{opt}$'s MISE in equation (7) given by

$$\begin{aligned} \text{MISE} \left(\tilde{f}_{K,h}^{opt} \right) &= \frac{1}{n} \left(1 - \frac{2}{3}h\right)^2 \sum_{x \in \mathbb{Z}} f(x) (1 - f(x)) + \frac{4}{9} h^2 f^2(x) + o\left(\frac{h^2}{n}\right) \\ &= \frac{1}{n} \left(1 - \frac{4}{6}h\right) \sum_{x \in \mathbb{Z}} f(x) (1 - f(x)) + \frac{4}{9} h^2 + o\left(\frac{h^2}{n}\right). \quad (11) \end{aligned}$$

Thus, we get $\text{MISE}(\tilde{f}_{K,h}^{opt}) \rightarrow 0$ when $n \rightarrow \infty$ and $h = h(n) \rightarrow 0$, which indicates that the estimator $\tilde{f}_{K,h}^{opt}$ is consistent in the sense of MISE. The expression of $\text{MISE}(\tilde{f}_{K,h}^{opt})$ can be asymptotically compared to that of the frequency estimator \tilde{f}_D using the Dirac type kernel, which is given by

$$\text{MISE}(\tilde{f}_D) = \frac{1}{n} \sum_{x \in \mathbb{Z}} f(x)(1 - f(x)) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Remark 2 An approximate \tilde{h}_{opt} of optimal bandwidth parameter

$$h_{opt} = \arg \min_{h > 0} \text{MISE}(\tilde{f}_{K,h}^{opt})$$

can be expressed by minimising the leading term of $\text{MISE}(\tilde{f}_{K,h}^{opt})$ in equation (11). That results in an approximate optimal bandwidth

$$\tilde{h}_{opt} = \frac{3}{4n} \sum_{x \in \mathbb{Z}} f(x)(1 - f(x)),$$

which depends on n and tends to 0 at the speed n^{-1} . One could also investigate an other approximate of h_{opt} by using another expansion of MISE as in equation (7), which involves finite differences of the pmf f . However, these various approximates of h_{opt} would deserve to be investigated in order to improve the accuracy of non-parametric kernel estimator. In this work, the classical cross-validation procedure is used in applications for selecting the bandwidth parameter.

4.1 Comparison with discrete triangular symmetric kernel estimator

4.1.1 Comparison of discrete kernels

We compare two discrete symmetric kernels: optimal (with $k = 1$) and triangular (Example 4 with $p = 1$) ones (Figure 2). We recall that discrete associated kernels must attribute the probability mass closest to one at the target point x as h goes to 0 (equation 1), while having a smoothing parameter $h > 0$ to take into account the probability mass at points $y \in \mathbb{Z} \setminus \{x\}$ in the neighbourhood of x . Thus, the following proposition compares modal probability and variance of discrete symmetric optimal and triangular kernels.

Proposition 2 *Consider any fixed $x \in \mathbb{Z}$ and $h \in (0, 1)$. As $h \rightarrow 0$, the modal probability and variance of discrete symmetric optimal (with $k = 1$) and triangular (with $p = 1$) kernels satisfy:*

$$\Pr(\mathcal{K}_{p;x,h} = x) \leq \Pr(\mathcal{K}_{x,h}^{opt} = x) \quad \text{and} \quad \text{Var}(\mathcal{K}_{p;x,h}) \geq \text{Var}(\mathcal{K}_{x,h}^{opt}).$$

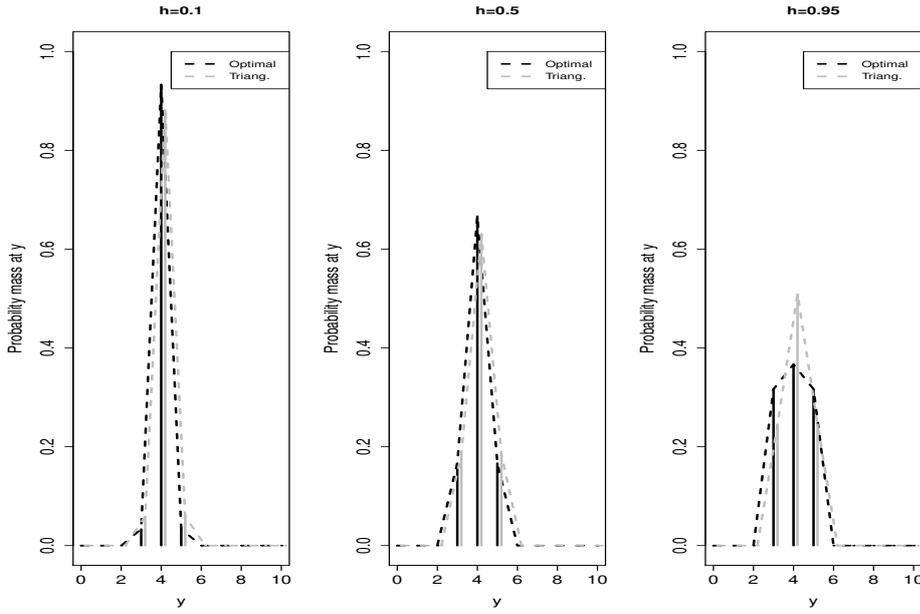


Fig. 2 Comparison of discrete optimal symmetric kernel with $k = 1$ and discrete symmetric triangular kernel with $p = 1$ at the target $x = 4$ for h -values $\in \{0.1, 0.5, 0.95\}$ on the support $\{0, 1, \dots, 10\}$.

The proof of proposition is postponed to the appendices.

Figure 2 illustrates that the ranking proposed in Proposition 2 is not always available for all h -values. For $h = 0.95$, the probability mass at target $x = 4$ of discrete symmetric triangular kernels is smaller than that of discrete optimal symmetric kernel. Without loss of generality, one can assume that a threshold bandwidth $h_0 > 0$ exists such that, for $h > h_0$, the discrete symmetric triangular kernels attribute the largest probability mass at target $x \in \mathbb{Z}$. Conversely, the previous remark implies that a threshold sample size n_0 exists such that for $n < n_0$ the discrete symmetric triangular kernels can attribute more probability mass at $x \in \mathbb{Z}$ than the discrete optimal symmetric kernel (and reciprocally), since the smoothing parameter $h = h(n)$ is linked to the sample size n such that $h \rightarrow 0$ when $n \rightarrow \infty$. The main question thus remains to find the threshold h_0 -value (or reciprocally the threshold n_0 -sample size).

Remark 3 A similar result as in Proposition 2 can be easily proved when comparing the modal probability and variance of discrete optimal symmetric kernel with $k = 1$ to those of discrete kernels in Examples 1 and 2. However, the comparison of discrete optimal symmetric kernel to the (variance of) discrete kernel in Example 3 depends on x .

4.1.2 Comparison of MISE of non-parametric kernel estimators

We now propose to compare $\tilde{f}_{K,h}$'s MISE using the two previous kernels. To this end, $\tilde{f}_{K,h}$'s variance in equation (4) can be majored as follows:

$$\text{Var}\left(\tilde{f}_{K,h}(x)\right) \leq \frac{1}{n} \text{Var}\left(K_{x,h}(X_1)\right) \leq \frac{1}{n} \mathbb{E}\left(K_{x,h}^2(X_1)\right).$$

The previous inequality traduces that as n increases, the variance term tends to 0 since it is penalised by the factor $1/n$. As $n \rightarrow \infty$ and $h \rightarrow 0$, the decrease in $\tilde{f}_{K,h}$'s variance term leads to considering mainly the influence of $\tilde{f}_{K,h}$'s bias term (given in equation (6)) on $\text{MISE}(\tilde{f}_{K,h})$, such that we obtain

$$\begin{aligned} \text{MISE}\left(\tilde{f}_{K,h}\right) &= \frac{1}{4} \sum_{x \in \mathbb{Z}} \left(\text{Var}(\mathcal{K}_{x,h}) f^{(2)}(x)\right)^2 + O\left(\frac{1}{n}\right) + o(h^2) \\ &= \text{AMISE}\left(\tilde{f}_{K,h}\right) + O\left(\frac{1}{n}\right) + o(h^2), \end{aligned} \quad (12)$$

where AMISE denotes the approximate MISE. We are now able to formulate the following proposition.

Proposition 3 *Consider any fixed $x \in \mathbb{Z}$ and $h \in (0, 1)$. As $n \rightarrow \infty$ and $h \rightarrow 0$, AMISE of the non-parametric estimators $\tilde{f}_{K,h}$ using discrete symmetric optimal (K^{opt}) and triangular (T) kernels satisfy:*

$$\text{AMISE}(\tilde{f}_{K^{opt},h}) \leq \text{AMISE}(\tilde{f}_{T,h}).$$

Proof. Let us consider the expression of AMISE in equation (12). Proposition 3 is a consequence of the ranking of variances of discrete symmetric optimal and triangular kernels in Proposition 2. ■

5 Simulations

This section illustrates the performance of the discrete optimal kernel estimator $\tilde{f}_{K,h}^{opt}$ on simulated sample sizes on the support \mathbb{N} . A comparison was conducted with non-parametric estimators using discrete triangular symmetric kernels with $p = 1$ and the binomial kernel $B_{x,h}$ on $\mathcal{S}_x = \{0, 1, \dots, x + 1\}$ such that

$$B_{x,h}(y) = \frac{(x+1)!}{y!(x+1-y)!} \left(\frac{x+h}{x+1}\right)^y \left(\frac{1-h}{x+1}\right)^{x+1-y}, \quad x \in \mathbb{N}, \quad h \in (0, 1].$$

These two kernels are among the most efficient for modelling count data distributions with medium and large sample sizes. We conducted Monte Carlo simulations to compare the discrete kernel estimators using mean values of their bias, variance and global error, but also to investigate effects of sample sizes. To measure the performance of estimator $\tilde{f}_{K,h}$ in (2), we used the mean MISE

Table 1 Mean and standard deviation (sd) of h_{cv} -values for discrete kernel estimators of count data of sample sizes $n \in \{15, 25, 50, 100, 250, 500\}$ simulated from Poisson pmf $\mathcal{P}(\mu)$ with $\mu = 2$.

Sample size n	Binomial kernel estimator		Triangular kernel estimator ($p = 1$)		Optimal kernel estimator ($k = 1$)	
	mean	sd	mean	sd	mean	sd
15	0.36	0.341	1.73	0.985	0.73	0.389
25	0.27	0.292	1.77	1.030	0.71	0.394
50	0.17	0.160	1.88	1.206	0.68	0.407
75	0.11	0.055	1.86	1.239	0.65	0.409
100	0.09	0.032	1.64	1.248	0.59	0.409
250	0.07	0.011	0.67	0.730	0.47	0.393
500	0.06	0.009	0.21	0.092	0.38	0.364

of $\tilde{f}_{K,h}$ over 500 replicates of sample size $n = \{15, 25, 50, 75, 100, 250, 500\}$ such that

$$\overline{\text{MISE}}(\tilde{f}_{K,h}) = \frac{1}{500} \sum_{i=1}^{500} \text{MISE}_i(\tilde{f}_{K,h}),$$

with MISE_i being the global squared error of the $\tilde{f}_{K,h}$ calculated after each replicate i of count data. The cross-validation procedure was satisfying for these aspects, and choosing a different bandwidth-choice procedure did not modify trends in the results. For each simulation, the smoothing bandwidth was found as $h_{cv} = \arg \min_{h>0} CV(h)$ with

$$CV(h) = \sum_{x \in \mathbb{N}} \left(\frac{1}{n} \sum_{i=1}^n K_{x,h}(X_i) \right)^2 - \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i} K_{X_i,h}(X_j)$$

being the cross-validation criterion ((Kokonendji and Senga Kiessé, 2011)).

Samples were simulated by randomly generating count data from a Poisson pmf $f = \mathcal{P}(\mu)$ with $\mu = 2$. Table 1 presents the descriptive statistics of h_{cv} -values such that $h_{cv} = h_{cv}(n)$ tends 0 as n goes to ∞ . Due to having a smoothing parameter defined on the interval (0,1), the binomial and optimal kernels estimators had mean h_{cv} -values smaller than those of the discrete symmetric triangular kernel estimator considered. Nevertheless, h_{cv} -values from the estimators using discrete binomial and symmetric triangular kernels went more quickly to 0 than those from the estimator using discrete optimal symmetric kernel, as the sample size n increased (see also Figure 3). Then, Table 2 presents the means integrated variance ($\overline{\text{IVar}}$), integrated squared bias ($\overline{\text{IBias}}$) and MISE ($\overline{\text{MISE}}$) of estimator $\tilde{f}_{K,h}$ using the various discrete kernels studied. For the n -values considered, the discrete optimal symmetric kernel estimator was generally ranked first while the Dirac kernel estimator was ranked last, in the sense of $\overline{\text{MISE}}$. However, for larger sample size ($n = 500$), the Dirac kernel estimator was ranked third, better than the binomial kernel estimator. Moreover, one can see that $\overline{\text{MISE}}$ of discrete symmetric optimal and triangular kernel estimators became closer as n increased.

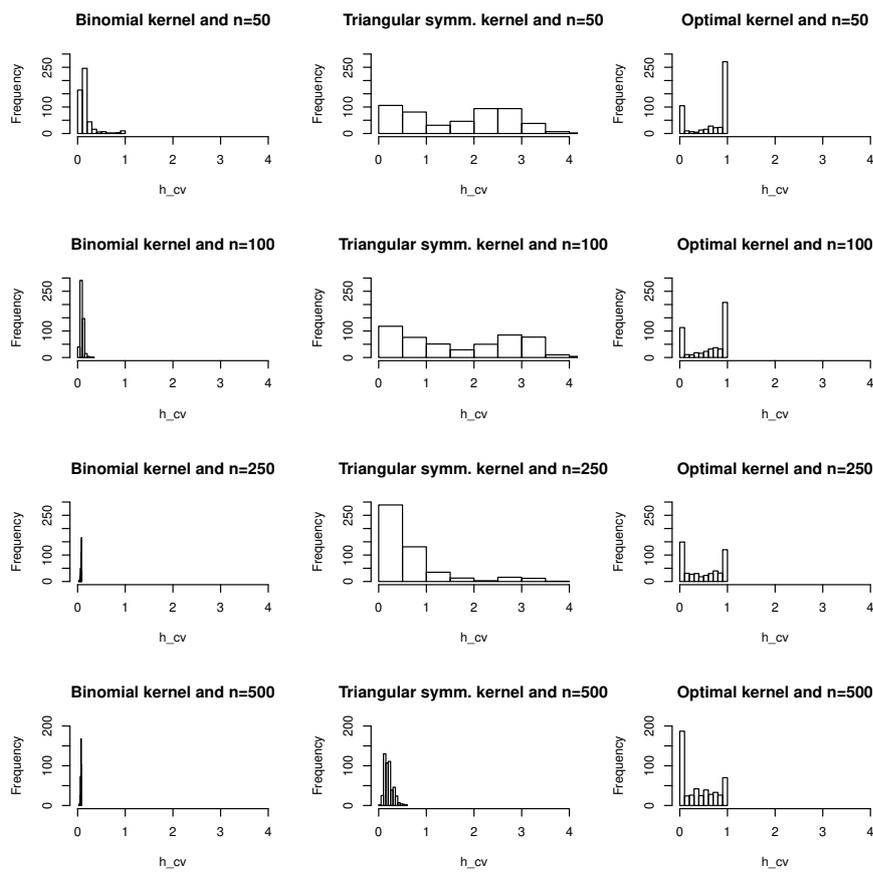


Fig. 3 Distribution of h_{cv} -values for discrete kernel estimators of count data of sample sizes $n \in \{50, 100, 250, 500\}$ simulated from Poisson pmf $\mathcal{P}(\mu)$ with $\mu = 2$.

Table 2 Results ($\times 10^3$) of average mean integrated squared error ($\overline{\text{MISE}}$), integrated squared bias ($\overline{\text{IBias}}$) and integrated variance ($\overline{\text{IVar}}$) for discrete kernel estimators of count data of sample sizes $n \in \{15, 25, 50, 100, 250, 500\}$ simulated from Poisson pmf $\mathcal{P}(\mu)$ with $\mu = 2$.

Sample size n	Dirac kernel estimator	Binomial kernel estimator		Triangular kernel estimator ($p = 1$)			Optimal kernel estimator ($k = 1$)			
	MISE	IBias	IVar	MISE	IBias	IVar	MISE	IBias	IVar	MISE
15	52.86	2.80	23.77	26.57	1.21	16.53	17.74	1.12	15.85	16.97
25	31.72	1.54	13.30	14.84	1.05	11.65	12.70	1.32	10.20	11.52
50	15.86	1.40	5.21	6.61	0.96	5.90	6.86	0.99	5.65	6.64
75	10.57	1.58	3.27	4.85	1.07	4.22	5.30	1.05	3.88	4.93
100	7.93	1.54	2.51	4.05	0.89	3.40	4.29	0.79	3.04	3.83
250	3.17	1.49	0.97	2.46	0.52	1.82	2.34	0.64	1.52	2.15
500	1.59	1.54	0.48	2.02	0.15	1.10	1.25	0.34	0.92	1.26

Table 3 Estimation of pre-adult development time (days) of spiraling whitefly observed on safou fruit-tree

Development time (days)	Frequency estimator	Bin. kern. estimator	Triang. kern. estimator ($p = 1$)	Optimal kern. estimator ($k = 1$)
SAFOU TREE				
30	28	24.21	27.15	26.51
31	21	22.82	21.53	21.94
32	11	12.96	11.30	11.54
h_{cv}		0.004	0.08	0.25
ISE		0.0059	0.0003	0.0009

6 Applications

We present applications on two real data sets. In these real cases there were few alternatives to assess the performance of non-parametric discrete kernel estimators $\hat{f}_{K,h}$ of empirical frequency, since the true count data distribution was unknown. We propose to use integrated squared error criterion based on the frequency estimator (\tilde{f}_D), which is an unsatisfactory estimator on its own. The ISE criterion, given as

$$\text{ISE}(h) = \sum_{x \in \mathbb{N}} \left(\tilde{f}_{K,h}(x) - \tilde{f}_D(x) \right)^2,$$

represents a descriptive degree-of-fit. For the applications cases, discrete non-parametric estimators using binomial, symmetric triangular kernel with $p = 1$ and optimal kernels are applied. The bandwidth parameter was selected using the cross-validation procedure.

6.1 Development of an insect pest

The real data sets were explanatory count variables describing development of an insect pest (spiralling white-fly, *Aleurodicus dispersus* Russel), which damages plants by sucking sap, decreasing photosynthesis activity and drying up leaves. This insect, originally from Central America and the Caribbean, is present in Congo-Brazzaville, and Congolese biologists were seeking to model its development. Thus, experimental plantations were established for several host plants, such as the fruit trees known as safou (*Dacryodes edulis*). Among other data collected, pre-adult development time was quantified as the number of days required for an insect to develop from egg to adult stages (Table 3). The sample size was $n = 60$.

Results in Table 3 showed that discrete symmetric triangular estimations were closest to empirical frequencies (in term of ISE), while the Binomial kernel estimations were further away from empirical frequencies. However, performances of the discrete symmetric triangular kernels and the optimal kernel were close. The main difference came from the underestimation of the number of days for the development time equal to 30.

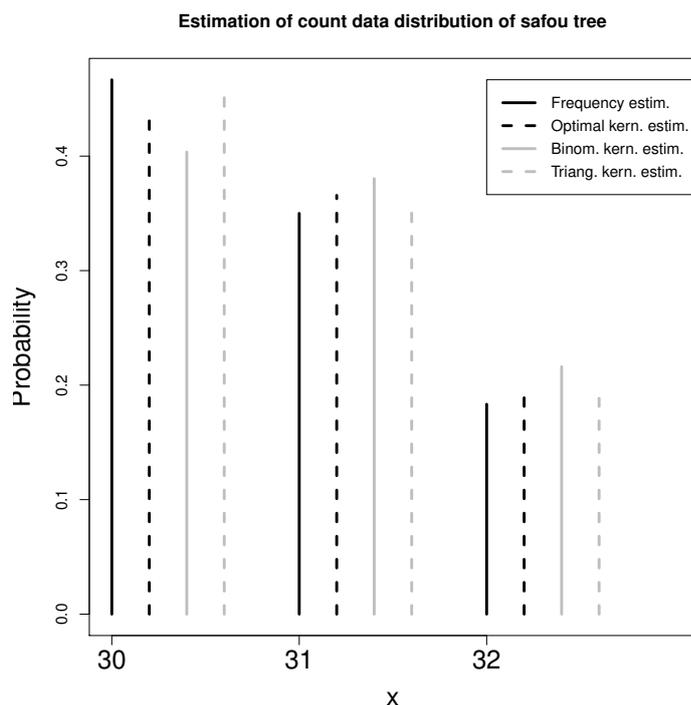


Fig. 4 Non-parametric kernel estimates of empirical frequency of count data from Table 3

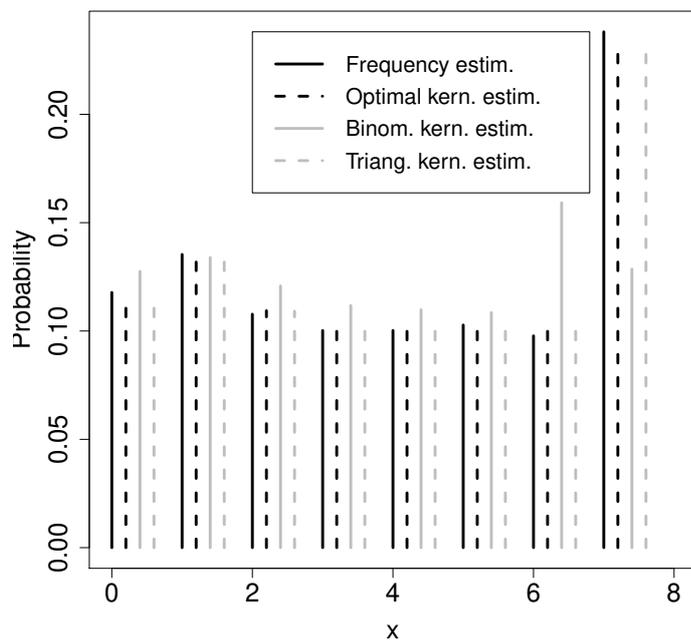
6.2 Daily alcohol consumption

A sociological experiment was carried out concerning the number of days per week in which alcohol was consumed (Alanko and Lemmens, 1996). 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. Table 4 presents the numerical results of fits obtained by means of the three discrete non-parametric kernel estimators previously considered.

Similar to results of simulations in the previous section, results in Table 4 showed that empirical frequencies, on the one hand, and estimations using discrete symmetric triangular and optimal kernels, on the other hand, were close for a large sample size. The main difference came from the frequency of the number of days equal to 7 which was more underestimated by using the discrete optimal kernel than the discrete symmetric triangular kernel with $p = 1$. The binomial kernel estimates stayed further away from empirical frequencies.

Table 4 Estimation of number of alcohol drinking days for week 1 from Alanko and Lemmens (1996)

Number of days per week	Frequency estimator	Bin. kern. estimator	Triang. kern. estimator ($p = 1$)	Optimal kern. estimator ($k = 1$)
0	47	50.86	46.46	46.36
1	54	53.43	54.02	54.03
2	43	48.19	43.53	43.63
3	40	44.58	40.39	40.46
4	40	43.81	40.35	40.41
5	41	43.30	41.26	41.31
6	39	63.51	40.65	40.95
7	95	51.29	92.29	91.80
h_{cv}		0.014	0.035	0.081
ISE		0.016	6.897×10^{-5}	9.640×10^{-5}

**Fig. 5** Non-parametric kernel estimates of empirical frequency of count data from Table 4

7 Concluding comments

We have derived a general form of discrete optimal symmetric kernels for non-parametric estimation of pmf. One simple case of our new optimal kernel has been particularly illustrated. This optimal kernel asymptotically converges to the Dirac type kernel and is competing with the other discrete kernels proposed in the literature. Properties of the general form of discrete optimal kernels should be deeply explored as well as its efficiency in various applications, such

as non-parametric multiple regressions (Somé and Kokonendji, 2016). Discrete optimal asymmetric kernels should also be explored.

A Appendices:Proofs

A.1 Proof of Theorem 1

The first constraint C_1 of the minimisation problem can be expressed as

$$\begin{aligned} & \sum_{y \in \mathcal{S}_x} \lambda_1 + \lambda_2 \sum_{y \in \mathcal{S}_x} y + \lambda_3 \sum_{y \in \mathcal{S}_x} z^2 = -2 \\ \Leftrightarrow & (2k+1)\lambda_1 + (2k+1)\lambda_2 x + \lambda_3 \left(x^2 + \sum_{j=1}^k (x+j)^2 + \sum_{j=1}^k (x-j)^2 \right) = -2 \\ \Leftrightarrow & (2k+1)\lambda_1 + (2k+1)\lambda_2 x + \lambda_3 \left((2k+1)x^2 + 2 \sum_{j=1}^k j^2 \right) = -2. \end{aligned} \quad (13)$$

The second constraint C_2 results successively in

$$\begin{aligned} & \lambda_1 \sum_{y \in \mathcal{S}_x} y + \lambda_2 \sum_{y \in \mathcal{S}_x} y^2 + \lambda_3 \sum_{y \in \mathcal{S}_x} y^3 = -2x \\ \Leftrightarrow & (2k+1)\lambda_1 x + \lambda_2 \left((2k+1)x^2 + 2 \sum_{j=1}^k j^2 \right) + \lambda_3 \left(x^3 + \sum_{j=1}^k (x+j)^3 + \sum_{j=1}^k (x-j)^3 \right) = -2x \\ \Leftrightarrow & (2k+1)\lambda_1 x + \lambda_2 \left((2k+1)x^2 + 2 \sum_{j=1}^k j^2 \right) + \lambda_3 \left((2k+1)x^3 + 6x \sum_{j=1}^k j^2 \right) = -2x. \end{aligned} \quad (14)$$

The third constraint C_3 can be expressed as

$$\begin{aligned} & \lambda_1 \sum_{y \in \mathcal{S}_x} y^2 + \lambda_2 \sum_{y \in \mathcal{S}_x} y^3 + \lambda_3 \sum_{y \in \mathcal{S}_x} y^4 = -2q \\ \Leftrightarrow & \lambda_1 \left((2k+1)x^2 + 2 \sum_{j=1}^k j^2 \right) + \lambda_2 \left((2k+1)x^3 + 6x \sum_{j=1}^k j^2 \right) + \lambda_3 \left(x^4 + \sum_{j=1}^k (x+j)^4 \right. \\ & \left. + \sum_{j=1}^k (x-j)^4 \right) = -2q \\ \Leftrightarrow & \lambda_1 \left((2k+1)x^2 + 2 \sum_{j=1}^k j^2 \right) + \lambda_2 \left((2k+1)x^3 + 6x \sum_{j=1}^k j^2 \right) + \lambda_3 \left((2k+1)x^4 \right. \\ & \left. + 12x^2 \sum_{j=1}^k j^2 + 2 \sum_{j=1}^k j^4 \right) = -2q. \end{aligned} \quad (15)$$

By multiplying Equation (13) by x then comparing with Equation (14), we get

$$\begin{aligned} & \lambda_3 x \left((2k+1)x^2 + 2 \sum_{j=1}^k j^2 \right) = 2\lambda_2 \sum_{j=1}^k j^2 + \lambda_3 \left((2k+1)x^3 + 6x \sum_{j=1}^k j^2 \right) \\ \Leftrightarrow & 2\lambda_2 \sum_{j=1}^k j^2 = \lambda_3 \left((2k+1)x^3 + 2x \sum_{j=1}^k j^2 \right) - \lambda_3 \left((2k+1)x^3 + 6x \sum_{j=1}^k j^2 \right) \\ \Leftrightarrow & \lambda_2 = -2x\lambda_3. \end{aligned} \quad (16)$$

Then, by multiplying Equation (14) by x then comparing with Equation (15), we get

$$\begin{aligned}
& -\lambda_1 \sum_{j=1}^k j^2 - 2x\lambda_2 \sum_{j=1}^k j^2 - 3x^2\lambda_3 \sum_{j=1}^k j^2 - \lambda_3 \sum_{j=1}^k j^4 = -2x^2 + 2q \\
\iff & -\lambda_1 \sum_{j=1}^k j^2 + \lambda_3 \left(x^2 \sum_{j=1}^k j^2 - \sum_{j=1}^k j^4 \right) = 2b \\
\iff & \lambda_1 = \frac{\lambda_3 \left(x^2 \sum_{j=1}^k j^2 - \sum_{j=1}^k j^4 \right) - 2b}{\sum_{j=1}^k j^2}, \tag{17}
\end{aligned}$$

where λ_2 was replaced by the expression in (16).

Finally, from the previous expression, we substitute λ_1 (Equation 17) and λ_2 (Equation 16) in Equation (13) such that:

$$\begin{aligned}
& (2k+1)\lambda_3 x^2 \sum_{j=1}^k j^2 - (2k+1)\lambda_3 \sum_{j=1}^k j^4 - 2(2k+1)b - 2(2k+1)\lambda_3 x^2 \sum_{j=1}^k j^2 \\
& + \lambda_3(2k+1)x^2 \sum_{j=1}^k j^2 + 2\lambda_3 \left(\sum_{j=1}^k j^2 \right)^2 = -2 \sum_{j=1}^k j^2 \\
\iff & \lambda_3 = \frac{2(2k+1)b - 2 \sum_{j=1}^k j^2}{2 \left(\sum_{j=1}^k j^2 \right)^2 - (2k+1) \sum_{j=1}^k j^4} \\
\iff & \lambda_3 = 60 \frac{6(2k+1)b - k(k+1)(2k+1)}{10k^2(k+1)^2(2k+1)^2 - 6k(k+1)(2k+1)^2(3k^2+3k-1)} \\
\iff & \lambda_3 = \frac{30}{k(k+1)(2k+1)} \frac{6b - k(k+1)}{-4k^2 - 4k + 3},
\end{aligned}$$

with the positive integer $k \geq 1$ and $6b < k(k+1)$ to insure the optimal kernel in equation (8) to be concave, i.e. $\lambda_3 > 0$, since $-4k^2 - 4k + 3 < 0$ for $k \geq 1$. One simple option is to choose $b = b(h) = hk(k+1)/6$, with $0 < h < 1$, such that

$$\begin{cases} \lambda_3 = 30(1-h)/((2k+1)(4k^2+4k-3)) \\ \lambda_2 = -60x(1-h)/((2k+1)(4k^2+4k-3)) \\ \lambda_1 = \lambda_3(x^2 - (3k^2+3k-1)/5) - 2h/(2k+1). \end{cases}$$

Finally, the kernel $K_{x,h}^{opt}(y) = -(1/2)(\lambda_1 + \lambda_2 y + \lambda_3 y^2)$ is non-negative ($K_{x,h}^{opt}(y) \geq 0$) if y belongs to the interval $I_k = [x - \frac{\sqrt{\delta}}{2\lambda_3}, x + \frac{\sqrt{\delta}}{2\lambda_3}]$, which can be expressed by

$$I_k = \left[x - \sqrt{\frac{3k^2+3k-1}{5} + \frac{h(4k^2+4k-3)}{15(1-h)}}, x + \sqrt{\frac{3k^2+3k-1}{5} + \frac{h(4k^2+4k-3)}{15(1-h)}} \right],$$

such that $K_{x,h}^{opt}(x - \frac{\sqrt{\delta}}{2\lambda_3}) = K_{x,h}^{opt}(x + \frac{\sqrt{\delta}}{2\lambda_3}) = 0$ and $\delta = \lambda_2^2 - 4\lambda_1\lambda_3 > 0$, for $k \geq 1$. The support $\mathcal{S}_x = \{x-k, \dots, x, \dots, x+k\}$ of $K_{x,h}^{opt}$ is included in I_k ($\mathcal{S}_x \subseteq I_k$), if $3/5(1-1/k) < h < 1$, for $k \geq 1$.

Hence, the discrete optimal symmetric kernels with $k \geq 1$ are defined by

$$K_{x,h}^{opt}(y) = -\frac{1}{2} \left(\lambda_3 \left(x^2 - \frac{3k^2+3k-1}{5} \right) - \frac{2h}{2k+1} \right) + \lambda_3 x y - \frac{1}{2} \lambda_3 y^2, \quad y \in \mathcal{S}_x, \quad h \in \left(\frac{3}{5} \left(1 - \frac{1}{k} \right), 1 \right).$$

■

A.2 Proof of Proposition 2

Firstly, to compare the modal probability of two kernels considered, we express the difference

$$\Pr(\mathcal{K}_{p;x,h} = x) - \Pr(\mathcal{K}_{x,h}^{opt} = x) = \frac{2h}{3}(1 - 3A(p)) + O(h^2).$$

That requires to develop the probability mass term $\Pr(\mathcal{K}_{p;x,h} = x)$ by using a Taylor expansion as $h \rightarrow 0$ such that

$$\begin{aligned} \Pr(\mathcal{K}_{p;x,h} = x) &= \frac{(p+1)^h}{C(p,h)} \\ &= \frac{1 + h \log(p+1) + o(h^2)}{1 + h((2p+1) \log(p+1) - 2 \sum_{k=1}^p \log(k)) + o(h^2)} \\ &= 1 - h \left(2p \log(p+1) - 2 \sum_{k=1}^p \log(k) \right) + O(h^2) \\ &= 1 - 2hA(p) + O(h^2), \end{aligned}$$

with $C(p,h) = (2p+1)(p+1)^h - 2 \sum_{k=1}^p k^h$. We show that the function $A(p)$ is increasing with respect to $p \in \mathbb{N} \setminus \{0\}$ and $3A(p)$ is more than 1. One has

$$\begin{aligned} A(p+1) - A(p) &= (p+1) \log(p+2) - p \log(p+1) - \log(p+1) \\ &= (p+1) \log\left(\frac{p+2}{p+1}\right) > 0. \end{aligned}$$

Thus, we get $A(p+1) \geq A(p)$ with $A(1) = \log(2) > 1/3$ such that $3A(p) > 1$, and $\Pr(\mathcal{K}_{p;x,h} = x) < \Pr(\mathcal{K}_{x,h}^{opt} = x)$.

Secondly, to compare the variance of the two kernels studied, one has

$$\begin{aligned} \text{Var}(\mathcal{K}_{p;x,h}) - \text{Var}(\mathcal{K}_{x,h}^{opt}) &= \frac{2h}{3} \left(\frac{p(2p^2 + 3p + 1)}{2} \log(p+1) - 3 \sum_{k=1}^p k^2 \log(k) - 1 \right) + O(h^2) \\ &= \frac{2h}{3} (V(p) - 1) + O(h^2), \end{aligned}$$

since the variance term of discrete symmetric triangular kernels is such that

$$\begin{aligned} \text{Var}(\mathcal{K}_{p;x,h}) &= \frac{1}{C(p,h)} \left(\frac{p(2p+1)(p+1)^{h+1}}{3} - 2 \sum_{k=1}^p k^{h+2} \right) \\ &= \left[\frac{p(p+1)(2p+1)}{3} (1 + h \log(p+1)) - 2 \sum_{k=1}^p k^2 (1 + h \log(k)) \right] \\ &\quad \times \left[1 + h \left((2p+1) \log(p+1) - 2 \sum_{k=1}^p \log(k) \right) \right]^{-1} + o(h^2) \\ &= \left(\frac{p(2p^2 + 3p + 1)}{3} \log(p+1) - 2 \sum_{k=1}^p k^2 \log(k) \right) h + O(h^2). \end{aligned}$$

As for the modal probability previously, we show that the function $V(p)$ is increasing with respect to $p \in \mathbb{N} \setminus \{0\}$ and more than 1. To this end, we successively express

$$\begin{aligned} V(p+1) - V(p) &= \frac{(p+1)(2(p+1)^2 + 3(p+1) + 1)}{2} \log(p+2) \\ &\quad - \frac{p(2p^2 + 3p + 1)}{2} \log(p+1) - 3(p+1)^2 \log(p+1) \\ &= \frac{(p+1)(2p^2 + 7p + 6)}{2} \log(p+2) \\ &\quad - \frac{(p+1)(p(2p+1) + 6(p+1))}{2} \log(p+1) \\ &= \frac{(p+1)(2p^2 + 7p + 6)}{2} \log(p+2) - \frac{(p+1)(2p^2 + 7p + 6)}{2} \log(p+1) \\ &= \frac{(p+1)(2p^2 + 7p + 6)}{2} \log\left(\frac{p+2}{p+1}\right) > 0. \end{aligned}$$

Hence, we obtain $V(p+1) \geq V(p)$ with $V(1) = 3 \log(2) > 1$ resulting in $V(p) > 1$, and $\text{Var}(\mathcal{K}_{p;x,h}) > \text{Var}(\mathcal{K}_{x,h}^{opt})$. ■

References

- Abadir KM, Lawford S (2004) Optimal asymmetric kernels. *Economic Letters* 83:61–68
- Adamowski K (1985) Nonparametric Kernel Estimation of Flood Frequencies. *Water Resources Research* 21(11):1585–1590
- Aitchison J, Aitken CGG (1976) Multivariate binary discrimination by the kernel method. *Biometrika* 63:413–420
- Alanko T, Lemmens PH (1996) Response effects in consumption surveys: an application of the beta-binomial model to self-reported drinking frequencies. *Journal of Official Statistics* 12:253–273
- Argawal RP, Bohner M (1999) Basic calculus on time scales and some of its applications. *Results in Mathematics* 35:3–22
- Davies TM, Hazelton ML (2010) Adaptive kernel estimation of spatial relative risk. *Statistics in Medicine* 29(23):2423–2437
- Durrieu G, Grama I, Pham QK, Tricot JM (2015) Nonparametric adaptive estimation of conditional probabilities of rare events and extreme quantiles. *Extremes* 18(3):437–478
- Epanechnikov VA (1969) Non-parametric estimation of a multivariate probability density. *Theory of Probability and its Applications* 14:153–158
- Hodges JL, Lehmann EL, et al. (1956) The efficiency of some nonparametric competitors of the t -test. *The Annals of Mathematical Statistics* 27(2):324–335
- Kokonendji CC, Senga Kiessé T (2011) Discrete associated kernel method and extensions. *Statistical Methodology* 8:497–516
- Kokonendji CC, Zocchi SS (2010) Extensions of discrete triangular distribution and boundary bias in kernel estimation for discrete functions. *Statistical and Probability Letters* 80:1655–1662
- Parzen E (1962) On estimation of a probability density function and mode. *The Annals of Mathematical Statistics* 33(3):1065–1076
- Rosenblatt M (1956) Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics* 27(3):832–837
- Schumaker LL (1996) *Spline Functions: Basic Theory*. Wiley, New York
- Senga Kiessé T (2017) On finite sample properties of nonparametric discrete asymmetric kernel estimators. *Statistics* 51(5):1046–1060
- Somé SM, Kokonendji CC (2016) Effects of associated kernels in nonparametric multiple regressions. *Journal of Statistical Theory and Practice* 10(2):456–471
- Taskin F, Zaim O (2000) Searching for a kuznets curve in environmental efficiency using kernel estimation. *Economics Letters* 68(2):217 – 223

- Tsybakov AB (2004) Introduction à l'estimation non-paramétrique. Springer, Paris
- Wand MP, Jones MC (1995) Kernel smoothing. Chapman and Hall, London
- Wang MC, Ryzin JV (1981) A class of smooth estimators for discrete distributions. *Biometrika* 68:301–309