STAPRO: 7443 Model 3Gsc pp. 1–11 (col. fig: NIL)

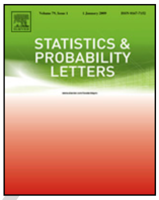
ARTICLE IN PRESS

Statistics and Probability Letters xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro



Q1 Performance of discrete associated kernel estimators through the total variation distance

Q2 Célestin C. Kokonendji, Davit Varron *

University of Franche-Comté, Laboratoire de Mathématiques de Besançon, UMR 6623 CNRS-UFC, France

ARTICLE INFO

Article history:
Received 3 June 2015
Received in revised form 15 October 2015
Accepted 15 October 2015
Available online xxxx

MSC:
primary 62G07
secondary 62G20
62G99

Keywords:
Concentration inequalities
Empirical processes
Probability mass function

ABSTRACT

We prove asymptotic results and concentration inequalities for a large class of discrete associated kernel estimators, under the total variation distance. We also propose a data driven bandwidth selection procedure aiming to minimize the total variation. Simulations are conducted.

1. Introduction and overview

Let $(\mathbb{T}, \mathcal{B}, c)$ be a measured space, where $\mathbb{T}$ is countable, $\mathcal{B}$ is the σ-algebra of all subsets of $\mathbb{T}$, and c is the counting measure on $\mathbb{T}$. Given an independent identically distributed [i.i.d.] sample $(X_1, \ldots, X_n)$ taking values in $\mathbb{T}$, we are interested in estimating the probability mass function [p.m.f.] $f : x \mapsto \mathbb{P}(X = x)$ where X stands for a generic random variable having the common distribution of the X_i. The natural estimator of f is the empirical p.m.f., namely Q3

$$f_n : x \mapsto \mathbb{P}_n(\{x\}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{x\}}(X_i), \tag{1.1}$$

where $\mathbb{I}_A$ denotes the indicator function of any given set $A \subset \mathbb{T}$. Recently, Kokonendji et al. (2007) (see also Kokonendji and Senga Kiessé, 2011) introduced the *discrete associated kernel* density estimator, extending their definition to the possible use of a multivariate bandwidth parameter: let $p \geq 1$ be an integer, let $\mathcal{D} \subset \mathbb{R}^p$ be a set containing 0, and let $\mathcal{K} := \left\{K_{x,h}(\cdot),\ x \in \mathbb{T},\ h \in \mathcal{D}\right\}$ be a collection of p.m.f. on $\mathbb{T}$. For a bandwidth parameter $h \in \mathcal{D}$, define:

$$g_{n,h}(x) := f_{n,h}(x)\left(\int_{\mathbb{T}} f_{n,h}(u)dc(u)\right)^{-1}, \tag{1.2}$$

where

$$f_{n,h}(x) := \frac{1}{n}\sum_{i=1}^{n} K_{x,h}(X_i), \quad x \in \mathbb{T}. \tag{1.3}$$

* Correspondence to: Université de Franche-Comté, LMB 16 route de Gray, 25030 Besançon cedex, France. Tel.: +33 381 666 330; fax: +33 381 666 623.
E-mail address: davit.varron@univ-fcomte.fr (D. Varron).

http://dx.doi.org/10.1016/j.spl.2015.10.008
0167-7152/© 2015 Published by Elsevier B.V.

Please cite this article in press as: Kokonendji, C.C., Varron, D., Performance of discrete associated kernel estimators through the total variation distance. Statistics and Probability Letters (2015), http://dx.doi.org/10.1016/j.spl.2015.10.008

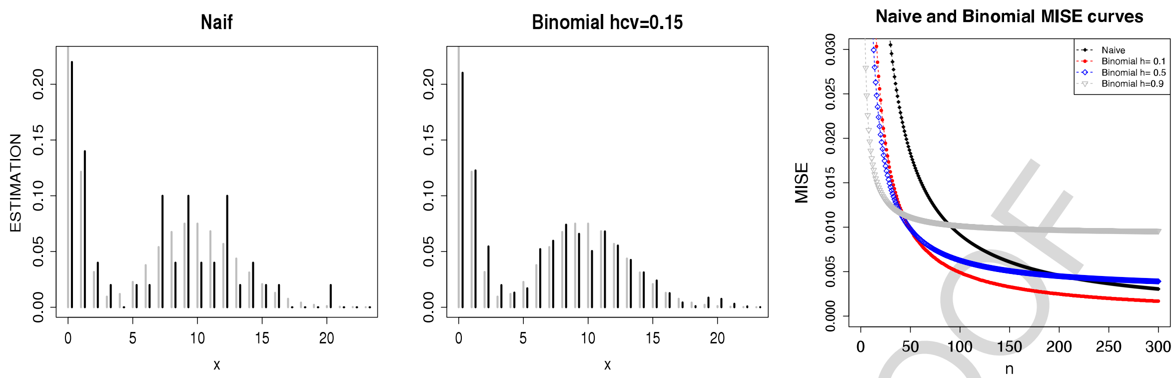


Fig. 1. On the left and center: Two discrete smoothings using empirical (or naive) and binomial kernel estimators of simulated data ($n = 50$) from $f = 0.4\mathcal{P}(0.5) + 0.6\mathcal{P}(10)$ in grey. On the right: MISE curves for Dirac and binomial kernels with $f = 0.3\mathcal{P}(0.6) + 0.7\mathcal{P}(9)$.

1.1. *On the normalizing constant*

Due to the general nature of $\mathcal{K}$, the normalization by

$$C_{n,h} := \int_{\mathbb{T}} f_{n,h}(x)dc(x) = \frac{1}{n}\sum_{i=1}^{n} Y_{i,h}, \quad \text{with } Y_{i,h} := \int_{\mathbb{T}} K_{x,h}(X_i)dc(x), \tag{1.4}$$

may be necessary in the sense that $C_{n,h}$ is not necessarily a.s. equal to 1. However, simulation studies show that $C_{n,h}$ only slightly oscillates around 1 for many discrete associated kernels (see e.g. Kokonendji and Senga Kiessé (2011)). Such a phenomenon is also observed for continuous *associated* kernels, as gamma and beta kernels of Chen (1999a,b) (see also Igarashi and Kakizawa, 2015; Malec and Schienle, 2014; Markovich, 2015). The *first contribution* of this paper is to provide a theoretical explanation of the observed smallness of those oscillations (see Theorem 2.1).

1.2. *About asymptotic properties of $g_{n,h}$*

In Abdous and Kokonendji (2009), when $\mathbb{T} \subset \mathbb{Z}$ and $\mathcal{D} := [0, 1]$, a first study on the asymptotic properties, as $n \to \infty$ and $h \to 0$, of the non normalized version $f_{n,h}$ in (1.3) shows that their pointwise consistency holds as soon as each discrete associated kernel $K_{x,h}$ converges, when $h \to 0$, toward the Dirac distribution δ_x in the following sense:

$$\forall k \in \{1, 2\}, \forall x \in \mathbb{T}, \quad \lim_{h\to 0} \int y^k K_{x,h}(y)dc(y) = x^k, \quad \text{and} \quad \overline{\lim_{h\to 0}}\, \mathbb{E}\left(\left(K_{x,h}(X)\right)^3\right) < \infty. \tag{1.5}$$

Besides that pointwise strong consistency, the authors also established the asymptotic normality of $f_{n,h}(x)$ for fixed x, with asymptotic variance $f(x)(1-f(x))$; see Abdous and Kokonendji (2009, Theorems 2.4 and 2.5). The main message of those first results is that, on the asymptotic point of view, the pointwise performances of those associated kernels are comparable to that of the empirical p.m.f. f_n. The *second contribution* of the present paper is to complete the picture by providing asymptotic results and concentration inequalities for the *total variation distance* between $g_{n,h}$ and f, namely the random variable

$$TV(g_{n,h}, f) := \sup_{A\subset\mathbb{T}} \left| \int_A g_{n,h}dc - \int_A fdc \right| = \frac{1}{2}\|g_{n,h} - f\|_1,$$

where, for two c-integrable functions f_1 and f_2, we write

$$\|f_1 - f_2\|_1 := \int_{\mathbb{T}} |f_1 - f_2|dc.$$

Those results are stated in Section 2.1 and show that the asymptotic performances of $g_{n,h}$ are comparable to those of f_n in regard with the total variation.

1.3. *On the small sample outperforming of $g_{n,h}$*

The unavoidable question pointed by Section 1.2 is then "*why use discrete associated kernel estimators* instead of f_n?" A partial answer comes from simulations studies showing that $g_{n,h}$ seems to approximate f better than f_n does. The following figures (Fig. 1), drawn from Kokonendji and Senga Kiessé (2011), compare the barplots of the "naive" estimator f_n and of $g_{n,h}$ with a binomial kernel as well as their mean integrated square errors (MISE).

Please cite this article in press as: Kokonendji, C.C., Varron, D., Performance of discrete associated kernel estimators through the total variation distance. Statistics and Probability Letters (2015), http://dx.doi.org/10.1016/j.spl.2015.10.008