



On discrete Epanechnikov kernel functions

Chi-Yang Chu^a, Daniel J. Henderson^{b,*}, Christopher F. Parmeter^c

^a Department of Economics, National Taipei University, New Taipei, Taiwan

^b Department of Economics, Finance and Legal Studies, University of Alabama, Tuscaloosa, AL, USA

^c Department of Economics, University of Miami, Miami, FL, USA



ARTICLE INFO

Article history:

Received 20 October 2016

Received in revised form 7 July 2017

Accepted 16 July 2017

Available online 27 July 2017

Keywords:

Cross-validation

Discrete kernel

Panel data

Smoothing

ABSTRACT

Least-squares cross-validation is commonly used for selection of smoothing parameters in the discrete data setting; however, in many applied situations, it tends to select relatively small bandwidths. This tendency to undersmooth is due in part to the geometric weighting scheme that many discrete kernels possess. This problem may be avoided by using alternative kernel functions. Specifically, discrete versions (both unordered and ordered) of the popular Epanechnikov kernel do not have rapidly decaying weights. The analytic properties of these kernels are contrasted with commonly used discrete kernel functions and their relative performance is compared using both simulated and real data. The simulation and empirical results show that these kernel functions generally perform well and in some cases demonstrate substantial gains in terms of mean squared error.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

An intuitive approach to estimate a univariate discrete probability (mass) function is to use the sample frequency of occurrence as the estimator of a cell probability (i.e., frequency approach). However, when the number of cells is close to or even greater than the sample size (the data are sparse), the frequency approach does not work well due to many zero counts (Simonoff, 1996). In this case, applied researchers often resort to a smoothing approach, which introduces bias but can dramatically lower mean squared error (MSE). In this paper, we focus on the kernel smoothing approach where the underlying density $p(x)$ is estimated by $\hat{p}(x) = \frac{1}{n} \sum_{i=1}^n l(\cdot)$, with a kernel function $l(\cdot)$ appropriate for smoothing discrete data. Existing discrete kernel functions date back to Aitchison and Aitken (1976), Habbema et al. (1978), Titterton (1980), Wang and vanRyzin (1981), and Aitken (1983). More recently, Li and Racine (2003) propose kernel functions for smoothing both unordered and ordered discrete data.

The kernel function's ability to smooth data hinges on the bandwidth (or smoothing parameter). How this bandwidth is selected is of the utmost importance in applied work and least-squares cross-validation (LSCV) has proven a popular approach when discrete data are present given the lack of simple rule-of-thumb or plug-in bandwidths (see Chu et al., 2015 for some recent work in this direction). However, in many applied situations, LSCV tends to select a relatively small bandwidth relative to the theoretical optimum (undersmoothing), particularly when discrete data are sparse (e.g., see Asparoukhov and Krzanowski (2001) or Coppejans (2003). One explanation for this problem is that many ordered discrete kernel functions possess geometrically decaying weighting schemes, leading to a rapid decline in the weights used to smooth the data (Rajagopalan and Lall, 1995). Adding to this line of reasoning, Chu et al. (2015) show that for an ordered discrete kernel function with geometric weighting structure, the optimal bandwidth, in terms of the mean summed squared error (MSSE) criterion, is a real root from a polynomial, with the order of the polynomial being determined by the number of cells.

* Correspondence to: Box 870224, Tuscaloosa, AL 35487-0224, USA

E-mail address: djhender@cba.ua.edu (D.J. Henderson).

The main insight from this relationship is that the optimal bandwidth is inversely related to the order of the polynomial, potentially compounding the small bandwidth problem.

These issues also occur in kernel regression estimation. For example, [Henderson and Kumbhakar \(2006\)](#) note that in their longitudinal/panel application, capturing unobserved heterogeneity through an unordered discrete variable (with respect to the cross-sectional dimension) results in a relatively small bandwidth. In this case, the regression estimator essentially uses only T (time) observations for each cross-sectional unit. It is likely that this problem is pervasive in papers using nonparametric methods in the presence of longitudinal data where the cross-sectional specific heterogeneity is treated as an unordered discrete variable.

Although different methods have been proposed to resolve the issue of undersmoothing, most are modifications of existing error criterion and are typically designed mainly for continuous variables (for example, see [Härdle et al., 1988](#), [Chiu, 1990](#), [Hart and Yi, 1998](#), [Hurvich et al., 1998](#), [Hall and Robinson, 2009](#)) or involve sample splitting ([Li et al., 2016](#)). Unlike existing studies, we attempt to use alternative discrete kernel functions in conjunction with the LSCV criterion.

[Rajagopalan and Lall \(1995\)](#) develop an ordered discrete version of the [Epanechnikov \(1969\)](#) kernel function which does not possess a geometric weighting scheme, providing sufficient smoothing in the presence of sparse data. Unfortunately, applied researchers who adopt kernel smoothing methods are largely unaware of this kernel function (an exception is [Guerra et al., 1997](#)) which motivates us to attempt to further its application. Specifically, we detail [Rajagopalan and Lall's \(1995\)](#) ordered discrete Epanechnikov kernel function and propose an unordered discrete Epanechnikov kernel function.

In a similar vein, [Kokonendji et al. \(2007\)](#) develop a so-called triangular probability mass function and use it as an ordered discrete kernel function. Their triangular kernel function does not impose a geometric weighting structure, but is less relevant for our discussion here as this kernel function is designed for use with count data with excess zeros and the function consists of two parameters which adds additional complications to bandwidth selection.

For both the unordered and ordered discrete Epanechnikov kernel functions, we derive the MSSE of the kernel density estimator (probability mass function). Further, we demonstrate that a sufficient condition for asymptotic normality of both the kernel density and regression estimators is satisfied by this new kernel, namely by establishing a second-order approximation of the discrete kernels proposed here, similar to that used by [Li and Racine \(2003\)](#).

The results here are unique relative to the continuous data setting where it is well known that kernel choice is ancillary to bandwidth choice. The discrete kernel seems to play a more important role. Given that the asymptotic bias and variance of kernel density and regression estimators are independent of the discrete kernel used to smooth the data, this becomes a finite sample issue; this topic is germane to study given that in the continuous only case it is relatively easy to assess efficiency loss through the use of a particular kernel relative to the optimal. Our goal here is to study the impact that the choice of discrete kernel has in a rigorous fashion through a variety of analytic, simulated and real data settings.

Here we examine the discrete Epanechnikov kernel functions versus [Aitchison and Aitken's \(1976\)](#), [Wang and vanRyzin's \(1981\)](#), and [Li and Racine's \(2003\)](#) kernel functions in simulations and empirical examples. For this set of kernel functions, the simulation results show the discrete Epanechnikov kernel functions generally perform well. We find that a researcher is generally no worse off and sometimes better off using a discrete Epanechnikov kernel in density estimation. However, the researcher is significantly worse off when using the ordered discrete Epanechnikov kernel in the (continuous) conditional density setting when there is an irrelevant ordered variable present. In both cross-sectional and longitudinal data regression, the researcher appears to be no worse off using the unordered discrete Epanechnikov kernel and is sometimes strictly better off. However, when the data are sparse, the Wang and van Ryzin or Li and Racine kernel performs better than the ordered discrete Epanechnikov kernel. This result is surprising as the ordered discrete Epanechnikov kernel was designed for sparse data in the density setting. It appears that these properties do not translate to the regression setting. In the case of ordered discrete kernels with longitudinal data, we find no substantial differences across kernel functions in our simulations. Our empirical examples largely mimic the simulation results except in the longitudinal data setting where we find both substantial gains and losses for the discrete Epanechnikov kernels.

The remainder of this paper is organized as follows: Section 2 presents the ordered discrete Epanechnikov kernel, develops the unordered discrete Epanechnikov kernel, compares the analytic properties of these kernel functions with those commonly used in the literature and presents the asymptotic properties of density and regression estimators using these kernels. Section 3 shows the finite sample performance via simulations. Section 4 provides several empirical illustrations and Section 5 concludes.

2. Discrete Epanechnikov kernel functions

For the case of a continuous random variable X , [Epanechnikov \(1969\)](#) shows that the MSE optimal second-order kernel function is

$$k(\psi_X) = \begin{cases} a\psi_X^2 + b & \text{if } |\psi_X| \leq 1 \\ 0 & \text{if } |\psi_X| > 1, \end{cases} \quad (1)$$

where $-a = b = 0.75$, $\psi_X = \frac{x-x}{h}$ and h is the bandwidth. [Rajagopalan and Lall \(1995\)](#) extend this set-up to an ordered, discrete random variable X . The discrete version of the optimal second-order kernel $l(\cdot)$ is required to satisfy two conditions: (A) $\sum_{x=x-h}^{x+h} l(\psi_X) = 1$ and (B) $\sum_{x=x-h}^{x+h} l(\psi_X)\psi_X = 0$. Condition (A) is the discrete counterpart of requiring a kernel function to integrate to 1, while Condition (B) is the discrete counterpart of having a symmetric kernel with zero mean. The constants

Download English Version:

<https://daneshyari.com/en/article/4949204>

Download Persian Version:

<https://daneshyari.com/article/4949204>

[Daneshyari.com](https://daneshyari.com)