Kernel-type density estimation on the unit interval

By M.C. JONES

Department of Statistics, The Open University, Walton Hall, Milton Keynes MK7 6AA, U.K.

m.c.jones@open.ac.uk

and D.A. HENDERSON

School of Mathematics & Statistics, Newcastle University, Newcastle upon Tyne NE1 7RU, U.K.

d.a.henderson@ncl.ac.uk

Summary

We consider kernel-type methods for estimation of a density on $[0, 1]$ which eschew explicit boundary correction. Our starting point is the successful implementation of beta kernel density estimators of Chen (1999). We propose and investigate two alternatives. For the first, we reverse the roles of estimation point x and datapoint X_i in each summand of the estimator. For the second, we provide kernels that are symmetric in x and X; these kernels are conditional densities of bivariate copulas. We develop asymptotic theory for the new estimators and compare them with Chen's in a substantial simulation study. We also develop automatic bandwidth selection in the form of 'rule-of-thumb' bandwidths for all three estimators. We find that our second proposal, that based on 'copula kernels', seems particularly competitive with the beta kernel method of Chen in integrated squared error performance terms. Advantages include its greater range of possible values at 0 and 1, the fact that it is a bona fide density, and that the individual kernels and resulting estimator are comprehensible in terms of a direct single picture (as is ordinary kernel density estimation on the line).

Some key words: Boundary behaviour; Copula; Kernel density estimation.

1. Introduction

Let $X_1, ..., X_n$ be a random sample from a univariate distribution with density f having support $I \equiv [0, 1]$. Suppose it is desired to estimate f at each point $0 < x < 1$ by means of a kernel-type estimator

$$\widehat{f}(x) = n^{-1} \sum_{i=1}^{n} K(x, X_i; h) \tag{1}$$

where h is the smoothing parameter, called the bandwidth. The usual version of $\widehat{f}$ for density estimation on the whole real line $\mathcal{R}$ takes $K(x, X; h) = h^{-1} K_R(h^{-1}(x - X))$ where K_R is a symmetric unimodal probability density function with support $\mathcal{R}$ or some finite interval such as $[-1, 1]$. This is, however, not available for estimation on the unit interval without correction for boundary effects, although many boundary correction schemes exist by now (e.g., Müller, 1991, Jones, 1993, Cheng *et al.*, 1997, Zhang *et al.*, 1999, Karunamuni & Alberts, 2005). See Silverman (1986), Wand & Jones (1995) and Simonoff (1996) for general background on kernel density estimation.

In important and insightful work, Chen (1999) introduces and develops — and in Chen (2000a,b, 2002) widely applies and adapts — kernel-type estimators of the form (1) tailored to the unit interval by choosing $K(x, X; h)$ to be a density *in* X with support I. Specifically, in his estimator $\hat{f}_1$, Chen takes K to be K_1, say, a beta density in X with parameters $(x/h^2) + 1$ and $\{(1 - x)/h^2\} + 1$, where we have replaced Chen's smoothing parameter $b > 0$ by h^2 to better fit with our later development. While it may seem 'obvious' that one might base kernel-type density estimators for support I on 'beta kernels', details of how properly to do so are by no means obvious and a successful implementation of this idea proved elusive until Chen's work. Background details of Chen's work are provided in Section 2.

In this paper, we investigate two novel approaches to the problem of kernel-type density estimation on I (plus some variations thereon) which are also, by and large, 'direct' and not explicitly making corrections for boundary effects. Each will be seen to have some practical merit, especially the second of these two, as does the best version of the estimators proposed by Chen (1999).

As our starting point, we note that the form of Chen's K_1 is such that (i) the kernel estimators displayed in Chen's (1999) Fig. 1 for use at several values of x cannot be superimposed and averaged out to provide a density estimate based on those values of x as if they were datapoints, and (ii) Chen's

estimator does not integrate to one. This is because K_1 is not a density *in* x. A first natural remedy for both points (i) and (ii) is to swop the roles of x and X_i in the summands of the beta kernel estimator: this is considered in our first new estimator, discussed in Section 3.

An attractive property of the standard real line kernel K_R is that it is symmetric in x and X and is therefore a density *in both* x *and* X. No 'kernel' considered in Sections 2 or 3 has this property. So, in Section 4, we ask that K be symmetric in x and X so that the ordinary kernel density estimator's indifference between consideration as a function of estimation point x or datapoint X is maintained. As well as the argument that such an estimator might be regarded as the 'fullest' analogue of the standard kernel density estimator, this proves to afford particularly good behaviour at the boundaries. The new requirement leads to an intimate connection to bivariate copulas i.e. symmetric bivariate distributions with uniform marginals (e.g. Joe, 1997, Nelsen, 2006). Indeed, our new kernel will be the conditional density of a (certain type of) symmetric bivariate copula. Moreover, the bandwidth proves to be intimately related to the dependence parameter of the copula. We will focus our efforts on one particular choice of copula-based kernels, that corresponding to the Gaussian copula, which may be regarded as a 'natural' adaptation of Gaussian kernels to the unit interval.

It should be noted that both of the new density estimators considered in this paper are bona fide probability densities. This is by no means the default in sophisticated kernel-type density estimation, one or other of the properties of nonnegativity and integration to unity (the latter in the case of Chen's estimator) often being sacrificed for otherwise improved performance.

The estimators of Section 2 to 4 are compared in a substantial simulation study in Section 5. In this section, the bandwidth is computed optimally (in integrated squared error terms) for each dataset, an option unavailable in practice. In these terms of practical *potential*, the estimators of Sections 2 and 4 will be seen to be preferred to that of Section 3. Practical bandwidth selection rules based on the 'rule-of-thumb' approach are then developed in Section 6.1 and their performance assessed in further simulations in Section 6.2. Our concluding remarks (Section 7) summarise the outcomes of this paper and look forward to similar developments for closely related topics.

For theoretical purposes, assume throughout this paper that f has two continuous derivatives. As is usual in kernel density estimation (e.g. Wand & Jones, 1995, Section 2.5), asymptotic results correspond to the smoothing parameter $h = h(n) \to 0$ as $n \to \infty$ in such a way that $nh \to \infty$.

2. Chen's estimators and their asymptotic biases and variances

We will rename Chen's (1999) first density estimator $\widehat{f}_1$ as $\widehat{f}_{C1}$ and note that it is given by

$$\widehat{f}_{C1}(x) = \frac{1}{nB((x/h^2)+1, \{(1-x)/h^2\}+1)} \sum_{i=1}^{n} X_i^{x/h^2}(1-X_i)^{(1-x)/h^2} \quad (2)$$

where $B(\cdot,\cdot)$ is the beta function. Chen (1999) shows that the asymptotic bias of $\widehat{f}_{C1}(x)$ is

$$\text{ABias}\{\widehat{f}_{C1}(x)\} = h^2\left\{(1-2x)f'(x) + \frac{1}{2}x(1-x)f''(x)\right\}.$$

Note that the asymptotic bias is of order h^2 for all values of x, showing that $\widehat{f}_{C1}$ is, in bias terms, asymptotically free of boundary effects. (Boundary effects usually show up as terms of $O(h)$, or even sometimes $O(1)$, in the asymptotic bias.) This boundary bias behaviour is, however, achieved at the expense of an increased boundary variance:

$$\text{AVar}\{\widehat{f}_{C1}(x)\} = \begin{cases} \frac{f(x)}{2nh\sqrt{\pi}\sqrt{x(1-x)}} & \text{if } x/h^2 \text{ and } (1-x)/h^2 \to \infty, \\ \frac{f(x)\Gamma(2\kappa+1)}{2nh^2 4^\kappa \Gamma^2(\kappa+1)} & \text{if } x/h^2 \to \kappa \text{ or } (1-x)/h^2 \to \kappa\ , \end{cases}$$

$\kappa \geq 0$. This is a somewhat unusual boundary effect albeit with increased variance only over a particularly small boundary region (of $O(h^2)$). Chen notes, however, that "the impact of the increased variance near the boundary on the mean integrated squared error [MISE] is negligible" and the practical importance of this asymptotic result is arguable. (For more on asymptotic theory for $\widehat{f}_{C1}$, see Bouezmarni & Rolin, 2003.)

With a view to removing (most of) the dependence of the asymptotic bias on $f'(x)$, Chen (1999) also proposed a modified estimator, his $\widehat{f}_2$, which we shall call $\widehat{f}_{C2}$. Introduce the function

$$\rho_h(x) = 2h^4 + \frac{5}{2} - \sqrt{4h^8 + 6h^4 + \frac{9}{4} - x^2 - (x/h^2)}.$$

Then,

$$\widehat{f}_{C2}(x) = \begin{cases} \frac{1}{nB(x/h^2,(1-x)/h^2)} \sum_{i=1}^n X_i^{(x/h^2)-1}(1-X_i)^{\{(1-x)/h^2\}-1} & \text{if } x \in [2h^2, 1-2h^2], \\ \frac{1}{nB(\rho_h(x),(1-x)/h^2)} \sum_{i=1}^n X_i^{\rho_h(x)-1}(1-X_i)^{\{(1-x)/h^2\}-1} & \text{if } x \in [0, 2h^2), \\ \frac{1}{nB(x/h^2,\rho_h(1-x))} \sum_{i=1}^n X_i^{(x/h^2)-1}(1-X_i)^{\rho_h(1-x)-1} & \text{if } x \in (1-2h^2, 1]. \end{cases}$$

The asymptotic bias of $\widehat{f}_{C2}(x)$ differs between boundary and interior regions, but is always of order h^2:

$$\text{ABias}\{\widehat{f}_{C2}(x)\} = \begin{cases} \frac{1}{2}h^2 x(1-x)f''(x) & \text{if } x \in [2h^2, 1-2h^2], \\ h^2\xi_h(x)f'(x) & \text{if } x \in [0, 2h^2), \\ -h^2\xi_h(1-x)f'(x) & \text{if } x \in (1-2h^2, 1]. \end{cases}$$

where $\xi_h(x) = (1-x)\{\rho_h(x) - (x/h^2)\}/\{1 + h^2\rho_h(x) - x\} = O(1)$. Also,

$$\text{AVar}\{\widehat{f}_{C2}(x)\} = \text{AVar}\{\widehat{f}_{C1}(x)\} \quad \text{for } x/h^2 \text{ and } (1-x)/h^2 \to \infty$$

while the multiplier of $(nh^2)^{-1}$ in the case of $x/h^2 \to \kappa$ or $(1-x)/h^2 \to \kappa$, $\kappa \geq 0$, "has a slightly different form" (Chen, 1999, p.136).

For fixed $h, X_1, ..., X_n > 0$,

$$\widehat{f}_{C1}(0) = \frac{1}{n}\left(1 + \frac{1}{h^2}\right)\sum_{i=1}^{n}(1-X_i)^{1/h^2} \text{ and } \widehat{f}_{C2}(0) = \frac{1}{nh^2}\sum_{i=1}^{n}(1-X_i)^{(1/h^2)-1}$$

suggesting a propensity, in practice, for estimating $f(0)$ by a non-zero value in all cases (and similarly when $x = 1$). It is, therefore, the case that these estimators are better in situations where $f(0) > 0$ than when $f(0) = 0$; see Section 5.

While it is perhaps unfortunate that the simplicity of $\widehat{f}_{C1}$ has been sacrificed in $\widehat{f}_{C2}$ and, indeed, that explicit boundary correction has been resorted to even when using beta kernels, it is the case that $\widehat{f}_{C2}$ is a successful estimator of f on $[0, 1]$. Chen (1999) shows that it has smaller asymptotic MISE than $\widehat{f}_{C1}$ and that its simulation performance is also superior and "is a serious competitor with the existing density estimators"; $\widehat{f}_{C2}$ will also play an important role in this paper and will prove to be a difficult estimator to improve upon.

3. Swopping the roles of x and X

3·1. Our first new estimator

Swopping the roles of x and X_i in (2) yields the new estimator

$$\widehat{f}_{Xx}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{x^{X_i/h^2}(1-x)^{(1-X_i)/h^2}}{B((X_i/h^2)+1, \{(1-X_i)/h^2\}+1)}. \tag{3}$$

Clearly, $\int_0^1 \widehat{f}_{Xx}(x)dx = 1$. The beta densities which comprise $\widehat{f}_{Xx}$ are shown in Fig. 1 for the case $h = \sqrt{0.1}$ and the following pseudo-dataset of 11 points:

$$0.1, 0.16, 0.25, 0.34, 0.42, 0.5, 0.58, 0.66, 0.75, 0.84, 0.9.$$

The bolder line on Fig. 1 is the resulting $\widehat{f}_{Xx}$, formed simply by averaging the individual curves. Note that there is no comparable single picture for either $\hat{f}_{C1}$ or $\hat{f}_{C2}$; for those estimators, a different such picture is required for each point of evaluation x. However, for fixed $h, X_1, ..., X_n > 0$, $\widehat{f}_{Xx}(0) = \widehat{f}_{Xx}(1) = 0$. This propensity towards zero boundary values is alleviated a little by the observation that non-zero values arise when both x and X_i approach the boundary. For example, if $x = X_i \to 0$, the ith summand of $\widehat{f}_{Xx}$ is $n^{-1}(1 + h^{-2})$.

* * * Fig. 1 about here * * *

3·2. Theory: asymptotic bias and variance

From (3), for $x \neq 0, 1$,

$$\begin{aligned} E\{\widehat{f}_{Xx}(x)\} &= \int_0^1 f(z) \exp\left[(z/h^2)\log x + \{(1-z)/h^2\}\log(1-x)\right. \\ &- \left.\log B((z/h^2)+1, \{(1-z)/h^2\}+1)\right] dz. \end{aligned}$$

By Stirling's approximation,

$$\begin{aligned} \log B((z/h^2)+1, \{(1-z)/h^2\}+1) &\simeq (1/h^2)\{z\log z + (1-z)\log(1-z)\} + \log h \\ &+ \frac{1}{2}\log 2\pi + \frac{1}{2}\log z(1-z) + \frac{h^2(z^2-z+1)}{12z(1-z)} \end{aligned}$$

and so

$$E\{\widehat{f}_{Xx}(x)\} \simeq \frac{1}{h\sqrt{2\pi}} \int_0^1 g(z) \exp\left[\frac{z}{h^2}\log\frac{x}{z} + \frac{(1-z)}{h^2}\log\frac{(1-x)}{(1-z)} - \frac{h^2(z^2-z+1)}{12z(1-z)}\right] dz$$

where $g(z) \equiv f(z)/\sqrt{z(1-z)}$. Making the usual kernel density estimation substitution of $w = (x-z)/h$ yields

$$E\{\widehat{f}_{Xx}(x)\} \simeq \frac{1}{\sqrt{2\pi}} \int_{-(1-x)/h}^{x/h} g(x-hw) \exp\left[-\frac{w^2}{2x(1-x)} + \frac{hw^3(2x-1)}{6x^2(1-x)^2}\right.$$

$$
\begin{aligned}
&\quad - \frac{h^2 w^4(3x^2-3x+1)}{12x^3(1-x)^3} - \frac{h^2(x^2-x+1)}{12x(1-x)}\Bigg] dw \\
&\simeq \frac{1}{\sqrt{2\pi}} \int_{-(1-x)/h}^{x/h} g(x-hw) \exp\left\{-\frac{w^2}{2x(1-x)}\right\} \Bigg[1 + \frac{hw^3(2x-1)}{6x^2(1-x)^2} \\
&\quad - \frac{h^2 w^4(3x^2-3x+1)}{12x^3(1-x)^3} - \frac{h^2(x^2-x+1)}{12x(1-x)} + \frac{h^2 w^6(2x-1)^2}{72x^4(1-x)^4}\Bigg] dw.
\end{aligned}
$$

Making the further substitution $v = w/\sqrt{x(1-x)}$ and writing ϕ for the standard normal density function gives

$$
\begin{aligned}
E\{\widehat{f}_{Xx}(x)\} &\simeq \sqrt{x(1-x)} \int_{-\sqrt{(1-x)/x}/h}^{\sqrt{x/(1-x)}/h} g\left(x - h\sqrt{x(1-x)}v\right) \phi(v) \Bigg[1 + \frac{hv^3(2x-1)}{6\sqrt{x(1-x)}} \\
&\quad - \frac{h^2 v^4(3x^2-3x+1)}{12x(1-x)} - \frac{h^2(x^2-x+1)}{12x(1-x)} + \frac{h^2 v^6(2x-1)^2}{72x(1-x)}\Bigg] dv.
\end{aligned}
$$

Now, (very) close to the boundary, when $x/h^2 \to \kappa$ or when $(1-x)/h^2 \to \kappa$, $\kappa > 0$, $E\{\widehat{f}_{Xx}(x)\} \simeq f(x)\Phi(\sqrt{\kappa})$ where Φ is the standard normal distribution function. There is, therefore, an $O(1)$ boundary bias. In the interior of $[0,1]$,

$$
\begin{aligned}
E\{\widehat{f}_{Xx}(x)\} &\simeq \sqrt{x(1-x)}\Big[g(x) + \frac{1}{2}h^2 x(1-x)g''(x) - \frac{1}{2}h^2 g'(x)(2x-1) \\
&\quad - \frac{h^2 g(x)(3x^2-3x+1)}{4x(1-x)} - \frac{h^2 g(x)(x^2-x+1)}{12x(1-x)} + \frac{5h^2 g(x)(2x-1)^2}{24x(1-x)}\Bigg] \\
&= f(x) - \frac{1}{2}h^2\sqrt{x(1-x)}(2x-1)g'(x) + \frac{1}{2}h^2\{x(1-x)\}^{3/2}g''(x) \\
&\quad - h^2\frac{g(x)}{8\sqrt{x(1-x)}}
\end{aligned}
$$

and a little further manipulation shows that

$$
\mathrm{ABias}(\widehat{f}_{Xx}(x)) = \frac{1}{2}h^2 x(1-x)f''(x) \qquad \text{if } x/h^2 \text{ and } (1-x)/h^2 \to \infty.
$$

This interior asymptotic bias is simple, attractive and the same as the interior asymptotic bias of $\widehat{f}_{C2}$ (and holds for a slightly wider definition of 'interior').

The asymptotic variance of $\widehat{f}_{Xx}(x)$ is $n^{-1}\{E(J^2(X_i)\} - E\{\widehat{f}_{Xx}(x)\}^2$ where $J(X_i)$ is the ith summand in (3). A development parallel to that above leads

to

$$\begin{aligned} E\{J^2(X_i)\} &\simeq \frac{1}{h2\pi}\int_{-(1-x)/h}^{x/h} G(x-hw)\exp\left\{-\frac{w^2}{x(1-x)}\right\} \\ &\simeq \frac{\sqrt{x(1-x)}}{h2\sqrt{\pi}}\int_{-\sqrt{2(1-x)/x}/h}^{\sqrt{2x/(1-x)}/h} G\left(x-h\sqrt{x(1-x)}v/\sqrt{2}\right)\phi(v)dv \end{aligned}$$

where $G(x) \equiv f(z)/\{z(1-z)\}$. We therefore arrive at

$$\mathrm{AVar}\{\widehat{f}_{Xx}(x)\} = \begin{cases} \frac{f(x)}{2nh\sqrt{\pi}\sqrt{x(1-x)}} & \text{if } x/h^2 \text{ and } (1-x)/h^2 \to \infty, \\ \frac{f(x)\Phi(\sqrt{2\kappa})}{2nh^2\sqrt{\pi\kappa}} & \text{if } x/h^2 \to \kappa \text{ or } (1-x)/h^2 \to \kappa, \\ \frac{f(x)}{4nh^{1+2m}\sqrt{\pi\kappa_m}} & \text{if } x/h^m \to \eta \text{ or } (1-x)/h^m \to \eta\ , \end{cases}$$

$\kappa, \eta > 0$, $m > 2$. This is broadly parallel with the case of Chen's estimators: an $o((nh)^{-1})$ inflated variance very close to the boundary, precisely the same asymptotic variance as $\widehat{f}_{C1}$ and $\widehat{f}_{C2}$ in the interior.

However, the deleterious effects of the density estimate at the boundary having a propensity towards zero values will show through into practical performance (Section 5) and so we next pursue a second alternative approach.

4. Copula-based kernels

4·1. General considerations

We have already argued, in Section 1, for $K(x, X; h)$ to be the conditional density $c(x|X)$, or equivalently $c(X|x)$, where c, which depends on h, is the density of a symmetric copula. Further, it is natural to expect the kernel used at $1-X$ to be a reflected version of the kernel used at X, and hence therefore for that used at $X = 0.5$ to be symmetric in X; ditto with x replacing X. This demands reflective symmetry (as well as ordinary symmetry) of the copula: $c(x, y) = c(y, x) = c(1-x, 1-y) = c(1-y, 1-x)$.

Some families of copulas interpolate between the case of independence and the Fréchet upper bound which corresponds to perfect positive dependence; in simple cases, a single 'dependence' parameter θ controls that interpolation, typically increasing from $\theta = 0$ at independence. This, too, is just what we want. If we take h to be a simple decreasing function of θ such that $h = 0$ at the Fréchet upper bound, this corresponds precisely to kernels reducing

to Dirac delta functions as $h \to 0$, and hence to no smoothing. On the other hand, when h is large, corresponding to $\theta = 0$, all kernels are uniform on [0,1] and a uniform distribution is fitted to the data, the ultimate case of oversmoothing.

While $K(0, X; h) = K(x, 0; h)$ certainly need not be zero, neither should it be infinite for $x, X, h > 0$; any such infinite value would persist into the density estimate. Absolute continuity of the copula is also a strict requirement for us, as is unimodality of the resulting kernels. A certain amount of continuous differentiability, at least to the second derivative, is desirable to avoid artefactual jumps or kinks in the density estimate, while tractability will help computationally.

There remain just a few families of copulas which satisfy the above: see e.g. Joe (1997, Section 5.1). We concentrate on what we think is the most appealing one, the Gaussian copula, in the next section (and the remainder of the paper), although some interesting alternatives will also be mentioned briefly.

4·2. Our second new estimator and other copula-based kernels

The copula associated with the bivariate normal distribution with correlation ρ is given by

$$c(x, y; \rho) = \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{[\rho^2\{\Phi^{-1}(x)\}^2 - 2\rho\Phi^{-1}(x)\Phi^{-1}(y) + \rho^2\{\Phi^{-1}(y)\}^2]}{2(1-\rho^2)} \right) \tag{4}$$

where Φ^{-1} is the standard normal quantile function. As well as Joe (1997), see Sungur (1990), Klaassen & Wellner (1997), Song (2000) and Demarta & McNeil (2005). We use this only for $\rho \geq 0$ and set $h^2 = 1 - \rho$. Then our Gaussian copula-based density estimator has the form

$$\begin{aligned} \widehat{f}_{GC}(x) &= \frac{1}{n}\sum_{i=1}^{n} c(x, X_i; 1-h^2) \\ &= \frac{1}{nh\sqrt{2-h^2}} \exp\left[-\frac{(1-h^2)^2\{\Phi^{-1}(x)\}^2}{2h^2(2-h^2)} \right] \\ &\times \sum_{i=1}^{n} \exp\left(-\frac{(1-h^2)}{2h^2(2-h^2)} \left[(1-h^2)\{\Phi^{-1}(X_i)\}^2 - 2\Phi^{-1}(x)\Phi^{-1}(X_i) \right] \right). \end{aligned} \tag{5}$$

The copula (4) is shown in Fig. 2 for $\rho = 0.9$ and hence $h = \sqrt{0.1}$. The associated kernel functions are shown in Fig. 3 for the pseudo-dataset introduced in Section 3.1. The solid line on Fig. 3 is the resulting density estimate which, like the estimate in Fig. 1, is formed simply by averaging the individual curves. Figs 1 and 3, which use the same value of h, are pretty similar. There appears to be a slightly greater degree of smoothing in the latter towards the centre and a slightly greater degree of smoothing in the former towards either end.

* * * Figs 2 and 3 about here * * *

Since, for fixed $h, X_1, ..., X_n > 0$, $\widehat{f}_{GC}(0) = \widehat{f}_{GC}(1) = 0$, it at first glance seems that we have the same problem as with $\widehat{f}_{Xx}$. However, for $x = X_i$, the ith summand in $\widehat{f}_{GC}(x)$ is

$$\frac{1}{nh\sqrt{2-h^2}} \exp\left[\frac{\rho}{1+\rho}\{\Phi^{-1}(x)\}^2\right] \to \infty,$$

even for fixed h. In practice, such greater limiting values translate into a more widespread distribution of values of $\widehat{f}_{GC}(0)$ compared with $\widehat{f}_{Xx}(0)$.

Copulas associated with elliptically symmetric distributions more generally (Fang et al., 2002, Frahm et al., 2003, Demarta & McNeil, 2005) are also good candidates to form the basis for kernels on I. In kernel density estimation on $\mathcal{R}$, the most popular kernels are probably the Gaussian density and several symmetric beta (or Pearson Type II) densities on $[-1, 1]$, notably the uniform, "Epanechnikov" and "biweight" kernels which correspond to both beta parameters, a say, equalling 1, 2 and 3, respectively. The copulas based on the bivariate elliptically symmetric beta (or bivariate Pearson Type II) distribution now form 'natural' analogues of the ordinary symmetric beta kernels for use on $[0, 1]$. Adapting formulae in Example 2.3 of Fang et al. (2002), we have $c(x, y; \rho) = b(Q_a(x), Q_a(y))$ where

$$b(u,v) = \frac{\Gamma^2(a)}{\Gamma(a-(1/2))\Gamma(a+(1/2))} \frac{\{1-(1-\rho^2)^{-1}(u^2+v^2-2\rho uv)\}^{a-(3/2)}}{\sqrt{1-\rho^2}\ (1-u^2)^{a-1}(1-v^2)^{a-1}}$$

and $Q_a(x)$ is the quantile function of the Beta(a, a) distribution on $[-1, 1]$. Again, set $h^2 = 1 - \rho$. Like their real-line counterparts, the resulting kernels on I are (further) restricted in their support. Another copula with the

right properties but not based on an elliptically symmetric distribution is the Plackett (1965) copula:

$$p(x,y) = \frac{(\eta+1)(1+\eta(x+y-2xy))}{\{(1+\eta(x+y))^2 - 4\eta(\eta+1)xy\}^{3/2}},$$

taking $h = 1/(1+\eta)$ and $\eta > 0$. Neither this nor the beta copula-based kernels appear to have any great advantage over the Gaussian copula-based kernels in practice, so we do not pursue them further here. (We implemented the Plackett-based estimator in our simulations and found its performance to be similar to, but overall a little inferior to, that of the Gaussian-based estimator.)

4·3. Theory: asymptotic bias and variance

Asymptotic bias *manipulations* are, in this case, much more similar to those of Chen (1999). First, note that

$$E(\widehat{f}_{GC}(x)) = \int_0^1 c(x,z;1-h^2)f(z)dz = E(f(Z)),$$

where $Z \sim c(z|x;1-h^2)$, and that

$$E(f(Z)) \simeq f(x) + (\mu_x - x)f'(x) + [\{\sigma_x^2 + (\mu_x - x)^2\}/2]f''(x)$$

where μ_x and σ_x^2 are the mean and variance of the distribution of Z. Now specialise to the case of the Gaussian copula. Then, $Z = \Phi(Y)$ and $Y \sim N((1-h^2)\Phi^{-1}(x), h^2(2-h^2))$ so that

$$\begin{aligned}
\mu_x &\simeq \Phi\{(1-h^2)\Phi^{-1}(x)\} + \{h^2(2-h^2)/2\}\phi'\{(1-h^2)\Phi^{-1}(x)\} \\
&\simeq x - h^2\Phi^{-1}(x)\phi(\Phi^{-1}(x)) + h^2\phi'(\Phi^{-1}(x)) \\
&= x - 2h^2\Phi^{-1}(x)\phi(\Phi^{-1}(x)).
\end{aligned}$$

Also,

$$\sigma_x^2 \simeq h^2(2-h^2)\left[\Phi'\{(1-h^2)\Phi^{-1}(x)\}\right]^2 \simeq 2h^2\phi^2(\Phi^{-1}(x)).$$

It follows that

$$\begin{aligned}
\mathrm{ABias}(\widehat{f}_{GC}(x)) &\simeq -2h^2\Phi^{-1}(x)\phi(\Phi^{-1}(x))f'(x) + h^2\phi^2(\Phi^{-1}(x))f''(x) \\
&= h^2(pf')'(x)
\end{aligned}$$

where $p(x) = \phi^2(\Phi^{-1}(x))$. This holds for all $x \in [0,1]$.

It is also the case that

$$\begin{aligned}\int_0^1 c^2(x,z;1-h^2)f(z)dz &= E\{c(x,Z;1-h^2)f(Z)\} \\ &\simeq c(x,x;1-h^2)f(x) \\ &= \frac{1}{h\sqrt{2-h^2}}\exp\left[\frac{(1-h^2)}{(2-h^2)}\{\Phi^{-1}(x)\}^2\right]f(x)\end{aligned}$$

and hence that

$$\mathrm{AVar}\{\widehat{f}_{GC}(x)\} = \begin{cases} \frac{f(x)}{nh2\sqrt{\pi}\phi\{\Phi^{-1}(x)\}} & \text{if } x/h^m \text{ and } (1-x)/h^m \to \infty, \\ \frac{f(x)}{nh^{1+2m}\sqrt{2}\eta^2} & \text{if } x/h^m \to \eta \text{ or } (1-x)/h^m \to \eta\ , \end{cases}$$

$\eta, m > 0$. We have used the facts that $\Phi^{-1}(x) \sim -\sqrt{-2\log x}$ and $\Phi^{-1}(1-x) \sim \sqrt{-2\log(1-x)}$ for $x \to 0$.

4·4. A theoretical comparison

We have already stressed that the asymptotic bias and variance properties of $\widehat{f}_{C2}$ and $\widehat{f}_{Xx}$ are very similar and now we add that those of $\widehat{f}_{GC}$ are of very similar *form* too. A precise comparison is possible in terms of asymptotic MSE at the point $x_0 = 1/2$ (where $\widehat{f}_{C1}$ has the same properties as $\widehat{f}_{C2}$ and $\widehat{f}_{Xx}$ too). In fact,

$$\mathrm{ABias}^2\{\widehat{f}_{C1}(x_0)\} = \mathrm{ABias}^2\{\widehat{f}_{C2}(x_0)\} = \mathrm{ABias}^2\{\widehat{f}_{Xx}(x_0)\} = \frac{1}{64}h^4\{f''(x_0)\}^2$$

while

$$\mathrm{ABias}^2\{\widehat{f}_{GC}(x_0)\} = \frac{1}{4\pi^2}h^4\{f''(x_0)\}^2.$$

The former is smaller than the latter. Also,

$$\mathrm{AVar}\{\widehat{f}_{C1}(x_0)\} = \mathrm{AVar}\{\widehat{f}_{C2}(x_0)\} = \mathrm{AVar}\{\widehat{f}_{Xx}(x_0)\} = (nh\sqrt{\pi})^{-1}f(x_0)$$

while

$$\mathrm{AVar}\{\widehat{f}_{GC}(x_0)\} = (nh\sqrt{2})^{-1}f(x_0).$$

The former is, again, smaller than the latter and so the Gaussian copula estimator performs less well at x_0 than do any of the estimators of Sections 2 and

3. To quantify this further, with asymptotic MSE of estimator j, say, being of the usual kernel density estimation form $B_j\{f''(x_0)\}^2h^4 + V_jf(x_0)/(nh)$, the estimator-specific multipliers of the otherwise common optimal bandwidth formulae are $h_0(x_0) = (V_j/B_j)^{1/5}$ and the estimator-specific multipliers of the optimal MSEs are $M_0(x_0) = (B_jV_j^4)^{1/5}$. We have that

$$h_0\{\widehat{f}_{C1}(x_0)\} = h_0\{\widehat{f}_{C2}(x_0)\} = h_0\{\widehat{f}_{Xx}(x_0)\} = (64/\sqrt{\pi})^{1/5} = 2.049$$

and $h_0\{\widehat{f}_{GC}(x_0)\} = (2\sqrt{2}\pi^2)^{1/5} = 1.946$. This chimes with what was said in Section 4.2 about the relative degrees of smoothing in the centres of Figs 1 and 3. However,

$$M_0\{\widehat{f}_{C1}(x_0)\} = M_0\{\widehat{f}_{C2}(x_0)\} = M_0\{\widehat{f}_{Xx}(x_0)\} = (64\pi^2)^{-1/5} = 0.275$$

and $M_0\{\widehat{f}_{GC}(x_0)\} = (16\pi^2)^{-1/5} = 0.363$; the estimators of Sections 2 and 3 asymptotically outperform $\widehat{f}_{GC}$ at $x_0 = 1/2$ by an MSE factor of $4^{-1/5} = 0.758$.

5. Simulation study I

In our initial simulation study, reported in this section, we divorce the issue of potential quality of estimator per se from that of bandwidth selection by comparing estimator performances optimised over choice of bandwidth. The latter choice would be unavailable in practice; the effects of (one version of) bandwidth selection will be explored in Section 6.

* * * Table 1 and Fig. 4 about here * * *

We compare performance of the four estimators $\widehat{f}_{C1}$, $\widehat{f}_{C2}$, $\widehat{f}_{Xx}$ and $\widehat{f}_{GC}$ for samples simulated from a testbed of some 16 densities on $[0, 1]$. These densities are listed in Table 1 and depicted in Fig. 4. For each density, 1000 datasets of size $n = 50$ and $n = 500$ were simulated. All density estimates were evaluated on an equally spaced grid of 999 points: 0.001(0.001)0.999. The (global) ISE of each $\widehat{f}$ was approximated by $\mathrm{ISE}(\widehat{f}) \simeq 0.001 \sum_{j=1}^{999} \left\{\widehat{f}(j/1000) - f(j/1000)\right\}^2$. The density estimates were evaluated over appropriate grids of values of $h > 0$ and the value of h that gave the minimum ISE for each density/estimator was calculated. 'ISEs' in this section therefore refer to ISEs minimised over h and represent a version of 'best case' performance for each density estimator. We only present results for the $n = 50$

case here because, qualitatively, results for $n = 500$ are similar. These are given in Table 2 which includes simulated values of MISE$\times 10^3$ (and their standard errors) for each estimator along with their median percentage reductions in ISE compared with that of $\widehat{f}_{C2}$. Minimum values of MISE and values not significantly different from those minima are highlighted in bold type. Also in Table 2, the line labelled 'All' gives the results averaged over all 16 densities. A graphical depiction of the ISE results is given in Fig. 5.

* * * Table 2 and Fig. 5 about here * * *

Now, from Table 2, it can be seen that $\widehat{f}_{C2}$ either 'wins' or 'shares the lead' on 8 (out of 16) occasions while the same is true for $\widehat{f}_{Xx}$ on 9 occasions; $\widehat{f}_{GC}$, on the other hand, wins only twice (and $\widehat{f}_{C1}$ once). However, in terms of the average MISE over all 16 densities, the ranking from best to worst is $\widehat{f}_{GC}$, $\widehat{f}_{C2}$, $\widehat{f}_{C1}$, $\widehat{f}_{Xx}$! The explanation is that while $\widehat{f}_{C2}$ and $\widehat{f}_{Xx}$ are 'best' for quite a wide range of densities each, their performance for densities for which they are not optimal can be very far from optimal (especially in the case of $\widehat{f}_{Xx}$). On the other hand, while $\widehat{f}_{GC}$ is rarely best it is rarely hugely suboptimal either. The consistent improvement of Chen's estimator $\widehat{f}_{C2}$ over his initial estimator $\widehat{f}_{C1}$ seems to be confirmed.

Consideration of the nature of the densities which each estimator prefers appears to come down largely to their boundary behaviour. To investigate this further, we provide Table 3 in which the quality of estimation of each of the 16 densities of Table 1 at the specific points $x = 0.01$ and $x = 0.99$ near each boundary is investigated. MSEs (and standard errors) are given there for the case $n = 500$, the larger overall sample size being taken to provide a more reasonable amount of data near the boundaries. We also use the same bandwidth values as in Table 2 and do not attempt to make bandwidth selection specific to estimation at 0.01 or 0.99; this is because we are really more interested in the overall quality of density estimation on $(0, 1)$ than very specifically in density estimation at a boundary. An increased number of 'equally good winning' estimators is observed. (In an attempt to overcome possible problems caused by skewed sampling distributions, bootstrap standard errors were calculated in addition to the normal theory ones provided here, but with almost no effect on the qualitative results.)

* * * Table 3 about here * * *

The situation proves to be fairly clear. First, $\widehat{f}_{GC}$'s two ('easy') wins occur for the two densities (5 and 10) with asymptotes at one or both boundaries,

the latter being coped with by $\widehat{f}_{GC}$ much better than by the other estimators. Second, best performances near boundaries where the densities have finite non-zero values are shared mostly between the two versions of Chen's estimator. Third, when the density is zero at the boundary, $\widehat{f}_{Xx}$ wins most, but both $\widehat{f}_{C2}$ and $\widehat{f}_{GC}$ win sometimes too. Overall — for this particular combination of boundary density values — $\widehat{f}_{C1}$ and $\widehat{f}_{GC}$ win, the former being (surprisingly) in front of $\widehat{f}_{C2}$ by dint largely of better performance for densities 5 and 10. It seems to us that $\widehat{f}_{Xx}$ cannot be recommended for general use on $[0, 1]$ because — as mentioned at the end of Section 4 — it really cannot cope very well with non-zero densities at boundaries.

6. Bandwidth selection

6·1. Rules-of-thumb

In this section, we explore the development and use of the simplest form of asymptotics-based bandwidth selection methodology, the 'rule-of-thumb' (or reference distribution) bandwidth. This consists in choosing h to minimise estimated versions of the asymptotic weighted mean integrated squared error (WMISE) of each estimator, obtained by replacing f by a suitable parametric fitted distribution. Weightings are introduced here to ensure finiteness of integrals. In fact, this bandwidth selection methodology is: (a) driven by global and to a large extent 'central', as opposed to boundary, estimation quality; and, as with all rule-of-thumb bandwidth selectors, (b) tends to yield a fairly high degree of smoothing which may not fare well for complicated multimodal densities.

The rule-of-thumb bandwidth for $\widehat{f}_{C2}$ and $\widehat{f}_{Xx}$ (and $\widehat{f}_{C1}$, although we do not pursue this case) is taken to minimise

$$\mathrm{WMISE}_1 \equiv \frac{1}{4}h^4 \int_0^1 z^5(1-z)^5(f_1'')^2(z)dz + \frac{\int_0^1 z^{5/2}(1-z)^{5/2}f_1(z)dz}{2nh\sqrt{\pi}}; \quad (6)$$

here, WMISE_1 arises from squared bias of the form $h^4z^2(1-z)^2(f'')^2(z)/4$, variance $f(z)/\{2nh\sqrt{\pi}\sqrt{z(1-z)}\}$ and weight function $z^3(1-z)^3$, while f_1 is the reference distribution. We take the latter to be the beta distribution with parameters a and b estimated by their method of moments estimators $\widehat{a}$ and $\widehat{b}$. The resulting formulae for the rule-of-thumb bandwidth h_1 is given in the Appendix.

Similarly, for $\widehat{f}_{GC}$,

$$\text{WMISE}_2 \equiv h^4 \int_0^1 p^{1/2}(z)\{(pf_2')'\}^2(z)dz + \frac{1}{2nh\sqrt{\pi}} \tag{7}$$

where $p(z) = \phi^2(\Phi^{-1}(z))$ as in Section 4.3. In this case, we take the weight function to be $p^{1/2}(z)$ and the reference distribution to have density

$$f_2(z) = \frac{\widehat{\sigma}^{-1}\phi(\widehat{\sigma}^{-1}(\Phi^{-1}(z) - \widehat{\mu}))}{\phi(\Phi^{-1}(z))}.$$

This distribution is that of $X = \Phi(Y)$ where $Y \sim N(\mu, \sigma^2)$; the parameters are therefore estimated by the sample mean and standard deviation of $Y_1 = \Phi^{-1}(X_1), ..., Y_n = \Phi^{-1}(X_n)$. Density f_2 is one of many natural generalisations of the transformation approach of Johnson (1949) which the first author of this paper is currently exploring. The resulting bandwidth formula (see the Appendix) turns out to be the simple

$$h_2 = \widehat{\sigma}\left\{2\widehat{\mu}^2\widehat{\sigma}^2 + 3(1-\widehat{\sigma}^2)^2\right\}^{-1/5} n^{-1/5}. \tag{8}$$

6·2. Simulation study II

We now incorporate the rule-of-thumb bandwidth selections in place of the ISE-optimal bandwidths in each of the estimators $\widehat{f}_{C2}$, $\widehat{f}_{Xx}$ and $\widehat{f}_{GC}$ and look at their performance on the simulated data from Section 5; the test concentrates on the difficult $n = 50$ case. Results are given in Table 4, where median % reductions are given relative to $\widehat{f}_{C2,h_1}$ (in an obvious notation), and depicted in Fig. 6.

* * * Table 4 and Fig. 6 about here * * *

'Winners' on this occasion are $\widehat{f}_{C2}$ 5 times, $\widehat{f}_{Xx}$ 6 times and $\widehat{f}_{GC}$ 6 times (the lead is shared for density 16). There is a contrast here with the earlier results, suggesting that $\widehat{f}_{GC}$ with its rule-of-thumb bandwidth selection is more competitive with $\widehat{f}_{C2}$ and $\widehat{f}_{Xx}$ with their rule-of-thumb bandwidth selections than it was in the ISE-optimal bandwidth case. In terms of average MISE over all 16 densities, however, the performances of $\widehat{f}_{GC}$ and $\widehat{f}_{C2}$ are now comparable whereas $\widehat{f}_{GC}$ 'won' (fairly narrowly) previously! Patterns of

results remain much the same within many of the densities of the simulation testbed (compare Figs 5 and 6), levels of ISE and MISE necessarily being somewhat higher in Table 4/Fig. 6 than in Table 2/Fig. 5, with the increases being most marked for $\widehat{f}_{Xx}$.

There are four exceptional densities for which $\widehat{f}_{GC}$ has taken over as best in the rule-of-thumb bandwidth case. Two of these, numbers 2 and 14, are the clearest bi- and tri-modal densities in the collection. Results for all three estimators are bad in these cases because the rule-of-thumb is selecting bandwidths that are too large; indeed, the multimodality feature is being obscured in the corresponding density estimators. When $n = 500$ (results not shown), the estimates do much better reflect the bimodality in Density 2 (though in oversmoothed form) but they do not capture the trimodality in Density 14. For that density, $\widehat{f}_{GC}$ tends to suggest two modes, but that is better than the other estimators. The two densities for which rule-of-thumb bandwidth selection has profited $\widehat{f}_{GC}$ in the context of good estimation quality are numbers 7 and 11.

We also briefly compared our results with those of a standard kernel density estimator with rule-of-thumb bandwidth (Silverman, 1986) applied to probit-transformed versions of the data (and then transformed back) and found that it did not compete well except for a small number of densities including those with bi- and tri-modal structure.

7. Concluding remarks

The general good performance of Chen's (1999) main beta kernel estimator, which we have called $\widehat{f}_{C2}$, has been confirmed, but we believe that we have proposed an equally, if not more, attractive alternative in $\widehat{f}_{GC}$. A practical advantage of the latter is that $\widehat{f}_{GC}(0) \geq 0$ while $\widehat{f}_{C2}(0) > 0$ (similarly for values at 1). Conceptually, $\widehat{f}_{GC}$ shares the property of the kernel density estimator on $\mathcal{R}$ of its components begin symmetric in x and X_i and hence: (i) integrates to unity; and (ii) is understandable in terms of a direct single picture like Fig. 3. Both estimators perform well in our simulations. The third estimator, $\widehat{f}_{Xx}$, while proving particularly good when boundary effects do not come into play (!), is more generally not competitive with the other two because of its propensity towards $\widehat{f}_{Xx}(0) = \widehat{f}_{Xx}(1) = 0$. (However, it might therefore have a role in the rather unorthodox approach to kernel density estimation of Clements, Hurn & Lindsay, 2003.)

We have provided novel rule-of-thumb bandwidth selectors for our estimators. $\widehat{f}_{C2}$ and $\widehat{f}_{GC}$ continue to perform well using such practical bandwidth selectors and, indeed, the practical performance of $\widehat{f}_{GC}$ suffers a little less than that of $\widehat{f}_{C2}$ when using such bandwidths, making $\widehat{f}_{GC}$ that bit more attractive again. Of course, there remains scope for much further research on the bandwidth question. Sophisticated plug-in bandwidth selectors (e.g. Sheather & Jones, 1991) provide a natural avenue down which to travel. (Chen, 1999, briefly utilised cross-validation with $\widehat{f}_{C2}$, did not trust its results and opted for subjective adjustment in one case out of three.)

The fact that a beta density with parameters a and b scaled to the interval $[0, b]$ tends to a unit gamma density with parameter a as $b \to \infty$ yields a natural analogue of $\widehat{f}_{C1}$ and $\widehat{f}_{C2}$ for data $Y_1, ..., Y_n$ on $\mathcal{R}^+ = [0, \infty)$: see Chen (2000a). (Scaillet, 2004, proposes alternative inverse Gaussian and reciprocal inverse Gaussian kernels in the same framework.) Swopping the roles of point of estimation y and datapoint Y_i in Chen's (2000a) estimator $\widehat{f}_1$ immediately yields, as analogue of $\widehat{f}_{Xx}$, an alternative gamma kernel density estimator of the form

$$\widehat{f}_{Yy}(y) = \frac{1}{n} e^{-y/h^2} \sum_{i=1}^{n} \frac{y^{Y_i/h^2}}{h^{2\{(Y_i/h^2)+1\}} \Gamma((Y_i/h^2)+1)}.$$

This can be shown to have theoretical properties similar to those of Chen's estimator (in a way analogous to how Section 3.2 and Section 2 are related to one another in this paper). However, we do not pursue it here because we have $\widehat{f}_{Yy}(0) = 0$. We have been unable to come up with an analogous estimator to $\widehat{f}_{GC}$ on $\mathcal{R}^+$ based on limiting or similar arguments. An alternative — available to anyone with a good estimator on I — would be to transform using the odds ratio i.e. set $X_i = Y_i/(1+Y_i)$, $i = 1, ..., n$, estimate the density of the X's by $\widehat{f}_{GC}$ and transform back by using $(1+y)^{-2} \widehat{f}_{GC}(y/(1+y))$.

Finally, the potential impact of this work on kernel-based nonparametric regression with a predictor variable on finite support is considerable but would require a major new project to investigate fully. See Chen (2000b, 2002) for work in this area utilising his beta kernel approach.

Appendix

Derivation of rule-of-thumb bandwidths

Let a, b and μ, σ denote the estimated parameters of f_1 and f_2, respectively. Writing $I_1 = \int_0^1 z^5(1-z)^5(f_1'')^2(z)dz$ and $J_1 = \int_0^1 z^{5/2}(1-z)^{5/2} f_1(z)dz$,

from (6), $h_1 = \{J_1/(2n\sqrt{\pi}I_1)\}^{1/5}$. It is immediate that

$$J_1 = B(a+5/2, b+5/2)/B(a,b).$$

Second,

$$\begin{aligned} B(a,b)f_1''(z) &= z^{a-3}(1-z)^{b-3}\left\{(a-1)(a-2)(1-z)^2 - 2(a-1)(b-1)z(1-z)\right. \\ &\qquad \left. + (b-1)(b-2)z^2\right\}. \end{aligned}$$

It follows that

$$\begin{aligned} B^2(a,b)I_1 &= \int_0^1 z^{2a-1}(1-z)^{2b-1}\left\{(a-1)(a-2)(1-z)^2\right. \\ &\qquad \left. - 2(a-1)(b-1)z(1-z) + (b-1)(b-2)z^2\right\}^2 dz \\ &= (a-1)^2(a-2)^2 B(2a, 2b+4) \\ &\quad - 4(a-1)^2(a-2)(b-1)B(2a+1, 2b+3) \\ &\quad + 2(a-1)(b-1)(3ab-4a-4b+6)B(2a+2, 2b+2) \\ &\quad - 4(a-1)(b-1)^2(b-2)B(2a+3, 2b+1) \\ &\quad + (b-1)^2(b-2)^2 B(2a+4, 2b) \\ &= \frac{4\Gamma(2a)\Gamma(2b)}{\Gamma(2a+2b+4)}\left[(a-1)^2(a-2)^2(2b+3)(b+1)(2b+1)b\right. \\ &\qquad -8(a-1)^2(a-2)(b-1)a(b+1)(2b+1)b \\ &\qquad +2(a-1)(b-1)(3ab-4a-4b+6)(2a+1)a(2b+1)b \\ &\qquad -8(a-1)(b-1)^2(b-2)(a+1)(2a+1)ab \\ &\qquad \left.+(b-1)^2(b-2)^2(2a+3)(a+1)(2a+1)a\right]. \end{aligned}$$

For h_2,

$$pf_2'(z) = -\sigma^{-2} f_2(z)\phi(\Phi^{-1}(z))\left\{(1-\sigma^2)\Phi^{-1}(z) - \mu\right\}.$$

Then,

$$\{pf_2'\}'(z) = \sigma^{-4} f_2(z)\left[(1-\sigma^2)\{\Phi^{-1}(z)\}^2 + \mu(\sigma^2-2)\Phi^{-1}(z) + \mu^2 - \sigma^2(1-\sigma^2)\right]$$

and so, with $I_2 \equiv \int_0^1 p^{1/2}(z)\{(pf_2')'\}^2(z)dz$,

$$I_2 = \frac{1}{\sigma^8}\int_{-\infty}^{\infty}\left\{\frac{1}{\sigma}\phi\left(\frac{w-\mu}{\sigma}\right)\right\}^2\left\{(1-\sigma^2)w^2 + \mu(\sigma^2-2)w + \mu^2 - \sigma^2(1-\sigma^2)\right\}^2 dw$$

$$= \frac{1}{\sigma^9 2\sqrt{\pi}} \int_{-\infty}^{\infty} \phi_{\sigma/\sqrt{2}}(w-\mu) \left((1-\sigma^2)^2 w^4 + 2\mu(1-\sigma^2)(\sigma^2-2)w^3 \right.$$
$$+ \left[\mu^2(\sigma^2-2)^2 + 2(1-\sigma^2)\left\{ \mu^2 - \sigma^2(1-\sigma^2) \right\} \right] w^2$$
$$\left. + 2\mu(\sigma^2-2)\{\mu^2 - \sigma^2(1-\sigma^2)\}w + \{\mu^2 - \sigma^2(1-\sigma^2)\}^2 \right) dw.$$

Now, for the particular normal distribution involved,

$$E(W) = \mu, \qquad E(W^2) = \frac{1}{2}\sigma^2 + \mu^2,$$

$$E(W^3) = \frac{3}{2}\sigma^2\mu + \mu^3, \qquad E(W^4) = \frac{3}{4}\sigma^4 + 3\sigma^2\mu^2 + \mu^4.$$

So,

$$I_2 = \frac{1}{\sigma^9 2\sqrt{\pi}} \left((1-\sigma^2)^2 \left(\frac{3}{4}\sigma^4 + 3\sigma^2\mu^2 + \mu^4 \right) + 2\mu^2(1-\sigma^2)(\sigma^2-2) \left(\frac{3}{2}\sigma^2 + \mu^2 \right) \right.$$
$$+ \left[\mu^2(\sigma^2-2)^2 + 2(1-\sigma^2)\left\{ \mu^2 - \sigma^2(1-\sigma^2) \right\} \right] \left(\frac{1}{2}\sigma^2 + \mu^2 \right)$$
$$\left. + 2\mu^2(\sigma^2-2)\{\mu^2 - \sigma^2(1-\sigma^2)\} + \{\mu^2 - \sigma^2(1-\sigma^2)\}^2 \right)$$
$$= \frac{1}{\sigma^5 8\sqrt{\pi}} \left\{ 2\mu^2\sigma^2 + 3(1-\sigma^2)^2 \right\}.$$

Formula (8) follows since $h_2 = (8n\sqrt{\pi}I_2)^{-1/5}$.

References

Bouezmarni, T. & Rolin, J.T. (2003). Consistency of the beta kernel density function estimator. *Canad. J. Statist.* **31**, 89–98.

Chen, S.X. (1999). Beta kernel estimators for density functions. *Comput. Statist. Data Anal.* **31**, 131–45.

Chen, S.X. (2000a). Probability density function estimation using gamma kernels. *Ann. Inst. Statist. Math.* **52**, 471–80.

Chen, S.X. (2000b). Beta kernel estimators for regression curves. *Statistica Sinica* **10**, 73–91.

Chen, S.X. (2002). Local linear smoothers using asymmetric kernels. *Ann. Inst. Statist. Math.* **54**, 312–23.

Cheng, M.Y., Fan, J. & Marron, J.S. (1997). On automatic boundary corrections. *Ann. Statist.* **25**, 1691–708.

Clements, A., Hurn, S. & Lindsay, K. (2003). Mobius-like mappings and their use in kernel density estimation. *J. Amer. Statist. Assoc.* **98**, 993–1000.

Demarta, S. & McNeil, A.J. (2003). The t copula and related copulas. *Int. Statist. Rev.* **73**, 111–29.

Fang, H.B., Fang, K.T. & Kotz, S. (2002). The meta-elliptical distributions with given marginals. *J. Multivariate Anal.* **82**, 1–16.

Frahm, G., Junker, M. & Szimayer, A. (2003). Elliptical copulas: applicability and limitations. *Statist. Probab. Lett.* **63**, 275–86.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman and Hall.

Johnson, N.L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika* **36**, 149–76.

Jones, M.C. (1993). Simple boundary correction for kernel density estimation. *Statist. Comput.* **3**, 135–46.

Karunamuni, R.J. & Alberts, T. (2005). A generalized reflection method of boundary correction in kernel density estimation. *Canad. J. Statist.* **33**, 497–509.

Klaassen, C.A.J. & Wellner, J.A. (1997). Efficient estimation in the bivariate normal copula model: normal margins are least favourable. *Bernoulli* **3**, 55–77.

Müller, H.G. (1991). Smooth optimum kernel estimators near endpoints. *Biometrika* **78**, 521–30.

Nelsen, R.B. (2006). *An Introduction to Copulas*, Second Edition. New York: Springer.

Plackett, R.L. (1965). A class of bivariate distributions. *J. Amer. Statist. Assoc.* **60**, 516–22.

Scaillet, O. (2004). Density estimation using inverse and reciprocal inverse Gaussian kernels. *J. Nonparametric Statist.* **16**, 217–26.

Sheather, S.J. & Jones, M.C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *J. Roy. Statist. Soc. Ser. B* **53**, 683–90.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Simonoff, J.S. (1996). *Smoothing Methods in Statistics*. New York: Springer.

Song, P.X.K. (2000). Multivariate dispersion models generated from Gaussian copula. *Scand. J. Statist.* **27**, 305–20.

Sungur, E.A. (1990). Dependence information in parametrized copulas. *Commun. Statist.— Simul. Comput.* **19**, 1339–60.

Wand, M.P. & Jones, M.C. (1995). *Kernel Smoothing*. London: Chapman and Hall.

Zhang, S., Karunamuni, R.J. & Jones, M.C. (1999). An improved estimator of the density function at the boundary. *J. Amer. Statist. Assoc.* **94**, 1231–41.

Table 1. *The densities in the simulation testbed*

	$f(x), x \in [0,1]$	Description
1	$140x^3(1-x)^3$	$\text{Beta}(4,4)$
2	$1120\left[x^3(1-2x)^3 I_{\{x\le 1/2\}} + 8\left(x-\frac{1}{2}\right)^3(1-x)^3 I_{\{x\ge 1/2\}}\right]$	$\frac{1}{2}\text{Beta}_{[0,1/2]}(4,4) + \frac{1}{2}\text{Beta}_{[1/2,1]}(4,4)$
3	$3x^2$	$\text{Beta}(3,1)$
4	$\frac{3}{2}\{x^2 + (1-x)^2\}$	$\frac{1}{2}\text{Beta}(3,1) + \frac{1}{2}\text{Beta}(1,3)$
5	$\{\pi\sqrt{x(1-x)}\}^{-1}$	$\text{Beta}(\frac{1}{2},\frac{1}{2})$
6	$\frac{231}{463}(1+3x)^5(1-x)^5$	Truncated $\text{Beta}_{[-1/3,1]}(6,6)$
7	$2e^{-2x}(1-e^{-2})^{-1}$	Truncated Exponential(2)
8	$\frac{2240}{1759}\left\{1-\left(x-\frac{1}{2}\right)^2\right\}^3$	Truncated $\text{Beta}_{[-1/2,3/2]}(4,4)$
9	$\frac{35}{16}(1-x^2)^3$	Truncated $\text{Beta}_{[-1,1]}(4,4)$
10	$2\{\pi\sqrt{x(2-x)}\}^{-1}$	Truncated $\text{Beta}_{[0,2]}(1/2,1/2)$
11	$2e^{-2x^2}\left[\sqrt{2\pi}\left\{\Phi(2)-\frac{1}{2}\right\}\right]^{-1}$	Truncated $2\phi(2x)$
12	$\frac{1}{2} + 280\left(2x-\frac{1}{2}\right)^3\left(\frac{3}{2}-2x\right)^3 I_{\{1/4\le x\le 3/4\}}$	$\frac{1}{2}\text{Beta}(1,1) + \frac{1}{2}\text{Beta}_{[1/4,3/4]}(4,4)$
13	$294x(1-x)^{19} + 33x^9(1-x)$	$\frac{7}{10}\text{Beta}(2,20) + \frac{3}{10}\text{Beta}(10,2)$
14	$102060\left[\sum_{i=1}^3\left\{x-\frac{(i-1)}{3}\right\}^3 \times\left(\frac{i}{3}-x\right)^3 I_{\{(i-1)/3\le x\le i/3\}}\right]$	$\frac{1}{3}\sum_{i=1}^3 \text{Beta}_{[(i-1)/3,i/3]}(4,4)$
15	$c(x, 0.7; 0.7)$	Gaussian copula
16	$5e^{-\lvert x-\frac{1}{2}\rvert}(1-e^{-5})^{-1}$	Truncated Laplace$(1/2, 1/10)$

In the table, $I_{\{A\}}$ is the indicator function which equals 1 if A is true and 0 otherwise, $c(x,y;\rho)$ is specifically the *Gaussian* copula given by (4), $\text{Beta}(a,b)$ denotes the beta density with parameters a and b, $\text{Beta}_{[c,d]}(a,b)$ denotes the beta density rescaled to the interval $[c,d]$ and "truncated" means truncated to $[0,1]$.

Table 2. *Simulation results for density estimation over I when $n = 50$. The first line for each density gives the simulated $1000\times$MISE (with standard error in parentheses) where, for each simulated dataset, h is chosen to minimise its ISE. The second line gives the median percentage reduction in ISE compared with that of $\widehat{f}_{C2}$ (negative values correspond, of course, to worse performance). Figures in bold denote values of MISE that are within two standard errors of the minimum MISE for that density*

OVERLEAF

Density	Estimator			
	$\widehat{f}_{C1}$	$\widehat{f}_{C2}$	$\widehat{f}_{Xx}$	$\widehat{f}_{GC}$
1	52.13 (1.28)	40.53 (0.97)	**35.49** (0.91)	46.00 (1.14)
	−26.80	0.00	13.13	−13.58
2	111.65 (2.03)	**102.42** (1.82)	**102.10** (1.85)	112.90 (1.96)
	−7.99	0.00	0.83	−10.35
3	44.02 (1.26)	**33.49** (1.14)	94.94 (1.95)	58.11 (1.38)
	−30.49	0.00	−278.66	−77.41
4	30.84 (0.51)	**29.44** (0.53)	82.53 (1.32)	31.45 (0.46)
	−4.12	0.00	−190.39	−3.99
5	241.09 (2.69)	263.03 (2.49)	333.01 (3.20)	**133.66** (2.80)
	8.08	0.00	−27.78	53.25
6	53.43 (1.37)	44.99 (1.18)	**33.95** (0.78)	60.16 (1.41)
	−20.83	0.00	18.72	−33.13
7	36.57 (1.11)	**28.58** (0.95)	77.84 (1.50)	40.29 (1.10)
	−24.91	0.00	−255.37	−43.26
8	28.44 (0.53)	26.35 (0.54)	**5.40** (0.10)	27.32 (0.49)
	−7.44	0.00	77.47	−1.62
9	33.85 (0.93)	**21.35** (0.78)	64.98 (1.49)	51.52 (1.11)
	−91.39	0.00	−254.76	−220.27
10	226.69 (3.29)	237.46 (2.77)	299.72 (3.58)	**103.91** (3.26)
	2.90	0.00	−27.14	66.70
11	26.90 (0.87)	**18.77** (0.72)	54.98 (1.18)	33.98 (0.84)
	−54.65	0.00	−265.14	−124.41
12	99.81 (1.81)	99.36 (1.83)	**95.42** (1.71)	135.60 (2.20)
	−1.50	0.00	1.65	−29.73
13	189.14 (3.35)	186.66 (3.11)	**152.21** (3.34)	170.40 (3.37)
	−2.93	0.00	13.11	10.25
14	167.15 (2.48)	**159.08** (2.33)	**159.48** (2.35)	179.16 (2.49)
	−4.40	0.00	−0.01	−12.67
15	44.58 (0.92)	39.10 (0.77)	**25.30** (0.86)	34.22 (0.87)
	−12.61	0.00	35.37	14.34
16	**128.64** (2.43)	**125.08** (2.37)	**124.67** (2.34)	134.62 (2.50)
	−2.67	0.00	0.40	−6.48
All	94.68 (0.74)	90.98 (0.76)	108.88 (0.87)	**84.58** (0.63)
	−7.82	0.00	−2.78	−12.09

Table 3. *Simulation results for estimation of density at points near boundary for* $n = 500$. *The table gives estimates of* $1000 \times$ *MSE (with standard errors in parentheses) for* $x = 0.01$ *and* $x = 0.99$. h *is still chosen, for each simulated dataset, to minimise overall ISE. Figures in bold denote values of MSE that are within two standard errors of the minimum MSE for that density.* †*The values for Density 6,* $x = 0.99$*, are* $\mathbf{1.487 \times 10^{-3}}$ (8.388×10^{-4}), 1.003×10^{-3} (4.344×10^{-4}), $\mathbf{1.984 \times 10^{-4}}$ (1.357×10^{-4}) *and* $\mathbf{2.107 \times 10^{-6}}$ (2.014×10^{-6}) *for* $\widehat{f}_{C1}$, $\widehat{f}_{C2}$, $\widehat{f}_{Xx}$ *and* $\widehat{f}_{GC}$, *respectively*

OVERLEAF

	x	Estimator			
		$\widehat{f}_{C1}$	$\widehat{f}_{C2}$	$\widehat{f}_{Xx}$	$\widehat{f}_{GC}$
1	0.01	0.084 (0.011)	0.064 (0.007)	**0.018** (0.004)	**0.023** (0.014)
	0.99	0.085 (0.012)	0.067 (0.009)	**0.019** (0.005)	**0.055** (0.031)
2	0.01	0.322 (0.044)	**0.125** (0.025)	**0.128** (0.026)	**0.312** (0.133)
	0.99	0.313 (0.052)	**0.138** (0.033)	**0.134** (0.031)	**0.327** (0.149)
3	0.01	0.101 (0.008)	0.274 (0.016)	**0.006** (0.003)	**0.021** (0.006)
	0.99	**68.997** (3.022)	**72.128** (3.041)	251.339 (10.123)	139.059 (5.574)
4	0.01	**30.549** (1.285)	**31.462** (1.285)	150.171 (4.752)	47.047 (1.948)
	0.99	**32.175** (1.411)	**32.159** (1.402)	156.305 (5.317)	51.618 (2.163)
5	0.01	264.498 (11.557)	835.462 (32.774)	667.592 (26.936)	**165.822** (7.280)
	0.99	299.808 (12.773)	822.084 (34.023)	674.043 (29.993)	**176.373** (7.555)
6	0.01	**30.507** (1.565)	**28.578** (1.573)	36.308 (1.125)	64.576 (3.524)
	0.99†	**0.001** (0.001)	0.001 (0.000)	**0.000** (0.000)	**0.000** (0.000)
7	0.01	**55.273** (2.283)	**60.657** (2.155)	255.215 (7.994)	84.127 (3.398)
	0.99	**6.678** (0.306)	**6.052** (0.277)	19.702 (0.700)	14.752 (0.662)
8	0.01	**15.807** (0.711)	**16.083** (0.707)	31.108 (0.242)	25.929 (1.301)
	0.99	**18.610** (0.837)	**18.974** (0.829)	30.564 (0.238)	29.997 (1.497)
9	0.01	35.187 (1.354)	**19.159** (1.002)	231.672 (7.731)	120.006 (5.206)
	0.99	0.056 (0.005)	0.457 (0.021)	**0.003** (0.002)	**0.011** (0.008)
10	0.01	411.634 (18.819)	1296.837 (47.484)	1073.052 (45.339)	**252.764** (12.230)
	0.99	69.380 (3.288)	73.218 (3.583)	81.589 (4.074)	**36.657** (1.697)
11	0.01	20.572 (0.842)	**12.014** (0.610)	220.065 (5.750)	66.094 (2.652)
	0.99	**6.762** (0.330)	9.745 (0.337)	11.588 (0.416)	11.687 (0.613)
12	0.01	**32.319** (1.539)	43.412 (2.257)	**35.757** (1.553)	100.275 (5.282)
	0.99	**32.029** (1.442)	41.002 (1.849)	**33.590** (1.329)	95.345 (4.856)
13	0.01	453.246 (19.042)	**206.258** (8.579)	**175.697** (7.466)	369.517 (19.012)
	0.99	36.247 (1.830)	**23.364** (1.134)	**20.008** (0.944)	41.570 (2.206)
14	0.01	0.629 (0.082)	**0.240** (0.042)	**0.297** (0.050)	**0.469** (0.195)
	0.99	0.935 (0.113)	**0.510** (0.082)	**0.544** (0.080)	1.524 (0.393)
15	0.01	3.455 (0.151)	3.066 (0.144)	**0.506** (0.038)	1.745 (0.175)
	0.99	99.816 (2.823)	74.783 (2.280)	**16.727** (0.601)	41.799 (2.392)
16	0.01	**4.581** (0.481)	**4.863** (0.491)	**4.503** (0.433)	10.227 (1.345)
	0.99	**5.063** (0.385)	**5.733** (0.417)	**5.256** (0.381)	16.611 (2.162)
All	0.01	**84.923** (2.167)	159.910 (4.608)	180.131 (4.108)	**81.810** (1.791)
	0.99	**42.310** (1.053)	73.776 (2.652)	81.338 (2.417)	**41.087** (0.837)

Table 4. *Simulation results for density estimation over I when $n = 50$ and rule-of-thumb bandwidth selectors are utilised in the estimators. The first line for each density gives the simulated* $1000\times$*MISE (with standard error in parentheses). The second line gives the median percentage reduction in ISE compared with that of $\widehat{f}_{C2,h_1}$ (negative values correspond, of course, to worse performance). Figures in bold denote values of MISE that are within two standard errors of the minimum MISE for that density*

OVERLEAF

Density	Estimator		
	$\widehat{f}_{C2,h_1}$	$\widehat{f}_{Xx,h_1}$	$\widehat{f}_{GC,h_2}$
1	47.94 (1.11)	**45.24** (1.12)	53.95 (1.24)
	0.00	7.35	−16.89
2	595.93 (1.20)	513.78 (1.28)	**460.64** (1.46)
	0.00	14.27	23.08
3	**64.05** (2.33)	166.94 (3.05)	73.05 (1.59)
	0.00	−243.87	−25.94
4	**40.09** (0.83)	122.19 (1.69)	42.49 (0.68)
	0.00	−280.24	−9.48
5	362.91 (2.16)	513.47 (3.01)	**169.00** (3.21)
	0.00	−42.77	57.83
6	55.11 (1.31)	**43.19** (0.96)	68.87 (1.50)
	0.00	14.39	−25.24
7	67.21 (1.96)	139.71 (2.27)	**53.21** (1.31)
	0.00	−176.18	1.87
8	34.96 (0.67)	**23.53** (0.65)	35.08 (0.67)
	0.00	37.19	1.68
9	**45.46** (1.68)	101.05 (1.88)	65.30 (1.51)
	0.00	−213.47	−77.64
10	373.94 (2.66)	526.96 (3.57)	**151.56** (3.86)
	0.00	−43.29	66.30
11	63.29 (1.52)	91.89 (1.36)	**44.98** (1.05)
	0.00	−61.36	8.86
12	181.90 (2.87)	**167.05** (2.17)	305.08 (3.54)
	0.00	4.54	−67.99
13	**342.47** (3.09)	382.82 (4.96)	737.03 (5.09)
	0.00	−15.24	−124.39
14	636.44 (0.48)	625.93 (0.69)	**591.16** (0.66)
	0.00	0.89	7.40
15	60.71 (1.29)	**36.53** (0.97)	42.64 (1.02)
	0.00	38.64	31.14
16	**183.64** (3.00)	**178.11** (2.86)	234.59 (3.82)
	0.00	2.87	−23.51
All	**197.25** (1.63)	229.90 (1.69)	**195.54** (1.78)
	0.00	−11.57	−2.09

Fig. 1. The basic beta kernel functions in (3) using $h = \sqrt{0.1}$ (dashed lines) for the eleven pseudo-datapoints listed in the text, and the estimator $\widehat{f}_{Xx}$ of their density based on the pseudo-datapoints (solid curve) which arises from averaging these kernel functions.

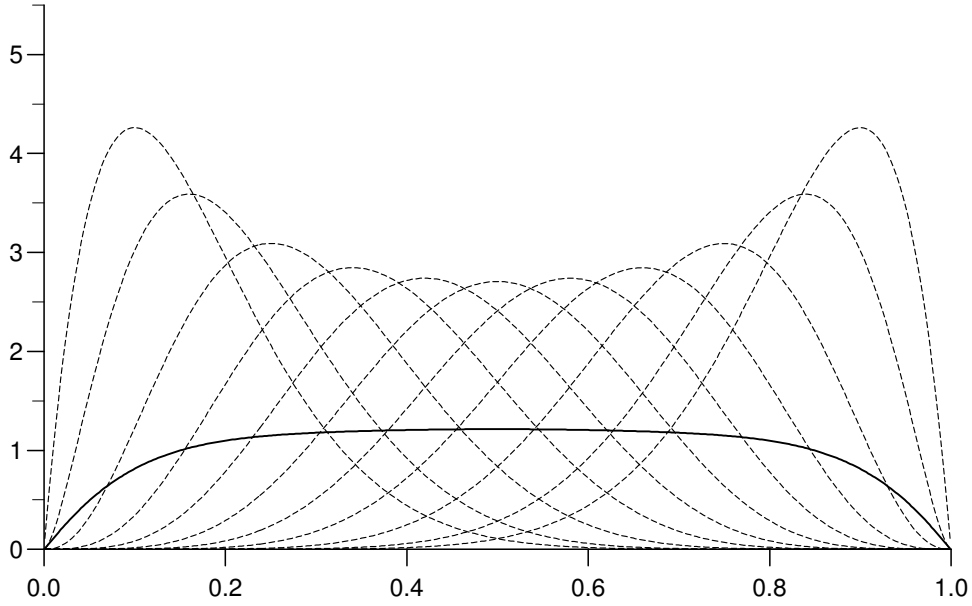

Fig. 2. The Gaussian-based copula for $\rho = 0.9$.

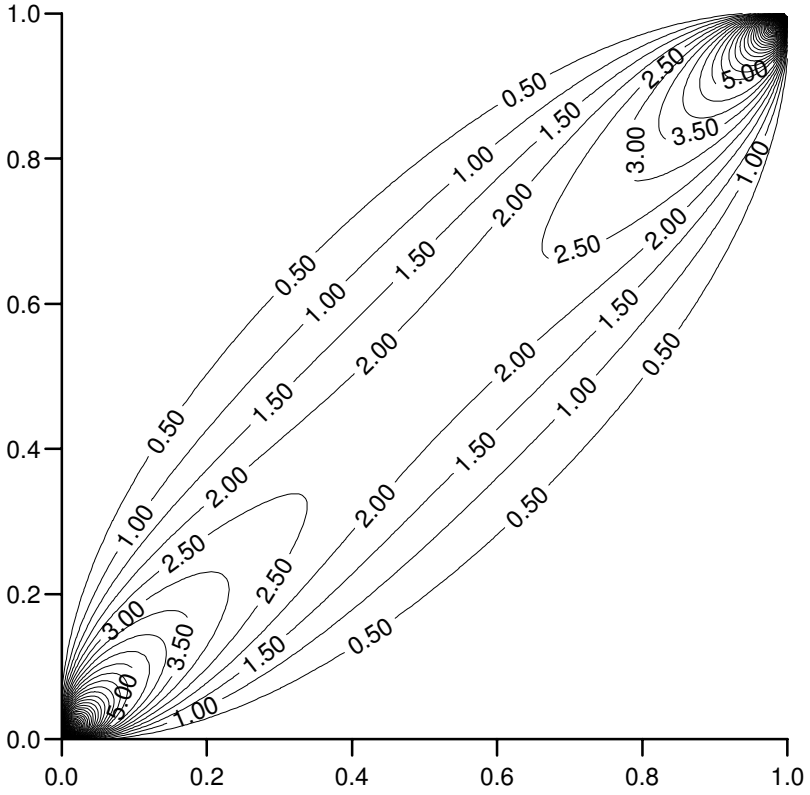

Fig. 3. The kernel functions $c(x, X_i; 0.9)$ (dashed lines) associated with the Gaussian-based copula for the eleven pseudo-datapoints listed in the text, and the estimator of the density based on the pseudo-datapoints (solid curve) which arises from averaging the kernel functions.

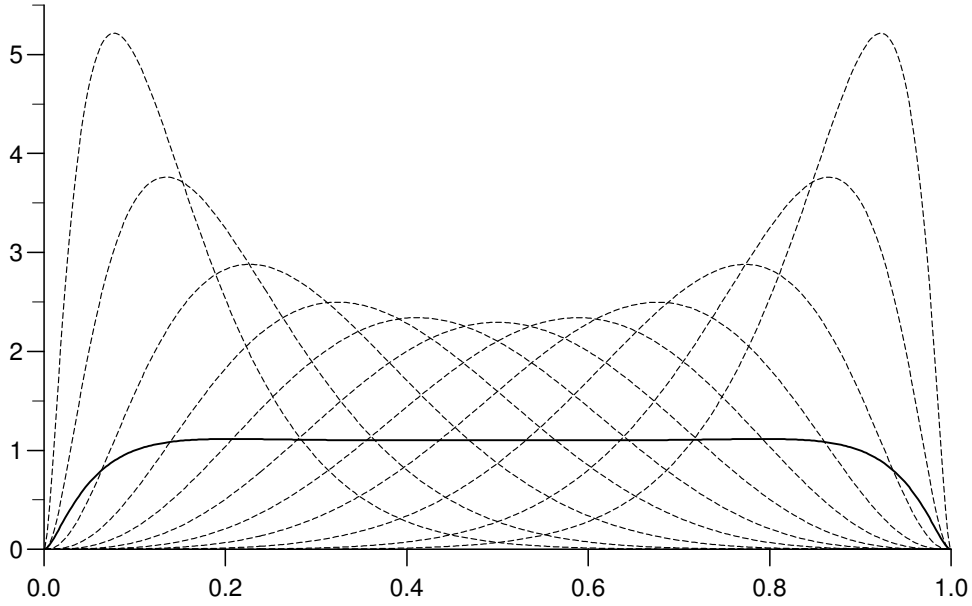

Fig. 4. Densities used in the simulation study. Their formulae are given in Table 1.

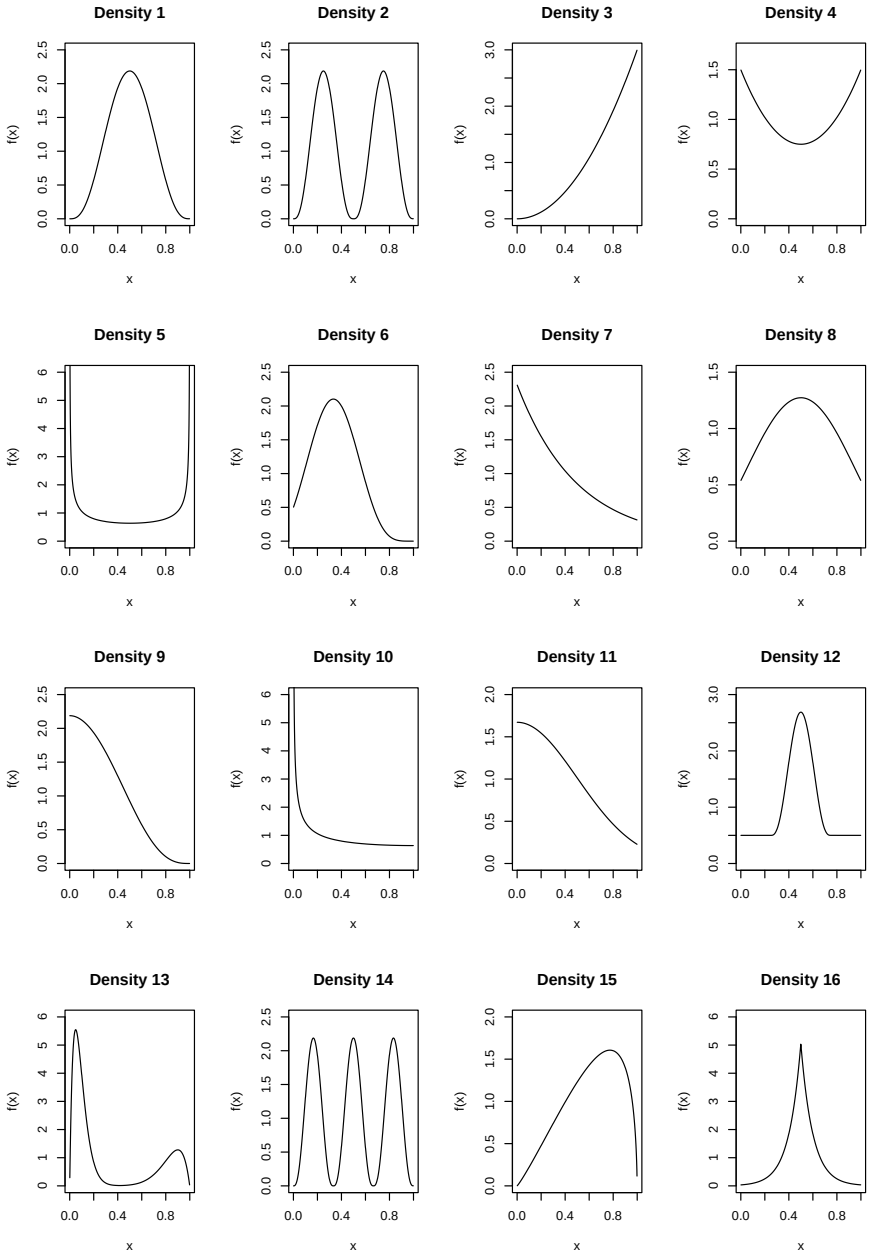

Fig. 5. Boxplots of ISEs corresponding to the simulation results with ISE-optimal bandwidths summarised in Table 2 ($n = 50$).

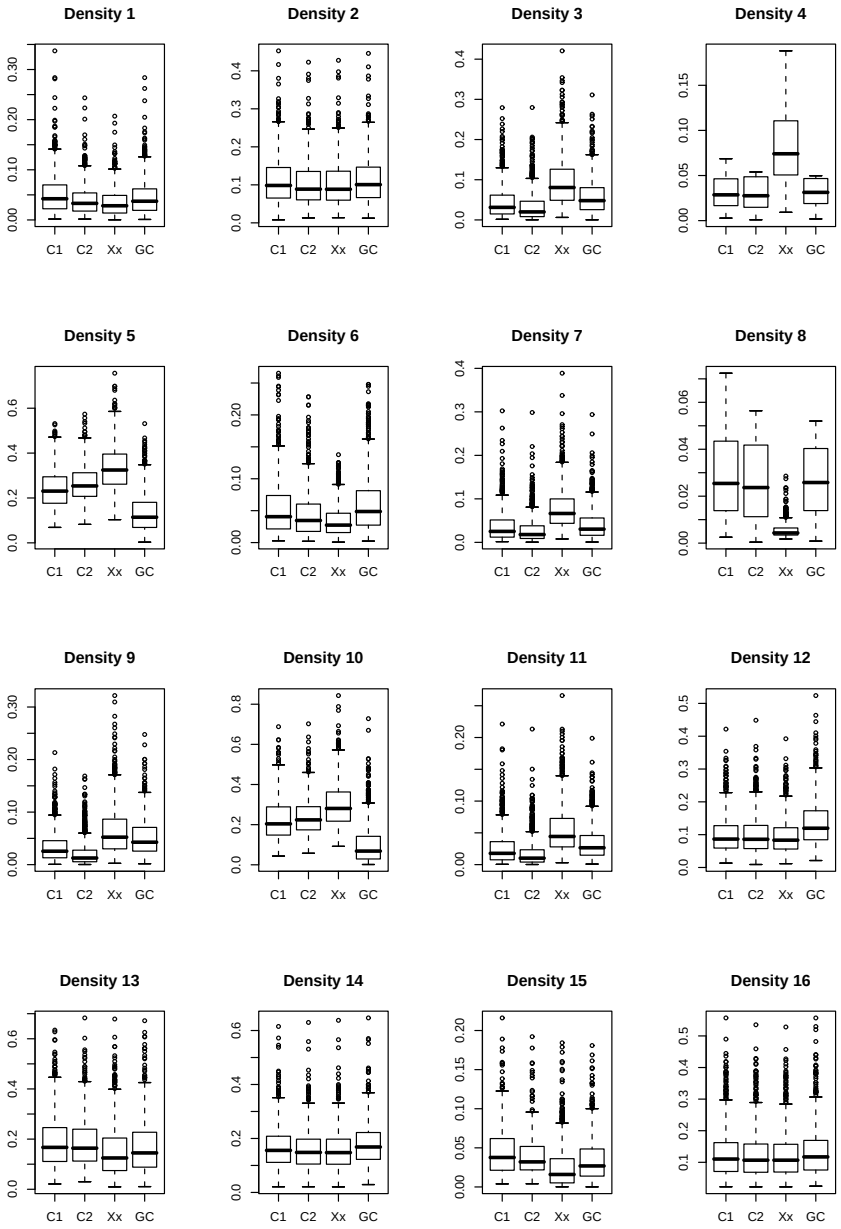

Fig. 6. Boxplots of ISEs corresponding to the simulation results with rule-of-thumb bandwidths summarised in Table 4 ($n = 50$).