isid/ms/2015/1
January 08, 2015
http://www.isid.ac.in/~statmath/index.php?module=Preprint

Smoothing Parameter Selection for Nonparametric Density Estimation for Length-biased Data: A Bayesian Perspective

Yogendra P. Chaubey, Jun Li and Isha Dewan

Indian Statistical Institute, Delhi Centre
7, SJSS Marg, New Delhi–110 016, India

Smoothing Parameter Selection for Nonparametric Density Estimation for Length-biased Data: A Bayesian Perspective

Yogendra P. Chaubey
Department of Mathematics and Statistics
Concordia University
1455 de Maisonneuve Boulevard West
Montreal, Quebec H3G 1M8, Canada
E-mail: yogen.chaubey@concordia.ca

Jun Li
School of Mathematics and Statistics
Nanjing Audit University
86 West Yushan Road
Pukou District, Nanjing, China
E-mail: lijun@nau.edu.cn

Isha Dewan
Stat Math Unit
Indian Statistical Institute
7, S. J. S. Sansanwal Marg
New Delhi - 110 016, India
E-mail: isha@isid.ac.in

Abstract

Nonparametric estimation of densities defined over non-negative observations using asymmetric kernels is of special interest as it has potential to remove the spill-over effect at the boundary. One important problem in this context is the selection of the smoothing parameter. The purpose of the this note is to review some recent work on the application of Bayes criterion for this purpose and investigate its application in the context of length biased data.

1 Introduction

In many applications the recorded observation may be assumed to have the probability density function $g(x)$, that is of the form

$$g(x) = \mu_w^{-1} w(x) f(x), \quad x \in \mathbb{R}^+, \tag{1.1}$$

where $f(x)$ is the original density, $w(x)$ is a non-negative known function called the weighting function,

$$\mu_w = 1/E_g(1/w(X)) \tag{1.2}$$

with $X \sim g(.)$. Patil and Rao [17] cite several examples including those generated by PPS (probability proportional to size) sampling scheme (that is common in sample surveys), damage models and sub-sampling [see also Rao [18], Patil and Rao [16] and Rao [19]]. Here, we concentrate on the case when $w(x) = x$, a situation known as giving length-biased data and where typically the observations are non-negative. Thus we consider a random sample $\{X_1, \ldots, X_n\}$ be n nonnegative independent and identically distributed (i.i.d.) random variables (r.v.) having a continuous probability density function (pdf) $g(x), x \in \mathbb{R}^+ = [0, \infty)$ given by

$$g(x) = \mu^{-1} x f(x), \quad x \in \mathbb{R}^+ \tag{1.3}$$

where $\mu = (E_g(X^{-1}))^{-1}$ is the harmonic mean of X with the pdf $g(.)$, and estimation of $f(x), x \in \mathbb{R}^+$ itself is of central importance. Here it is tacitly assumed that $(0 <) < \mu < \infty$. Note that $\mu \leq E_g(X) = \mu^{-1} \int_0^\infty x^2 f(x) dx$, or equivalently, $\mu^2 \leq E_f(Y^2)$ where Y has the pdf f. Thus, assuming that Y has a finite 2^{nd} moment insures that $\mu < \infty$ (even finite first moment of Y does so).

Bhattacharyya *et al.* [4] studied the kernel density estimator for $f(x)$ obtained by using the corresponding estimator for $g(x)$ and the relation (1.3), replacing the unknown value μ by its harmonic mean estimator as proposed by Cox [11]. Cox [11] also gave a direct estimator of $F(x)$ that has been used in proposing an alternative density estimator by Jones [13]. In some aspects, Jones estimator performs much better than Bhattacharyya *et al.* estimator [see also Wu and Mao [22]]. However, in general the kernels used are symmetric around zero the resulting density estimators may put a positive mass out side the support of $f(.)$ *i.e.* $\mathbb{R}^+ = [0, \infty)$. This may also produce a large bias in the estimators for x near zero. This problem has long been recognized in density estimation in the context of *i.i.d.* data [see Silverman [21]], however, it becomes more pertinent for the length-biased data, as the observations are necessarily non-negative.

In order to overcome these problems while dealing with non-negative data, several methods estimating underlying density for non-negative random variables, specially using asymmetric kernels have been proposed in literature. Bagai and Prakasa Rao [3] proposed replacing the symmetric kernel k by a pdf k^* with non-negative support. This certainly avoids the problem of positive mass in the negative region; however, only the first r order statistics are used for estimating $f(x)$, where $X_{(r)} < x \leq X_{(r+1)}$, $X_{(i)}$ denoting the i^{th} order statistic that is an undesirable feature. Chaubey and Sen [7] proposed a density estimator as the derivative of a smooth version of the *edf* by adapting the so called Hille's smoothing lemma, which, in contrast to the proposal of Bagai and Prakasa Rao [3], uses the whole data. This has been adapted to the length-biased set-up by Chaubey, Sen and Li [9]. While trying to develop smooth density estimators based on *asymmetric kernels*, Chen [10] proposed the use of gamma density, whereas Scaillet [20] proposed using inverse Gaussian and reciprocal inverse Gaussian densities. Another class of asymmetric kernel estimators in the *i.i.d.* context has been proposed recently by Chaubey, *et al.* [6] that is motivated by generalization of the estimator in Chaubey and Sen [7]. In a more recent paper Chaubey and Li [5] adapted the above smoothing estimator to the length-biased density estimation and proposed two new asymmetric kernel estimators, one based on the usual empirical distribution function and the other one based on the Cox's estimator.

Kulasekera and Padgett [14] motivate these estimators, although in the context of censored data, by the general form of the nonparametric density estimator given by

$$\hat{f}(x) = \int k(x; u, h) d\hat{F}_n(u) du \tag{1.4}$$

and find the following inverse Gaussian (IG) choice of the kernel (with mean u and dispersion parameter $1/h$) useful:

$$k_{KP}(x; u, h) = \frac{1}{\sqrt{2\pi h x^3}} \mathrm{e}^{-(\frac{1}{2h})\frac{(x-u)^2}{u^2 x}}. \tag{1.5}$$

Kulasekera and Padgett [14] note that the estimator $\hat{f}$ based in the above kernel is a true density while that using the reciprocal IG is not.

In these methods, selection of smoothing parameter may be numerically challenging. In this paper we explore the idea of using Bayesian method of smoothing parameter selection, inspired by the paper of Kulasekera and Padgett [14]. In this approach, we contrast some alternative choices of asymmetric kernels.

Along with some preliminary notions, the proposed smooth estimators of $f(.)$ are given in Section 2. Section 3 is devoted to determination of the smoothing parameter using a Bayesian perspective along with various alternative choices of asymmetric kernels and Section 4 addresses data based determination of the hyper parameters α and β. Section 5 contains a numerical study comparing various choices of the smoothing parameters where as the final section presents conclusions and discussion of the new methodology while illustrating it on a real data set.

2 Preliminary Notions and Smooth Density Estimators

There are basically two strategies to be used to obtain the smooth estimator of $f(x)$. One is based on a smooth version Cox's estimator $F_n(x)$ that is given below:

$$F_n(x) = \frac{\sum_{i=1}^{n} \frac{1}{X_i} I\{X_i \leq x\}}{\sum_{i=1}^{n} \frac{1}{X_i}}. \tag{2.1}$$

The other one is to first estimate $g(x)$, that is obtained as the derivative of the smooth version of the empirical distribution function

$$G_n(x) = \frac{1}{n} \sum_{i=1}^{n} I\{X_i \leq x\}, \tag{2.2}$$

and then making some adjustments to obtain the estimator of underlying density $f(x)$. Chaubey and Li [5] and Jones [13] found that the density estimators based on smoothing of $F_n(x)$ have better performance. Thus the density estimators we consider in the sequel are based on F_n. These are of the form

$$\hat{f}_h(x) = \sum_{i=1}^{n} s_i k(x; X_i, h) \tag{2.3}$$

where

$$s_i = \frac{1/X_i}{\sum_i (1/X_i)}, \tag{2.4}$$

k is a kernel function and h is a smoothing parameter. This expression is motivated by the general form of the nonparametric density estimators given by (1.4).

In the sequel, we will explore an alternative form of $k(x; u, h)$ following Chaubey and Li [5], that is required to be a density with respect to the argument u. This is based on the following smooth estimator of the distribution function $F(x)$:

$$\begin{aligned} \widetilde{F}_n(x) &= \int_0^\infty F_n(t) dQ_{h_n}(t/x) && (2.5) \\ &= 1 - \frac{\sum_{i=1}^n \frac{1}{X_i} Q_{h_n}(\frac{X_i}{x})}{\sum_{i=1}^n \frac{1}{X_i}}, && (2.6) \end{aligned}$$

where $Q_h(.)$ is a distribution function with mean 1 and variance h^2. Thus a smooth estimator of f obtained by differentiating $\tilde{F}_n(x)$ is given by

$$\tilde{f}_n(x) = \frac{\frac{1}{x^2} \sum_{i=1}^n q_{h_n}(\frac{X_i}{x})}{\sum_{i=1}^n \frac{1}{X_i}} \tag{2.7}$$

where $q_{h_n}(t) = \frac{d}{dt} Q_{h_n}(t)$. This estimator is of the same form as in (2.3) with $k(x; u, h)$ given by

$$k(x; u, h) = \frac{u}{x^2} q_h(u/x). \tag{2.8}$$

As noted in [6], (2.7) may not be defined at $x = 0$, except in cases where $\lim_{x \to 0} \tilde{f}_n(x)$ exists. Modifications have been discussed in Chaubey and Li [5] [see also Chaubey, Sen and Sen [8]]. However, this will not be pursued further in this paper and we may think of the methods discussed here for estimating $f(x)$ for $x > 0$ only. The following two alternative choices of $k(x; u, h)$ will be investigated here. The first one is obtained by choosing $q_h(.)$ to be the inverse Gaussian distribution with parameters $\mu = 1$ and $\lambda = 1/h$, *i,e.*

$$q_h(u) = \sqrt{\frac{1}{2\pi h u^3}} \exp\left\{ -\frac{1}{2hu} (u-1)^2 \right\}$$

This provides the following choice of $k(x; u, h)$:

$$k_{CL}(x; u, h) = \frac{u}{x^2} q_h\left(\frac{u}{x}\right) = \frac{1}{\sqrt{2\pi h u x}} \exp\{-\frac{1}{2hux}(x-u)^2\} \tag{2.9}$$

which represents a density of IG with respect to argument x as well as with respect to argument u; for fixed x $k_{CL}(x; u, h)$ is the density of $IG(\mu, \lambda)$ with $\mu = x, \lambda = x/h$ and it represents the density of reciprocal IG with $\mu = 1/u$ and $\lambda = 1/hu$ for fixed u. Thus u is the mode of k_{CL} for fixed x, as opposed to the mean in case of k_{KP}.

The second choice is obtained from the density of the so-called Co-Gaussian (CoG) distribution that is related to the IG family of distributions (see Awadalla [2], §2.2). The probability density function of CoG-distribution with mode μ and dispersion h is given by

$$g(x; \mu, h) = \sqrt{\frac{2}{\pi h}} \exp\left\{ -\frac{1}{2h} \left(x - \frac{\mu^2}{x} \right)^2 \right\}.$$

Thus we consider $q_h(u)$ generated from the CoG density with mode 1, *i.e.* we take

$$q_h(u) = \sqrt{\frac{2}{\pi h}} \exp\left\{-\frac{1}{2h}\left(u - \frac{1}{u}\right)^2\right\}. \tag{2.10}$$

This gives, in the spirit of (2.8) another kernel given by

$$k_{CG}(x; u, h) = \frac{u}{x^2} q_h(u/x) = \sqrt{\frac{2}{\pi h}} \frac{u}{x^2} \exp\left\{-\frac{1}{2hx^2}\left(u - \frac{x^2}{u}\right)^2\right\}. \tag{2.11}$$

Remark 2.1. The mean of CoG distribution with mode μ and dispersion h is given by

$$\mu_1' = \sqrt{\frac{2}{\pi h}} \mu^2 \mathrm{e}^{(\mu^2/h)} K_1(\mu^2/h),$$

where $K_\nu(z)$ represents the modified Bessel function of the second kind of order ν (see Awadalla [2], Eq. (2.2.4)). Also, for large values of z, we have approximately (see Abramowitz and Stegun [1])

$$K_1(z) \approx (\frac{\pi}{2z})^{1/2} \mathrm{e}^{-z}.$$

Thus, for small values of h, the mean and mode of CoG coincide, and hence k_{CG} follows the property required in (2.8).

The following theorem is about strong consistency of $\hat{f}_h$ using the CL kernel, though we believe that it will hold under other kernels considered here.

Theorem 2.1 *If $f(x)$ is Lipschitz continuous on $[0, \infty)$ and $\int_0^\infty \frac{1}{x} f(x)dx < \infty$, then as $n \to \infty$, $h \to 0$ and $hn(\log n)^{-(1+\theta)} \to \infty (\theta > 0)$,*

$$|\hat{f}_h(x) - f(x)| \to 0 \quad a.s.$$

for any $x > 0$.

Proof: Note that

$$\begin{aligned} \left|\hat{f}_h(x) - f(x)\right| &= \left|\hat{f}_h(x) - \int_0^\infty k(x; u, h)dF(x) + \int_0^\infty k(x; u, h)dF(x) - f(x)\right| \\ &\leq \left|\hat{f}_h(x) - \int_0^\infty k(x; u, h)dF(x)\right| \\ &\quad + \left|\int_0^\infty k(x; u, h)dF(u) - f(x)\right|. \end{aligned} \tag{2.12}$$

The strong consistency of density estimator $\hat{f}_h(x)$ can be established if the two terms above tend to zero under the conditions of the theorem. First of all, we consider the first term.

Using integration by parts, the first term above can be written as

$$
\begin{aligned}
\left|\hat{f}_h(x) - \int_0^\infty k(x;u,h)dF(u)\right| &= \left|\int_0^\infty k(x;u,h)dF_n(u) - \int_0^\infty k(x;u,h)dF(u)\right| \\
&= \left|k(x;u,h)[F_n(u) - F(u)]\right|_0^\infty \\
&\quad - \int_0^\infty [F_n(u) - F(u)]d_u k(x;u,h)\Big| \\
&= \left|\int_0^\infty [F_n(u) - F(u)]d_u k(x;u,h)\right| \\
&\leq \sup_{t\geq 0} |F_n(t) - F(t)| \left|\int_0^\infty |d_u k(x;u,h)\right|. \qquad (2.13)
\end{aligned}
$$

We can show that

$$
\sup_{t\geq 0} |F_n(t) - F(t)| = o\left(\frac{1}{\sqrt{n(\log n)^{-(1+\theta)}}}\right) \qquad (2.14)
$$

(see [9]).

On the other hand, note that $k(x;u,h) = \frac{u}{x^2} q_h(u/x)$ where $q_h(x)$ is a IG density with $\mu = 1$ and $\lambda = 1/h$. After some tedious algebra, we have

$$
\begin{aligned}
\int_0^\infty |d_u k(x;u,h)| &= \frac{1}{x}\int_0^\infty \left|\left(\frac{t^2-1}{2ht}\right) - \frac{1}{2}\right| q_h(t)dt \\
&= O\left(\frac{1}{xh}\int_0^\infty |t-1|\, q_h(t)dt\right) \\
&= O\left(\frac{1}{\sqrt{h}}\right) \qquad (2.15)
\end{aligned}
$$

By (2.14) and (2.15), we can have that the first term in (2.12)

$$
\left|\int_0^\infty [F_n(u) - F(u)]d_u k(x;u,h)\right| \to 0 \quad a.s.
$$

under the conditions of the Theorem 2.1.

Now we consider the second term in (2.12).

$$\begin{aligned}
\left|\int_0^\infty k(x;u,h)dF(u) - f(x)\right| &= \left|\int_0^\infty \frac{u}{x^2} q_h(u/x) f(u)du - f(x)\right| \\
&= \left|\int_0^\infty tq_n(t)f(xt) - \int_0^\infty f(x)tq_h(t)dt\right| \\
&= \left|\int_0^\infty [f(xt) - f(x)]tq_h(t)dt\right| \\
&\leq \int_0^\infty |f(xt) - f(x)|\, tq_h(t)dt \\
&\leq xL \int_0^\infty |t-1|\, tq_h(t)dt \\
&\leq xL\sqrt{\int_0^\infty (t-1)^2 q_h(t)}\sqrt{\int_0^\infty t^2 q_h(t)dt} \\
&= O(\sqrt{h}), \qquad (2.16)
\end{aligned}$$

which tends to zero as $h \to 0$ and the proof is complete. □

Remark 2.2. A similar theorem is offered in Kulasekera and Padgett [14], however, there the only condition on h assumed is that it goes to zero. We believe that the condition $hn(\log n)^{-(1+\theta)} \to \infty$ imposed here is also needed in their theorem. The basic reason for this is that the absolute value under the integral sign is not properly accounted for in Kulasekera and Padgett [14].

In the next section we discuss the Bayesian method for selecting the smoothing parameter h and computational aspects.

3 Bayesian Estimation of h

Consider

$$f_h(x) = \int k(x;t,h)dF(t)$$

and let $\xi(h)$ represent a prior on h then the posterior of h at the point x is given by

$$\xi(h|x) = \frac{f_h(x)\xi(h)}{\int f_h(x)\xi(h)dh}.$$

In the spirit of empirical Bayes methodology, using the estimate $\hat{f}_h(x)$ in place of $f_h(x)$, the posterior of h based on the data $\mathcal{X}$ may be estimated by

$$\hat{\xi}(h|x,\mathcal{X}) = \frac{\hat{f}_h(x)\xi(h)}{\int \hat{f}_h(x)\xi(h)dh}.$$

Then under the squared error loss, the Bayes estimator of h is given by

$$\begin{aligned}
\tilde{h} &= \frac{\int h\hat{f}_h(x)\xi(h)dh}{\int \hat{f}_h(x)\xi(h)dh} \\
&= \frac{\sum s_i \int hk(x;X_i,h)\xi(h)dh}{\sum s_i \int k(x;X_i,h)\xi(h)dh}.
\end{aligned}$$

We follow Kulasekera and Padgett [14] to obtain the explicit expressions for the Bayes estimator of h under the two choices of kernels mentioned earlier. We will employ the inverted gamma prior with parameters α and β, with the density

$$\xi(h) = \frac{1}{\beta^\alpha \Gamma(\alpha) h^{\alpha+1}} \mathrm{e}^{-1/\beta h}$$

KP kernel: In this case, the (estimated) posterior density is given by

$$\hat{\xi}(h|x, X_1, ..., X_n) = \frac{\sum_{i=1}^n (s_i/h^{\alpha^*+1}) \mathrm{e}^{-1/\beta_i^* h}}{\Gamma(\alpha^*) \sum_{i=1}^n s_i (\beta_i^*)^{\alpha^*}} \tag{3.1}$$

where $\alpha^* = \alpha + 1/2$ for $\alpha > 1/2$ and

$$\beta_i^* = \left[\frac{1}{\beta} + \frac{(x - X_i)^2}{2xX_i^2} \right]^{-1}$$

Then the Bayes estimator of h is given by

$$\tilde{h}(x) = \frac{\sum s_i \beta_i^{*(\alpha^*-1)}}{(\alpha^* - 1) \sum s_i \beta_i^{*\alpha^*}}. \tag{3.2}$$

If we were to use an improper prior

$$\xi(h) \propto \frac{1}{h^2}, \tag{3.3}$$

then the proper posterior density can be obtained to be

$$\hat{\xi}(h|x, X_1, ..., X_n) = \frac{2}{\sqrt{\pi}} \frac{\sum_{i=1}^n s_i h^{-2/5} \mathrm{e}^{-1/\beta_i^{**} h}}{\sum_{i=1}^n s_i (\beta_i^{**})^{3/2}}$$

and the resulting estimator of h is given by

$$\tilde{h}(x) = 2 \frac{\sum s_i \beta_i^{*1/2}}{\sum s_i \beta_i^{*3/2}} \tag{3.4}$$

where

$$\beta_i^{**} = \left[\frac{(x - X_i)^2}{2xX_i^2} \right]^{-1}.$$

CL kernel: In this case, the (estimated) posterior density using the gamma prior is given by

$$\hat{\xi}(h|x, X_1, ..., X_n) = \frac{\sum_{i=1}^n (s_i/X_i^{1/2} h^{\alpha^*+1}) \mathrm{e}^{-1/\tilde{\beta}_i h}}{\Gamma(\alpha^*) \sum_{i=1}^n s_i (\tilde{\beta}_i)^{\alpha^*}/X_i^{1/2}}, \tag{3.5}$$

where

$$\tilde{\beta}_i = \left[\frac{1}{\beta} + \frac{(x - X_i)^2}{2xX_i} \right]^{-1}.$$

This results in the following estimator of h :

$$\tilde{h}(x) = \frac{\sum (s_i/X_i^{(1/2)})\tilde{\beta}_i^{(\alpha^*-1)}}{(\alpha^*-1)\sum (s_i/X_i^{(1/2)})\tilde{\beta}_i^{\alpha^*}}. \tag{3.6}$$

For the improper prior as given in Eq. 3.3, the estimate of h is given by

$$\tilde{h}(x) = 2\frac{\sum (s_i/X_i^{1/2})\tilde{\beta}_i^{*1/2}}{\sum (s_i/X_i^{1/2})\tilde{\beta}_i^{*3/2}} \tag{3.7}$$

where

$$\tilde{\beta}_i^* = \left[\frac{(x-X_i)^2}{2xX_i}\right]^{-1}.$$

Co-Gaussian Kernel: Using the co-Gaussian kernel $k_{CG}(x; u, h)$, the Bayesian estimator of h under the gamma prior is given by

$$\tilde{h}(x) = \frac{\sum s_i X_i \tilde{\beta}_i {*}^{(\alpha^*-1)}}{(\alpha^*-1)\sum s_i X_i \tilde{\beta}_i^{\alpha^*}}. \tag{3.8}$$

where

$$\tilde{\beta}_i^* = \left[\frac{1}{\beta} + \frac{(X_i - x^2/X_i)^2}{2x^2}\right]^{-1}.$$

And that with improper prior is given by

$$\tilde{h}(x) = 2\frac{\sum s_i X_i \hat{\beta}_i {*}^{*1/2}}{\sum s_i X_i \tilde{\beta}_i {*}^{*3/2}} \tag{3.9}$$

where

$$\tilde{\beta}_i^* = \left[\frac{(X_i - x^2/X_i)^2}{2x^2}\right]^{-1}.$$

Remark 3.1. Another choice of the asymmetric kernel related to the co-Gaussian kernel, in the spirit of KP kernel is to choose the co-Gaussian density in x with mode u and dispersion h as given by

$$k_{CG*}(x; u, h) = \sqrt{\frac{2}{\pi h}} \mathrm{e}^{-(1/2h)(x-u^2/x)^2}. \tag{3.10}$$

This gives the estimator of h under the gamma prior as

$$\tilde{h}(x) = \frac{\sum s_i \tilde{\beta}_i {*}^{(\alpha^*-1)}}{(\alpha^*-1)\sum s_i \tilde{\beta}_i^{\alpha^*}}. \tag{3.11}$$

where

$$\tilde{\beta}_i^* = \left[\frac{1}{\beta} + \frac{(x - X_i^2/x)^2}{2}\right]^{-1}.$$

And that with improper prior is given by

$$\tilde{h}(x) = 2\frac{\sum s_i \hat{\beta}_i {*}^{*1/2}}{\sum s_i \tilde{\beta}_i {*}^{*3/2}} \tag{3.12}$$

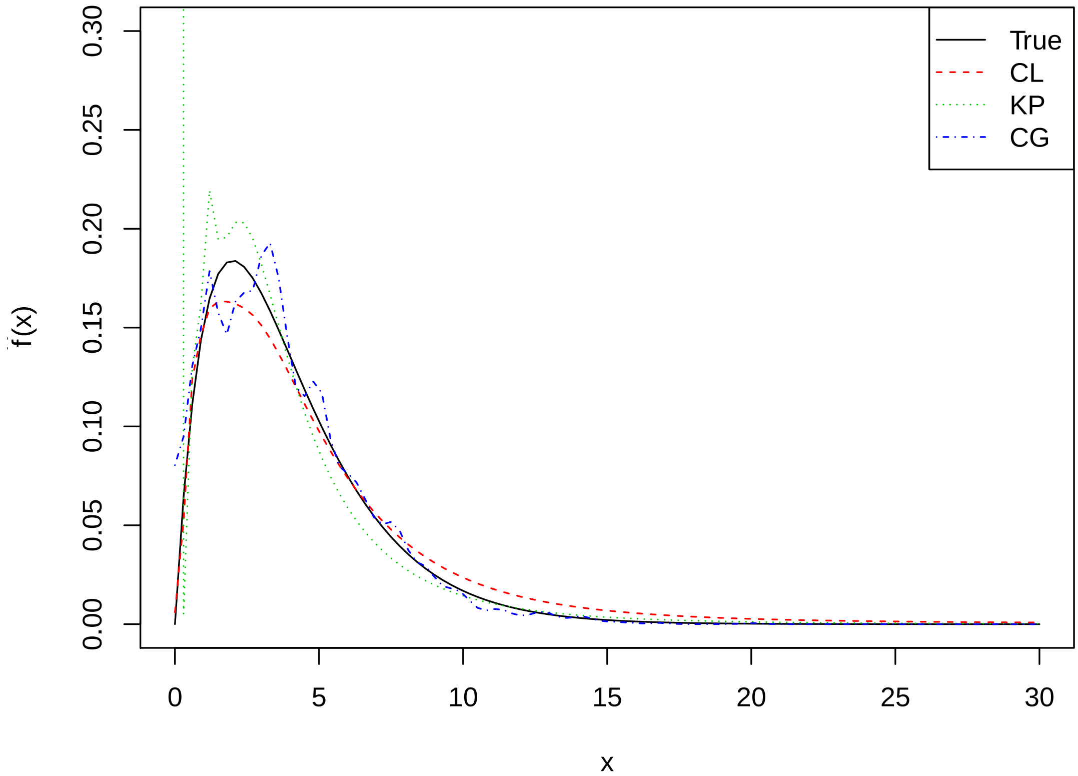


Figure 1: Smooth Estimators of the Length-biased $\chi^2(4)$ density, $n = 500$.

where

$$\tilde{\beta}_i^* = \left[\frac{(x - X_i^2/x)^2}{2}\right]^{-1}.$$

Figure 1 gives a qualitative idea of the nature of the two estimators using KP and CL kernels; the CG kernel is used in simulations but CG* kernel has been not considered any further after some preliminary investigation. It seems that the CL kernel has an edge over the other kernels considered here. The parameters α and β are obtained directly by minimizing the inetgrated squared error as described in the next section. Note that Kulasekara and Padgett [14] choose a value of $\beta > 3$ and then choose α so that the average values of h calculated at a range of values equals $1/(\alpha - 1)\beta$.

Remark 3.2. As mentioned in Remark 2.2, the mean and the mode of k_{CG*} coincide for small h. In practice, however, h may not be small enough for this approximation to hold.

Remark 3.3. Other choice of $k(x; u, h)$, for example an inverse-gamma kernel may also be used. We have not investigated these, however, we expect similar performance using such kernels as with the choice of kernels considered here.

Remark 3.4. The condition in Theorem 2.1, will be satisfied for the choice of $\alpha \to \infty$.

4 Optimal selection of parameters α and β

Here we describe the process of selecting h by developing a data-driven selection for parameters α and β. One of popular criteria that gives a measure of accuracy between any density estimator $\hat{f}(x)$ and the true density $f(x)$ is

$$\begin{aligned}\mathrm{ISE}(\hat{f}, f) &= \int_0^\infty [\hat{f}(x) - f(x)]^2 dx \\ &= \int_0^\infty \hat{f}^2(x)dx - 2\int_0^\infty \hat{f}(x)f(x)dx \\ &\quad + \int_0^\infty f^2(x)dx, \qquad (4.1)\end{aligned}$$

which is referred as integrated squared error(ISE). An optimal choice of bandwidth could be obtained by making the value of (4.1) as small as possible. However, since the true density $f(x)$ is unknown, (4.1) is not a practical objective function for optimization. So some modifications must be made. Ignoring the last (constant) term in (4.1) and replacing the second term with the leave-one-out estimator, an objective function for practical optimization arises as an approximation of (4.1), which is called as unbiased crossed-validation (UCV) criterion.

In the length-biased scheme, the objective function is given by

$$\mathrm{UCV} = \int_0^\infty \hat{f}^2(x)dx - 2\sum_{i=1}^n \hat{f}_{-i}(X_i)/Z_i, \qquad (4.2)$$

where $\hat{f}_{-i}(x)$ represents the density estimator built on data set excluding X_i and $Z_i = \sum_{j \neq i} \dfrac{X_i}{X_j}$ [see Wu and Mao (1996) and Chaubey *et al.* (2012)].

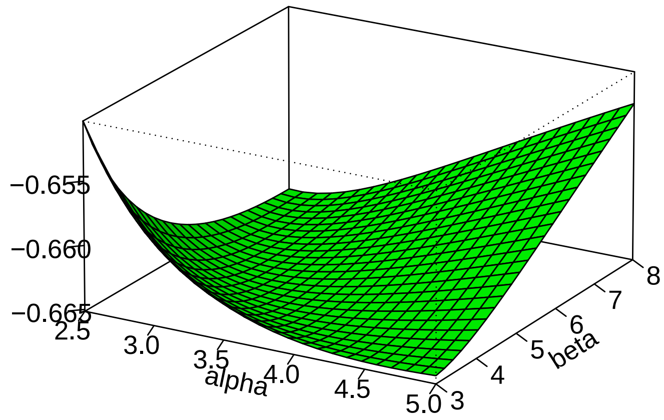


Figure 2: Plot of UCV function.

Note that the Bayesian method gives local bandwidth choice as h depends on x. After the local bandwidth h is plugged into (4.2), the UCV becomes a function of parameters α and β. In this case, the optimum choices of α and β will be obtained by minimizing the UCV function (4.2), which would definitively lead to the optimal selection of Bayesian bandwidth under criterion (4.1). The plot in Figure 2 is an example of a figure of UCV function with independent variables α and β. The surface of UCV function is concave down, which means that the minimum value of UCV function exists.

5 Numerical studies

We compare KP estimator with CL and CG estimator based on a simulation study that considers a class of Weibull densities:

$$f(x) = \theta x^{\theta-1} \exp\{-x^{\theta}\} I\{x \geq 0\}$$

with $\theta = 0.5, 1, 1.5$ and the followings are their figures.

We use the UCV method as mentioned earlier to choose the global parameters α and β in the KP, CL and CG priors. Then we compare their performance based on ratios $R_{KP,CL}$ and $R_{KP,CG}$, where

$$R_{KP,CL} = \frac{\text{EMSE}(\hat{f}_{KP}(x))}{\text{EMSE}(\hat{f}_{CL}(x))}$$

and

$$R_{KP,CG} = \frac{\text{EMSE}(\hat{f}_{KP}(x))}{\text{EMSE}(\hat{f}_{CG}(x))}$$

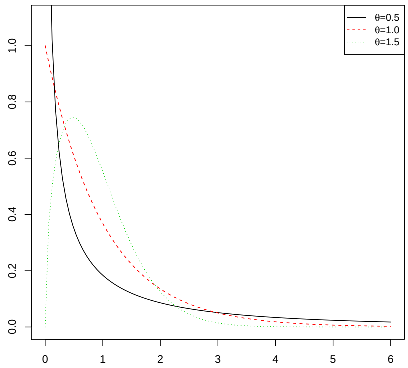
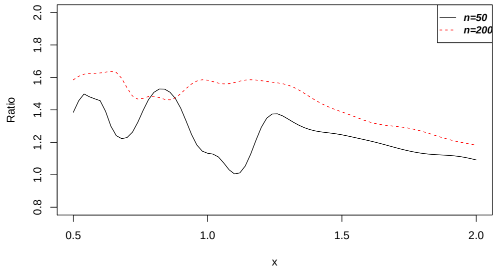


Figure 3: Plot of various underlying pdf's

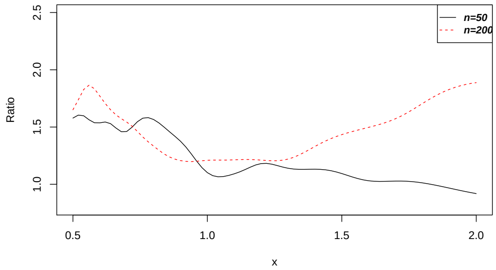
for a range of the values of x;, here the estimated MSE(EMSE) for the density estimator $\hat{f}(x)$ is defined as

$$\text{EMSE}(\hat{f}(x)) = \sum_{i=1}^{N}[\hat{f}(x) - f(x)]^2/N,$$

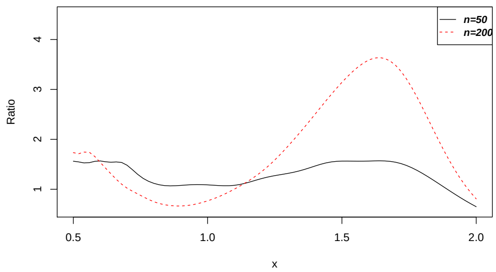
f being the true density function $f(x)$ and N being the number of simulations. Here we use 500 repetitions to estimate MSE. The ratios $R_{KP,CL}$ and $R_{KP,CG}$ are plotted over a range of $x-$values for different values of θ in Figure 4.



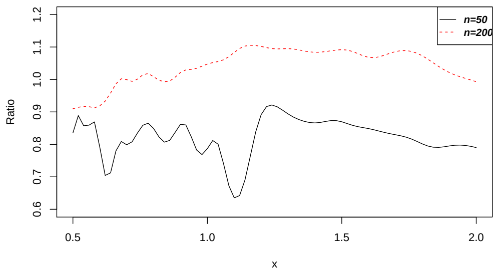
(a) $R_{KP,CL}$ for $\theta = 0.5$, Solid line: $n = 50$, Dotted line: $n = 200$.



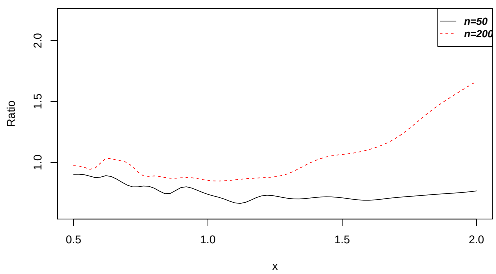
(b) $R_{KP,CL}$ for $\theta = 1.0$, Solid line: $n = 50$, Dotted line: $n = 200$.



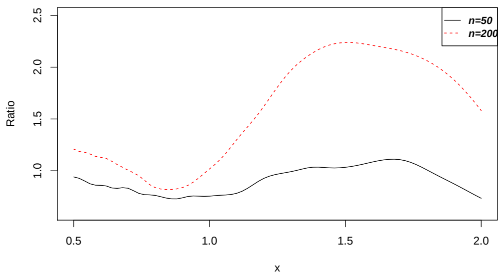
(c) $R_{KP,CL}$ for $\theta = 1.5$, Solid line: $n = 50$, Dotted line: $n = 200$.



(d) $R_{KP,CG}$ for $\theta = .5$, Solid line: $n = 50$, Dotted line: $n = 200$.



(e) $R_{KP,CG}$ for $\theta = 1.0$, Solid line: $n = 50$, Dotted line: $n = 200$.



(f) $R_{KP,CG}$ for $\theta = 1.5$, Solid line: $n = 50$, Dotted line: $n = 200$.

Figure 4: Plots of the Ratios $R_{KP,CL}$ and $R_{KP,CG}$ for various values of θ.

6 Conclusions and Discussion

For smaller values of x as well as in tails, CL kernel outperforms the KP kernel; KP kernel may have a slight advantage over the CL kernel in middle range, but this advantage disappears for larger values of n. The co-Gaussian kernel may be a bit better than KP kernel only for larger sample sizes and large values of x. The co-Gaussian kernel was later abandoned from further investigation realizing that the resulting $k_{CG}(x; u, h)$ may neither be a density in x nor in u. However, when the mean of the generating kernel $q_h(u)$ is close to one [that would be true for small values of h,], $q_h(u)$ is a density with approximate mean 1 and the CG kernel will behave as a density in u with mean x.

The co-Gaussian kernel CG* that is given by (3.10) have not been investigated here, as

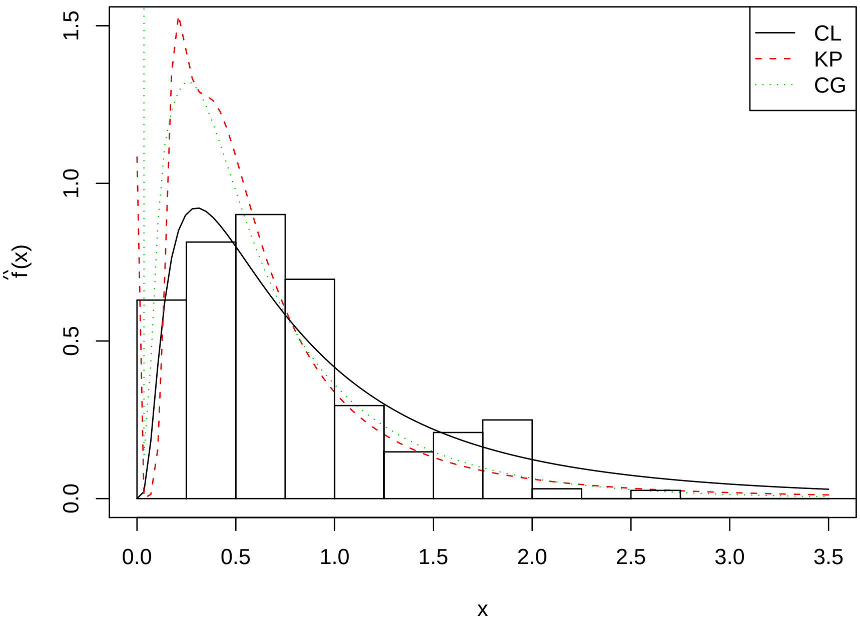


Figure 5: Plots of density estimators and histogram for 46 shrubs.

a preliminary investigation showed that it does not seem to be any better qualitatively than using the KP kernel.

We also apply the density estimators consider here to real length biased data constituting widths of 46 shrubs given in Muttlak and McDonald [15]. These estimators are plotted in Figure 5. Their behavior in extreme tails is similar, however KP and CG estimators do not perform well at the edge.

Acknowledgements

A major portion of this research was carried out while Y. Chaubey visited the Indian Statistical Institute - Delhi Center. The academic facilities accorded there are gratefully acknowledged. The partial support of this research through a Discovery Grant from NSERC, Canada to Y. Chaubey is also acknowledged. The authors are also thankful to Professor H.L. Koul for his comments on an earlier draft.

References

[1] Abramowitz, M. and Stegun, I.A. (1965). *Handbook of Mathematical Functions*, Dover, New York.

[2] Awadalla, S. (2011). *Some Contributions to the Theory and Applications of R-Symmetry*, Ph.D. Thesis written under the supervison of Professor Govind S. Mudholkar and Professor Michael P. McDermott, University of Rochester, Rochester, NY (USA).

[3] Bagai, I. and Prakasa Rao, B.L.S. (1995). Kernel Type Density Estimates for Positive Valued Random Variables. *Sankhyā* **A57**, 56-67.

[4] Bhattacharyya, B.B., Franklin, L.A. and Richardson, G.D. (1988). A Comparison of Nonparametric Unweighted and Length-Biased Density Estimation of Fibres. *Communications in Statistics - Theory and Methods*, **17**, 3629-3644.

[5] Chaubey, Y.P. and Li, J. (2012). Asymmetic kernel density estimator for length biased data. In *Contemporary Topics in Mathematics and Statistics with Applications,* Vol. 1, (Ed.: A. Adhikari, M.R. Adhikari and Y.P. Chaubey), Pub: Asian Books Pvt. Ltd (New Delhi), to appear.

[6] Chaubey, Y.P., Li, J., Sen, A. and Sen, P.K. (2012). A new smooth density estimator for non-negative random variables. *Journal of the Indian Statistical Association* **50**, 83-104.

[7] Chaubey, Y.P. and Sen, P.K. (1996). On Smooth Estimation of Survival and Density Function. *Statistics and Decision*, **14**, 1-22.

[8] Chaubey, Y. P., Sen, A. and Sen, P. K. (2007). A New Smooth Density Estimator for Non-Negtive Random Variables. *Technical Report No. 01/07, Department of Mathematics and Statistics, Concordia University, Montreal.*

[9] Chaubey, Y.P., Sen, P.K. and Li, J. (2010). Smooth Density Estimation for Length-biased Data. *Journal of the Indian Society of Agricultural Statistics*, **64**, 145-155.

[10] Chen, S.X. (2000). Probability Density Function Estimation Using Gamma Kernels. *Annals of the Institute of Statistical Mathematics*, **52**, 471-480.

[11] Cox, D.R. (1969). Some Sampling Problems in Technology. In *New Developments in Survey Sampling*, 506-527 (Eds.: N.L. Johnson and H. Smith,) John Wiley, New York.

[12] Feller, W. (1966). *An Introduction to Probability Theory and Application II* , John Wiley, New York.

[13] Jones, M.C. (1991). Kernel Density Estimation for Length Biased Data. *Biometrika,* **78,** 511-519.

[14] Kulasekera, K. B. and Padgett, W. J. (2006). Bayes Bandwidth Selection in Kernel Density Estimation with Censored Data. *Journal of Nonparametric Statistics*, **18**, 129-143.

[15] Muttlak, H. A. and McDonald, L. L. (1990) *Ranked set sampling with size-biased probability of selection. Biometrics*, 46, 435-446.

[16] Patil, G. P. and Rao, C. R. (1977). Weighted distributions: a survey of their applications. In *Applications of Statistics.* P. R. Krishnaiah (ed.), North Holland Publishing Company, 383-405.

[17] Patil, G. P. and C. R. Rao (1978). Weighted Distributions and Size-Biased Sampling with Applications to Wildlife Populations and Human Families. *Biometrics,* **34** 179-189.

[18] Rao, C. R. (1965). On Discrete Distributions Arising out of Methods of Ascertainment. In *Classical and Contagious Discrete Distribution*, 320-332 (Ed. G. P.Patil,) Pergamon Press and Statistical Publishing Society, Calcutta.

[19] Rao, C. R. (1977). A Natural Example of A Weighted Binomial Distribution. *American Statistician,* **31,** 24-26.

[20] Scaillet, O. (2004). Density Estimation Using Inverse and Reciprocal Inverse Gaussian Kernels. *Journal of Nonparametric Statistics*, **16**, 217-226.

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

[22] Wu, C. O. and Mao, A. Q. (1996). Minimax Kernel for Density Estimation with Biased Data. *Annals of Institute of Statistical Mathematics*, **48**, 451-467.