Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 64(2) 2010 145-155

Smooth Density Estimation for Length-biased Data

Yogendra P. Chaubey[1*], Pranab K. Sen[2] and Jun Li[1]
[1]*Concordia University, Montreal, Quebec, Canada*
[2]*University of North Carolina at Chapel Hill, Chapel Hill, NC, USA*

Received 25 January 2010; Revised 25 May 2010; Accepted 16 June 2010

SUMMARY

Biased data frequently arise in applications concerning wild life and human populations as indicated in the comprehensive article by Patil and Rao (1978). Such data follow densities that are proportional to the original population density. Here we provide a non-parametric estimator of the density that is based on the smoothing of the Cox's (1969) estimator using Poisson weights. The new method that is appropriate for nonnegative data is contrasted with some estimators in literature based on nonparametric kernel smoothing. Based on simulation studies, it is shown that the new estimator fares better in terms of the Mean Integrated Squared Error (MISE) compared to kernel based estimators. The asymptotic consistency and normality of the new estimator is also established under standard regularity conditions.

Keywords: Length-biased density, Weighted distribution, Nonparametric density estimation.

1. INTRODUCTION

When an observation following a stochastic model is recorded, its distribution may not be the same as the original distribution, unless every observation is given equal chance of being recorded. In many applications the recorded observation may be assumed to have the probability density function $g(x)$, that is of the form

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

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 $X \sim g(\cdot)$. Patil and Rao (1978) 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 (1965), Patil and Ord (1975), Patil and Rao (1977) and Rao (1977)]. Such models are appropriate in many applied fields including agriculture, ecology and forestry. As one could easily expect, not incorporating the knowledge of the weighting function in any inference procedure dealing with such data may seriously invalidate the underlying study. Ricker (1969) highlights such concerns of bias on estimates of growth, mortality, production, and yield in the context of fisheries study that are equally valid in other contexts.

Our aim in this article is to study the problem of estimation of the original density $f(x)$ given a set of n nonnegative, independent and identically distributed (i.i.d.) observations $X_1, ..., X_n$ having a length-biased continuous probability density function (pdf) $g(x)$, $x \in \mathbb{R}^+ = [0, \infty)$. Here we assume for the rest of the paper that $w(x) = x$, in which case the density $g(x)$ is known as the length-biased density and μ_w becomes the harmonic mean of $X \sim g(x)$, i.e. $\mu^{-1} = E_g(1/X)$. The treatment in this paper can be easily modified to incorporate the general weight function. Further, it is tacitly assumed that $(0 <)\mu < \infty$. Note that

$$\mu \le E_g(X) = \mu^{-1} \int_0^\infty x^2 f(x) dx \tag{1.3}$$

or equivalently, $\mu^2 \le E_f(Y^2)$ where Y has the pdf f. Thus, assuming that Y has a finite 2nd moment insures that $\mu < \infty$ (even finite first moment of Y does so).

*Corresponding author : Y.P. Chaubey
E-mail address : chaubey@alcor.concordia.ca

In a length-biased sampling scheme, though the observable r.v.s are X_i with pdf $g(\cdot)$, the primary focus of statistical conclusions is on the characteristics of the pdf $f(\cdot)$, and in this context estimation of $f(x)$, $x \in \mathbb{R}^+$ itself is often of central importance. A crude way of estimating $f(\cdot)$ [see Bhattacharyya *et al*. 1988] is to estimate $g(\cdot)$, possibly by a *kernel-method,* and μ separately and thereby consider a plugged-in estimator of $f(\cdot)$. As such, we let

$$f_n(x) = \hat{\mu}_n (nx)^{-1} \sum_{i=1}^{n} h_n^{-1} k\left(\frac{x - X_i}{h_n}\right), \; x \in \mathbb{R}^+ \qquad (1.4)$$

where in analogy with (1.2),

$$\hat{\mu}_n = \left(\int_0^\infty x^{-1} \mathrm{d}G_n(x)\right)^{-1} = n \Big/ \left(\sum_{i=1}^{n} X_i^{-1}\right) \qquad (1.5)$$

and $k(\cdot)$ is a suitable *kernel* function. The presence of x^{-1} in (1.4) may generally induce heavier bias near $x \downarrow 0$, especially when $f(0) = 0$ (Jones 1991). Motivated by the seminal work of Cox (1969), Jones (1991) advocated an alternative estimator

$$\hat{f}_n(x) = \hat{\mu}_n \left(n^{-1} \sum_{i=1}^{n} \frac{1}{h_n X_i} k\left(\frac{x - X_i}{h_n}\right)\right), \; x \in \mathbb{R}^+ \qquad (1.6)$$

and observed that the mean integrated squared error (MISE) of $\hat{f}_n(x)$ in (1.6), at least for large sample sizes, is smaller than that of $\tilde{f}_n(x)$ in (1.4). Wu and Mao (1996) showed that under the *minimax criterion,* asymptotically, the MSE of $\hat{f}_n(\cdot)$ is smaller than that of $f_n(\cdot)$.

If the kernel function $k(y)$ is symmetric around zero, as is the usual case, while the support of $f(\cdot)$ is $\mathbb{R}^+ = [0, \infty)$, for x near zero (*viz.,* $0 < x < h_n$, $h_n \to 0$ as $n \to \infty$), there will be large bias in these estimates [see Fig. 1]. This phenomenon is also observed in conventional kernel estimates of a density function near a finite end point of the support, but the presence of X_i^{-1} in (1.6) [or (1.4)] makes it even more pronounced in the length-biased case. As such, alternate smoothing procedures have been advocated for the conventional case, and in the present study, some of these will be used in the length-biased case.

For non-negative r.v.s, smoothed estimation of probability density and other functionals has been advocated by a number of workers. In this context, Chaubey and Sen (1996) incorporated a celebrated smoothing lemma, known as the *Hille's Lemma* (see Lemma 2.1) and exhibited its utility and comparative performance. This approach uses the weights generated by a Poisson distribution in smoothing the empirical distribution function in the traditional *i.i.d*. setting. Since, the Poisson distribution is defined for non-negative values, the new approach offers a prime advantage in removing the boundary bias at the lower end point of the distribution (here $x = 0$), that is inherent in the kernel method as seen in Fig. 1. The present study extends this line of attack to the length-biased sampling scheme.

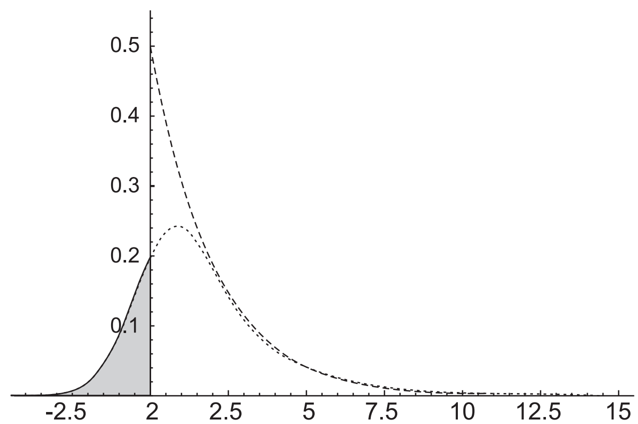

Fig. 1. Probability Density function of χ_2^2 and Jones kernel estimator. Solid line represents true density and dashed line represents the estimated density.

Along with some preliminary notions, the proposed smooth estimator of $f(\cdot)$ is formulated in Section 2. Asymptotic properties are studied in Section 3. Section 4 is devoted to numerical simulation studies and a real example. The concluding Section 5 deals with some useful discussion. The proofs are relegated to an Appendix.

2. PRELIMINARY NOTIONS AND A NEW SMOOTH DENSITY ESTIMATOR

Cox (1969) proposed the following version of the empirical distribution function for estimating $F(\cdot)$ in the length-biased setup:

$$F_n(x) = \frac{\sum_{i=1}^{n} \frac{1}{X_i} I\{X_i \le x\}}{\sum_{i=1}^{n} \frac{1}{X_i}} \qquad (2.1)$$

It can be shown to be the nonparametric maximum likelihood estimator of $F(x)$ in this context. Obviously, this estimator has jump discontinuities and therefore is not amenable to provide an estimator of the corresponding pdf in case $F(\cdot)$ is assumed to be absolutely continuous. It is therefore desirable to have a continuous version, that can at least be differentiable. For this purpose we incorporate the Hille's Lemma (see Feller (1966), pp. 219) as given below:

Lemma 2.1. If $u(x)$ is a bounded, continuous function on $\mathbb{R}^+$, then, as $\lambda \uparrow \infty$,

$$e^{-\lambda x} \sum_{k \geq 0} u(k/\lambda)(x\lambda)^k/k! \to u(x)$$

uniformly in any finite interval **J** contained in $\mathbb{R}^+$.

It can be readily seen that substitution of $F_n(\cdot)$ in place of $u(\cdot)$ provides a continuous version of the Cox's estimator,

$$\tilde{F}_n(x) = \sum_{k \geq 0} p_k(x\lambda_n) F_n(k/\lambda_n) \tag{2.2}$$

where $p_k(u) = e^{-u} \frac{u^k}{k!}$ and λ_n is such that, as $n \to \infty$, $\lambda_n \to \infty$.

The above estimator has all the properties of a distribution function as can be readily demonstrated. Hence taking its derivative provides a valid estimator of the underlying pdf $f(.)$ given by

$$\tilde{f}_n(x) = \lambda_n \sum_{k \geq 0} p_k(x\lambda_n) \left[F_n\left(\frac{k+1}{\lambda_n}\right) - F_n\left(\frac{k}{\lambda_n}\right) \right] \tag{2.3}$$

The role of Poisson weights here is very similar to the kernel method of density estimation method and λ_n controls the smoothness of the estimator. Since, the weights are concentrated around the value of x, the local properties of the function to be smoothed are somewhat preserved. In the following section we study the asymptotic properties of this estimator. We see that the error of the smooth estimator with respect to the raw estimator converges very fast to zero as $n \to \infty$. For simplicity of notation, we will use the notation E and V for the expectation and variance respectively with respect to the density g, unless there is any ambiguity.

3. ASYMPTOTIC PROPERTIES OF THE ESTIMATORS

3.1 Asymptotic Properties of $\tilde{F}_n(x)$

First note that by the Kolmogorov Strong Law of Large of Numbers (see Loève (1977) pp. 251), it is easy to show that, as $n \to \infty$, $F_n(x) \xrightarrow{a.s.} F(x)$. Since $F_n(x)$ is nondecreasing, we have

$$\sup_{x \in R^+} |F_n(x) - F(x)| \xrightarrow{a.s.} 0 \tag{3.1}$$

By (3.1), following the proof of Theorem 3.1 of Chaubey and Sen (1996), we can obtain the uniform strong convergence of $\tilde{F}_n(x)$.

Theorem 3.1. Let $0 < E(X^{-1}) < \infty$, and $F(x)$ be continuous (a.e.), $\lambda_n \to \infty$, then, as $n \to \infty$,

$$\|\tilde{F}_n(x) - F(x)\| = \sup_{x \in R^+} \{|\tilde{F}_n(x) - F(x)|\} \xrightarrow{a.s.} 0 \tag{3.2}$$

Remark 3.1: In theorem 3.1 of Chaubey and Sen (1996), additional condition on λ_n, namely that $n^{-1}\lambda_n \to 0$ is assumed that is not required for the above theorem to hold. It may be noted that the estimator in Chaubey and Sen (1996) uses truncated Poisson weights, where such a condition may be necessary.

Next, we discuss the weak convergence of the estimator in (2.2). Following along the lines of the proof of Theorem 3.2 in Chaubey and Sen (1996) using Lemma 6.1 (in the Appendix) with $b_n = n^{-\frac{1}{2}} (\log n)^{\frac{1+\theta}{2}}$ [see also the treatment in Sen (1984)], we establish the following theorem.

Theorem 3.2. If $E(X^{-2}) < \infty$, $\lambda_n \to \infty$, and $n^{-1}\lambda_n \to 0$, $f(x)$ is absolutely continuous with bounded derivative $f'(x)$, then for some $\delta > 0$, as $n \to \infty$,

$$\|\tilde{F}_n(x) - F_n(x)\| = O(n^{-3/4}(\log n)^{1+\delta}) \text{ a.s. } \forall x \in \mathbb{R}^+ \tag{3.3}$$

Using Delta-Method, it is easy to show the weak convergence of raw estimator (2.1). That is

$$\sqrt{n}\,(F_n(x) - F(x)) \xrightarrow{\mathcal{D}} N(0, \delta^2(x))$$

where $\delta^2(x) = \mu[\int_0^x \frac{1}{t} f(t)\, dt - 2F(x) \int_0^x \frac{1}{t} f(t)\, dt + \bar{\mu}F^2(x)]$ and $\bar{\mu} = \int_0^\infty \frac{f(t)}{t} dt$. Since we have

$\tilde{F}_n(x) - F(x) = \tilde{F}_n(x) - F_n(x) + F_n(x) + F(x)$, by Theorem 3.2, we immediately obtain the weak convergence of smooth estimator (2.2). In this regard, we have the following theorem.

Theorem 3.3. If $E(X^{-2}) < \infty$, $\lambda_n \to \infty$, and $n^{-1}\lambda_n \to 0$, $f(x)$ is absolutely continuous with bounded derivative $f'(x)$, then, as $n \to \infty$,

$$\sqrt{n}(\tilde{F}_n(x) - F(x)) \xrightarrow{\mathcal{D}} N(0, \delta^2(x)) \qquad (3.4)$$

where

$$\delta^2(x) = \mu\left[\int_0^x \frac{1}{t} f(t)\mathrm{d}t - 2F(x)\int_0^x \frac{1}{t} f(t)\mathrm{d}t + \bar{\mu}F^2(x)\right]$$

Theorems 3.1 and 3.3 show that the smooth estimator preserves the convergence properties of the raw estimator under suitable choice of the smoothing parameter.

3.2 Asymptotic Properties of $\tilde{f}_n(x)$

The strong convergence of the smooth density estimator given in (2.3) follows along the same lines as in the conventional case.

Theorem 3.4. If $E(X^{-2}) < \infty$, $f'(x)$ is bounded and $\lambda_n = O(n^{\alpha})$ for some $0 < \alpha < 1$, then, as $n \to \infty$,

$$\left\| \tilde{f}_n(x) - f(x) \right\| \xrightarrow{a.s.} 0 \qquad (3.5)$$

We obtain the weak convergence of $\tilde{f}_n$, under the assumption that $f'(x)$ satisfies a Lipschitz order α condition, *i.e.* for some $\alpha > 0$, there exits a finite positive K, such that

$$|f'(s) - f'(t)| \le K|s - t|^{\alpha}, \text{ for every } t, s \in \mathbb{R}^+ \qquad (3.6)$$

Theorem 3.5. If $E(X^{-2}) < \infty$, $\lambda_n = O(n^{2/5})$ (non-stochastic) and (3.6) holds, then, for a compact set $\mathcal{C} \subset \mathbb{R}^+$,

$$\{(n^{2/5}[\tilde{f}_n(x) - f(x)] - \frac{1}{2\delta^2} f'(x)), x \in \mathcal{C}\} \xrightarrow{\mathcal{D}} \mathcal{G}$$

where $\mathcal{G}$ denotes the Gaussian process with covariance function $\gamma_x^2 \delta_{xs}$ where $\gamma_x^2 = \frac{\mu}{2}(2\pi x^3)^{-1/2} f(x)\delta$, $\delta_{xs} = 0$ for $x \neq s$ and 1 for $x = s$ and $\delta = \lim_{n\to\infty}(n^{-/5}\lambda_n^{1/2})$.

The proofs of the above theorems are relegated to the Appendix.

Remark 3.2: If $\lambda_n = cn^h$ and (3.6) holds, then, by the proofs of Theorem 3.4 and 3.5, we have

$$Bias^2(\tilde{f}_n(x)) \approx c^{-2}(f'(x)/2)^2 n^{-2h} \qquad (3.7)$$

and

$$V(\tilde{f}_n(x)) \approx \frac{\mu}{2}\sqrt{\frac{c}{2\pi x^3}} f(x) n^{\frac{h}{2}-1} \qquad (3.8)$$

then we have

$$MSE(\tilde{f}_n(x)) \approx c^{-2}(f'(x)/2)^2 n^{-2h} + \frac{\mu}{2}\sqrt{\frac{c}{2\pi x^3}} f(x) n^{\frac{h}{2}-1} \qquad (3.9)$$

Remark 3.3: When $\lambda_n = cn^{2/5}$, the bias and variance of the Poisson weights estimator (2.3) go to zero at the rate $O(n^{-4/5})$, which is the same as that in the case of the classical kernel estimator, however it requires the existence of higher order derivatives.

4. NUMERICAL AND SIMULATION STUDIES

In this section, we will compare the mean integrated squared error (MISE) of the three different density estimators by simulation, where the MISE for the estimator $f_n(x)$ is given by

$$MISE = E\int (f_n(x) - f(x))^2 \mathrm{d}x$$

Note that $\int (f_n - f(x))^2 \mathrm{d}x$ is known as the integrated squared error (ISE) of the estimator $\hat{f}_n(x)$. We discuss below the method of selecting the smoothing parameters required in computing each of the estimators.

4.1 Choice of Smoothing Parameters

Choice of Smoothing Parameter λ_n : For Poisson estimator, we minimize the following biased cross-validation function to obtain the optimal choice of λ_n,

$$BCV(\lambda_n) = \lambda_n^{-2}\int_0^{\infty} (\tilde{f}_n'(x)/2)^2 dx + \sqrt{\lambda_n}\frac{\mu^2}{\sqrt{8\pi n}} MCE_n \qquad (4.1)$$

where $MCE_n = \frac{1}{n}\sum_{i=1}^n X_i^{-5/2}$, which is a Monte-Carlo estimator of $\int_0^{\infty} \frac{f(x)}{\mu x^{3/2}} \mathrm{d}x$. Since μ is unknown, we substitute μ with its estimator $\hat{\mu}_n$ as given in (1.5) for computation. The alternative unbiased cross validation criterion is found to be too variable, and therefore is not suitable for large scale automated computation. In general, for the large samples, the two criteria yield similar results.

Choice of the Bandwidth for the Kernel Estimator: For Jones and Bhattacharyya *et al.* estimators, we minimize the following unbiased cross-validation function to obtain the optimal choice of parameter h_n,

$$UCV(h_n) = \int_0^\infty \hat{f}_n^2(x)\mathrm{d}x - 2\sum_{i=1}^{n} \hat{f}_{n-1}(X_i, h_n; D_i)/Z_i \quad (4.2)$$

where $\hat{f}_{n-1}(\cdot, h_n; \mathcal{D}_i)$ denotes the estimator built on data set $\mathcal{D}_i$ which consists of original data set $\mathcal{D}$ except for X_i and $Z_i = \sum_{j \neq i} \frac{X_i}{X_j}$.

For the subsequent numerical studies we have used the R-package for statistical analysis and its subroutine optimise for minimization of the cross-validation criteria [see Ihaka and Gentleman (1996)]. For other numerical details the reader may refer to Chaubey and Sen (2009).

4.2 Simulation for χ_2^2 Distribution

We simulate length biased sample from a χ_2^2 distribution with sample sizes 30, 50, 100, 200, 300, 500, and 1000. We illustrate the three density estimators from such a sample of size 200 in Fig. 2, that shows that Poisson weights estimator can be very accurate at the boundary $x = 0$.

Using the smoothing parameters as described above we compute the ISE of each estimator for a simulated sample. These values are averaged over 1000 replications producing an approximate value of MISE. The results are plotted in Fig. 3) that shows the Poisson weights estimator to have lower MISE than the other two estimators.

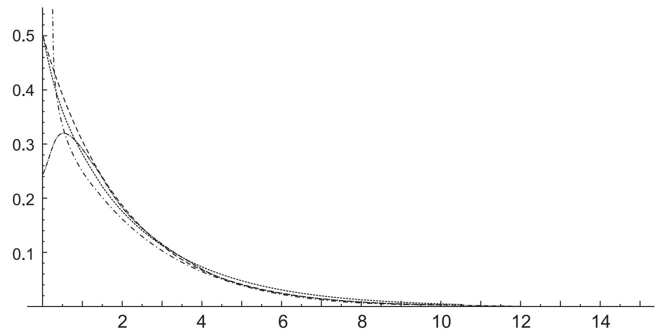

Fig. 2. Plots of Density Function and Estimators, Sample Size = 200, Solid Line = True (exponential) Density, Dash Dot Dash = Bhattacharyya *et al.* estimator (h_n = 0.5), Short Dash = Poisson Estimator (λ_n = 1.6), Long Dash = Jones Estimator (h_n = 0.7)

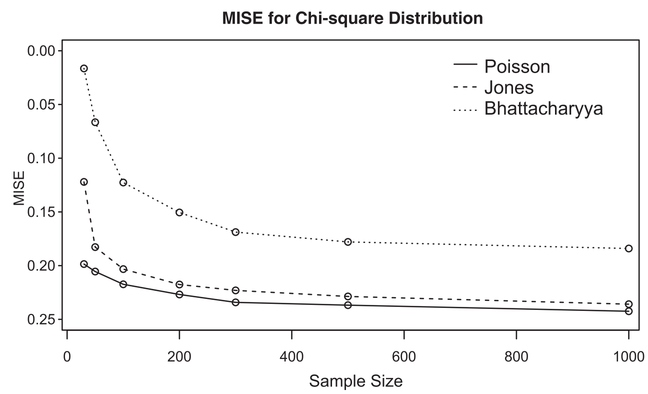

Fig. 3. Plots of MISE for length-biased χ_2^2 density.

4.3 Simulation for Some Standard Distributions

We also follow the procedure described above for the following standard distributions.

(i) Chi-square Distribution

$$f(x) = \frac{1}{2^{\alpha/2}\Gamma(\alpha/2)} x^{\alpha/2-1} \exp(-x/2) I\{x > 0\}$$

(ii) Lognormal Distribution

$$f(x) = \frac{1}{\sqrt{2\pi x}} \exp\{-(\log x - \mu)^2/2\}\ I\{x > 0\}$$

(iii) Weibull Distribution

$$f(x) = \alpha x^{\alpha-1} \exp(-x^{\alpha})\ I\{x > 0\}$$

(iv) Mixtures of Two Exponential Distribution

$$f(x) = [\pi \frac{1}{\theta_1} \exp(-x/\theta_1)$$

$$+ (1 - \pi) \frac{1}{\theta_2} \exp(-x/\theta_2)] \; I\{x > 0\}$$

The methods of generating corresponding length biased data are given by, respectively,

(i′) $X \sim \chi^2(\alpha + 2)$;

(ii′) $X = e^Y$ where $Y \sim N(1 + \mu, 1)$;

(iii′) $X = Y^{1/\alpha}$ where $Y \sim \Gamma(1 + \frac{1}{\alpha}, 1)$;

(iv′) $X = \pi Y_1 + (1 - \pi) Y_2$, where $Y_1 \sim \Gamma(2, \theta_1)$ and $Y_2 \sim \Gamma(2, \theta_2)$.

Resulting values of MISE are plotted in Fig. 4-7, respectively.

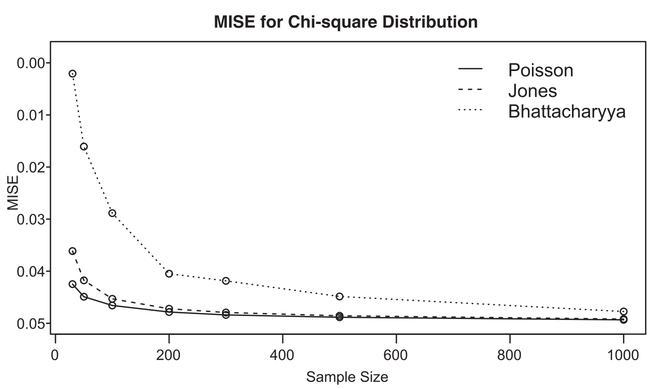

Fig. 4. Plots of MISE for χ^2_6

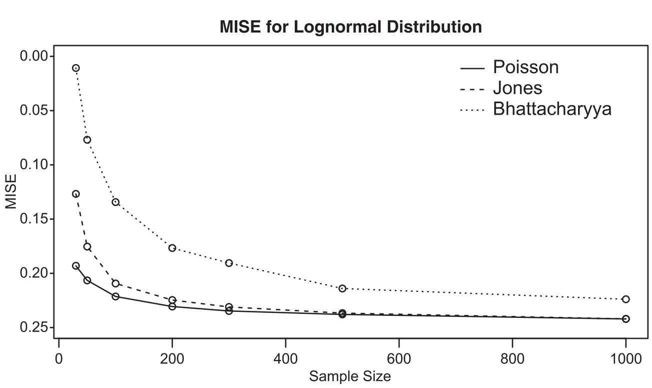

Fig. 5. Plots of MISE for Lognormal distribution with $\mu = 0$

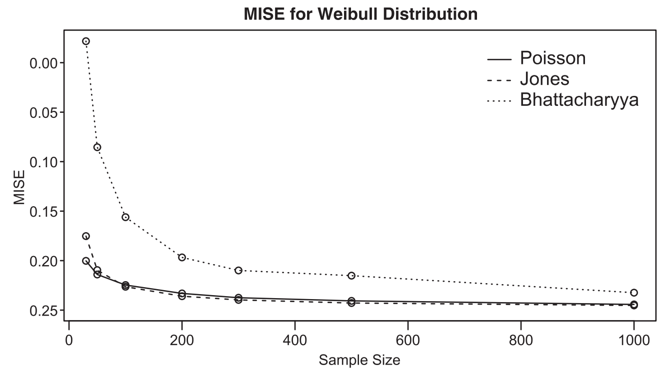

Fig. 6. Plots of MISE for Weibull distribution with $\alpha = 2$

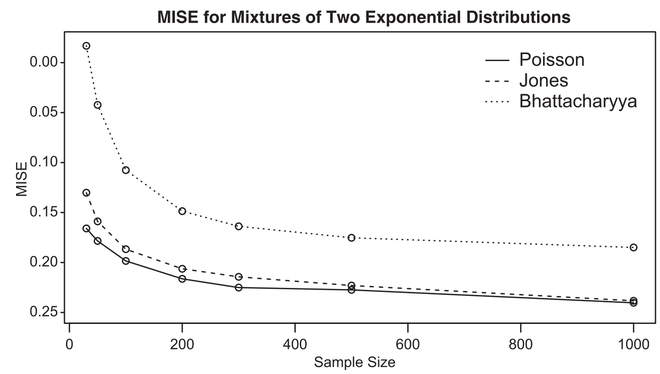

Fig. 7. Plots of MISE for mixtures of two exponential distributions with $\pi = 0.4$, $\theta_1 = 2$ and $\theta_2 = 1$

4.4 A REAL DATA EXAMPLE

In this section, we apply the three density estimators to real length biased data constituting widths of 46 shrubs given in Muttlak and McDonald (1990). The smoothing parameters λ_n required for the Poisson weights estimator and h_n required for kernel based estimators are obtained by minimizing the cross-validation functions given in (4.1) and (4.2). The values of these smoothing parameters along with cross-validation criteria values are summarized in Table 1. In Fig. 8, we plot these estimators along with the probability histogram. The smoothed density resembles closely to a Weibull density that has been used earlier for fitting these data (see Chaubey and Yang (2007)).

Table 1. Parameter and UCV

	Poisson	Jones	Bhattacharyya
Smoothing	λ_n	h_n	h_n
Parameter	6.186	0.2459	0.2222
UCV	–0.6727	–0.6536	–0.6519

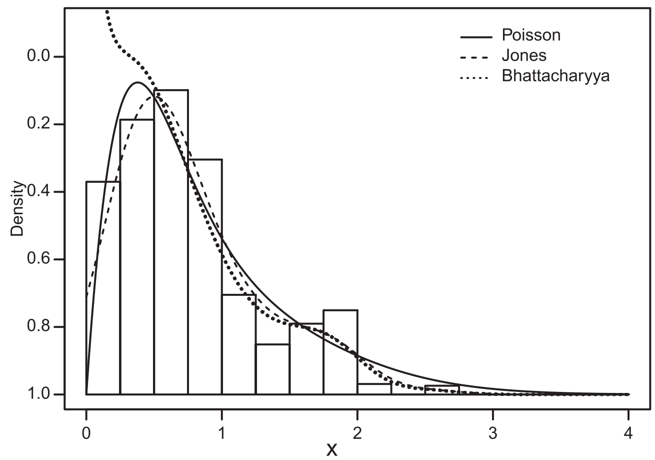

Fig. 8. Plots of histogram and estimators

From Table 1, we can see that the unbiased cross-validation function of Poisson weights estimator has the smallest value among the three estimators. Fig. 8 shows that all the three estimators have no quantitative difference in the region which is far from the origin. However, the three estimators perform much differently near the origin. Bhattacharyya *et al.* estimator tends to infinity very fast near $X = 0$, where as Jones estimator has positive value at $x = 0$ and the Poisson weights estimator gives value zero at the origin. This feature is apparent in some parametric models fitted to the above data (see Chaubey and Yang (2007)). Furthermore, for the integral of different estimators as obtained by numerical methods over the range $(0, \infty)$, only the integral of Poisson weights estimator is 1 (see Table 2).

Table 2. Integral and absolute error

Method	Integral	Absolute Error
Poisson	1.000	$< 7.3 \times 10^{-6}$
Jones	0.9560	$< 5.1 \times 10^{-5}$
Bhattacharyya *et al.*	1.394	$< 7.0 \times 10^{-5}$

5. DISCUSSION AND CONCLUSIONS

Based on the above simulation studies, we might draw the following conclusions.

1. When the value of variable is close to zero, Bhattacharyya *et al.* estimator diverges very fast. It seems that this estimator has an order $O\left(\frac{1}{x}\right)$ and its graph looks like a vertical line near the lower boundary. As a result this estimator produces the largest MISE among the three estimators, even with larger sample sizes.
2. In some cases, the MISE of Bhattacharyya *et al.* estimator is mainly determined by the huge bias near the boundary. Therefore, even with increased sample size, its MISE may be hard to improve. However, the MISE's of both the Jones and Poisson weights estimators decrease with increasing sample size.
3. When $f(0) \neq 0$, Poisson weights estimator usually produces much smaller MISE than Jones estimator. Our computation shows that the two estimators are both very accurate at the points that are away from boundary. So we believe that the decrease of Poisson weights estimator's MISE is due to the fact that it reduces the bias at the boundary.
4. When $f(0) = 0$, Poisson weights and Jones estimator usually have similar MISE. Technically, in our examples, when sample size is small, Poisson weights estimator performs better than Jones estimator. When sample size is large, two estimators have almost the same MISE. This is likely due to the fact that the MISE's of both the estimators have the same asymptotic order $O(n^{-4/5})$.

6. APPENDIX: PROOFS OF THEOREMS 3.4 AND 3.5

First, we will introduce an important lemma, which plays a critical role in the proof of strong consistency of $\tilde{f}_n(x)$.

Lemma 6.1. If $E(X^{-2}) < \infty$ and $f'(t)$ is bounded, then

$$\sup_{t \in \mathbb{R}^+} \sup_{|\beta| \le b_n} \{|F_n(t+\beta) - F_n(t) - F(t+\beta) + F(t)|\}$$
$$= O(b_n^{\frac{1}{2}} n^{-\frac{1}{2}} (\log n)^{1+\theta}) \ a.s.,$$

where θ (> 0) is arbitrary.

In order to prove Lemma 6.1, we need the following two lemmas. For convenience, we denote

$$x_i(t, \beta) = \frac{\mu}{X_i} I\,\{\min(t, t+\beta) < X_i \le \max(t, t+\beta)\}$$
$$-|F(t+\beta) - F(t)|\ (i = 1, ..., n) \qquad (6.1)$$

Lemma 6.2. If $E(X^{-2}) < \infty$, then, for any $t \geq 0$ and $t + \beta \geq 0$,

$$\frac{1}{n}\sum_{i=1}^{n} x_i(t, \beta) = O(n^{-\frac{1}{2}} (\log n)^{\frac{(1+\theta)}{2}}) \ a.s. \quad (6.2)$$

Proof of Lemma 6.2: In order to prove the lemma, we need the Kolmogorov's Proposition *A* in *M*. Loève (pp. 250), that is stated below.

If the integrable r.v.'s X_n are independent, then

$$\sum \frac{\sigma^2 X_n}{b_n^2} < \infty,\ b_n \uparrow \infty, \text{ entails } \frac{S_n - ES_n}{b_n} \xrightarrow{a.s.} 0$$

where $S_n = \sum_{i=1}^{n} X_i$ and $\sigma^2 X_i$ means the variance of X_i.

Under the assumption $E(X^{-2}) < \infty$, for any $t \geq 0$ and $t + \beta \geq 0$, we have

$$\sum_{n=1}^{\infty} \frac{\sigma^2 x_n(t, \beta)}{(n^{1/2}(\log n)^{(1+\theta)/2})^2} \leq \sum_{n=1}^{\infty} \frac{E(X_n^{-2})}{(n^{1/2}(\log n)^{(1+\theta)/2})^2} < \infty$$

By the above proposition, we have

$$\frac{\sum_{i=1}^{n} x_i(t, \beta)}{n^{1/2}(\log n)^{(1+\theta)/2}} \xrightarrow{a.s.} 0 \quad (6.3)$$

Further

$$\frac{1}{n}\sum_{i=1}^{n} x_i(t, \beta) = n^{-\frac{1}{2}}(\log n)^{\frac{(1+\theta)}{2}} \frac{\sum_{i=1}^{n} x_i(t, \beta)}{n^{\frac{1}{2}}(\log n)^{\frac{(1+\theta)}{2}}} \quad (6.4)$$

hence, by (6.3) and (6.4), we obtain the desired result.

Lemma 6.3. If $E(X^{-2}) < \infty$ and $f'(t)$ is bounded, then there exits $d > 0$ such that, for any $t \geq 0$, $0 < b_n^{-1} < O(n^{1-\gamma})$ $(0 < \gamma < 1)$, $-b_n < \beta < b_n$, $D = b_n^{1/2} n^{1/2} (\log n)^{1+\theta}$, we have

$$p\{|\sum_{i=1}^{n} x_i(t, \beta)| > 2dD\} \leq O(n^{-4}) \quad (6.5)$$

The order $O(n^{-4})$ does not depend on t and β.

Proof of Lemma 6.3: First we should verify several facts. For any $\delta > 2$, we have

$$E|\sum_{i=1}^{p} x_i(t, \beta)|^{\delta} \leq (p(\log n)^{1+\theta})^{\delta/2} \quad (6.6)$$

since

$$E|\sum_{i=1}^{p} x_i(t, \beta)|^{\delta} = p^{\delta} E|\frac{1}{p}\sum_{i=1}^{p} x_i(t, \beta)|^{\delta}$$

and, by Lemma 6.2

$$\frac{1}{p}\sum_{i=1}^{p} x_i(t, \beta) = O(p^{-1/2}(\log p)^{(1+\theta)/2}) \ a.s.$$

Also,

$$E(x_1(t, \beta))^2 = E(\frac{\mu^2}{X_1^2} I\{\min(t, t+\beta) < X_i \leq \max(t, t+\beta)\}) - |F(t+\beta) - F(t)|^2$$

$$= \left|\int_t^{t+\beta} \frac{\mu f(x)}{x} dx\right| - |F(t+\beta) - F(t)|^2$$

$$= O(|\beta|) \quad (6.7)$$

The conclusion of the last step follows because $E(X_1^{-2}) < \infty$ and that $f'(x)$ is bounded. Since, $|f(x)/x| = |f'(\eta)| < M$, ($\eta \in (0, x)$ and M is finite), the first term of (6.7) has an order O($|\beta|$). And since $f(x)$ is, the second term of (6.7) has an order $O(\beta^2)$.

Thus, using (6.7) and the independence of $x_i(t, \beta)$ $(i = 1, ..., n)$, we can also establish (2.7) in Lemma 2.1 of Babu and Singh (1978), that is

$$E(\xi_1^2) \leq O(pb_n) \quad (6.8)$$

Substituting (2.4) in Lemma 2.1 of Babu and Singh (1978) with (6.6), taking $\delta = 60/\gamma$ and $p = [n^{\gamma/2}]$, and following the proof of Lemma 2.1 of Babu and Singh (1978), we can obtain the result.

Remark: The second term $\exp(-8D^2 n^{-1} b_n^{-1})$ in (2.1) of Babu and Singh (1978) disappears in our inequality, because under our choice of D, this term is of much smaller than $O(n^{-4})$.

Proof of Lemma 6.1: Let

$$H_n(t, \beta) = F_n(t+\beta) - F_n(t) - F(t+\beta) + F(t)$$

Since $F_n(t+\beta) - F_n(t)$ can be expanded as

$$F_n(t+\beta) - F_n(t) = \frac{1}{n}\sum_{i=1}^{n}\frac{\mu}{X_i}\{t < X_i \le t+\beta\}$$

$$-[F(t+\beta) - F(t))]\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\mu}{X_i} - 1\right)$$

$$+ O([F(t+\beta) - F(t)]\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\mu}{X_i} - 1\right)) \ a.s. \tag{6.9}$$

we have

$$| H_n(t, \beta) | \le J_{n1}(t, \beta) + J_{n2}(t, \beta) + O(J_{n2}(t, \beta)) \ a.s. \tag{6.10}$$

where

$$J_{n1}(t, \beta) = \frac{1}{n}\left|\sum_{i=1}^{n} x_i(t,\beta)\right| \tag{6.11}$$

and

$$J_{n2}(t, \beta) = \left|[F(t+\beta) - F(t)]\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\mu}{X_i} - 1\right)\right| \tag{6.12}$$

For (6.11), first we consider that t is fixed. Using Lemma 6.3, following the proof of Lemma 1 of Bahadur (1966), we can claim that

$$\sup_{|\beta|\le b_n} \{|J_{n1}(t, \beta)|\} = O(b_n^{\frac{1}{2}} n^{-\frac{1}{2}} (\log n)^{1+\theta}) \ a.s.$$

Furthermore, since $O(b_n^{\frac{1}{2}} n^{-\frac{1}{2}} (\log n)^{1+\theta})$ does not depend on t and $f'(t)$ is bounded, using the same technique as in Sen and Ghosh (1971), we can extend the result for t to the whole real line, that is

$$\sup_{t\in \mathbb{R}^+} \sup_{|\beta|\le b_n} \{|J_{n1}(t, \beta)|\} = O(b_n^{\frac{1}{2}} n^{-\frac{1}{2}} (\log n)^{1+\theta}) \text{ a.s.} \tag{6.13}$$

Using $t = 0$ and $\beta \to +\infty$ in Lemma 6.2, we have

$$\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\mu}{X_i} - 1\right) = O(n^{-1/2}(\log n)^{(1+\theta)/2}) \ a.s. \tag{6.14}$$

Since $f(t)$ is bounded (because $E(X_1^{-2}) < \infty$) as well, we have

$$\sup_{t\in \mathbb{R}^+} \sup_{|\beta|\le b_n} | F(t+\beta) - F(t) | = O(b_n) \tag{6.15}$$

For (6.12), by (6.14) and (6.15), we have

$$\sup_{t\in \mathbb{R}^+} \sup_{|\beta|\le b_n} \{(|J_{n2}(t, \beta)|\} = O(b_n n^{-\frac{1}{2}} (\log n)^{(1+\theta)/2}) \ a.s. \tag{6.16}$$

By (6.10), (6.13) and (6.16), we can establish the Lemma 6.1.

After the preparatory material, we can prove Theorem 3.4.

Proof of Theorem 3.4. By the proof of Theorem 4.1 of Chaubey and Sen (1996), we just need to show that, when t belongs to some finite interval $[0, C]$, we have (3.5), since we can deliberately choose C such that when t belongs to interval $(C, +\infty)$, $\tilde{f}_n(t)$ and $f(t)$ can both be made suffciently small.

We can write

$$\tilde{f}_n(x) = \lambda_n\{\sum_{k\ge 0} p_k(x\lambda_n)[F(\frac{k+1)}{\lambda_n}) - F(\frac{k}{\lambda_n})]$$

$$+ \sum_{k\ge 0} p_k(x\lambda_n)[F_n(\frac{k+1}{\lambda_n}) - F_n(\frac{k}{\lambda_n})$$

$$-F(\frac{k+1}{\lambda_n}) + F(\frac{k}{\lambda_n})]\}$$

$$= T_{n1}(x) + T_{n2}(x) \tag{6.17}$$

Using Lemma 6.1 by taking $b_n = 1/\lambda_n$, we have

$$\sup_{k\ge 0}\{|F_n(\frac{k+1}{\lambda_n}) - F_n(\frac{k}{\lambda_n}) - F(\frac{k+1}{\lambda_n}) + F(\frac{k}{\lambda_n})|\}$$

$$= O(\lambda_n^{-\frac{1}{2}} n^{-\frac{1}{2}} (\log n)^{1+\theta}) \ a.s. \tag{6.18}$$

By (6.18) and the fact that $\sum_{k\ge 0} p_k(x\lambda_n) = 1$, we have

$$\sup_{x\in \mathbb{R}^+} \{| T_{n2}(x) |\} = O(\lambda_n^{1/2} n^{-1/2} (\log n)^{1+\theta}) \ a.s. \tag{6.19}$$

which tends to 0 almost surely as $n \to \infty$ provided that $\lambda_n = O(n^{\alpha})(0 < \alpha < 1)$.

At the same time, according to the proof of Theorem 4.1 of Chaubey and Sen (1996), under the assumption of boundedness of $f'(x)$, we have

$$\sup_{t\in[0,C]} \{|T_{n1}(x) - f(x)|\} \to 0 \ a.s. \qquad (6.20)$$

By (6.19) and (6.20), we obtain the theorem. The proof is complete.

Proof of Theorem 3.5: By (3.6), we have

$$\tilde{f}_n(x) = f(x) + \frac{1}{2\lambda_n} f'(x) + T_{n2}(x) + O(\lambda_n^{-1-\alpha}) \qquad (6.21)$$

Using Taylor's expansion which is similar to (6.9), we can write

$$T_{n2}(x) = \lambda_n \sum_{k\geq 0} p_k(x\lambda_n)\{(\frac{1}{n}\sum_{i=1}^{n}\frac{\mu}{X_i} I\{\frac{k}{\lambda_n} < X_i \leq \frac{k+1}{\lambda_n}\}$$

$$-[F(\frac{k+1}{\lambda_n}) - F(\frac{k}{\lambda_n})])\}$$

$$-\lambda_n \sum_{k\geq 0} p_k(x\lambda_n) \ \{[F(\frac{k+1}{\lambda_n}) - F(\frac{k}{\lambda_n})](\frac{1}{n}\sum_{i=1}^{n}\frac{\mu}{X_i} - 1)\}$$

$$+O(\frac{1}{n}\sum_{i=1}^{n}\frac{\mu}{X_i} - 1)$$

$$= T_{n3}(x) - T_{n4}(x) + O(\frac{1}{n}\sum_{i=1}^{n}\frac{\mu}{X_i} - 1) \ a.s. \qquad (6.22)$$

For the leading term $T_{n3}(x)$, following the proof of Theorem 4.2 of Chaubey and Sen (1996), we can show that

$$V(T_{n3}(x)) \approx \frac{\mu}{2}(2\pi x^3)^{-1/2} f(x) \ (\lambda_n^{1/2}/n) \qquad (6.23)$$

and, for $s \neq t$, as $n \to \infty$,

$$Cov[T_{n3}(s), T_{n3}(t)] = O(\frac{1}{n}) \qquad (6.24)$$

Moreover, since $T_{n4}(x) = O(\frac{1}{n}\sum_{i=1}^{n}\frac{\mu}{X_i} - 1) = O(n^{-1/2}(\log n)^{(1+\theta)/2})$, the order of $T_{n2}(x)$ is determined by that of $T_{n3}(x)$.

From (6.21), we can see that the asymptotic normality of $T_{n2}(x)$ leads to the asymptotic normality of $\tilde{f}_n(x)$. By (6.21), (6.22), (6.23) and (6.24), following the proof of Theorems 4.1 and 4.2 of Chaubey and Sen (1996), we can complete the proof of the theorem.

ACKNOWLEDGEMENT

Thanks are due to to Professor Raj S. Chhikara for his efficient handling of the paper and to the referee for his critical reading and constructive comments. The research of the first author was partially supported from the authors Discovery Grant from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

Babu, G.J. and Singh, K. (1978). On deviations between empirical and quantile process for mixing random variables. *J. Multivariate Anal.,* **8**, 532-549.

Bahadur, R.R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.,* **37**, 577-580.

Bhattacharyya, B.B., Franklin, L.A. and Richardson, G.D. (1988). A comparison of nonparametric unweighted and length-biased density estimation of fibres. *Comm. Statist.,* **A17**, 3629-3644.

Chaubey, Y.P. and Sen, P.K. (1996). On smooth estimation of survival and density functions. *Statist. Decision,* **14**, 1-22.

Chaubey, Y.P., Sen, A. and Sen, P.K. (2007). A new smooth density estimator for non-negative random variables. *Technical Report No. 01/07*, Department of Mathematics and Statistics, Concordia University, Montreal.

Chaubey, Y.P. and Yang, M.J. (2007). Inference for length and area biased exponentiated Weibull distribution. *J. Statist. Studies,* **26**, 19-27.

Chaubey, Y.P. and Sen, P.K. (2009). On the selection of smoothing parameter in Poisson smoothing of histogram estimator: Computational aspects. *Pak. J. Statist.,* **25(4)**, 385-401.

Cox, D.R. (1969). Some sampling problems in technology. In: *New Developments in Survey Sampling*. N. L. Johnson and H. Smith (eds.). John Wiley, New York, 506-527.

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

Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics. *J. Comput. Graph. Statist.,* **5**, 299-314.

Jones, M.C. (1991). Kernel density estimation for length biased data. *Biometrika,* **78**, 511-519.

Loève, M. (1977). *Probability Theory I ,* 4th Edition. Springer-Verlag, New York.

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

Patil, G.P. and Ord, J.K. (1975). On size-biased sampling and related form-invariant weighted distributions. *Sankhyā*, **B38**, 48-61.

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.

Patil, G.P. and Rao, C.R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics,* **34**, 179-189.

Rao, C.R. (1965). On discrete distributions arising out of methods of ascertainment. In: *Classical and Contagious Discrete Distribution.* G.P. Patil (ed.), Pergamon Press and Statistical Publishing Society, Calcutta, 320-332.

Rao, C.R. (1977). A natural example of a weighted binomial distribution. *Amer. Statist.,* **31**, 24-26.

Ricker, W.E. (1969). Effects of size-selective mortality and sampling bias on estimates of growth, mortality, production, and yield. *J. Fish. Res. Board Canada,* **26**, 479541.

Sen, P.K. and Ghosh, M. (1971). On bounded length sequential confidence intervals based on one sample rank order statistic. *Ann. Math. Statist.,* **42**, 189-203.

Sen, P.K. (1984). On asymptotic representation of reduced quantiles in sampling from length-biased distribution. *Cal. Stat. Assoc. Bull.,* **33**, 59-67.

Wu, C.O. and Mao, A.Q. (1996). Minimax kernels for density estimation with biased data. *Ann. Inst. Statist. Math.,* **48**, 451-467.