Taylor & Francis
Taylor & Francis Group

Adaptive quadratic functional estimation of a weighted density by model selection

Athanasia Petsa and Theofanis Sapatinas*

Department of Mathematics and Statistics, University of Cyprus, Nicosia, Cyprus

We consider the problem of estimating the integral of the square of a probability density function f on the basis of a random sample from a weighted distribution. Specifically, using model selection via a penalized criterion, an adaptive estimator for $\int f^2$ based on weighted data is proposed for probability density functions which are uniformly bounded and belong to certain Besov bodies. We show that the proposed estimator attains the minimax rate of convergence that is optimal in the case of direct data. Additionally, we obtain the information bound for the problem of estimating $\int f^2$ when weighted data are available and compare it with the information bound for the case of direct data. A small simulation study is conducted to illustrate the usefulness of the proposed estimator in finite sample situations.

Keywords: adaptive minimax estimation; Besov bodies; Haar basis; projection estimators; quadratic functional estimation; weighted distributions

1. Introduction

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (i.i.d.) random variables with cumulative distribution function (c.d.f.) F and probability density function (p.d.f.) f with respect to the Lebesgue measure on the real line $\mathbb{R} = (-\infty, \infty)$. In practice, it sometimes happens that direct data are not available. There are several settings that lead to weighted data sets. Weighted distributions are used in statistics to model sampling in the presence of selection bias. Observations which do not have an equal chance of being selected lead to this sampling scheme which can be described in the following way: let $Y_1, Y_2, \ldots, Y_n$ be i.i.d. random variables from a weighted distribution with p.d.f. g_w given by

$$g_w(y) = \frac{w(y)f(y)}{\mu_w}, \tag{1}$$

where the weight function w satisfies $w(y) > 0$ for all $y \in \mathbb{R}$ and $\mu_w = \mathbb{E}(w(X)) < \infty$ (see, e.g., [1]). (The restriction $w(y) > 0$ for all $y \in \mathbb{R}$ is necessary for identifiability reasons; this constraint guarantees that g_w is indeed a p.d.f.; see, e.g., [2,3].) When the probability that an

---

*Corresponding author. Email: t.sapatinas@ucy.ac.cy; fanis@ucy.ac.cy

observation is selected is proportional to its size, i.e., when $w(y) = y$, length-biased data arise. Inter-event time data, the visibility bias in aerial survey techniques, the quality control problem of estimating fibre length distribution and sampling from queues or telephone networks are some examples of settings where weighted data arise (see, e.g., [2,4]).

Cox [4] proposed an estimator of F given by

$$\tilde{F}(y) = n^{-1}\hat{\mu}_w \sum_{i=1}^{n} w^{-1}(Y_i)\mathbb{I}_{(-\infty, y]}(Y_i),$$

where $\hat{\mu}_w = n(\sum_{i=i}^{n} w^{-1}(Y_i))^{-1}$ and $\mathbb{I}_A(y) = 1$ if $y \in A$ and 0 otherwise. Hence, this estimator can be interpreted as the empirical distribution function for weighted data. Vardi [2,3] showed that $\tilde{F}$ is the non-parametric maximum likelihood estimator of F for this case, and that $\hat{\mu}_w$ is a $\sqrt{n}$-consistent estimator of μ_w. Kernel estimators of f for weighted data from model (1) were proposed by Bhattacharyya *et al.* [5] and Jones [6], while their multivariate extensions were considered in [7]. Asymptotic properties of these estimators were considered in [8,9], for a Hölder class of p.d.f.'s. A Fourier series estimator of f for weighted data from model (1) was proposed by Jones and Karunamuni [10], while a transformation-based estimator was suggested by El Barmi and Simonoff [11]. Efromovich [12] suggested a blockwise shrinkage estimator of f for weighted data from model (1) and showed that this estimator is sharp minimax, that is, the proposed estimator attains the optimal constant and minimax rate of convergence for a Sobolev class of p.d.f.'s. Additionally a second-order sharp minimax estimator for F, via a projection on trigonometric bases, and of f by differentiation, for an analytic class of c.d.f.'s, was derived in [13].

Let X be a random variable with c.d.f. F and p.d.f. f with respect to Lebesgue measure on the real line $\mathbb{R}$, and let $f \in L^2(\mathbb{R})$ (the space of squared-integrable functions on $\mathbb{R}$). The aim is to estimate $\int f^2$, assuming f belongs to some smooth class of p.d.f.'s. This functional appears, e.g., in the Pitman efficacy of the Wilcoxon signed-rank statistic, in rank tests based on residuals in the linear model and in the asymptotic variance of the Hodges–Lehmann estimator (see, e.g., [14–16]). Additionally, an estimate of this quantity can be used in test statistics based on the L^2-distance (see, e.g., [17,18]). If independent direct realizations $X_1, X_2, \ldots, X_n$ of X are available, then optimal solutions to this problem are well known. Bickel and Ritov [19] proposed an estimator of $\int (f^{(k)})^2$, where $f^{(k)}$ is the kth derivative of f, for p.d.f.'s satisfying the Hölder condition on $f^{(m)}$ with smoothness parameter α. Although their estimator is asymptotically efficient when $m + \alpha > 2k + 1/4$ and asymptotically optimal for $k < m + \alpha \leq 2k + 1/4$, it is non-adaptive since it depends on unknown parameters. Birgé and Massart [20] proposed non-adaptive $\sqrt{n}$-consistent estimators for functionals of the form $\int \phi(f, f', \ldots, f^{(k)}, \cdot)$, for f belonging to some smooth class of p.d.f.'s with smoothness parameter s satisfying $s \geq 2k + 1/4$, and proved that $\int \phi(f, f', \ldots, f^{(k)}, \cdot)$ cannot be estimated at a rate faster than $n^{-4(s-k)/(4s+1)}$ if $s < 2k + 1/4$. Laurent [21,22] extended these results and built non-adaptive and asymptotically efficient estimators of more general functionals.

Recently, Laurent [23] used model selection methods to construct an adaptive and asymptotically optimal (in the minimax sense) estimator of $\int f^2$ for p.d.f.'s which are uniformly bounded and belong to certain Besov bodies. In this paper, we consider the problem of estimation of $\int f^2$ based on a weighted sample from model (1). An estimate of this functional could be used when statistical procedures developed for direct data (e.g., tests based on L^2-distance) are adapted to weighted data. By modifying the method used by Laurent [23] for the case of direct data and borrowing also ideas from [6,10,24] in order to take into account the selection bias, we construct an adaptive estimator of $\int f^2$.

The paper is organized as follows. Section 2 describes the method used to construct an adaptive estimator for $\int f^2$ based on a weighted sample, using model selection via a penalized criterion.

We show that the proposed estimator attains the minimax rate of convergence that is optimal in the direct data case for p.d.f's which are uniformly bounded and belong to certain Besov bodies, under the assumption that the biasing function $w(y)$ is bounded away from 0 and ∞. In Section 2, using the theory of Ibramigov and Khasminski [25], we derive the information bound for the problem of estimating $\int f^2$ when weighted data are available. A comparison with the information bound given for the case of direct data (see, e.g., [19,21,26]) leads to the conclusion that model sampling in the presence of selection bias can either improve or worsen the information bound in the problem of estimating $\int f^2$. In Section 3, a small simulation study is conducted to illustrate the usefulness of the proposed estimator in finite sample situations and to compare it with a simple projection estimator. Finally, the Appendix contains some auxiliary statements and proofs of the theoretical results stated in Section 2.

2. Estimation of $\int f^2$ using weighted data by model selection

We consider below the problem of estimating $\theta = \int f^2$ based on a weighted sample from model (1) for p.d.f's which are uniformly bounded and belong to a certain Besov body. The approach adapted modifies the method used by Laurent [23] for the case of direct data, also borrowing ideas from [6,10,24] in order to take into account the selection bias.

2.1. An adaptive estimator of $\int f^2$ using weighted data

We project f onto the space generated by the constant piecewise functions on the intervals $(k/D, (k+1)/D]$, $k \in \mathbb{Z}$. The projection of f onto this space is given by

$$f_D = \sum_{k \in \mathbb{Z}} \alpha_{k,D}\, p_{k,D},$$

where $p_{k,D} = \sqrt{D}\mathbb{I}_{(k/D,(k+1)/D]}$ and $\alpha_{k,D} = \int f p_{k,D}$. It is easy to see that

$$\tilde{\theta}_D = \frac{\mu_w^2}{n(n-1)} \sum_{\substack{1 \le i,j \le n \\ i \ne j}} \sum_{k \in \mathbb{Z}} \frac{p_{k,D}(Y_i)}{w(Y_i)} \frac{p_{k,D}(Y_j)}{w(Y_j)} \tag{2}$$

is an unbiased estimator for $\theta_D = \int f_D^2$.

The following assumption is used to prove Theorems 2.1 and 2.2 below.

ASSUMPTION 2.1 *Let w be a real-valued function satisfying $0 < w_1 \le w(y) \le w_2 < \infty$ for all $y \in \mathbb{R}$.*

Under Assumption 2.1, and for uniformly bounded densities f, i.e., $\|f\|_\infty = \sup_{y \in \mathbb{R}} |f(y)| \le M$ for some finite constant $M > 0$, it is easy to check that

$$\mathbb{E}(\tilde{\theta}_D - \theta)^2 \le \left\{ (\theta_D - \theta)^2 + C(M, w) \left(\frac{D}{n^2} + \frac{1}{n} \right) \right\},$$

where $C(M, w)$ is an absolute constant depending on M, w_1 and w_2. According to the ideas presented in [27], the optimal choice of D should minimize the quantity $\theta - \theta_D + \sqrt{D}/n$ or,

equivalently, maximize the quantity $\theta_D - \sqrt{D}/n$. Therefore, we consider the following estimator:

$$\tilde{\theta} = \sup_{D \in \mathcal{D}_n} (\tilde{\theta}_D - \text{pen}(D)), \tag{3}$$

where $\text{pen}(D)$ is given by

$$\text{pen}(D) = \frac{\kappa}{n} \sqrt{(\tilde{\theta}_D + 1) D \log(D + 1)}, \tag{4}$$

for some constant $\kappa > 0$. However, μ_w is unknown in practice and therefore $\tilde{\theta}_D$ must be replaced by

$$\hat{\theta}_D = \frac{\hat{\mu}_w^2}{n(n-1)} \sum_{\substack{1 \le i, j \le n \\ i \ne j}} \sum_{k \in \mathbb{Z}} \frac{p_{k,D}(Y_i)}{w(Y_i)} \frac{p_{k,D}(Y_j)}{w(Y_j)}, \tag{5}$$

where $\hat{\mu}_w$ is the $\sqrt{n}$-consistent estimator of μ_w (see Section 1). Therefore, a natural adaptive estimator for $\theta = \int f^2$ is given by the penalized estimator

$$\hat{\theta} = \sup_{D \in \mathcal{D}_n} (\hat{\theta}_D - \text{pen}_u(D)), \tag{6}$$

where $\text{pen}_u(D)$ is given by

$$\text{pen}_u(D) = \frac{\kappa}{n} \sqrt{(\hat{\theta}_D + 1) D \log(D + 1)}.$$

In the later sections, the notation C is used for absolute constants whose values may vary from one line to another. The dependency of a constant on some parameter or the bounds of the weight function is implied in the following way: for example, $C(\alpha, R, M)$ denotes an absolute constant depending on α, R and M, while $C(w)$ denotes an absolute constant depending on w_1 and w_2.

2.2. *Upper bounds for the L^2-risk*

Let $\phi(x) = \mathbb{I}_{[0,1)}(x)$ and $\psi(x) = \mathbb{I}_{[0,1/2)}(x) - \mathbb{I}_{[1/2,1)}(x)$, and for any $j \in \mathbb{N}$, $k \in \mathbb{Z}$, let

$$\phi_{j,k}(x) = 2^{j/2} \mathbb{I}_{[0,1)}(2^j x - k) \quad \text{and} \quad \psi_{j,k}(x) = 2^{j/2}[\mathbb{I}_{[0,1/2)}(2^j x - k) - \mathbb{I}_{[1/2,1)}(2^j x - k)].$$

Then, the functions $\{\phi_{J,k}, \psi_{j,k} : j \ge J, k \in \mathbb{Z}\}$ form an orthonormal basis for $L^2(\mathbb{R})$, which is the well-known Haar basis of $L^2(\mathbb{R})$. Therefore, any f can be represented (in the L^2-sense) by a Haar series as

$$f = \sum_{k \in \mathbb{Z}} \alpha_{J,k}(f) \phi_{J,k} + \sum_{j=J}^{\infty} \sum_{k \in \mathbb{Z}} \beta_{j,k}(f) \psi_{j,k},$$

where $\alpha_{j,k}(f) = \int f \phi_{j,k}$ and $\beta_{j,k}(f) = \int f \psi_{j,k}$.

Now let $\mathcal{F}(\alpha, R, M)$ be the class of p.d.f's f which are uniformly bounded by some finite constant $M > 0$, and the sequence of coefficients onto the Haar basis belongs to the following Besov body:

$$\mathcal{B}_{\alpha,2,\infty}(R) = \left\{ f \mid \beta(f) = (\beta_{j,k})_{j \ge J, k \in \mathbb{Z}}, \sum_{k \in \mathbb{Z}} \beta_{j,k}^2 \le R^2 2^{-2j\alpha}, \forall j \ge J \right\}$$

for some finite constants $\alpha, R > 0$; that is, we consider the class of p.d.f's

$$\mathcal{F}(\alpha, R, M) = \{ f \mid \beta(f) \in \mathcal{B}_{\alpha,2,\infty}(R), \|f\|_\infty \le M \}. \tag{7}$$

THEOREM 2.1 *Let* $Y_1, Y_2, \ldots, Y_n$ *be i.i.d. random variables from a weighted distribution with p.d.f.* g_w *given by Equation* (1), *with weight function* w *being continuous and satisfying Assumption* 2.1. *Consider the class of p.d.f's* $\mathcal{F}(\alpha, R, M)$ *defined by Equation* (7) *with* $\alpha > 0$, *and let*

$$\mathcal{D}_n = \left\{ D \mid D \in \mathbb{N}, D \leq \frac{n^2}{\log^3(n)} \right\}. \tag{8}$$

There exists some constant $\kappa_0 > 0$ *such that if* $\mathrm{pen}_u(D)$ *is given by Equation* (7) *for all* $D \in \mathcal{D}_n$ *with* $\kappa \geq \kappa_0$ *then, there exists some* $n_0 := n_0(\alpha, R, M, w)$ *such that* $\hat{\theta}$, *given by Equation* (6), *satisfies the following inequalities for* $n \geq n_0$:

- *for* $\alpha > 1/4$,

$$\sup_{f \in \mathcal{F}(\alpha, R, M)} \mathbb{E}(\hat{\theta} - \theta)^2 \leq \frac{C(\alpha, R, M, w)}{n},$$

- *for* $0 < \alpha \leq 1/4$,

$$\sup_{f \in \mathcal{F}(\alpha, R, M)} \mathbb{E}(\hat{\theta} - \theta)^2 \leq C(\alpha, R, M, w) \left(\frac{\sqrt{\log(nR^2)}}{n} \right)^{8\alpha/(1+4\alpha)}.$$

Remark 2.1 Theorem 2.1 gives a uniform bound of the mean squared error (MSE) of $\hat{\theta}$, leading to the conclusion that $\hat{\theta}$ is an adaptive and $\sqrt{n}$-consistent estimator of $\theta = \int f^2$, uniformly over $\mathcal{F}(\alpha, R, M)$ with $\alpha > 1/4$, and it also achieves the minimax rate of convergence $(\sqrt{\log n}/n)^{4\alpha/(1+4\alpha)}$ which is optimal (in the minimax sense) in the case of direct data when $0 < \alpha \leq 1/4$. The fact that the minimax rate of convergence that is optimal in the case of direct data can also be attained in the presence of selection bias is consistent with analogous results for density estimation (see, e.g., [8,9,12]) and distribution estimation (see, e.g., [13]).

Remark 2.2 The estimator $\hat{\theta}$ can be used in tests of the null hypothesis $H_0 : f = f_0$, based on the L^2-distance in order to estimate $\int (f - f_0)^2$, in the case of weighted data. More precisely, under the assumptions of Theorem 2.1, the L^2-distance of f and f_0 can be estimated by $\hat{\theta} - \int f_0^2 - 2\hat{\mu}_w \sum_{i=1}^n f_0(Y_i)/w(Y_i)$ at the minimax rate of convergence that is optimal in the case of direct data.

Remark 2.3 The simple projection estimator $\hat{\theta}_n$ (i.e. $\hat{\theta}_D$ given in Equation (5) with $D = n$) can be shown to be uniformly $\sqrt{n}$-consistent for all $f \in \mathcal{F}(\alpha, R, M)$ with $\alpha \geq 1/4$. On the other hand, in addition, the penalized estimator given by Equation (6) also attains the minimax rate of convergence that is optimal in the case of direct data for all $f \in \mathcal{F}(\alpha, R, M)$ with $0 < \alpha \leq 1/4$. Furthermore, it performs better in finite sample situations, as it will be illustrated in Section 3.

Remark 2.4 If $\beta(f) \in \mathcal{B}_{\alpha,2,\infty}(R)$ with $\alpha > 1/2$, then f is uniformly bounded and the restriction $\|f\|_\infty \leq M$ is not needed in the definition of $\mathcal{F}(\alpha, R, M)$ (see, e.g., inequality (8.15) of Proposition 8.3 in [28]).

Remark 2.5 An upper bound of the constant κ_0 can be found using the explicit constants given in [29]. In practice, a simulation study, analogous to the one used by Birgé and Rozenholc [30] in density estimation, is needed to calibrate κ_0.

Remark 2.6 The assumption $0 < w_1 \leq w(y) \leq w_2 < \infty$, $y \in [0, 1]$, is very common in density estimation for weighted data (see, e.g., [9,12,24]). The only difference when compared with

Assumption 2.1 considered above is that we require $0 < w_1 \leq w(y) \leq w_2 < \infty$ for all y, in order to cover the case of densities with non-compact support. In fact, it is sufficient to require $0 < w_1 \leq w(y) \leq w_2 < \infty$ for all y in the support of the density f. Below, we report some examples that arise in practical settings leading to weighted data with weight function satisfying Assumption 2.1.

(i) Let $1 - w(y)$ be a proportion of the frequency of the variable X that is missing (see, e.g., [12]). Then, weighted data from model (1) arise. Let $w(y) = 0.9y + 0.1$ for $0 \leq y \leq 1$ and $w(y) = 1$ for $y \geq 1$. The missing proportion decreases in the interval [0, 1] and remains 0 for $y > 1$. A generalization of this weight function is $w(y) = cy + b$, for $0 < y < (1 - b)/c$ and $w(y) = 1$ for $y > (1 - b)/c$, $0 < c, b < 1$, where the missing proportion decreases in the interval [0, c] and remains 0 for $y > c$.

(ii) Line transect sampling is another example where weighted data arise (see, e.g., [12]). If we are interested in estimating the abundance of plants or animals of a particular species in a given region, we can use line transects. This essentially means that an observer moves along fixed paths and includes the sighted clusters of objects of interest in the sample. It is obvious that larger clusters have a larger probability to be included in the sample. An appropriate weight function would be $w(y) = cy + b$, for $0 < y < (1 - b)/c$ and $w(y) = 1$ for $y > (1 - b)/c$, $0 < c, b < 1$.

(iii) The purpose of a photographic survey described by Patil [31] was to estimate the abundance of the deep-sea red crab. The data were analyzed using the composite weight function of the form $w(y) = (a + by)v(y, \theta)$, where the sighting function $v(y, \theta)$ represented the sighting-distance bias that is usually bounded away from zero.

(iv) In meta-analysis we study the publication-selection bias and the heterogeneity that might exist among different studies. Appropriate weight functions that have been found include (a) half-normal model $w(y) = \exp[-\beta p^2(y)]$ and (b) negative exponential model $w(y) = \exp[-\beta p(y)]$, where $p(y)$ is the P-value when the resulting test statistic takes the value y (see, e.g., [32]).

2.3. *The information bound for estimation of $\int f^2$ using weighted data*

Theorem 2.2 below provides the information bound for the problem of estimating $\theta = \int f^2$ when weighted data are available. For some finite $M > 0$ let $\mathcal{H}$ be a class of p.d.f.'s defined by

$$\mathcal{H} = \{f | f \in L^2(\mathbb{R}), \|f\|_\infty \leq M\}.$$

THEOREM 2.2   *Let f be a member of $\mathcal{H}$. Then, the information bound, $I_w(f)$, for the estimation of $\theta = \int f^2$ using a weighted sample given by Equation (1), with w satisfying Assumption 2.1, is given by*

$$I_w(f) = 4\,\mu_w \int \frac{f^3}{w} - 4\left(\int f^2\right)^2.$$

Remark 2.7   The information bound, $I_d(f)$, for the estimation of $\theta = \int f^2$ based on a direct sample (see, e.g., [19,21,26]) equals

$$I_d(f) = 4\int f^3 - 4\left(\int f^2\right)^2.$$

It is easy to see that for any uniform distribution $U(a, b)$, with $a < b$, $I_w(f)$ is no smaller than $I_d(f)$ since

$$\mu_w \int \frac{f^3}{w} = \frac{1}{(b-a)^2}d(f, w) = d(f, w)\int f^3 \geq \int f^3,$$

where $d(f, w) = \mu_w \int f/w \geq 1$, by Jensen's inequality, with equality if and only if $w \equiv 1$ (see, e.g., [13]). However, there are cases where $I_w(f)$ is (strictly) smaller than $I_d(f)$. For example, let $w(y) = 1 - 0.9y$ for all $y \in (0, 1)$. Let f be the p.d.f. of a Beta distribution with parameters $\alpha = 1$ and $\beta = 3$. Then, using numerical integration (performed in R, version 2.4.0), or by direct calculations, we can compute

$$\mu_w \int \frac{f^3}{w} = 3.4209404 \quad \text{and} \quad \int f^3 = \frac{27}{7},$$

thus concluding that $I_w(f)$ is smaller than $I_d(f)$. The above observations lead to the conclusion that model sampling in the presence of selection bias can either improve or worsen the information bound in the problem of estimating $\theta = \int f^2$. Analogous conclusions regarding density estimation based on weighted data can be found in [4,13].

Remark 2.8 Theorem 2.2 has the following implication. If an estimator, say T_n, of θ based on a weighted sample given by Equation (1), with w satisfying Assumption 2.1 and f belonging to $\mathcal{H}$, satisfies

$$\sqrt{n}(T_n - \theta) \longrightarrow N(0, I_w(f)) \text{ in distribution} \quad \text{and} \quad \lim_{n \to \infty} n\mathbb{E}(T_n - \theta)^2 = I_w(f),$$

then T_n is asymptotically efficient (see, e.g., [21]).

3. Simulations

We present a small simulation study to illustrate the usefulness of the proposed estimator in finite sample situations. We use the weight function

$$w(y) = \begin{cases} 1.10^{-40} & \text{if } y < 1.10^{-40}, \\ y & \text{if } 1.10^{-40} \leq y \leq 40, \\ 40 & \text{if } y > 40, \end{cases}$$
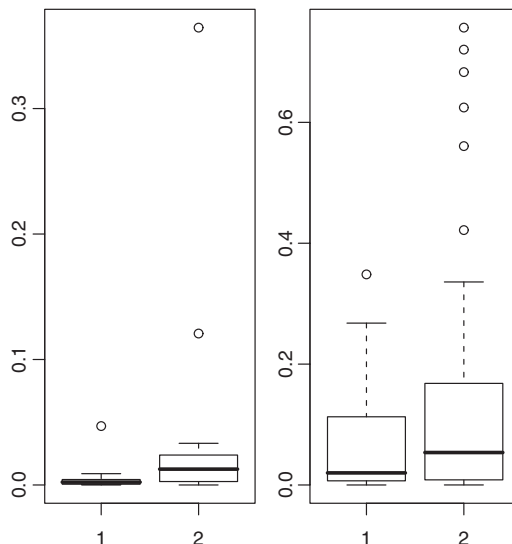

Figure 1. MSE over 50 replications of a weighted sample of size $n = 50$ generated as in Cases(I) and (II).

and five different distributions, i.e., (I) χ^2-distribution with 3 degrees of freedom, (II) Beta distribution with parameters $\alpha = 3$ and $\beta = 1$, (III) Beta distribution with parameters $\alpha = 5$ and $\beta = 4$, (IV) Beta distribution with parameters $\alpha = 5$ and $\beta = 2$, and (V) Gamma distribution with parameters $\alpha = 3$ and $\lambda = 1$.

In each case, $M = 50$ samples of size $n = 50$ and 100 were used in order to construct the boxplots of the MSE. For brevity, we just present the boxplots of MSE for $n = 50$. For the proposed estimator we set $\kappa = 2$. In Figures 1 and 2, we compare the proposed estimator with a simple projection estimator described in Remark 2.3. The boxplot on the right represents the MSE of the projection estimator, while the boxplot on the left represents the MSE of the proposed estimator. Obviously, the proposed estimator outperforms the projection estimator in all cases.
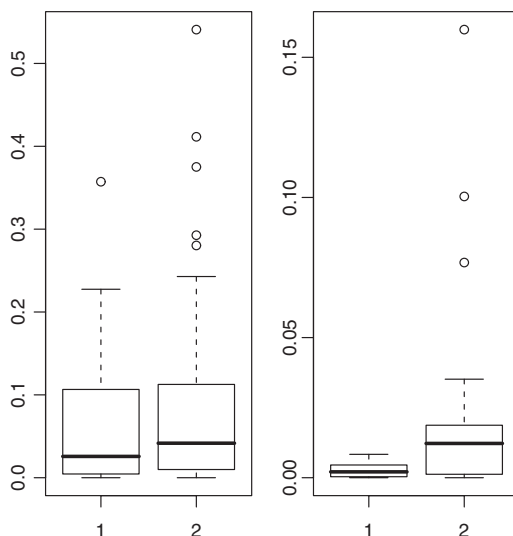

Figure 2. MSE over 50 replications of a weighted sample of size $n = 50$ generated as in Cases (III) and (V).
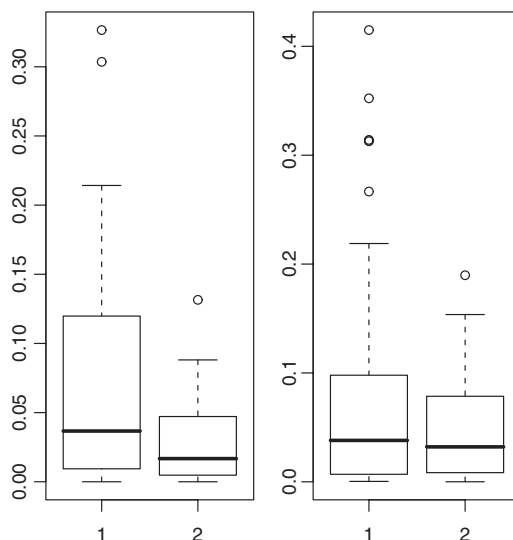

Figure 3. MSE over 50 replications of a weighted sample of size $n = 50$ generated as in Cases (II) and (IV).

Although not reported here, the proposed estimator is still better than the projection estimator for larger sample sizes. In Figure 3, we compare the MSE of pseudoestimator (3) with the MSE of the proposed estimator, in Cases (II) and (IV). The boxplot on the right represents the MSE of the pseudoestimator, while the boxplot on the left represents the MSE of the proposed estimator. Obviously, as it is expected, the estimation of μ_w deteriorates the quality of the estimator.

Acknowledgements

References

[1] G.P. Patil, C.R. Rao, and M. Zelen, *Weighted distributions*, in *Encyclopedia of Statistical Sciences*, S. Kotz and N.L. Johnson, eds., (1988), Wiley, New York, Vol. 6, pp. 664–667.

[2] Y. Vardi, *Nonparametric estimation in the presence of length bias*, Ann. Statist. 10 (1982), pp. 616–620.

[3] Y. Vardi, *Empirical distribution in selection bias model (with discussion)*, Ann. Statist. 13 (1985), pp. 178–205.

[4] D.R. Cox, *Some sampling problems in technology*, in *New Developments in Survey Sampling*, N.L. Johnson and H. Smith, eds., Wiley, New York, 1969, pp. 506–527.

[5] B.B. Bhattacharyya, L.A. Franklin, and G.D. Richardson, *A comparison of nonparametric unweighted and length-biased density estimation of fibres*, Commun. Stat. Ser. A 17 (1988), pp. 3629–3644.

[6] M.C. Jones, *Kernel density estimation for length biased data*, Biometrika 78 (1991), pp. 511–519.

[7] I.A. Ahmad, *On multivariate kernel estimation for samples from weighted distributions*, Statist. Probab. Lett. 22, (1995), pp. 121–129.

[8] C.O. Wu, *Minimax kernel density estimators with length biased data*, Math. Methods Statist. 4 (1995), pp. 56–80.

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

[10] M.C. Jones and R.J. Karunamuni, *Fourier series estimation for length biased data*, Aust. J. Stat. 39 (1997), pp. 57–68.

[11] H. El Barmi and J.S. Simonoff, *Transformation-based density estimation for weighted distributions*, J. Nonparametr. Stat. 12 (2000), pp. 861–878.

[12] S. Efromovich, *Density estimation for biased data*, Ann. Statist. 32 (2004), pp. 1137–1161.

[13] S. Efromovich, *Distribution estimation for biased data*, J. Statist. Plann. Inference 124 (2004), pp. 1–43.

[14] J.C. Aubuchon, and T.P. Hettmansperger, *A note on the estimation of the integral $f^2(x)$*, J. Statist. Plann. Inference, 9 (1984), pp. 321–331.

[15] D.C. Draper, *Rank-based robust analysis of linear models (with discussion)*, Statist. Sci. 3, (1988), pp. 239–271.

[16] Y. Ritov and P.J. Bickel, *Achieving information bounds in non and semiparametric models*, Ann. Statist. 18 (1990), pp. 925–938.

[17] M. Fromont and B. Laurent, *Adaptive goodness of fit tests in a density model*. Ann. Statist. 34 (2006), pp. 680–720

[18] C. Butucea, *Goodness-of-fit testing and quadratic functional estimation from indirect observations*, Ann. Statist. 35 (2007), pp. 1907–1930.

[19] P.J. Bickel and Y. Ritov, *Estimating integrated squared density derivatives: Sharp best order of convergence estimates*, Sankhyā Ser. A 50 (1988), pp. 381–393.

[20] L. Birgé and P. Massart, *Estimation of integral functionals of a density*, Ann. Statist. 23 (1995), pp. 11–29.

[21] B. Laurent, *Efficient estimation of integral functionals of a density*, Ann. Statist. 24 (1996), pp. 659–681.

[22] B. Laurent, *Estimation of integral functionals of a density and its derivatives*, Bernoulli 3 (1997), pp. 181–211.

[23] B. Laurent, *Adaptive estimation of a quadratic functional of a density by model selection*, ESAIM: Probab. Stat. 9 (2005), pp. 1–18.

[24] E. Brunel, F. Comte, and A. Guilloux, *Nonparametric density estimation in the presence of bias and censoring*, Test 18 (2009), pp. 166–194.

[25] I.A. Ibragimov and R.Z. Khasminskii, *Asymptotically normal families of distributions and efficient estimation*, Ann. Statist. 19 (1991), pp. 1681–1724.

[26] J. Pfanzagl, *Contributions to a General Asymptotic Statistical Theory*, Lecture Notes in Statistics, Vol. 13, Springer-Verlag, New York, 1982.

[27] B. Laurent and P. Massart, *Adaptive estimation of a quadratic functional by model selection*, Ann. Statist. 28 (2000), pp. 1302–1338.

[28] W. Härdle, G. Kerkyacharian, D. Picard, and A. Tsybakov, *Wavelets, Approximations and Statistical Applications*, Lecture Notes in Statistics, Vol. 129, Springer-Verlag, New York, 1998.

[29] C. Houdré and P. Reynaud-Bouret, *Exponential inequalities, with constants, for U-statistics of order two*, in *Stochastic Inequalities and Applications*, E. Giné, C. Houdré, and D. Nualart, eds., Basel, Birkhäuser, 2003, pp. 55–69.

[30] L. Birgé and Y. Rozenholc, *How many bins should be put in a regular histogram*, ESAIM: Probab. Stat. 10 (2006), pp. 24–45.

[31] G.P. Patil, *Weighted distributions*, in *Encyclopedia of Environmetrics*, A.H. El-Shaarawi and W.W. Piegorsch, eds., Vol. 4, John Wiley and Sons, Chichester, 2002, pp. 2369–2377.

[32] G.P. Patil and C. Taillie, *Probing encountered data, meta analysis and weighted distributions methods*, in *Statistical Data Analysis and Inference*, Y. Dodge, ed., Elsevier, Amsterdam, 1989, pp. 317–345.

[33] A. Petsa, *Contributions to wavelet methods in nonparametric statistics*. Ph.D. Thesis, Department of Mathematics and Statistics, University of Cyprus, Cyprus, 2009.

Appendix Proofs

For the detailed proofs we refer to [33].

A.1. *Proof of Theorem* 2.1

The proof of Theorem 2.1 is broken into several parts. We first prove a lemma and three propositions which are used in the proof of Theorem 2.1.

Let

$$U_n(H_D) = \mu_w^2 \sum_{\substack{1 \leq i, j \leq n \\ i \neq j}} \frac{H_D(Y_i, Y_j)}{w(Y_i) w(Y_j)} \tag{A1}$$

and

$$H_D(x, y) = \sum_{k \in \mathbb{Z}} \left(p_{k,D}(x) - \frac{\alpha_{k,D} w(x)}{\mu_w} \right) \left(p_{k,D}(y) - \frac{\alpha_{k,D} w(y)}{\mu_w} \right). \tag{A2}$$

In order to prove Lemma A1, we use an exponential inequality for U-statistics of order 2 with constants, obtained by Houdré and Reynaud-Bouret [29].

LEMMA A.1 *Let $Y_1, Y_2, \ldots, Y_n$ be i.i.d. random variables from a weighted distribution with p.d.f. g_w given by Equation (1), where f belongs to $L^2(\mathbb{R})$ and the weight function w is continuous and satisfies Assumption 2.1. There exist some positive constants κ_1, κ_2 and κ_3 for which the following inequality holds*:

$$\mathbb{P} \left\{ \frac{1}{n(n-1)} |U_n(H_D)| > \frac{1}{n-1} \left[\kappa_1 \sqrt{D\theta_D t} + \kappa_2 \|f\|_\infty t + \frac{\kappa_3 D t^2}{n} \right] \right\} \leq 5.6 \exp(-t).$$

Proof Let

$$g(x, y) = \frac{H_D(x, y) \mu_w^2}{w(x) w(y)},$$

and let

$$A_1^2 = n(n-1) \mathbb{E}(g(Y_1, Y_2)^2),$$

$$A_2 = \sup \left\{ \left| \mathbb{E}(\sum_{i=2}^{n} \sum_{j=1}^{i-1} g(Y_i, Y_j) \alpha_i(Y_i) b_j(Y_j)) \right| : \sum_{i=2}^{n} \mathbb{E}\left(\alpha_i^2(Y_i)\right) \leq 1, \sum_{j=1}^{n-1} \mathbb{E}\left(b_j^2(Y_j)\right) \leq 1 \right\},$$

$$A_3^2 = n \sup_x \left[\mathbb{E}(g(Y_1, x)^2) \right],$$

$$A_4 = \sup_{x,y} |g(x, y)|.$$

Using arguments similar to those used by Laurent [23], it is easy to see that the following inequalities hold:

$$A_1 \leq C_1(w) \sqrt{n(n-1) D\theta_D}, \quad A_2 \leq C_2(w) n \|f\|_\infty, \tag{A3}$$

$$A_3 \leq C_3(w) \|f\|_\infty D, \quad A_4 \leq C_4(w) D, \tag{A4}$$

where $C_1(w) = \sqrt{2\left(1 + w_2^2/w_1^2\right)}$, $C_2(w) = w_3^2/w_1^3 + 1 + 2(w_2/w_1)$, $C_3(w) = \sqrt{w_3^3/w_1^3 + w_2/w_1}$ and $C_4(w) = 1 + w_2^2/w_1^2 + 2(w_2/w_1)$. Using inequalities (A3) and (A4), we can deduce from Theorem 3.4 of [29] that

$$\mathbb{P} \left\{ |U_n(H_D)| > \frac{1}{n-1} \left[\kappa_1 \sqrt{D\theta_D t} + \kappa_2 \|f\|_\infty t + \frac{\kappa_3 D t^2}{n} \right] \right\} \leq 5.6 \exp(-t),$$

where $\kappa_1 = C_1(\epsilon_0, w)$, $\kappa_2 = C_3(\epsilon_0, w) + C_2(\epsilon_0, w)$, $\kappa_3 = C_3(\epsilon_0, w) + C_4(\epsilon_0, w)$, $C_1(\epsilon_0, w) = 4(1 + \epsilon_0)^{3/2} C_1(w)$, $C_2(\epsilon_0, w) = 2n(\epsilon_0) C_2(w)$, $C_3(\epsilon_0, w) = 2\beta(\epsilon_0) C_3(w)$ and $C_4(\epsilon_0, w) = 2\gamma(\epsilon_0) C_4(w)$ and ϵ_0 is a fixed positive number. This completes the proof of Lemma A1. ∎

Proposition A1 provides a risk bound for the pseudo-estimator $\tilde{\theta}$, given by Equation (3), when M is unknown.

PROPOSITION A.1 *Let $Y_1, Y_2, \ldots, Y_n$ be i.i.d. random variables from a weighted distribution with p.d.f. g_w given by Equation (1), with weight function w being continuous and satisfying Assumption 2.1. Consider the class of functions satisfying $\|f\|_\infty \leq M$ with M unknown. Let $\theta = \int f^2$ and $\mathcal{D}_n$ be defined by Equation (8). There exists some constant $\kappa_0 > 0$ such that if $\mathrm{pen}(D)$ is given by Equation (4) for all $D \in \mathcal{D}_n$, then there exists some $n^* := n^*(\alpha, R, M, w)$ such that $\tilde{\theta}$, given by Equation (3), satisfies the following inequality for all $n \geq n^*$ and for all $\kappa \geq \kappa_0$:*

$$\mathbb{E}\left\{\tilde{\theta} - \theta - \frac{2}{n}\sum_{i=1}^n\left(\frac{f(Y_i)\mu_w}{w(Y_i)} - \theta\right)\right\}^2 \leq C(w)\inf_{D \in \mathcal{D}_n}\left[\|f_D - f\|_2^4 + \frac{D(M+1)\log(D+1)}{n^2}\right] + \frac{C(M, w)}{n^2}.$$

Proof Let

$$P_n(h_D) = \frac{1}{n}\sum_{i=1}^n \frac{h_D(Y_i)\mu_w}{w(Y_i)} - \int h_D f$$

$$= \frac{1}{n}\sum_{i=1}^n \mu_w\frac{2(f_D(Y_i) - f(Y_i))}{w(Y_i)} - \int 2(f_D - f)f,$$

and let $H_D(x, y)$ and $U_n(H_D)$ be defined by Equations (A1) and (A2), respectively. Also note that the following decomposition holds:

$$U_n(H_D) + P_n(h_D) - \int (f - f_D)^2 = \tilde{\theta}_D - \theta - \frac{2}{n}\sum_{i=1}^n\left(\frac{\mu_w f(Y_i)}{w(Y_i)} - \theta\right).$$

Let

$$V_D = U_n(H_D) + P_n(h_D) - \int (f - f_D)^2 - \mathrm{pen}(D).$$

Hence, it is easy to check that

$$\tilde{\theta} - \theta - \frac{2}{n}\sum_{i=1}^n\left(\frac{\mu_w f(Y_i)}{w(Y_i)} - \theta\right) = \sup_{D \in \mathcal{D}_n}(V_D).$$

Let $A = \{\omega \in \omega : \tilde{\theta}_D + 1/2 \geq \theta_D, \ \forall D \in \mathcal{D}_n\}$. We first obtain an upper bound for

$$\mathbb{E}\left[\left(\tilde{\theta} - \theta - \frac{2}{n}\sum_{i=1}^n\left(\frac{f(Y_i)\mu_w}{w(Y_i)} - \theta\right)\right)^2 \mathbb{I}_A\right].$$

The following inequality holds:

$$\mathbb{E}\left[|\sup_{D \in \mathcal{D}_n} V_D|^2 \mathbb{I}_A\right] \leq \sum_{D \in \mathcal{D}_n}\mathbb{E}\left[(V_D)_+^2\mathbb{I}_A\right] + \inf_{D \in \mathcal{D}_n}\mathbb{E}\left[(V_D)_-^2\right].$$

Let

$$C_0(M) = \inf\left\{D \in \mathbb{N} : \frac{D}{\log(D+1)} \geq M^2\right\}.$$

If

$$\mathrm{pen}(D) \geq u_D(\sqrt{2}y_D) + \frac{10y_D M w_2}{3nw_1}$$

holds on A, where $u_D(t) = (2\kappa_1\sqrt{D\theta_D t})/n + (2\kappa_2 Mt)/n + (2\kappa_3 Dt^2)n^2$, then the following inequality holds:

$$\mathbb{P}\left\{\left(V_D > u_D(\sqrt{2}t) + \frac{10Mtw_2}{3nw_1}\right)\bigcap A\right\} \leq 6.6\exp(-y_D - t).$$

Using arguments similar to those used by Laurent [23], we get

$$\sum_{D \in \mathcal{D}_n} \mathbb{E}((V_D)^2_+ \mathbb{I}_A) \le \sum_{D \in \mathcal{D}_n, D \le C_0(M)} \mathbb{E}((V_D)^2_+ \mathbb{I}_A) + \sum_{D \in \mathcal{D}_n, D \ge C_0(M)} \mathbb{E}((V_D)^2_+ \mathbb{I}_A) \le \frac{C(M, w)}{n^2}$$

and

$$\mathbb{E}((V_D)^2_-) \le C(w) \left[\frac{D(M+1)x_D}{n^2} + \frac{M^2}{n^2} + \|f_D - f\|^4_2 \right].$$

Now it remains to find an upper bound for

$$\mathbb{E} \left\{ \left(\tilde{\theta} - \theta - \frac{2}{n} \sum_{i=1}^n \left(\frac{f(Y_i)\mu_w}{w(Y_i)} - \theta \right) \right)^2 \mathbb{I}_{A^c} \right\}.$$

We first obtain an upper bound for $P\{A^c\}$. Following the lines of Proposition 2 in [23] and using the fact that $|\mathcal{D}_n| \le n^2/\log^3 n$, one obtains

$$\mathbb{P}(A^c) \le |\mathcal{D}_n| \frac{C(M, w)}{n^8} \le \frac{C(M, w)}{n^6} \quad \text{for all } n \ge n' := \max(n_{01}, n_{02}).$$

It is easy to see that the following inequalities hold:

$$\tilde{\theta}_D \le \frac{2Dw_2^2}{w_1^2}, \quad \text{pen}(D) \le C(w)n, \quad |\tilde{\theta}| \le C(w)n^2 \quad \text{for all } n \ge 3.$$

$$\theta = \int f^2 \le \|f\|_\infty \int f \le M,$$

$$\left| \frac{2}{n} \sum_{i=1}^n \left(\frac{f(Y_i)\mu_w}{w(Y_i)} - \theta \right) \right| \le 2M \left(1 + \frac{w_2}{w_1} \right),$$

$$\left(\tilde{\theta} - \theta - \frac{2}{n} \sum_{i=1}^n \left(\frac{f(Y_i)\mu_w}{w(Y_i)} - \theta \right) \right)^2 \le C(M, w)n^4 \quad \text{for all } n \ge 3.$$

Therefore, we arrive at

$$\mathbb{E} \left[\left(\tilde{\theta} - \theta - \frac{2}{n} \sum_{i=1}^n \left(\frac{f(Y_i)\mu_w}{w(Y_i)} - \theta \right) \right)^2 \mathbb{I}_{A^c} \right] \le C(M, w)n^4 \mathbb{P}(A^c)$$

$$\le \frac{C(M, w)}{n^2} \quad \text{for all } n \ge n^* = \max(3, n').$$

thus completing the proof of Proposition A1.   ∎

PROPOSITION A.2   *Let $Y_1, Y_2, \ldots, Y_n$ be defined as in Proposition A1. Consider the smooth class of p.d.f.'s $\mathcal{F}(\alpha, R, M)$ defined by Equation (7). Let $\tilde{\theta}$ be defined as in Proposition A1. For any $\alpha > 0$, $R > 0$ and $M > 0$, there exist some $\kappa_0 > 0$ and some integer $n^{**} := n^{**}(\alpha, R, M)$ such that the following inequality holds for all $n \ge n^{**}$ and all $\kappa \ge \kappa_0$:*

$$\sup_{f \in \mathcal{F}(\alpha, R, M)} \mathbb{E} \left\{ \tilde{\theta} - \theta - \frac{2}{n} \sum_{i=1}^n \left(\frac{f(Y_i)\mu_w}{w(Y_i)} - \theta \right) \right\}^2 \le C(\alpha, w)(RM^\alpha)^{4/(1+4\alpha)} \left(\frac{\sqrt{\log(nR^2)}}{n} \right)^{8\alpha/(1+4\alpha)}.$$

Furthermore, for all $\alpha > 0$, $R > 0$ and $M > 0$, there exists some integer $n_1 := n_1(\alpha, R, M)$ such that the following inequality holds for all $n \ge n_1$:

$$\sup_{f \in \mathcal{F}(\alpha, R, M)} \mathbb{E}(\tilde{\theta} - \theta)^2 \le C(\alpha, w)(RM^\alpha)^{4/(1+4\alpha)} \left(\frac{\sqrt{\log(nR^2)}}{n} \right)^{8\alpha/(1+4\alpha)} + C(w) \frac{M^2}{n}.$$

Proof   Let

$$J_n = \left\lfloor \log_2 \left(\frac{n^2 R^4}{(M+1)\log(nR^2)} \right)^{1/(1+4\alpha)} \right\rfloor + 1.$$

Following the steps of the proof of Theorem 1 in [23], one can easily complete the proof of Proposition A2.   ∎

Proposition A3 shows that $\mathbb{E}\left(\sup_{D \in \mathcal{D}_n}(\hat{\theta}_D - \theta_D)\right)^4$ is uniformly bounded for all $f \in \mathcal{F}(\alpha, R, M)$.

PROPOSITION A.3   *Let $Y_1, Y_2, \ldots, Y_n$ and the smooth class of p.d.f's $\mathcal{F}(\alpha, R, M)$ be defined as in Proposition A2. $\tilde{\theta}_D$ is given by Equation* (2). *Then the following inequality holds:*

$$\mathbb{E}\left(\sup_{D \in \mathcal{D}_n}(\tilde{\theta}_D - \theta_D)\right)^4 \le C(M, w, R, \alpha).$$

Proof   It is easy to see that the following inequalities hold:

$$\mathbb{E}\left(|\sup_{D \in \mathcal{D}_n}(\tilde{\theta}_D - 2\theta_D)|^4\right) = \mathbb{E}\left\{|\sup_{D \in \mathcal{D}_n}(\tilde{\theta}_D - 2\theta_D)|^4 \mathbb{I}_F\right\} + \mathbb{E}\left\{|\sup_{D \in \mathcal{D}_n}(\tilde{\theta}_D - 2\theta_D)|^4 \mathbb{I}_{F^c}\right\}$$

$$\le c\sum_{D \in \mathcal{D}_n}\mathbb{E}((V_{D2})_+^4) + c\inf_{D \in \mathcal{D}_n}\mathbb{E}((V_{D2})_-^4) + \inf_{D \in \mathcal{D}_n}\mathbb{E}((V_{D1})_-^4)$$

$$+ c\left\{\sup_{D \in \mathcal{D}_n}\left[u_D(\sqrt{2}y_D) + \frac{8My_Dw_2}{3w_1n}\right]\right\}^4,$$

where $F = \{\tilde{\theta}_D - 2\theta_D \ge 0 \text{ for some } D \in D_n\}$, $V_{D1} = \tilde{\theta}_D - 2\theta_D$, $V_{D2} = \tilde{\theta}_D - 2\theta_D - u_D(\sqrt{2}y_D) - 8My_Dw_2/3w_1n$ and $y_D = 4\log(D + 1)$. Additionally, using

$$\mathbb{P}\left(P_n(2f_D) - \|f_D\|^2 > \frac{8Mw_2(t + y_D)}{3w_1n}\right) \le \exp(-t - y_D)$$

and

$$\mathbb{P}\left(V_{D2} > u_D(\sqrt{2}t) + \frac{8Mw_2t}{3w_1n}\right) \le 6.6\exp(-t - y_D)$$

we get

$$\mathbb{E}\left\{(V_{D2})_+^4\right\} \le C(M, w)\exp(-y_D)\left[\frac{D^4}{n^8} + \frac{D^2}{n^4} + \frac{1}{n^4}\right],$$

and

$$\sum_{D \in \mathcal{D}_n}\mathbb{E}\left\{(V_{D2})_+^4\right\} \le C(M, w). \tag{A5}$$

Now,

$$\mathbb{E}\left((V_{D1})_-^4\right) \le c\mathbb{E}\left((U_n(H_D)^4\right) + c\mathbb{E}\left((P_n(2f_D))^4\right) + c\|f_D\|_2^8,$$

$$\mathbb{E}\left((V_{D2})_-^4\right) \le c\mathbb{E}\left((U_n(H_D)^4\right) + c\mathbb{E}\left((P_n(2f_D))^4\right) + c\|f_D\|_2^8 + c\left(u_D(\sqrt{2}y_D) + \frac{8My_Dw_2}{3nw_1}\right)^4.$$

Using the inequalities

$$\mathbb{E}\left((U_n(H_D)^4\right) \le C(M, w)\left[\frac{D^2}{n^4} + \frac{D^4}{n^8} + \frac{1}{n^4}\right],$$

$$\mathbb{E}\left((P_n(2f_D))^4\right) \le \frac{C(M, w)}{n^2},$$

$$\|f_D\|_2^8 \le M^8, \quad \text{and}$$

$$\left\{\sup_{D \in \mathcal{D}_n}\left[u_D(\sqrt{2}y_D) + \frac{8My_Dw_2}{3w_1n}\right]\right\}^4 \le C(M, w), \tag{A6}$$

we arrive at

$$\inf_{D \in \mathcal{D}_n}\mathbb{E}\left((V_{D1})_-^4\right) + c\inf_{D \in \mathcal{D}_n}\mathbb{E}\left((V_{D2})_-^4\right) \le C(M, w). \tag{A7}$$

Finally, inequalities (A5)–(A7) complete the proof of Proposition A3.

Now we are in a position to prove Theorem 2.1. By using the elementary inequality $(a + b)^2 \leq 2a^2 + 2b^2$ and splitting on complementary events B and B^c, where

$$B = \left\{ \omega \mid \sup_{D \in \mathcal{D}_n} \left(\tilde{\theta}_D - \mathrm{pen}(D) \right) \geq \sup_{D \in \mathcal{D}_n} \left(\hat{\theta}_D - \mathrm{pen}_u(D) \right) \right\},$$

one can check that

$$\mathbb{E}(\tilde{\theta} - \hat{\theta})^2 \leq 4\mathbb{E} \left[\sup_{D \in \mathcal{D}_n} (\tilde{\theta}_D - \hat{\theta}_D) \right]^2 + 4\mathbb{E} \left[\sup_{D \in \mathcal{D}_n} (\mathrm{pen}_u(D) - \mathrm{pen}(D)) \right]^2$$

$$+ 4\mathbb{E} \left[\sup_{D \in \mathcal{D}_n} (\hat{\theta}_D - \tilde{\theta}_D) \right]^2 + 4\mathbb{E} \left[\sup_{D \in \mathcal{D}_n} (\mathrm{pen}(D) - \mathrm{pen}_u(D)) \right]^2. \tag{A8}$$

Additionally, using the Cauchy–Schwartz inequality, one arrives at

$$\mathbb{E} \left\{ \left[\sup_{D \in \mathcal{D}_n} (\hat{\theta}_D - \tilde{\theta}_D) \right]^2 \right\} = \mathbb{E} \left\{ \frac{(\hat{\mu}_w^2 - \mu_w^2)^2}{\mu_w^4} \left[\sup_{D \in \mathcal{D}_n} (\tilde{\theta}_D) \right]^2 \right\}$$

$$\leq C(w) \mathbb{E} \left\{ (\hat{\mu}_w - \mu_w)^2 \left[\sup_{D \in \mathcal{D}_n} (\tilde{\theta}_D - 2\theta_D) \right]^2 \right\} + C(w, M) \mathbb{E}(\hat{\mu}_w - \mu_w)^2$$

$$\leq C(w) \sqrt{\mathbb{E}(\hat{\mu}_w - \mu_w)^4} \sqrt{\mathbb{E} \left\{ \sup_{D \in \mathcal{D}_n} (\tilde{\theta}_D - 2\theta_D) \right\}^4} + C(w, M) \mathbb{E}(\hat{\mu}_w - \mu_w)^2.$$

Similarly, we can show that

$$\mathbb{E} \left[\sup_{D \in \mathcal{D}_n} (\tilde{\theta}_D - \hat{\theta}_D) \right]^2 \leq C(w) \sqrt{\mathbb{E}(\hat{\mu}_w - \mu_w)^4} \sqrt{\mathbb{E} \left\{ \sup_{D \in \mathcal{D}_n} (\tilde{\theta}_D - 2\theta_D) \right\}^4} + C(w, M) \mathbb{E}(\hat{\mu}_w - \mu_w)^2. \tag{A9}$$

Moreover, one can easily see that

$$\mathbb{E} \left[\sup_{D \in \mathcal{D}_n} (\mathrm{pen}(D) - \mathrm{pen}_u(D)) \right]^2 \leq C\kappa^2 \mathbb{E} \left[\sup_{D \in \mathcal{D}_n} (\tilde{\theta}_D - \hat{\theta}_D) \right]^2 \tag{A10}$$

and

$$\mathbb{E} \left[\sup_{D \in \mathcal{D}_n} (\mathrm{pen}_u(D) - \mathrm{pen}(D)) \right]^2 \leq C\kappa^2 \mathbb{E} \left[\sup_{D \in \mathcal{D}_n} (\hat{\theta}_D - \tilde{\theta}_D) \right]^2. \tag{A11}$$

It is also easy to check that

$$\mathbb{E}(\hat{\mu}_w - \mu_w)^2 \leq \frac{C(w)}{n} \quad \text{and} \quad \mathbb{E}(\hat{\mu}_w - \mu_w)^4 \leq \frac{C(w)}{n^2} \tag{A12}$$

(see, e.g., [13,24]). Inequalities (A8)–(A12), together with Proposition A3, lead to

$$\mathbb{E}(\hat{\theta} - \tilde{\theta})^2 \leq \frac{C(w)}{n} \quad \text{for all} \quad n \geq n_0 := \max(n^*, 3)$$

which, together with Proposition A2, completes the proof of Theorem 2.1. ∎

A.2. *Proof of Theorem 2.2*

Let g_ν be a sequence of p.d.f.'s such that $\|g_\nu - g_0\|_2 \longrightarrow 0$ as $n \longrightarrow \infty$, where $g_0 = wf_0/\mu_0$ and $\mu_0 = \int f_0 w$. Let $\mu_w = \int f w$. We are going to determine the Fréchet derivative of the functional $\theta = \int g^2 \mu_w^2/w^2$ at a point g_0, where

$g_0 = wf_0/\mu_0$ with f_0 belonging to the class of p.d.f's $\mathcal{H}$. It is easy to see that the following equalities hold:

$$\theta(g_\nu) = \int \frac{\mu_0^2 g_\nu^2}{w^2} = \int \frac{\mu_0^2 g_0^2}{w^2} + \int \frac{\mu_0^2(g_\nu - g_0)^2}{w^2} + 2\int \frac{\mu_0^2 g_0(g_\nu - g_0)}{w^2} \tag{A13}$$

and

$$\int \frac{\mu_0^2 g_0(g_\nu - g_0)}{w^2} = \int \frac{\mu_0^2 g_0 g_\nu}{w^2} - \theta(g_0) = \int g_\nu \left[\frac{\mu_0^2 g_0}{w^2} - \theta(g_0)\right]. \tag{A14}$$

Additionally, one observes that the following inequality holds:

$$\int \frac{\mu_0^2(g_\nu - g_0)^2}{w^2} \le \frac{w_2^2}{w_1^2} \|g_\nu - g_0\|_2^2 = o(\|g_\nu - g_0\|_2). \tag{A15}$$

Using (A13)–(A15), one obtains that

$$\theta(g_\nu) = \theta(g_0) + 2\int g_\nu \left[\frac{\mu_0^2 g_0}{w^2} - \theta(g_0)\right] + o(\|g_\nu - g_0\|_2)$$

$$= \theta(g_0) + 2\int (g_\nu - g_0) \left[\frac{\mu_0^2 g_0}{w^2} - \theta(g_0)\right] + o(\|g_\nu - g_0\|_2).$$

Therefore, the Fréchet derivative is given by $\theta^{'}(g_0) = 2[\mu_0^2 g_0/w^2 - \theta(g_0)]$. In what follows, $\langle \cdot, \cdot \rangle$ denotes the scalar product in $L^2(\mathbb{R})$. Following Ibragimov and Khasminskii [25], we consider the space orthogonal to the square root of the likelihood $s_0 = \sqrt{g_0}$, i.e.,

$$H = \left\{k \in L^2(\mathbb{R}) : \int ks_0 = 0\right\},$$

and the projection operator onto this space, i.e.,

$$P_H(t) = t - \left(\int ts_0\right) s_0.$$

Since $Y_1, Y_2, \ldots, Y_n$ are i.i.d. random variables, the family $\{P_g^n\}$ is locally asymptotically Gaussian at all points $g = wf/\mu_w$ with f belonging to $\mathcal{H}$, in the direction $H(g_0)$ with normalizing factor $A_n(g_0)$, where $A_n(g) = (1/\sqrt{n})(\sqrt{g_0})g$ (see, e.g., Example 2.2 of [25]). Let $K_n = \sqrt{n}\theta^{'}(g_0)A_n P_{H(g_0)}$, where $\theta^{'}(g_0) = 2[\mu_0^2 g_0/w^2 - \theta(g_0)]$.

Then

$$K_n(k) = K(k) = \int 2s_0 k \left[\frac{\mu_0^2 g_0}{w^2} - \theta(g_0)\right] - 2\int ks_0 \int g_0 \left[\frac{\mu_0^2 g_0}{w^2} - \theta(g_0)\right]$$

$$= \int k \left\{2s_0 \left[\frac{\mu_0^2 g_0}{w^2} - \theta(g_0)\right] - 2s_0 \int g_0 \left[\frac{\mu_0^2 g_0}{w^2} - \theta(g_0)\right]\right\}.$$

Therefore, $K_n(k) \longrightarrow K(k)$ weakly, where $K(k) = \langle h, k \rangle$ and

$$h = 2s_0 \left[\frac{\mu_0^2 g_0}{w^2} - \theta(g_0)\right] - 2s_0 \int g_0 \left[\frac{\mu_0^2 g_0}{w^2} - \theta(g_0)\right].$$

According to Theorem 4.1 of [25], for any estimator of θ, say T_n, and for any family of vicinities of g_0, say $\{V(g_0)\}$, we have

$$\inf_{V(g_0)} \liminf_{n \longrightarrow \infty} \sup_{g \in V(g_0)} n\,\mathbb{E}(T_n - \theta)^2 \ge \|h\|_2^2.$$

Hence, the information bound is given by

$$I_w(f) := \|h\|_2^2 = 4\int g_0 \left\{\frac{\mu_0^2 g_0}{w^2} - \theta(g_0) - \int g_0 \left[\frac{\mu_0^2 g_0}{w^2} - \theta(g_0)\right]\right\}^2$$

$$= 4\int \frac{f_0^3 \mu_0}{w} + 4\theta^2(g_0) - 8\theta(g_0)\int \frac{g_0^2 \mu_0^2}{w^2}$$

$$= 4\mu_0 \int \frac{f_0^3}{w} - 4\left(\int f_0^2\right)^2,$$

thus completing the proof of Theorem 2.2. ∎