𝔓𝔞𝔨𝔦𝔰𝔱𝔞𝔫 𝔍𝔬𝔲𝔯𝔫𝔞𝔩 𝔬𝔣 𝔖𝔱𝔞𝔱𝔦𝔰𝔱𝔦𝔠𝔰 𝔞𝔫𝔡 𝔒𝔭𝔢𝔯𝔞𝔱𝔦𝔬𝔫 ℜ𝔢𝔰𝔢𝔞𝔯𝔠𝔥

On Smoothed MWSD Estimation of Mixing Proportion

Satish Konda[1*], Mehra, K.L.[2], Ramakrishnaiah, Y.S.[3]

* Corresponding Author

1. Department of Statistics, YSR Vivekananda Government Degree College, Kadapa, Andhra Pradesh, India.
statishlaks@gmail.com , satishkonda1999@ysrvgdcvempalli.ac.in
2. Department of Mathematical and Statistical Sciences, University of Alberta, Canada. kmehra@ualberta.ca
3. Department of Statistics, Osmania University, India. ysrkou@gmail.com

Abstract

The problem considered in the present paper is that of smooth (nonparametric) estimation of mixing proportions in a mixture population of two (known) distributions represented by F(x) = pF_1(x) + (1-p)F_2(x), $-\infty < x < +\infty$, $0 < p < 1$. The two classes of 'smoothed' and 'unsmoothed' estimates studied in the paper are based on an iid sample of n observations from the mixture population, using the minimum weighted square distance (MWSD) methodology due to Wolfowitz (1953). Comparison of estimators is done based on their relative mean square errors (MSE's). The superiority of smoothed estimators over their corresponding unsmoothed counterparts is established theoretically as well as by conducting a small Monte-Carlo study that compares their resulting MSE's. Large sample properties such as a.s. rates of their convergence and asymptotic normality etc. are also established. The results proved are new in the literature.

Key Words: mixture of distributions, mixing proportions, smoothed MWSD estimation, mean square error, optimal band width

1. Introduction

Let $\{X_1, X_2, \ldots, X_n\}$ be independent observations from a mixture population represented by the mixture distribution function (d.f.)

$$F(x) = \sum_{v=1}^{m} p_v F_v(x), \quad -\infty < x < \infty, \tag{1.1}$$

a finite linear mixture of m d.f.s F_v, $v=1,2,\ldots,m$ with mixing proportions p_v's satisfying $0 < p_v < 1$ and $\sum_{v=1}^{m} p_v = 1$. In any mixture analysis, the main objective is to estimate as effectively and accurately as possible, the mixing proportions p_v, $v=1,2,\ldots,m$, along with possibly other parameters which may also be of interest or may simply be in the underlying model as nuisance parameters. This depends on the extent to which we have knowledge and are willing to make assumptions regarding the nature of component distributions. In situations where one can legitimately assume the component distributions to belong to a parametric family of densities or discrete distributions that depend on a finite number of real parameters, Maximum Likelihood (MLE) or Minimum Distance Estimation (MDE) methodologies (including that of the Minimum Hellinger Distance (MHD)) have been utilized in statistical literature to derive highly efficient estimates of the mixing proportions and other parameters of interest, (see Day(1969), Wolfe (1970), Hosmer (1973), Hall and Titterington (1984), Woodward et.al.(1984), Titterington, Smith and Makov (1985) respectively, Choi and Bulgren (1968), Choi (1969), McDonald (1971), Clarke and Heathcote (1994), Woodward, Whitney and Eslinger (1995) etc.).

In the nonparametric situation also, with little information about the component distributions beyond that they possess densities, Wu and Karunamuni (2009) employing Rosenblat-Parzen type density estimates coupled with the MHD estimation technique have derived asymptotically fully efficient estimate of mixing proportions (see also Wu and Karunamuni (2009) and Tang and Karunamuni (2013)).

However, when dealing with (continuous) mixture populations where one cannot legitimately or formally assume that the component distributions do possess densities or when the data recordings are available only in discrete or grouped forms, the validity of the preceding techniques becomes highly questionable. In such situations, it seems much safer to revert to the suggestion of Hall (1981) to base ones Minimum Distance Estimation procedures based on empirical distribution functions rather than the empirical density estimators.

Mixtures of distributions widely used in both biological and physical sciences. Many typical problems in which such mixtures occur have been well described in a series of research papers. Few among them are,

i.   In Oceanography, the interest of study is to measure characteristics in natural population of species (fish). So that the samples of species (fish) are taken to measure the characteristic of each species (fish) in the sample. However we can measure many characteristics such as weight, length etc. of fish which vary with age of species(fish). These characteristics have different distribution for each age group so that the population has a mixture of distributions.

ii.   In failure time analysis, it is desired to measure failure time of units in a population. For this purpose, samples of units are taken and failure time measured for each unit in the sample. However, failures are occurring due to different causes. The failure times have distinct distributions due to different causes so that the overall population has a mixture of distributions.

For more details of such examples, refer Choi and Bulgren(1968), Harris(1958), Blischke(1963), Fu(1968), Macdonald and Pitcher(1979), Odell and Basu (1976) and Bruni et al(1983) etc.

There are several methods of estimating mixing proportions discussed in the literature. Choi and Bulgren(1968) estimated the mixing measures of combination of known distributions by using Wolfowitz minimum distance method, which minimizes

$$\int (\tilde{F}_n(x)\text{-}F(x))^2 d\tilde{F}_n(x), \tag{1.2}$$

based on the usual unsmoothed standard empirical distribution function $\tilde{F}_n(x) = n^{-1}\sum_{i=1}^{n} I(X_i \leq x)$ of a random sample X_i, $1\leq i\leq n$. They investigated the asymptotic properties of their estimators such as strong consistency and asymptotic normality etc. Van Houwelingen (1974) pointed that the variance of Choi and Bulgren estimator is hard to compute and small sample properties of the estimator are difficult to evaluate. As pointed out in Hall (1981), methods based on nonparametric density estimators involve some significant drawbacks.

- specification of window width in kernel based estimators and their behavior which is very sensitive to the choice of window width parameter and also that
- their mean square errors converge at a slower rate than order n^{-1} .

To avoid these draw backs, Hall (1981) proposed the nonparametric (MWSD) estimators of mixing proportions in finite mixtures based on the usual empirical distribution function only. But did not make any attempt to derive small as well as large sample properties of these proposed estimators. In the present paper, we discuss estimators based on both the usual empirical distribution functions (e.d.f.) $\tilde{F}_n(x)$ and also the kernel based smoothed e.d.f. $\hat{F}_n(x)$ defined, respectively, by

$$\tilde{F}_n(x) = n^{-1}\sum_{i=1}^{n} I(X_i \leq x) \text{ and } \hat{F}_n(x)= n^{-1}\sum_{i=1}^{n} K(\frac{x - X_i}{a_n}), \tag{1.3}$$

$\{a_n\}$ being the smoothing bandwidth sequence satisfying $0< a_n \rightarrow 0$, $na_n \rightarrow \infty$, as $n \rightarrow \infty$, and K the distribution function corresponding to a known suitable kernel density k.

2. Smoothed Hall's Minimum Weighted Squared Distance Estimator of Mixing Proportions

In this section, we shall define and study the 'smoothed' version of Hall's (1981) Minimum Weighted squared Distance (MWSD) estimators based on smoothed e.d.f's defined below in (2.2).

The problem of estimation of mixing proportion p for the m = 2 case defined in (1.1) is studied in the present paper and the case m>2 would be considered seperately. Now, if F(x) = pF_1(x) + (1 - p) F_2(x),-∞<x<+∞, clearly we have

$$p = \frac{F(X)- F_2(X)}{F_1(X)- F_2(X)}, \quad \text{-∞<x<∞}, \tag{2.1}$$

so that if the d.f.s F_1 and F_2 were known, for each value of x one can have an estimator of p by substituting the smoothed (or unsmoothed) e.d.f. in place of F in the equation (2.1). Some suitable averaging seems to be in order to

come up with an overall appropriate estimator. This is provided by Hall's minimization of weighted squared distance between the smoothed e.d.f. $\hat{F}_n$ (or e.d.f $\tilde{F}_n$) and the mixture d.f. [F(x)=pF_1(x)+(1-p)F_2(x)], i.e., the squared distance $\int[\hat{F}_n(x)-(pF_1(x)+(1-p)F_2(x))]^2$ dW(x) w.r. to a suitable averaging d.f. W(x), -∞<x<∞.
We propose the smoothed MWSD estimator defined as follows:

Smoothed MWSD estimator $\hat{p}_n$: Smoothed MWSD estimator $\hat{p}_n$ of p based on the kernel empirical d.f. $\hat{F}_n$(x) as proposed is obtained by choosing p such that

$$\hat{\Delta}(p) = \int(\hat{F}_n(x) - F(x))^2 dW(x)$$
$$= \int[\hat{F}_n(x) - F_2(x) - p(F_1(x) - F_2(x))]^2 dW(x) \qquad (2.2)$$

is minimum, when F_j(x); j = 1,2 are known and W(x) is a suitable known weight function-discrete or continuous. Solving $\hat{\Delta}$(p) as $\frac{\partial \hat{\Delta}(p)}{\partial p} = 0$ for p and letting D = $\int d_{12}^2(x)dW(x)$, $d_{12}(x) = F_1(x) - F_2(x)$, we obtain the smoothed MWSD estimator $\hat{p}_n$ as,

$$D\hat{p}_n = \int(\hat{F}_n(x) - F_2(x))d_{12}(x)dW(x), \qquad (2.3)$$

Note that, Hall's (1981) unsmoothed version denoted by $\tilde{p}_n$ is obtained by replacing $\hat{F}_n$(x) in (2.3) by the usual e.d.f. $\tilde{F}_n$(x), so that

$$D\tilde{p}_n = \int(\tilde{F}_n(x) - F_2(x))d_{12}(x)dW(x). \qquad (2.4)$$

The object of the present investigation is to show that Hall's (1981) MWDE procedure can be improved by basing it on the smoothed e.d.f.s in place of the usual e.d.f.s. This is demonstrated below, generally speaking, for both cases when the averaging distribution W is discrete or continuous, but decidedly when W is discrete.

So far none of the researchers have attempted to compute the variance of such estimators. In the present work, the variances, exact MSEs and large sample properties such as strong consistency and asymptotic normality are established and also the superiority of smoothed estimator over the unsmoothed version in the sense of a smaller MSE. Our results are completely new in the literature.

Representations to MWSD estimators $\hat{p}_n$ and $\tilde{p}_n$
We first establish the representation to $\hat{p}_n$ in order to prove its asymptotic properties. Recall from (2.3),

$$\mathbf{D}\hat{p}_n = \int[\hat{F}_n(x) - F_2(x)]d_{12}(x)dW(x),$$

where $d_{12}(x) = F_1(x) - F_2(x)$ and from (2.1) for m=2

$$D(\hat{p}_n - p) = \int[\hat{F}_n(x) - F(x)]d_{12}(x)dW(x)$$
$$= \int[\hat{F}_n(x) - E\hat{F}_n(x) + E\hat{F}_n(x) - F(x)]d_{12}(x)dW(x)$$
$$= \int\hat{\lambda}_n(x)d_{12}(x)dW(x) + \int[E\hat{F}_n(x) - F(x)]d_{12}(x)dW(x)$$
$$=: \hat{\lambda}_n^* + C_n \qquad . \qquad (2.5)$$

Similarly, it is easy to show

$$D(\tilde{p}_n - p) = \tilde{\lambda}_n^* , \qquad (2.6)$$

where $\hat{\lambda}_n^* = \int\hat{\lambda}_n(x)d_{12}(x)dW(x)$, $\hat{\lambda}_n(x) = \hat{F}_n(x) - E\hat{F}_n(x)$ and $C_n = \int[E\hat{F}_n(x) - F(x)]d_{12}(x)dW(x)$
We shall use equation (2.5) to evaluate the MSE of the estimator $\hat{p}_n$ in both cases when the selected W(x) is a discrete or a continuous (averaging) distribution.

Lemma 2.1: Assume the following conditions on F and the kernel function k:

i. F(x) possesses at least three continuous derivatives;
ii. The kernel K satisfies $\mu_l(K) = \int t^l dK(t) = 0$, l=1,3,
    $\mu_{2l}(K) = \int t^{2l}dK(t) < \infty$, l=1,2, $\psi_j(K) = 2\int t^j K(t)dK(t) < \infty$, l=1,2;
iii. {a_n} satisfies $0 < a_n \to 0$, $na_n \to \infty$ and $na_n^3 \to 0$, as n→∞;
then

a. when W is continuous, Bias $\hat{p}_n = D^{-1} a_n^2 \xi_{n4c} + O(a_n^4)$,
    where $\xi_{n4c} = \frac{1}{2}\mu_2(K)\int_{-\infty}^{\infty} F^{(2)}(x)d_{12}(x)dW(x)$;
b. when W is discrete(see(2.11) below), Bias $\hat{p}_n = D^{-1}a_n^2\xi_{n4d} + O(a_n^4)$,
    where $\xi_{n4d} = \frac{1}{2}\mu_2(K)\sum_i F^{(2)}(x_i)d_{12}(x_i)p(x_i)$.

Proof: Recall from (2.5),

$$\text{Bias } \hat{p}_n = E\, \hat{p}_n - p = D^{-1}C_n \,.$$

Noting that $\mu_l(K) = \int t^l dK(t) = 0$ for $l=1,3$,

$$E\, K(\tfrac{x-X_i}{a_n}) = \int_{-\infty}^{\infty} K\,(\tfrac{x-u}{a_n})dF(u) = \int_{+\infty}^{-\infty} K(t)dF(x-a_n t)$$

$$= \int_{-\infty}^{\infty}[F(x)-a_n t\, F^{(1)}(x)+\tfrac{a_n^2 t^2}{2} F^{(2)}(x)-\tfrac{1}{6} a_n^3 t^3 F^{(3)}(x)+ \tfrac{1}{24} a_n^4 t^4 F^{(4)}(x)+O(a_n^5)]dK(t)$$

$$= F(x)+\tfrac{a_n^2}{2}\mu_2(K)\int_{-\infty}^{\infty} F^{(2)}(x)d_{12}(x)dW(x)+O(a_n^4), \tag{2.7}$$

so that from (2.5) and (2.7), we have

$$\text{Bias } \hat{p}_n =: D^{-1}\, a_n^2\, \xi_{n4c} + O(a_n^4), \tag{2.8}$$

where ξ_{n4c} is given in part (a) above. Hence the part (a) of Lemma follows. Similarly when the selected W(x) is a discrete d.f., then (2.8) becomes

$$\text{Bias } \hat{p}_n = D^{-1}\, a_n^2\, \xi_{n4d} + O(a_n^4),$$

where ξ_{n4d} is given in part (b) above. Hence part (b) of the Lemma is proved.

We now establish large sample MSEs of $\hat{p}_n$, $\tilde{p}_n$ and show the superiority of smoothed estimator $\hat{p}_n$ over the unsmoothed version $\tilde{p}_n$ in the sense of a lower MSE.

Variances of MWSD estimators: We first obtain the large smaple expressions for the variances of both MSWD estimators $\tilde{p}_n$ and $\hat{p}_n$ defined in (2.4) and (2.3), respectively, in the following Lemma 2.2 and Corollary 2.3.

Lemma 2.2: Under the conditions (i)-(iii) of Lemma 2.1 on F, the kernel function K and the bandwidth sequence $\{a_n\}$,

a)   when the weight W is a discrete d.f., we have

$$\sigma_{nd}^2 = \text{Var } \hat{p}_n =: D^{-2}[\tfrac{\xi_{n0d}}{n} - \tfrac{a_n}{n}\xi_{n1d} + \tfrac{a_n^2}{n}\,\xi_{n2d} + O(\tfrac{a_n^3}{n})]$$

where $\xi_{n0d}, \xi_{n1d}, \xi_{n2d}$ are defined below in (2.19);

b)   when the weight W is a continuous d.f., we have

$$\sigma_{nc}^2 = \text{Var } \hat{p}_n =: D^{-2}[\tfrac{\xi_{n0c}}{n} - \tfrac{a_n^2}{n}\,\xi_{n2c} + O(\tfrac{a_n^3}{n})]$$

where ξ_{n0c}, ξ_{n2c} are defined below in (2.35).

Proof: From Lemma 2.1 and (2.5)

$$nD^2 \text{Var}\hat{p}_n = nD^2 \text{Var}\hat{\lambda}_n^* = nD^2 \text{Var}\{\int [\,\hat{F}_n(x)-E\hat{F}_n(x)]\, d_{12}(x)dW(x)\}$$

$$= \tfrac{1}{n} E\, [\sum_{i=1}^n \int (\, K(\tfrac{x-X_i}{a_n}) - E\, K(\tfrac{x-X_i}{a_n}))\, d_{12}(x)dW(x)]^2 \,. \tag{2.9}$$

Let

$$E\, [\int \{ K(\tfrac{x-X_1}{a_n}) - E\, K(\tfrac{x-X_1}{a_n})\}d_{12}(x)dW(x)]^2 =: L_1 - L_2 \tag{2.9a}$$

To evaluate the quantities L_1, L_2, as $n \to \infty$, note that in view of (2.7), the last equation (2.9a) immediately yields

$$L_2 = [\int E[\, K(\tfrac{x-X_1}{a_n})]d_{12}(x)dW(x)]^2$$

$$= [\int \{F(x) + \tfrac{a_n^2}{2}\mu_2(K)F^{(2)}(x) + O(a_n^4)\}d_{12}(x)dW(x)]^2$$

$$= [\int F(x)d_{12}(x)dW(x)+\tfrac{a_n^2}{2}\mu_2(K)\int F^{(2)}(x)d_{12}(x)dW(x)+O(a_n^4)\}]^2$$

$$= [\int F(x)d_{12}(x)dW(x)]^2 + a_n^2\mu_2(K)\{\int F(x)d_{12}(x)dW(x)\}\{\int F^{(2)}(x)d_{12}(x)dW(x)\}+ O(a_n^4). \tag{2.10}$$

For the evaluation of L_1, we have to deal with two cases when W is discrete or of continuous type separately:

Let **W stand for a discrete d.f.** assigning positive probability $p(x_i)$ to a countable set of reals $\{x_i : x_i < x_{i+1},$ with $i \in C$, a countable set of integers with only a finite number of x_i's belonging to each compact interval. (For example, consider the Double Exponential distribution with density $f(x) = \tfrac{1}{2}e^{-|x|}$, $-\infty < x < \infty$ and sequence of discrete point sets $\{x_{nk}^* : x_{nk}^* = (\tfrac{k \mp 0.5}{n^{1/5}}),\ k=\ldots\text{-3,-2,-1,0,1,2,3}\ldots\} \subset$ Real line). $\tag{2.11}$

Then the expressions L_2 from (2.10) above, can be written as

$$L_2 = [\sum_i F(x_i)d_{12}(x_i)p(x_i)]^2 + a_n^2 \mu_2(K)[\sum_i F(x_i)d_{12}(x_i)p(x_i)][\sum_i F^{(2)}(x_i)d_{12}(x_i)p(x_i)] + O(a_n^4)$$

$$=: \xi_{nod}^{(1)} + a_n^2 \xi_{n2d}^{(1)} + O(a_n^4), \text{(say)} \tag{2.12}$$

and the expression L_1 from (2.9a) as

$$L_1 = \iint E\, K\left(\frac{x - X_1}{a_n}\right) K(\frac{y - X_1}{a_n}) d_{12}(x)\, d_{12}(y) dW(x) dW(y)$$

$$= \sum_j \sum_i E\{K\left(\frac{x_j - X_1}{a_n}\right) K(\frac{x_i - X_1}{a_n})\} d_{12}(x_j) p(x_j) d_{12}(x_i) p(x_i)$$

$$= \sum_j E\,[K^2\left(\frac{x_j - X_1}{a_n}\right) d_{12}^2(x_j) p^2(x_j) + 2\sum_j \sum_{i>j} E\{K\left(\frac{x_j - X_1}{a_n}\right) K(\frac{x_j - X_1}{a_n} + \frac{x_i - x_j}{a_n})\} d_{12}(x_j) p(x_j) d_{12}(x_i) p(x_i)$$

$$=: L_{11} + L_{12}, \text{ (say).} \tag{2.13}$$

Now in L_{11}, for large n,

$$E\,[K^2\left(\frac{x_j - X_1}{a_n}\right)] = \int K^2(t)\, dF(x_j - a_n t)$$

$$= \int_{-\infty}^{\infty} [F(x_j) - a_n t\, F^{(1)}(x_j) + \frac{a_n^2 t^2}{2} F^{(2)}(x_j) + O(a_n^3)] dK^2(t)$$

$$= F(x_j) - a_n \psi_1(K) F^{(1)}(x_j) + \frac{a_n^2}{2} \psi_2(K) F^{(2)}(x_j) + O(a_n^3), \tag{2.14}$$

with $\psi_l(K) = 2\int t^l\, K(t) dK(t)$, $l = 1,2$ and in L_{12} for j<i $\Leftrightarrow x_j < x_i$ and sufficiently large n, so that $(x_i - x_j) > 2a_n$,

$$E\,\{K\left(\frac{x_j - X_1}{a_n}\right) K(\frac{x_i - X_1}{a_n})\} = \int K(t) K\left(\frac{x_i - x_j}{a_n} + t\right) dF(x_j - a_n t)$$

$$= \int K(t)\, dF(x_j - a_n t)$$

$$= F(x_j) + \frac{a_n^2}{2} \mu_2(K) F^{(2)}(x_j) + O(a_n^4), \tag{2.15}$$

the last but one equality in (2.15) above following since for $(x_i - x_j) > 2a_n$, $K(\frac{x_i - x_j}{a_n} + t) \equiv 1$ identically. From (2.10), (2.14) and (2.15), it therefore follows that

$$L_{11} = \sum_j F(x_j) d_{12}^2(x_j) p^2(x_j) - a_n \psi_1(K) \sum_j F^{(1)}(x_j)\, d_{12}^2(x_j) p^2(x_j) + \frac{a_n^2}{2} \psi_2(K) \sum_j F^{(2)}(x_j)\, d_{12}^2(x_j) p^2(x_j)$$
$$+ O(a_n^3)$$
$$=: \xi_{nod}^{(2)} - a_n\, \xi_{n1d} + a_n^2 \xi_{n2d}^{(2)} + O(a_n^3) \qquad \text{(say),} \tag{2.16}$$

and

$$L_{12} = 2\sum_j \sum_{i>j} F(x_j) d_{12}(x_i) d_{12}(x_j) p(x_i) p(x_j) + a_n^2 \mu_2(K) \sum_j \sum_{i>j} F^{(2)}(x_j) d_{12}(x_i) d_{12}(x_j) p(x_i) p(x_j) + O(a_n^4)$$
$$=: \xi_{nod}^{(3)} + a_n^2 \xi_{n2d}^{(3)} + O(a_n^4) \qquad \text{(say).} \tag{2.17}$$

Substituting the expressions for L_2 and $L_1 = L_{11} + L_{12}$, from (2.12) and (2.13), (2.16)-(2.17), respectively, in (2.9), we obtain that, as $n \to \infty$,

$$nD^2 \text{Var}\,\hat{p}_n = [\sum_j F(x_j) d_{12}^2(x_j) p^2(x_j) + 2\sum_j \sum_{i>j} F(x_j) d_{12}(x_i) d_{12}(x_j) p(x_i) p(x_j)$$
$$- [\sum_j F(x_j) d_{12}(x_j) p(x_j)]^2 - a_n \psi_1(K) \sum_j F^{(1)}(x_j) d_{12}^2(x_j) p^2(x_j)$$
$$+ \frac{a_n^2}{2} [\psi_2(K) \sum_j F^{(2)}(x_j) d_{12}^2(x_j) p^2(x_j)$$
$$- 2\mu_2(K) [\sum_j F(x_j) d_{12}(x_j) p(x_j)][\sum_i F^{(1)}(x_i) d_{12}(x_i) p(x_i)]$$
$$+ 2\mu_2(K) \sum_j \sum_{i>j} F^{(2)}(x_j) d_{12}(x_i) d_{12}(x_j) p(x_i) p(x_j)]$$

$$= [\sum_j F(x_j)(1 - F(x_j)) d_{12}^2(x_j) p^2(x_j) + 2\sum_j \sum_{i>j} F(x_j)(1 - F(x_i)) d_{12}(x_i) d_{12}(x_j) p(x_i) p(x_j)]$$
$$- a_n \psi_1(K) \sum_j F^{(1)}(x_j) d_{12}^2(x_j) p^2(x_j) + \frac{a_n^2}{2} [\psi_2(K) \sum_j F^{(2)}(x_j) d_{12}^2(x_j) p^2(x_j)$$
$$- 2\mu_2(K) [\sum_j F(x_j) F^{(1)}(x_j) d_{12}^2(x_j) p^2(x_j)]$$
$$+ 2\mu_2(K) \sum_j \sum_{i>j} \{F^{(2)}(x_j)(1 - F(x_i)) - F^{(2)}(x_i) F(x_j)\}\, d_{12}(x_i) d_{12}(x_j) p(x_i) p(x_j)] + O(a_n^3),$$

$$= \xi_{n0d} - a_n \xi_{n1d} + a_n^2 \xi_{n2d} + O(a_n^3) \qquad ; \tag{2.18}$$

or $\sigma_{nd}^2 = \text{Var}\,\hat{p}_n =: D^{-2}[\frac{\xi_{n0d}}{n} - \frac{a_n}{n} \xi_{n1d} + \frac{a_n^2}{n} \xi_{n2d} + O(\frac{a_n^3}{n})]$, \qquad (2.19)

where, from (2.12) - (2.19),

$\xi_{n0d} = \xi_{nod}^{(1)} + \xi_{nod}^{(2)} = [\sum_j F(x_j)(1 - F(x_j)) d_{12}^2(x_j) p^2(x_j) + 2\sum_j \sum_{i>j} F(x_j)(1 - F(x_i)) d_{12}(x_i) d_{12}(x_j) p(x_i) p(x_j)]$,

$\xi_{n1d} = \psi_1(K) \sum_j F^{(1)}(x_j) d_{12}^2(x_j) p^2(x_j)$ and

$\xi_{n2d} = \xi_{n2d}^{(1)} + \xi_{n2d}^{(2)} + \xi_{n2d}^{(3)}$

$$= \frac{1}{2}[\psi_2(K) \sum_j F^{(2)}(x_j) d_{12}^2(x_j) p^2(x_j) - 2\mu_2(K)[\sum_j F(x_j) F^{(1)}(x_j) d_{12}^2(x_j) p^2(x_j)]$$
$$+ 2\mu_2(K) \sum_j \sum_{i>j} \{F^{(2)}(x_j)(1 - F(x_i)) - F^{(2)}(x_i) F(x_j)\} d_{12}(x_i) d_{12}(x_j) p(x_i) p(x_j)]. \tag{2.20}$$

Hence the part (a) of Lemma follows.

Let **W** be continuous: In the definition of the weighted (Wolfowitz) square distance if W possesses a density or is simply a **continuous d.f.**, the detailed evaluations for the MSE $(\hat{p}_n)$ show (see Theorem 2.4(b) below) that the situation is not as clear cut as in the case when W is discrete. Since in this continuous case, W assigns no weight to the terms involving sets [x=y] in integral evaluations (unlike the evaluations done for equations (2.12)-(2.19) in the discrete case), the MSE $(\hat{p}_n)$ evaluations, in the continuous d.f. W case, contain no negative term of the order a_n. Even in such a case, among the two terms of the order a_n^2 that we obtain, one is clearly negative and the other neither positive nor negative. So for large n, one may expect in most situations on the average, some reduction in the MSE $(\hat{p}_n)$ brought in by 'smoothing'.

Now we do the calculations for the continuous case: From (2.10), we already have

$$L_2 = [\int F(x)d_{12}(x)dW(x)]^2 + a_n^2 \,\mu_2(K)[F(x)d_{12}(x)dW(x)][\int F^{(2)}(y)d_{12}(y)dW(y)] + O(a_n^4), \qquad (2.21)$$

To calculate L_1, note from (2.13) that in this case,

$$L_1 = 2\iint_{y<x} E\{K(\tfrac{y-X_1}{a_n})\,K(\tfrac{x-X_1}{a_n})d_{12}(x)d_{12}(y)dW(x)dW(y), \qquad (2.22)$$

where for y<x

$$E\{K(\tfrac{y-X_1}{a_n})K(\tfrac{x-X_1}{a_n})\} = \int_{-\infty}^{\infty} K(\tfrac{y-u}{a_n})K(\tfrac{x-u}{a_n})dF(x)$$
$$= - \int_{-\infty}^{\infty} K(t)\, K(\tfrac{x-y}{a_n}+t)dF(y-a_nt)$$
$$= \int_{-\infty}^{\infty}[F(y)-a_nt F^{(1)}(y)+\tfrac{1}{2}a_n^2 t^2 F^{(2)}(y)+O(a_n^3)]d[K(\tfrac{x-y}{a_n}+t)K(t)]; \qquad (2.23)$$

substituting (2.23) in (2.22), we obtain

$$L_1 = J_1 - J_2 + J_3 + O(a_n^3), \qquad (2.24)$$

where, in view of $\int_{-\infty}^{\infty} d\,[K(\tfrac{x-y}{a_n}+t)K(t)] = [K(\tfrac{x-y}{a_n}+t)K(t)]_{-\infty}^{\infty} = 1$ for x > y,

$$J_1 = 2\iint_{y<x} F(y)d_{12}(x)d_{12}(y)dW(x)dW(y) =: \xi_{n0c} \text{ (say)}. \qquad (2.25)$$

Further, for evaluating J_2, J_3 and the later order terms note that the limits for the variable t, $-\infty<t<\infty$, are $-1\le t\le 1$ effectively. This is because for $\tfrac{x-y}{a_n} > 0$, when t>1, $K(\tfrac{x-y}{a_n}+t)K(t) \equiv 1$ identically so that $d[K(\tfrac{x-y}{a_n}+t)K(t)]=0$ for t>1; again for t<-1, $[K(\tfrac{x-y}{a_n}+t)K(t)]=0$, so that $d[K(\tfrac{x-y}{a_n}+t)K(t)]=0$ also for t<-1. Hence the effective limits for the variable t are given by $-1\le t\le 1$. Accordingly,

$$J_2=2a_n \iint_{y<x} F^{(1)}(y)\int_{-1}^{1} t\,d[K(t)K(\tfrac{x-y}{a_n}+t)]d_{12}(x)d_{12}(y)dW(x)dW(y)$$
$$=2a_n \int F^{(1)}(y)[\int_{y<x\le y+2a_n}\{\int_{-1}^{1} t\,d[K(\tfrac{x-y}{a_n}+t)K(t)]\}d_{12}(x)dW(x)]d_{12}(y)dW(y)$$
$$\quad +2a_n \int F^{(1)}(y)[\int_{y+2a_n<x}\{\int_{-1}^{1} t\,d[K(\tfrac{x-y}{a_n}+t)K(t)]\}d_{12}(x)dW(x)]d_{12}(y)dW(y)$$
$$= J_{21} + J_{22}, \qquad (2.26)$$

with $J_{22} = 0$; this is so since for evaluating J_{22}, the range of integration for x, (viz., $y+2a_n<x<\infty$) is such that $(\tfrac{x-y}{a_n})+t > 2 + t > 1$ (since $t \ge -1$), so that $K(\tfrac{x-y}{a_n} + t) \equiv 1$ identically in this range. This leads to $\int_{-1}^{1} t\,d[K(\tfrac{x-y}{a_n}+t)K(t) = \int_{-1}^{1} t\,dK(t) = 0$, which renders $J_{22} = 0$. To evaluate J_{21}, set $\tfrac{x-y}{a_n} = u$, so that the limits for u in the evaluation of J_{21} are $0\le u\le 2$. Thus,

$$J_{21} = 2a_n \int F^{(1)}(y)[\int_0^2\{t\,d[K(t+u)K(t)]d_{12}(y+a_nu)dW(y+a_nu)]d_{12}(y)dW(y)$$
$$= 2a_n^2 \int F^{(1)}(y) \int_{-1}^{1} t\, d[K(t)\int_0^2 K(u+t)du]d_{12}{}^2(y)W'(y)dW(y)+O(a_n^3)$$
$$=: a_n^2 \xi_{n2c}^{(1)} + O(a_n^3). \text{ (say)} \qquad (2.27)$$

To evaluate (2.27), note that for each t, $-1<t<1$,

$$\int_0^2 K(u+t)du = [u\,K(u+t)du]_0^2 - \int_0^2 K^{(1)}(u+t)du$$
$$= 2 - \int_t^1 (v-t)K^{(1)}(v)dv = 2 + E(t) + t(1-K(t)), \qquad (2.28)$$

where $E(t) = \int_{-1}^{t} vK^{(1)}(v)dv$ (so that E(1)=0) and therefore,

$$\int_{-1}^{1} t\, d\{K(t)\int_0^2 K(u+t)du\} = \int_{-1}^{1} t\,(\int_0^2 K(u+t)du)dK(t) + \int_{-1}^{1} t\,K(t)d(\int_0^2 K(u+t)du)$$
$$=\int_{-1}^{1} t\,[2+E(t) + t(1-K(t))]dK(t) + \int_{-1}^{1} t\,K(t)[tk(t)+(1-K(t))-tk(t)]dt$$
$$= \int_{-1}^{1} t\,E(t)dK(t) + \int_{-1}^{1} t^2 dK(t) -\int_{-1}^{1} t^2\, K(t)dK(t) + \int_{-1}^{1} tK(t)(1-K(t))dt$$
$$=\int_{-1}^{1} tE(t)dK(t) + \mu_2(K) -\int_{-1}^{1} t^2 K(t)dK(t) +\int_{-1}^{1} tK(t)(1-K(t))dt. \qquad (2.29)$$

Since $\int_{-1}^{1} tE(t)dK(t) = \int_{-1}^{1} E(t)d(t) = \frac{1}{2}E^2(t)]_{-1}^{1} = 0$ and $\int_{-1}^{1} tK(t)(1-K(t))dt = \int_{-1}^{1} t^2K(t) \, dK(t) - \frac{1}{2}\mu_2(K)$, it follows from (2.29) that $\int td\{K(t)\int_0^2 K(u+t)du\} = \frac{1}{2}\mu_2(K)$, which along with (2.26) implies that

$$J_2 = a_n^2 \, \mu_2(K) \int_{-\infty}^{\infty} F^{(1)}(y)d_{12}{}^2(y)W'(y)dW(y)$$

$$=: a_n^2 \xi_{n2c}^{(1)} + O(a_n^3). \text{ (say)} \tag{2.30}$$

Further, arguing as for J_3, we evaluate J_3 by noting that we can also write J_2 as

$$J_3 = a_n^2 \int F^{(2)}(y)[\int_{y<x<y+2a_n}[\int_{-1}^{1} t^2 d\{K(\frac{x-y}{a_n}+t)K(t)\}]d_{12}(x)dW(x)]d_{12}(y)dW(y)$$

$$+ a_n^2 \int F^{(2)}(y)[\int_{y+2a_n<x}[\int_{-1}^{1} t^2 d\{K(\frac{x-y}{a_n}+t)K(t)\}]d_{12}(x)dW(x)]d_{12}(y)dW(y)$$

$$=: J_{31} + J_{32}, \quad \text{(say)} \tag{2.31}$$

where following the same reasoning as for J_{22} and J_{21} respectively, we obtain

$$J_{32} = a_n^2 \int F^{(2)}(y)[\int_{y+2a_n<x}(\int_{-1}^{1} t^2 dK(t))d_{12}(x)dW(x)]d_{12}(y)dW(y)$$

$$= a_n^2\mu_2(K)\iint_{y<x} F^{(2)}(y)d_{12}(x)d_{12}(y)dW(x)dW(y)+O(a_n^3)$$

$$=: a_n^2 \, \xi_{n2c}^{(2)} + O(a_n^3) \tag{2.32}$$

and

$$J_{31} = O(a_n^3). \tag{2.33}$$

Thus, we obtain from (2.31)-(2.33) that

$$J_3 = a_n^2 \, \mu_2(K)\iint_{y<x} F^{(2)}(y)d_{12}(x)d_{12}(y)dW(x)dW(y)+O(a_n^3)$$

$$=: a_n^2 \, \xi_{n2c}^{(2)}+ O(a_n^3), \tag{2.34}$$

as n→∞. Combining (2.9)-(2.9a) with (2.21), (2.22), (2.29), (2.30), and (2.34), we obtain

$$nD^2Var \, \hat{p}_n = L_1 - L_2$$

$$= 2 \iint_{y<x} F(y)(1-F(x))d_{12}(x)d_{12}(y)dW(x)dW(y) - a_n^2\mu_2(K)[\int_{-\infty}^{\infty} F^{(1)}(y)d_{12}{}^2(y)W'(y)dW(y)$$

$$-\iint_{y<x}[F^{(2)}(y)(1-F(x)) -F^{(2)}(x)F(y)]d_{12}(x)d_{12}(y)dW(x)dW(y)] + O(a_n^3)$$

$$= \xi_{n0c} - a_n^2 \, \xi_{n2c} + O(a_n^3) \tag{2.35}$$

or $\sigma_{nc}^2 = Var \, \hat{p}_n =: D^{-2}[\frac{\xi_{noc}}{n} - \frac{a_n^2}{n} \, \xi_{n2c} + O(\frac{a_n^3}{n})]$, \tag{2.36}

where, from (2.25) – (2.36),

$\xi_{n0c} = \xi_{noc}^{(1)} + \xi_{noc}^{(2)} = 2\iint_{y<x} F(y)(1-F(x))d_{12}(x)d_{12}(y)dW(x)dW(y)$ and

$\xi_{n2c} = \xi_{n2c}^{(1)} - \xi_{n2c}^{(2)} = \mu_2(K)[\int_{-\infty}^{\infty} F^{(1)}(y)d_{12}{}^2(y)W'(y)dW(y)$

$$-\iint_{y<x}[F^{(2)}(y)(1-F(x))-F^{(2)}(x)F(y)]d_{12}(x)d_{12}(y)dW(x)dW(y)]. \tag{2.37}$$

Hence the proof of part (b) is complete.

Corollary 2.3: Under the condition (i) of Lemma 2.1,

a) when weight W is a discrete d.f., we have
   $Var \, \tilde{p}_n = D^{-2} \frac{\xi_{nod}}{n}$;
b) when weight W is a continuous d.f. we have,
   $Var \, \tilde{p}_n = D^{-2} \frac{\xi_{noc}}{n}$.

Proof: The proof follows exactly on the similar line of argument as for the proof of Lemma 2.2.

Calculations of Large Sample MSEs of $\hat{p}_n$ and $\tilde{p}_n$: We now obtain the expressions for the Mean Square Errors of the MWSD estimators $\tilde{p}_n$ and $\hat{p}_n$ defined in (2.4) and (2.3) in the following theorems when component distribution functions $F_j(x)$, j=1,2 are known.

Theorem 2.4: Under the conditions of Lemma 2.2,

a) when the weight W is a discrete d.f., we have,
   $MSE \, \hat{p}_n = D^{-2}[\frac{\xi_{nod}}{n} - \frac{a_n}{n}\xi_{n1d} + \frac{a_n^2}{n} \, \xi_{n2d} + a_n^4\xi_{n4d}^2] + O(\frac{a_n^3}{n})$,
   where $\xi_{nod}, \xi_{n1d}, \xi_{n2d}, \xi_{n4d}$ are defined in Lemmas 2.1 and 2.2.

b)   when the weight W is a continuous d.f., we have

$$\text{MSE } \hat{p}_n = D^{-2}[\frac{\xi_{noc}}{n} - \frac{a_n^2}{n}\xi_{n2c} + a_n^4\xi_{n4c}] + O(\frac{a_n^3}{n}),$$

where $\xi_{noc}, \xi_{n2c}, \xi_{n4c}$ are defined in Lemmas 2.1 and 2.2.

Proof: (a) From Lemmas 2.1 and 2.2,

$$\text{MSE } \hat{p}_n = \text{Var } \hat{p}_n + \text{Bias}^2 \hat{p}_n$$

$$= D^{-2}[\frac{\xi_{nod}}{n} - \frac{a_n}{n}\xi_{n1d} + \frac{a_n^2}{n}\xi_{n2d} + O(\frac{a_n^3}{n})] + D^{-2}a_n^4\xi_{n4d}^2 + O(a_n^4)$$

$$= D^{-2}[\frac{\xi_{nod}}{n} - \frac{a_n}{n}\xi_{n1d} + \frac{a_n^2}{n}\xi_{n2d} + a_n^4\xi_{n4d}^2] + O(\frac{a_n^3}{n}).$$

Hence part (a) Lemma follows. Similarly, for part (b) also

$$\text{MSE } \hat{p}_n = \text{Var } \hat{p}_n + \text{Bias}^2 \hat{p}_n$$

$$= D^{-2}[\frac{\xi_{noc}}{n} - \frac{a_n^2}{n}\xi_{n2c} + O(\frac{a_n^3}{n})] + D^{-2} a_n^4 \xi_{n4c}$$

$$= D^{-2}[\frac{\xi_{noc}}{n} - \frac{a_n^2}{n}\xi_{n2c} + a_n^4\xi_{n4c}] + O(\frac{a_n^3}{n}).$$

Hence part (b) of the Lemma follows.

Corollary 2.5: Under the conditions of Lemma 2.1 and Lemma 2.2,

$$\text{MSE } \tilde{p}_n = \text{Var } \tilde{p}_n$$

Proof: The proof follows exactly on the same line of argument as for the proof of Theorem 2.4:

When W is a discrete, $\text{MSE } \tilde{p}_n = \text{Var } \tilde{p}_n = D^{-2}\frac{\xi_{nod}}{n}$,

and when W is continuous, $\text{MSE } \tilde{p}_n = \text{Var } \tilde{p}_n = D^{-2}\frac{\xi_{noc}}{n}$,

where ξ_{n0d} and ξ_{n0c} are defined in (2.20) and (2.37).

Mean square comparisons of $\hat{p}_n$ and $\tilde{p}_n$:

We now establish the superiority of smoothed MWSD estimator $\hat{p}_n$ over its corresponding unsmoothed version $\tilde{p}_n$ in the sense of having lower MSE. Suppose the smoothness assumptions (i) made in Lemma 2.1 concerning the mixture distribution defined in (1.1) and the conditions (ii) and (iii) imposed therein on the density kernel function k and the bandwidth $\{a_n\}$ are also all satisfied. Then, corresponding to any suitable 'discrete' averaging distribution W, the 'smoothed' and 'unsmoothed' (Wolfowitz) minimum weighted square distance estimators $\hat{p}_n$ and $\tilde{p}_n$, respectively, given by (2.3) and (2.4), satisfy the following inequality by Theorem 2.4, namely,

$$\text{MSE } \hat{p}_n < \text{MSE } \tilde{p}_n$$

for sufficiently large n. We now compare the estimators $\tilde{p}_n$ and $\hat{p}_n$ in terms of their MSE. The percentage gain in precision of $\hat{p}_n$ over $\tilde{p}_n$ is

$$\frac{\text{MSE } \tilde{p}_n - \text{MSE } \hat{p}_n}{\text{MSE } \tilde{p}_n} \text{ X } 100$$

Remark 2.6: The result of the Theorem 2.4 would continue to hold even when the averaging discrete distribution W = $\{p(x_j): j \in C\}$ depends on the sample size n, i.e., for example, when W=W_n=$\{p(x_{nj}): j \in C\}$, provided the parameter $v_n = \min_j |x_{n(j+1)} - x_{nj}| \to 0$, as n→∞, at a rate slower than that of the bandwidth a_n→0; that is when $a_n = o(v_n)$ or equivalently when $(a_n/v_n) \to 0$, as n→∞. The distribution W_n then may converge to a continuous limiting distribution W.

3. Strong Consistency of $\hat{p}_n$ and $\tilde{p}_n$: We now establish the a.s. convergence of $\hat{p}_n$ and $\tilde{p}_n$ to p as n→∞ in the following Theorem:

Theorem 3.1: Under the conditions of Lemma 2.1, WP1

$$\hat{p}_n - p = O(\frac{\log n}{n})^{1/2} \text{ a.s., as n→∞.}$$

Proof: From (2.5),

$$(\hat{p}_n - p) D = \hat{\lambda}_n^* + C_n,$$

where C_n is defined in (2.5) and

$$\hat{\lambda}_n^* = n^{-1} \sum_{i=1}^n \hat{A}_i, \quad \hat{A}_i = \int (K(\frac{x - X_i}{a_n}) - EK(\frac{x - X_i}{a_n})) \, d_{12}(x) dW(x), \tag{3.1}$$

with

E $\hat{A}_i = 0$, $|\hat{A}_i| \leq 2$ a.s. and

E $\hat{A}_i^2$ = E $[\int \{K(\frac{x-X_i}{a_n}) - E\,K(\frac{x-X_i}{a_n})\}d_{12}(x)dW(x)]^2$.

From (2.9) and (2.19), if W(x) is a discrete d.f.,

$$\frac{1}{n}\sum_{i=1}^{n} E\,\hat{A}_i^2 = D^{-2}[\xi_{n0d} - a_n\xi_{n1d} + a_n^2\,\xi_{n2d} + O(\frac{a_n^3}{n})] \rightarrow D^{-2}\,\xi_{n0d}$$

$$= \sigma_{nd}^2 \text{ as n} \rightarrow \infty.$$

Now, applying Bernstein inequality to the mean of i.i.d. sequence $\{\hat{A}_i\}$,

$$P_n = P\,(\frac{1}{n}\sum_{i=1}^{n}\hat{A}_i > t) \leq \exp\,(-\frac{\frac{t^2}{2n}}{\frac{1}{n}\sum_{i=1}^{n}E\,\hat{A}_i^2 + \frac{4}{3}\frac{t}{n}}) = \exp(-\frac{\frac{t^2}{2n}}{\sigma_{nd}^2 + \frac{4}{3}\frac{t}{n}}); \qquad (3.2)$$

so that setting t = $(\frac{4\sigma_{nd}^2\log n}{n})^{\frac{1}{2}}$, the RHS of inequality (3.2) becomes

$$= \exp\,[\frac{-2\log n}{1+\frac{2}{3\sigma_{nd}}(\frac{4\log n}{n})^{\frac{1}{2}}}] < n^{-2}, \qquad (3.3)$$

with $\sum_{n\geq 1}^{\infty} P_n \leq \sum_{n\geq 1}^{\infty} n^{-2} < \infty$. By Borel – Cantelli lemma, we can conclude from (3.2) – (3.3) that,

$$\frac{1}{n}\sum_{i=1}^{n}\hat{A}_i \xrightarrow{a.s.} 0, \text{ as n} \rightarrow\infty \text{ i.e. } \hat{p}_n - p \xrightarrow{a.s.} O(\frac{\log n}{n})^{1/2}, \text{ as n} \rightarrow\infty.$$

In a similar way, strong consistency holds when W is a continuous d.f. with σ_{nd}^2 replaced by σ_{nc}^2 where $\sigma_{nc}^2 = D^{-2}\xi_{n0c}$ as defined in (2.36).

Corollary 3.2: Under the conditions of Lemma 2.1(i) on F, WP1

$$\tilde{p}_n - p = O(\frac{\log n}{n})^{1/2}, \text{ as n} \rightarrow\infty.$$

Proof: This follows exactly on the same line of argument as for the proof of the Theorem 3.1.

Asymptotic Normality of $\hat{p}_n$ and $\tilde{p}_n$:
Theorem 3.3: Under the conditions of Lemma 2.1,

a. when W is a discrete d.f.
$$\sqrt{n}\,(\hat{p}_n - p) \xrightarrow{L} N(0, \sigma_d^2),$$

where $\sigma_d^2 = \lim_{n\rightarrow\infty} n\,\sigma_{nd}^2$, σ_{nd}^2 is as defined in (2.19);

b. when W is a continuous d.f.
$$\sqrt{n}\,(\hat{p}_n - p) \xrightarrow{L} N(0, \sigma_c^2),$$

where $\sigma_c^2 = \lim_{n\rightarrow\infty} n\,\sigma_{nc}^2$, σ_{nc}^2 is as is defined in (2.36).

Proof: From (2.5),
$$\sqrt{n}\,(\hat{p}_n - p) = \sqrt{n}\,D^{-1}\,\hat{\lambda}_n^* + \sqrt{n}D^{-1}C_n, \qquad (3.4)$$
with, in view of (2.7),
$$\sqrt{n}D^{-1}C_n = \sqrt{n}D^{-1}[\frac{a_n^2}{2}\mu_2(K)\int_{-\infty}^{\infty} F^{(2)}(x)d_{12}(x)dW(x) + O(a_n^4)]$$
and
$$\sqrt{n}D^{-1}\hat{\lambda}_n^* = \sqrt{n}D^{-1}\int[\hat{F}_n(x) - E\hat{F}_n(x)]d_{12}(x)dW(x)$$
$$= \sqrt{n}D^{-1}\,n^{-1}\sum_{i=1}^{n}\hat{A}_i = D^{-1}\,n^{-1/2}\sum_{i=1}^{n}\hat{A}_i, \qquad (3.5)$$
with $\sum_{i=1}^{n}\hat{A}_i$ as the sum of i.i.d. r.v.'s. From Lemma 2.2(a) – (b), in view of (2.9) and (3.4) – (3.5), we have the expressions $n\text{Var}(\hat{p}_n) = D^{-2}\,n\text{Var}(\hat{\lambda}_n^*)$ given by $(n\sigma_{nd}^2)$ or $(n\sigma_{nc}^2)$, respectively, for the two cases when W is discrete or continuous d.f.'s.

When W is a discrete d.f.,
$$\sigma_{nd}^2 = D^{-2}[\frac{\xi_{n0d}}{n} - \frac{a_n}{n}\xi_{n1d} + \frac{a_n^2}{n}\xi_{n2d} + O(\frac{a_n^3}{n})],$$
where ξ_{n0d}, ξ_{n1d} and ξ_{n2d} are defined in (2.20);
when the weight W is a continuous d.f. we have,
$$\sigma_{nc}^2 = D^{-2}[\frac{\xi_{n0c}}{n} - \frac{a_n^2}{n}\xi_{n2c} + O(\frac{a_n^3}{n})].$$
Now applying Lindeberg-Levy CLT to the sum of i.i.d. sequence $\{\hat{A}_i\}$ in (3.5); it at once follows that

$$\sqrt{n}\,(\hat{p}_n - p) \xrightarrow{L} N(0, \sigma_d^2), \text{ as } n \to \infty, \tag{3.6}$$

where $\sigma_d^2 = \lim_{n\to\infty} n\sigma_{nd}^2$ when W is discrete.

Similarly, the Asymptotic Normality like (3.6) holds when W is continuous with $\sigma_c^2 = \lim_{n\to\infty} n\sigma_{nc}^2$. The proof is complete.

Corollary 3.4: Under the conditions of Lemma 2.1(i) on F,

a. when W is a discrete d.f.
$$\sqrt{n}\,(\tilde{p}_n - p) \xrightarrow{L} N(0, \sigma_d^2),$$

where $\sigma_d^2 = \lim_{n\to\infty} n\,\sigma_{nd}^2$;

b. when W is a continuous d.f.
$$\sqrt{n}\,(\tilde{p}_n - p) \xrightarrow{L} N(0, \sigma_c^2),$$

where $\sigma_c^2 = \lim_{n\to\infty} n\,\sigma_{nc}^2$.

Proof: The proof follows in the same manner as that of Theorem 3.3.

4. Optimal band width $a_{opt,n}$

The selection of optimal bandwidth 'a_n' in the MWSD estimator $\hat{p}_n$ is done in such a way that its MSE is minimum w.r.to a_n. From Theorem 2.4, when W is a discrete d.f,

$$\text{MSE } \hat{p}_n = D^{-2}\left[\frac{\xi_{nod}}{n} - \frac{a_n}{n}\xi_{n1d} + a_n^4\xi_{n4c}\right] + O\left(\frac{a_n^3}{n}\right) =: M_{nd};$$

the $a_{opt,n}$ is obtained by solving $\frac{dM_{nd}}{da_n} = 0$, i.e. $-\frac{\xi_{n1d}}{n} + 4a_n^3\xi_{n4d} = 0$, which yields

$$a_{opt,n} = \frac{\xi_{n1d}}{4\xi_{n4d}}\,n^{-1/3}.$$

When W is a continuous d.f

$$\text{MSE } \hat{p}_n = D^{-2}\left[\frac{\xi_{noc}}{n} - \frac{a_n^2}{n}\xi_{n2c} + a_n^4\xi_{n4c}\right] + O\left(\frac{a_n^2}{n}\right) =: M_{nc};$$

the $a_{opt,n}$ is obtained by solving $\frac{dM_{nc}}{da_n} = 0$, $-\frac{2a_n}{n}\xi_{n2c} + 4a_n^3\xi_{n4c} = 0$, which yields

$$a_{opt,n} = a_n = \frac{2\xi_{n2c}}{4\xi_{n4c}}\,n^{-1/2}.$$

5. Monte Carlo Simulation study

A simulation study is carried out to estimate mixing proportion p by $\hat{p}_n$ and $\tilde{p}_n$ when the two component distributions are known in both Normal and Exponential populations.

Table 5.1: Simulation results of $\hat{p}_n$ and $\tilde{p}_n$ for different sets N of sample size n with p = 0.3, 0.5, 0.7.

p	N	$F_1(x)=N(1,0.5^2)$, $F_2(x)=N(2,3^2)$				Efficiency	$F_1(x)=Exp(2)$, $F_2(x)=Exp(3)$				Efficiency
		$\tilde{p}_n$	$\hat{p}_n$	$\widehat{MSE}$			$\tilde{p}_n$	$\hat{p}_n$	$\widehat{MSE}$		
				$\tilde{p}_n$	$\hat{p}_n$				$\tilde{p}_n$	$\hat{p}_n$	
0.3 n=12	**10**	0.093	0.423	0.016	0.001	90.19	0.159	0.929	0.009	0.004	57.88
	25	0.06	0.461	0.008	0.002	68.76	0.133	0.933	0.009	0.003	66.30
	50	0.049	0.454	0.005	0.004	27.49	0.128	0.934	0.008	0.003	64.74
	75	0.048	0.446	0.004	0.004	4.11	0.120	0.931	0.008	0.003	59.89
	100	0.043	0.446	0.003	0.003	3.84	0.114	0.931	0.007	0.003	58.56
0.5 n=12	**10**	0.186	0.173	0.009	0.007	16.53	0.35	0.59	0.034	0.059	-78.06
	25	0.139	0.125	0.011	0.008	27.54	0.45	0.66	0.039	0.052	-35.58
	50	0.159	0.142	0.014	0.010	26.49	0.42	0.62	0.05	0.049	0.51
	75	0.160	0.139	0.012	0.009	23.49	0.43	0.63	0.054	0.048	10.29
	100	0.170	0.148	0.012	0.009	25.10	0.44	0.65	0.049	0.046	4.82
0.7 n=12	**10**	0.45	0.48	0.007	0.005	30.93	0.076	0.903	0.003	0.005	-51.96
	25	0.47	0.48	0.008	0.004	56.22	0.081	0.909	0.005	0.004	2.76
	50	0.42	0.45	0.011	0.004	65.91	0.074	0.922	0.004	0.004	3.99
	75	0.40	0.44	0.011	0.004	61.86	0.075	0.924	0.004	0.004	15.12
	100	0.39	0.44	0.010	0.004	64.19	0.076	0.931	0.005	0.003	28.51

Comments: The simulation results show $\widehat{MSE}$ for smoothed estimator $\hat{p}_n$ is less than that of unsmoothed estimator $\tilde{p}_n$ uniformly for all samples. So the smoothed estimator appears a better estimator in terms of MSE. The average gain in observed efficiency due to smoothing is lying between 3% to 90% for different N sets, each of size n.

References

1. Bernstein, S.N.(1946): The Theory of Probabilities, Gastehizdat Publishing House, Moscow.
2. Blischke, W.R.(1965): Mixtures of discrete distributions, Collected papers presented at the Int. Symp. Ed. By G.P.Patil, Oxford, Pergamom.
3. Bruni, C. et al(1985): On the inverse problem in eytofluorometry- Recovering DNA distribution from FMF date, Cell Bio. Phys. 5, 5-19.
4. Choi, K. & Bulgren, W. G.(1968): An estimation procedure for mixtures distributions, JRSS, Ser. B, Vol. 30, No.3, 444-460.
5. Choi, K.(1969): Estimators for the parameters of a finite mixture of distributions, Annals of Institute of Statistical Mathematics, 21, 107-116.
6. Clark and Heathcote(1994): Robust estimation of k-component univariate normal mixtures, Annals of Institute of Statistical Mathematics, Vol. 46, issue 1, 83-93.
7. Day, N.E.(1969): Estimating the components of a mixture of Normal distributions, Biometrika, Vol. 56, No.3, 463-474.
8. Fu, K.S.(1968): Sequential methods in Pattern Recognition and Machine learning, Academic Press, New York.
9. Hall, P.(1981): On the nonparametric estimation of mixture proportions, JRSS, Ser B, Vol. 43, No.2, 147-156.
10. Hall, P., and Titterington(1984): Efficient nonparametric estimation of mixture proportions, JRSS, Ser B, Vol. 46, No.3, 465-473.
11. Harris, E. K.(1958): On the probability of survival of bacteria in sea water. Biometrics, 14, 195-206.
12. Hosmer(1973): A comparison of iterative maximum likelihood estimates of the parameters of two normal distribution under three different types of samples, Biometrika, 29, 761-770.
13. Macdonal, P.D. and Pitcher, T.J.(1979): Age groups from size frequency data, J. Fish. Res. Bd. Cano. 36, 987-1008.
14. McDonald(1971): Comment on " An estimation procedure for mixtures of distributions by Choi and Belgren, JRSS, Ser B, Vol.33, 326-329.
15. Mehra K.L., Ramakrishnaiah Y.S., Konda Satish(2018): On smoothed mwsd estimation of mixing proportion presented in International conference on mathematical modeling, appllied analysis and computation on july 6-8, 2018 in JECRC University, Rajasthan, India.

16. Odell, P.L. and Basu, J.P. (1976): Concerning several methods for estimations crop acreages using remote sensing data, Commun. Statist. A 5, 1091-1114.

17. Tang and Karunamuni(2013): Minimum Distance Estimation in a finite mixture regression model, Jl. of Multivariate Analysis, Vol. 120, 185-204.

18. Titterington, D.M., Smith., A.F., and Makov, U.E.(1985): Statistical analysis of finite mixture distributions, Wiley publications.

19. Van Houwelingen, J.C. and Devries (1974): Minimax estimation of mixing proportion of two known distributions, Jl. American Stat. Assoc., 300-304.

20. Woodward et al(1984): A comparison of minimum distance and maximum likelihood estimation of a mixture proportion, Jl. of American Statistical Association, Vol. 79, 590-598.

21. Woodward, W.A., Whitney and Eslinger(1995): Minimum Hellinger Distance estimation of mixture proportion, Journal of Statistical Planning and Inference, Vol.48, Issue 3, 303-319.

22. Wolfe, J.H.(1970): Pattern clustering by multivariate mixture analysis, Multivariate Behavioral Research 5, 329-350.

23. Wolfowitz, J. (1953): Estimation by the minimum distance method, Annals of the Institute of Statistical Mathematics, Vol.5 (1), 9-23.

24. Wu, J. and Karunamuni, R.J.(2009): On Minimum Hellinger Distance estimation, Canadian Jl. of Statistics, 37, 514-533.