Ramakrishnaiah, Y. S.; Trivedi, Manish; Satish, Konda

Article

ON THE SMOOTHED PARAMETRIC ESTIMATION OF MIXING PROPORTION UNDER FIXED DESIGN REGRESSION MODEL

Statistics in Transition New Series

Provided in Cooperation with:
Polish Statistical Association

Suggested Citation: Ramakrishnaiah, Y. S.; Trivedi, Manish; Satish, Konda (2019) : ON THE SMOOTHED PARAMETRIC ESTIMATION OF MIXING PROPORTION UNDER FIXED DESIGN REGRESSION MODEL, Statistics in Transition New Series, ISSN 2450-0291, Exeley, New York, NY, Vol. 20, Iss. 1, pp. 87-102, https://doi.org/10.21307/stattrans-2019-005

This Version is available at:
https://hdl.handle.net/10419/207925

STATISTICS IN TRANSITION new series, March 2019
Vol. 20, No. 1, pp. 87–102, DOI 10.21307/stattrans-2019-005

ON THE SMOOTHED PARAMETRIC ESTIMATION OF MIXING PROPORTION UNDER FIXED DESIGN REGRESSION MODEL

Y. S. Ramakrishnaiah[1], Manish Trivedi[2], Konda Satish[3]

ABSTRACT

The present paper revisits an estimator proposed by Boes (1966) – James (1978), herein called BJ estimator, which was constructed for estimating mixing proportion in a mixed model based on independent and identically distributed (i.i.d.) random samples, and also proposes a completely new (smoothed) estimator for mixing proportion based on independent and not identically distributed (non-i.i.d.) random samples. The proposed estimator is nonparametric in true sense based on known "kernel function" as described in the introduction. We investigated the following results of the smoothed estimator under the non-i.i.d. set-up such as (a) its small sample behaviour is compared with the unsmoothed version (BJ estimator) based on their mean square errors by using Monte-Carlo simulation, and established the percentage gain in precision of smoothed estimator over its unsmoothed version measured in terms of their mean square error, (b) its large sample properties such as almost surely (a.s.) convergence and asymptotic normality of these estimators are established in the present work. These results are completely new in the literature not only under the case of i.i.d., but also generalises to non-i.i.d. set-up.

Key words: mixture of distributions, mixing proportion, smoothed parametric estimation, fixed design regression model, mean square error, optimal band width, strong consistency, asymptotic normality.

1. Introduction

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and not identically distributed (non-i.i.d.) random variables with continuous distribution functions (d.f.s) $\{F_i(x), 1 \le i \le n\}$. Let H(x) be a continuous cumulative distribution function (cdf) of mixture of component cdfs $H_1(x), \ldots, H_m(x)$, ($m \ge 2$) such that $H(x) = \sum_{j=1}^{m} p_j H_j(x)$, where $\{p_j;\ 1 \le j \le m\}$ is a set of mixing proportions satisfying (i) $0 < p_j < 1$, (ii) $\sum_{j=1}^{m} p_j = 1$. Let $\bar{H}_n(x) = n^{-1} \sum_{i=1}^{n} F_i(x) \rightarrow H(x)$ as $n \rightarrow \infty$ and $\bar{H}_j(x) = n_j^{-1} \sum_{i=1}^{n_j} F_{ji}(x) \rightarrow H_j(x)$,

[1] Faculty of Statistics, Osmania University in Hyderabad. E-mail: ysrkou@gmail.com.
[2] Faculty of Statistics, School of Sciences, Indira Gandhi National Open University in New Delhi. E-mail: Manish_trivedi@ignou.ac.in. ORCID ID: https://orcid.org/0000-0002-5790-6546.
[3] Faculty of Statistics, Aurora College in Hyderabad. E-mail: statishlaks@gmail.com. ORCID ID: https://orcid.org/0000-0003-4119-8773.

$n_j \to \infty$, j=1,2,…,m; H(x), H_j(x) are known d.f.s. The problem of estimation of mixing proportions p_j in a mixture

$$\mathbf{H(x)} = \boldsymbol{p_1 H_1}(x) + \boldsymbol{p_2 H_2}(x) + \ldots + \boldsymbol{p_m H_m}(x) \tag{1.1}$$

of m known distributions H_j(x) is investigated based on independent random samples of sizes n, n_j generated from the fixed design regression models

$$X_i = \beta t_i + \epsilon_i \text{ , } 1 \le i \le n, \ \epsilon_i \sim \text{i.i.d.F(x)} \tag{1.2}$$

$$X_{ji} = \beta_j t_{ji} + \epsilon_{ji}, \ 1 \le i \le n_j, \ j = 1, 2, \ldots, m, \ \epsilon_{ij} \sim \text{i.i.d. } F_j(x), \tag{1.3}$$

β's and t's are known reals satisfying the model conditions

$$\beta_j > 0, \ \sum_{i=1}^{n} t_i = 0 \text{ and } \frac{1}{n}\sum t_i^2 = o(n^{-1}). \tag{1.4}$$

Note that $t_i = \mp \frac{i}{n^{\delta}}$, i =∓1, ∓2,…, ∓n, $\delta \ge \frac{3}{2}$ fulfill (1.4) and $\bar{H}_n(x)=n^{-1}\sum_{i=1}^{n} F_i(x)$ = F(x) +O($\frac{1}{n}\sum t_i^2$) + … = H(x) + o(n^{-1}) and $\bar{H}_{n_j}(x)=n_j^{-1}\sum_{i=1}^{n_j} F_{ji}(x) \to H_j(x)$, j=1,2 as $n_j \to \infty$.

Mixture distributions have been used in a wide variety of numerous applications in such diversified fields as physics, chemistry, biology, social sciences and others. Many typical problems in which such mixtures occur have been well described in a series of research papers. Karl Pearson (1894) dealt with the application of normal mixtures to the theory of evolution, which considered the first paper in the mixtures of distributions. Acheson and McElwee (1951), who identified failures in an electronic tube in gaseous defects, mechanical defects and normal deterioration of the cathode. One can find the proportion of the population which will fail in each cause to redesign the system or to improve the methods of manufacturing process. Apart from this, it would be desirable to know the distribution of defectives for each cause. Mendenhall and Hader (1958) studied censored life testing as a mixed failure populations. They suggested an example that the engineer may identify the product as defective/failure and nondefective by two or more different types of causes. Hosmer (1973) studied characteristics such as sex, age, and length of halibut (fish). Odell and Basu (1976) applied them in the field of remote sensing to estimate the crop acreages from remote sensors on orbiting satellites.

We shall show some of the typical problems which were described in Choi and Bulgren (1968):

1. In fishery biology, it is often desired to measure certain characteristics in a natural population of fish. For this purpose samples of fish are taken and the desired trait is measured for each fish in the sample. However, many characteristics vary markedly with the age of the fish. Then, the trait has a distinct distribution for each age group so that the population has a mixture of distributions.
2. A geneticist analyses the inheritance of qualitative characters. In general, such characters vary continuously over some intervals of real numbers so that a given genotype may be able to produce phenotypic values over an interval of real numbers. Then, the phenotypic value which the geneticist observes has a mixture of distributions, each of which is given by a genotype.

3. In photographing the absorption spectrum of an ionized atom, we obtain a photograph of a constantly varying intensity distribution on the photographic plate, and not a series of discrete "lines". This phenomenon is caused by several effects (such as the Doppler effect) and it is accepted in spectroscopy that an intensity distribution whose graph can be approximated closely by that of a normal density function belongs to every theoretical "line". Then the graph of the whole spectrum section can be considered as a mixture of normal density functions.

Other references to mixed failure populations are given in papers by Davis (1952), Epstein (1953), Herd (1953), Steen and Wilde (1952), Everitt and Hand (1981), Titterington et al. (1985), McLachlan and Basford (1988), Lindsay (1995) and McLachlan and Peel (2000), Fu (1968-Pattern Recognition), Varli et al. (1975-Pattern Recognition), Clark (1976-Geology), Macdonald and Pitcher (1979-Fisheries), Bruni et al. (1985-Genetics), Merz (1980-Physics) and Christensen et al. (1980-Nuclear Physics).

The applications of finite mixture distributions describing mixture populations for non- i.i.d. sequence of variables are given below:

Area	**Characteristic X_i**	**Distribution function $F_i(x)$**
Survival Analysis	life time of components produced by i^{th} machine operated by i^{th} foreman.	Life time distribution of various products
Nutritional Studies	weight for age/height for age/ weight for height of i^{th} infant of i^{th} origin or group.	Distribution of weight for age/height of i^{th} infant
Fisheries	Fish length or weight of age of i^{th} fish.	Distribution of weight/length of i^{th} fish

The mixing model with two component populations becomes

$$\mathbf{H(x) = p}\boldsymbol{H_1}\mathbf{(x) + (1-p)}\boldsymbol{H_2}\mathbf{(x).}$$

Here, X_i is the characteristic with distribution function $F_i(x)$ assuming

$$\bar{F}_n(x) = n^{-1}\sum_{i=1}^{n} F_i(x) = \bar{H}_n(x) \rightarrow H(x)$$

$$\bar{F}_{jn}(x) = n^{-1}\sum_{i=1}^{n} F_{ji}(x) = \bar{H}_{jn}(x) \rightarrow H_j(x) \text{ as } n\rightarrow\infty$$

and n, n_1 and n_2 are independent random sample sizes from mixed and component populations selected in such a way that $n = n_1 + n_2$ with $n_1 = [\rho n]$, $n_2 = [(1-\rho)n]$, $0< \rho <1$.

More details on such examples can be found in Choi and Bulgreen (1968), Harris (1958), Blischke (1965), Fu (1968-Pattern Recognition), Varli et al. (1975-Pattern Recognition), Clark (1976-Geology), Macdonald and Pitcher (1979-Fisheries), Odell and Basu (1976-Remote sensing), Bruni et al. (1985-Genetics), Merz (1980-Physics) and Christensen et al. (1980-Nuclear Physics), etc.

i.i.d case: The mixing model for i.i.d. case is

$$F(x) = pG_1(x)+(1-p)G_2(x) \tag{1.5}$$

where F(x), $G_j(x)$; j=1,2 are cdfs of mixed and component populations respectively. The following estimator is studied in the literature.

Boes (1966)-James (1978) (BJ) estimator: Let $F_n(x)= n^{-1} \sum_{i=1}^{n} I(X_i \leq x)$ be the empirical distribution function of a random sample X_i, 1≤i≤n from a mixture of two known component distribution functions G_j, j=1,2. Boes (1966) proposed an estimator of p given by

$$\boldsymbol{p_{n,1}}(x) = \frac{\widetilde{F}_n(x)- G_2(x)}{G_1(x)-G_2(x)} \tag{1.6}$$

and shown as a minimax unbiased estimator, and derived the Cramer-Rao lower bound. James (1978) considered the problem of estimating the mixing proportion in a mixture of two known normal distributions. He studied the simple estimators based on (a) the number of observations lower than a fixed point r, (b) the numbers lower than s and greater than t, and (c) the sample mean. Van Houwelingen (1974) used Boes (1966) estimator to estimate the mixing proportion by using frequency densities and obtained the Cramer-Rao lower bound. Jayalakshmi (2002) used BJ estimator using kernel-based empirical distribution and established that smoothing improves efficiency when the component distributions are known.

In the present work, we extend the idea of estimation of mixing proportion p in two directions:

- The estimators based on kernel based empirical d.f. called smoothed estimations are proposed under regression models (1.2) - (1.3).
- The proposed parametric estimators are based on independent, but not identically distributed (non-i.i.d.) samples generated by the fixed design regression models described by (1.2)-(1.3).

The main object of the present paper is to confine attention to m=2 case in the model (1.1) and to construct parametric estimators when component distributions are known, based on the usual empirical and kernel-based distribution functions defined by

$$\widetilde{H}_n(x) = n^{-1} \sum_{i=1}^{n} I(X_i \leq x),\ \widehat{H}_n(x)= n^{-1} \sum_{i=1}^{n} K\left(\frac{x- X_i}{a_n}\right), \tag{1.7}$$

$\{a_n\}$ being the smoothing sequence satisfying $0< a_n \to 0$, $na_n \to \infty$ defined by

$$p_{n,1}(x) = \frac{\widetilde{H}_n(x)- H_2(x)}{H_1(x)- H_2(x)},\quad p_{n,2}(x) = \frac{\widehat{H}_n(x)- H_2(x)}{H_1(x)- H_2(x)}. \tag{1.8}$$

We study the small and large sample behaviour of the proposed parametric estimators and establish the superiority of smoothed estimator $p_{n,2}(x)$ over unsmoothed one $p_{n,1}(x)$ in the sense of minimum mean square error. The results of the present investigations for the non-i.i.d. sequences are completely new in the literature.

In section 2, we obtain the exact expressions for MSEs of the proposed estimators in order to establish the superiority of $p_{n,2}(x)$ over $p_{n,1}(x)$ for some fixed x. Furthermore, large sample behaviour, such as asymptotic normality and rates

of a.s. convergence of the proposed estimators is also established. In section 3, the crucial choice of smoothing parameter 'a_n' in kernel based estimator $p_{n,2}$(x) is discussed and its value is determined by employing minimum mean square criterion. Section 4 deals with establishing superiority of $p_{n,2}$(x) over $p_{n,1}$(x). Section 5 explains the small sample comparisons by Monte Carlo method based on the samples generated by regression models.

2. Asymptotics of $p_{n,1}$(x) and $p_{n,2}$(x)

We now present the properties of both estimators $p_{n,1}$(x) and $p_{n,2}$(x) under the fixed design regression model (1.3). The properties such as mean square errors (MSEs), rates of a.s. convergence, and asymptotic normality of $p_{n,1}$(x) and $p_{n,2}$(x) are established. We first consider the representations of the proposed estimators: from the mixing model,

$$\mathrm{H(x)} = \mathrm{p}H_1(\mathrm{x}) + (1\text{-}\mathrm{p})H_2(\mathrm{x}) \tag{2.1}$$

H(x) – H_2(x) = d_{12}(x)p for each x, where d_{12}(x)= H_1(x) - H_2(x)

A.S. Representations to $p_{n,1}$(x), $p_{n,2}$(x): from (1.8) and (2.1),

$$\mathrm{d_{12}(x)}[p_{n,1}(\mathrm{x})\text{-}\mathrm{p}] = \widetilde{\mathrm{H}}_{\mathrm{n}}(\mathrm{x}) - \mathrm{H(x)} = n^{-1}\sum_{i=1}^{n}(\mathrm{I}(X_i \leq \mathrm{x}) - F_i(\mathrm{x}) + F_i(\mathrm{x}) - \mathrm{H(x)})$$

$$=: n^{-1}\sum_{i=1}^{n} Z_{1i}(\mathrm{x}) + \tau_{n1}(\mathrm{x}) \tag{2.2}$$

where by (1.4) and by the assumption on regression model

$$\tau_{n1}(\mathrm{x}) = n^{-1}\sum_{i=1}^{n} F_i(\mathrm{x}) - \mathrm{H(x)} = \mathrm{o}(\mathrm{n}^{-1}\textstyle\sum_1^n t_i^2) = \mathrm{o}(\mathrm{n}^{-1}) \tag{2.3}$$

Similarly, from (1.8) and (2.1)

$$\mathrm{d_{12}(x)}[\mathrm{p}_{n,2}(\mathrm{x})\text{-}\mathrm{p}] = \widehat{\mathrm{H}}_{\mathrm{n}}(\mathrm{x}) - \mathrm{H(x)} = \widehat{\mathrm{H}}_{\mathrm{n}}(\mathrm{x}) - \mathrm{E}\,\widehat{\mathrm{H}}_{\mathrm{n}}(\mathrm{x}) + \mathrm{E}\,\widehat{\mathrm{H}}_{\mathrm{n}}(\mathrm{x}) - \mathrm{H(x)}$$

$$=: \mathrm{n}^{-1}\sum_{i=1}^{n}[K(\tfrac{\mathrm{x}-\mathrm{X_i}}{\mathrm{a_n}}) - \mathrm{E}\,K(\tfrac{\mathrm{x}-\mathrm{X_i}}{\mathrm{a_n}})] + \tau_{n2}(\mathrm{x})$$

$$= n^{-1}\sum_{i=1}^{n} Z_{2i}(\mathrm{x}) + \tau_{n2}(\mathrm{x}) \tag{2.4}$$

with τ_{n2}(x) = o(n^{-1}) as in (2.3).

2.1. Mean Square Error of $p_{n,1}$(x) and $p_{n,2}$(x)

We first consider the small sample property, i.e. MSEs of both estimators $p_{n,j}$(x), j=1,2 for each x in the following, which will help in establishing the

superiority of smoothed estimator over unsmoothed version based on a random sample of exact size n under the regression model.

Theorem 2.1. Let $\{X_i$: 1≤i≤n$\}$ be a sequence of non-i.i.d. random variables with corresponding sequence of uniformly continuous distribution functions $\{F_i$(x): 1≤i≤n$\}$. If $\{F_i\}$, the kernel density function k and $\{a_n\}$ in (1.7) satisfy

AI: i) F_i(x) is uniformly continuous distribution function with finite q[th] derivatives $F_i^{(q)}$(x) < ∞, 1 ≤ i ≤ n and

$$\bar{H}_n^{(q)}(x) = \frac{1}{n} \sum F_i^{(q)}(x),\ q = 2,4,6$$

ii) $\bar{H}_n$(x) = $n^{-1}\sum_{i=1}^{n} F_i(x) \to H(x)$ as $n \to \infty$

AII: i) The kernel function satisfies $\mu_{2j}(K) = \int_{-\infty}^{\infty} t^{2j}\, dK(t) \neq 0$ and

$$\mu_j(K) = \int_{-\infty}^{\infty} t^j\, dK(t) = 0 \text{ for } j=1,3,\ldots$$

ii) $\psi_j(K) = 2\int_{-\infty}^{\infty} t^j K(t)\, dK(t) < \infty$, j = 0, 1, 2, 3, 4

AIII: $\{a_n\}$ is a sequence of bandwidths such that

i) $0 < a_n \downarrow 0$; $na_n \to \infty$ as $n \to \infty$

ii) $na_n^4 \to 0$ as $n \to \infty$

then

$$\mathbf{MSE}\,[p_{n,2}(x)] = \mathbf{MSE}(p_{n,1}(x)) - d_{12}^{-2}(x)\left[\frac{a_n}{n}\bar{H}_n^{(1)}(x)\psi_1(K) - \frac{a_n^4}{4}\bar{H}_n^{(2)^2}(x)\mu_2^2(K)\right] + O\left(\frac{a_n^2}{n}\right) + o(a_n^4)$$

Proof: From (2.4),

$$d_{12}^2(x)\,nMSE[p_{n,2}(x)] = Var(n^{-1/2}\sum_{i=1}^{n} Z_{2i}(x)) + n\tau_{n2}^2 \qquad (2.5)$$

where σ_{n2}^2 = $Var(n^{-1/2}\sum_{i=1}^{n} Z_{2i}(x))$

$= n^{-1}\sum_{i=1}^{n} E\, Z_{2i}^2(x)$

$= n^{-1}\sum_{i=1}^{n} \sigma_{2i}^2$

$\sigma_{2i}^2 = E\,K^2\left(\frac{x-X_i}{a_n}\right) - E^2\,K\left(\frac{x-X_i}{a_n}\right)$

$= I_{1i} - I_{2i}^2$ (say)

where $I_{1i} = E\,K^2\left(\frac{x-X_i}{a_n}\right) = \int K^2\left(\frac{x-u}{a_n}\right) dF_i(u) = \int F_i(x-a_nt)\, dK^2(t)$

$= F_i(x)\int dK^2(t) - F_i^{(1)}(x)a_n\int t\,dK^2(t) + \frac{a_n^2}{2!}F_i^{(2)}(x)\int t^2 dK^2(t) - \frac{a_n^3}{3!}F_i^{(3)}(x)\int t^3 dK^2(t)$

$+ \frac{a_n^4}{4!}F_i^{(4)}(x)\int t^4 dK^2(t) + o(a_n^4)$

$$= F_i(x)\,\psi_0(K) - a_n F_i^{(1)}(x)\,\psi_1(K) + \frac{a_n^2}{2!} F_i^{(2)}(x)\,\psi_2(K) - \frac{a_n^3}{3!} F_i^{(3)}(x)\,\psi_3(K)$$

$$+ \frac{a_n^4}{4!} F_i^{(4)}(x)\,\psi_4(K) + o(a_n^4) \qquad (2.6)$$

where $\psi_0(K) = 2\int_{-\infty}^{\infty} K(t)dK(t) = 2\int_0^1 y\,dy = 1$, while

$$I_{2i} = E\,K\left(\frac{x - X_i}{a_n}\right) = \int F_i(x - a_n t)dK(t)$$

$$= F_i(x) + \frac{a_n^2}{2!} F_i^{(2)}(x) \int_{-\infty}^{\infty} t^2 dK(t) + \frac{a_n^4}{4!} F_i^{(4)}(x) \int_{-\infty}^{\infty} t^4 dK(t) + o(a_n^4)$$

$$=: F_i(x) + \frac{a_n^2}{2} F_i^{(2)}(x)\mu_2(K) + \frac{a_n^4}{4!} F_i^{(4)}(x)\,\mu_4(K) + o(a_n^4) \qquad (2.7)$$

From (2.5) and (2.7),

$$\sigma_{n2}^2 = n^{-1} \sum_{i=1}^n \{[F_i(x) - a_n F_i^{(1)}(x)\psi_1(K) + \frac{a_n^2}{2} F_i^{(2)}(x)\psi_2(K) - \frac{a_n^3}{3!} F_i^{(3)}(x)\psi_3(K)$$

$$+ \frac{a_n^4}{4!} F_i^{(4)}(x)\psi_4(K) + o(a_n^4)]$$

$$- (F_i(x) + \frac{a_n^2}{2} F_i^{(2)}(x)\mu_2(K) + \frac{a_n^4}{4!} F_i^{(4)}(x)\,\mu_4(K) + o(a_n^4))^2\}$$

$$= n^{-1} \sum_{i=1}^n F_i(x)(1 - F_i(x)) - a_n \bar{H}_n^{(1)}(x)\psi_1(K) + \frac{a_n^2}{2} \bar{H}_n^{(2)}(x)\psi_2(K)$$

$$- \frac{a_n^3}{6} \bar{H}_n^{(3)}(x)\,\psi_3(K) - a_n^2\, n^{-1} \sum_{i=1}^n F_i(x) F_i^{(2)}(x)\mu_2(K)$$

$$- \frac{a_n^4}{12} n^{-1} \sum_{i=1}^n F_i(x) F_i^{(4)}(x)\mu_4(K) + o(\frac{a_n^4}{n})$$

$$= n^{-1} \sum_{i=1}^n F_i(x)(1 - F_i(x)) - a_n \bar{H}_n^{(1)}(x)\psi_1(K)$$

$$+ a_n^2\, (\frac{1}{2} \bar{H}_n^{(2)}(x)\psi_2(K) - n^{-1} \sum_{i=1}^n F_i(x) F_i^{(2)}(x)\mu_2(K)) + O(a_n^3) \qquad (2.8)$$

and from (2.7),

$$\tau_{n2}^2 = (n^{-1} \sum_{i=1}^n E\,K\left(\frac{x - X_i}{a_n}\right) - H(x))^2$$

$$= [\bar{H}_n(x) - H(x) + \frac{a_n^2}{2} \bar{H}_n^{(2)}(x)\mu_2(K) + \frac{a_n^4}{4!} \bar{H}_n^{(4)}(x)\mu_4(K) + o(a_n^4)]^2$$

$$=: [\xi_{n,0}(x) + a_n^2\, \xi_{n,2}(x) + a_n^4 \xi_{n,4}(x) + o(a_n^4)]^2$$

$$= \xi_{n,0}(x)[\xi_{n,0}(x) + 2a_n^2 \xi_{n,2}(x) + 2a_n^4 \xi_{n,4}(x) + o(a_n^4)] + a_n^4 \xi_{n,2}^2(x) + o(a_n^4) \qquad (2.9)$$

$$= a_n^4 \xi_{n,2}^2(x) + O(\frac{a_n^2}{n})$$

in view of $\xi_{n,0}(x) = o(n^{-1})$, where $\xi_{n,2}(x) = \frac{1}{2} \bar{H}_n^{(2)}(x)\mu_2(K)$.

Thus, from (2.5), (2.8) and (2.9),

$$MSE(p_{n,2}(x)) = d_{12}^{-2}(x) n^{-1} [n^{-1} \sum_{i=1}^n F_i(x)(1 - F_i(x)) - a_n \bar{H}_n^{(1)}(x)\psi_1(K)$$

$$+ a_n^2 (\frac{1}{2} \bar{H}_n^{(2)}(x)\psi_2(K) - n^{-1} \sum_{i=1}^n F_i(x) F_i^{(2)}(x)\mu_2(K))]$$

$$+ d_{12}^{-2}(x)\xi_{n,0}(x)[\xi_{n,0}(x) + 2a_n^2\xi_{n,2}(x) + 2a_n^4\xi_{n,4}(x)]$$

$$+ a_n^4\ \xi_{n,2}^2(x) + o(\tfrac{a_n^4}{n})$$

$$= d_{12}^{-2}(x)[n^{-1}\textstyle\sum_{i=1}^{n} F_i(x)(1- F_i(x))/n - \frac{a_n}{n}\bar{H}_n^{(1)}(x)\psi_1(K) + a_n^4\ \xi_{n,2}^2(x)]$$

$$+ O(\tfrac{a_n^2}{n})$$

as $\xi_{n,0}(x) = \bar{H}_n(x) - H(x) = o(n^{-1})$ as $n \to \infty$

$$n^{-1}\textstyle\sum_{i=1}^{n} F_i(x)(1-F_i(x)) = n^{-1}\sum_{i=1}^{n} F_i(x) - n^{-1}\sum_{i=1}^{n} F_i^2(x)$$

$$= \bar{H}_n(x) - \bar{H}_n^2(x) - n^{-1}(\textstyle\sum_{i=1}^{n} F_i^2(x) - n\bar{H}_n^2(x))$$

$$= \bar{H}_n(x)(1-\bar{H}_n(x)) - (n^{-1}\textstyle\sum_{i=1}^{n}(F_i(x) - \bar{H}_n(x))^2$$

$$= \bar{H}_n(x)(1 - \bar{H}_n(x)) - V_{nF}(x) \qquad (2.10)$$

where $V_{nF}(x) = n^{-1}\sum(F_i(x) - \bar{H}_n(x))^2 > 0$ and by considering the terms containing $\frac{a_n^2}{n}$ as of higher order,

$$MSE(p_{n,2}(x)) = d_{12}^{-2}(x)[\tfrac{\bar{H}_n(x)[1-\bar{H}_n(x)]}{n} - \tfrac{V_{nF}(x)}{n}] - d_{12}^{-2}(x)[\tfrac{a_n}{n}\bar{H}_n^{(1)}(x)\psi_1(K)$$

$$+ \tfrac{a_n^4}{4}\bar{H}_n^{(2)^2}(x)\mu_2^2(K)] + O(\xi_{n,0}(x)\ a_n^2) + o(a_n^4)$$

Corollary 2.1: Under the conditions of Theorem 2.1 on $\{F_i(x)\}$,

$$MSE(p_{n,1}(x)) = d_{12}^{-2}(x)[\tfrac{\bar{H}_n(x)[1-\bar{H}_n(x)]}{n} - \tfrac{V_{nF}(x)}{n}] + O(n^{-2})$$

where $V_{nF}(x) = n^{-1}\sum(F_i(x) - \bar{H}_n(x))^2 > 0$

Proof: This proof follows exactly the similar line of argument as for the proof of Theorem 2.1.

2.2. Asymptotic Normality of $p_{n,1}(x)$ and $p_{n,2}(x)$

We now consider the limiting distribution of BJ estimators $p_{n,j}(x)$, j=1,2 of p using Lyapunov CLT to the sequence $\{Z_{2i}(x)\}$ of independent random variables in the following Theorem.

Theorem 2.2: Under the conditions AI – AIII on $\{F_i\}$, the kernel function k, and the sequence $\{a_n\}$ for each fixed x,

$$\sqrt{n}\ (p_{n,2}(x) - p) \xrightarrow{L} N(0, \tfrac{\tau^2}{d_{12}^2(x)}) \text{ as } n \to \infty$$

where $\tau^2 = \lim_{n\to\infty}[\bar{F}_n(x)(1 - \bar{F}_n(x)) - V_{nF}(x)]$, $V_{nF}(x) = n^{-1}\sum(F_i(x) - \bar{F}_n(x))^2 > 0$

Proof: Note from (2.4)

$$d_{12}(x)[p_{n,2}(x)-p] = n^{-1}\textstyle\sum_{i=1}^{n} Z_{2i}(x) + \tau_{n2}(x)$$

with $\tau_{n2}(x) = n^{-1}\sum_1^n$ [E K($\frac{x-X_i}{a_n}$) – H(x)]

$= \bar{H}_n(x) - H(x) + \frac{a_n^2}{2}\bar{H}_n^{(2)}(x)\,\mu_2(K) + o(a_n^2) \to 0$ as $n\to\infty$

$Z_{2i}(x) = K(\frac{x-X_i}{a_n}) - E\,K(\frac{x-X_i}{a_n})$, $|Z_{2i}(x)| \le 2\,\|K\| = M < \infty$

$\sigma_{2i}^2 = Var\ Z_{2i}(x)$

$\frac{s_{n2}^2}{n} = \sum\frac{\sigma_{2i}^2}{n} = n^{-1}\sum[F_i(x)(1-F_i(x)) - a_n F_i^{(1)}(x)\psi_1(K) + O(a_n^2)]$

$= \bar{H}_n(x)(1-\bar{H}_n(x)) - V_{nF}(x) - a_n\bar{H}_n^{(1)}(x)\psi_1(K) + O(a_n^2/n)$

$s_{n2}^2 = O(n)$

In order to apply Lyapunov CLT to the sequence $\{Z_{2i}(x)\}$, consider the Lyapunov condition

$$\frac{1}{s_{n2}^3}\sum_1^n E\,|Z_i|^3 = \frac{n}{s_{n2}^3}[n^{-1}\sum_1^n E\,|Z_{2i}(x)|^3]$$

$$= O(\frac{n}{n^{3/2}}) \to 0 \text{ as } n\to\infty.$$

Now, Lyapunov condition is satisfied, and Lyapunov CLT to the sequence $\{Z_{2i}(x)\}$ holds. As $\tau_{n2}(x)\to 0$ as $n \to \infty$

$$n^{-1/2}\sum_1^n Z_{2i}(x) \xrightarrow{L} N(0,1)$$

$$\frac{s_{n2}}{n^{1/2}} \to [H(x)(1-H(x)) - V(x)]^{1/2} = \tau$$

where $V(x) = \lim_{n\to\infty} V_F(x) = \lim_{n\to\infty} n^{-1}\sum(F_i(x) - \bar{F}_n(x))^2$

$$\text{i.e. } d_{12}(x)\sqrt{n}\,(p_{n,2}(x) - p) \xrightarrow{L} N(0, \tau^2)$$

$$\sqrt{n}\,(p_{n,2}(x) - p) \xrightarrow{L} N(0, (\tau/d_{12}(x))^2)$$

Thus, the Theorem is proved.

Corollary 2.2: Under the conditions of Theorem 2.1 on $\{F_i(x)\}$,

$$\sqrt{n}\,(p_{n,1}(x) - p) \xrightarrow{L} N(0, \tfrac{\tau^2}{d_{12}^2(x)}) \text{ as } n \to \infty$$

Proof: This proof follows exactly the similar line of argument as for the proof of Theorem 2.2. ■

2.3. Rates of strong convergence of $p_{n,1}(x)$ and $p_{n,2}(x)$

We now establish a.s. convergence of the BJ type estimators $p_{n,1}(x)$ and $p_{n,2}(x)$ defined in (1.8) under non-i.i.d. set-up in the following result:

Theorem 2.3: Under the conditions of Theorem 2.1,

I. $p_{n,2}(x) - p = O(\frac{\log n}{n})^{\frac{1}{2}}$ **a.s.**

II. $\hat{\lambda}_n(x) = \hat{H}_n(x) - H(x) = O(\frac{\log n}{n})^{\frac{1}{2}}$ **a.s. as n** $\rightarrow \infty$

Proof: Note that from (2.4) and (2.8)

$$E(Z_{2i}^2(x)) = \sigma_{2i}^2 = F_i(x)(1-F_i(x)) - a_n F_i^{(1)}(x)\psi_1(K) + O(a_n^2)$$

$$\leq \frac{1}{4} + |\psi_1(K)| = C < \infty$$

$$\sigma_{n2}^2 = \frac{1}{n}\sum E[Z_{2i}^2(x)] = \frac{\sum F_i(x)(1-F_i(x))}{n} - \frac{a_n\psi_1(K)\sum F_i^{(1)}(x)}{n} + O(a_n^2)$$

$$= \bar{H}_n(x)(1-\bar{H}_n(x)) - V_{nF}(x) - a_n\bar{H}_n^{(1)}(x)\psi_1(K) + O(a_n^2) < \infty$$

By applying Bernstein (1946) inequality to $\{Z_{2i}(x)\}$ with M=2,

$$P(n^{-1}\sum Z_{n2}(x) > t) \leq \exp\left(-\frac{\frac{nt^2}{2}}{C+\frac{2}{3}t}\right)$$

setting $t = \left(\frac{4C\log n}{n}\right)^{\frac{1}{2}}$

$$P(n^{-1}\sum Z_{n2}(x) > t) \leq \exp\left[-\frac{\frac{n4C\log n}{2n}}{C+\frac{2}{3}\left(\frac{4C\log n}{n}\right)^{\frac{1}{2}}}\right]$$

$$= \exp\left[-\frac{\frac{n4C\log n}{2n}}{C\left(1+\frac{2}{3C}\left(\frac{4\log n}{n}\right)^{\frac{1}{2}}\right)}\right]$$

$$= \exp\left[\frac{-2\log n}{1+\frac{2}{3C}\left(\frac{4\log n}{n}\right)^{\frac{1}{2}}}\right]$$

$\leq n^{-2}$ for sufficiently large.

$\Rightarrow \sum_{n\geq 1}^{\infty} P(\bar{Z}_{n2} > t) \leq \sum_{n\geq 1}^{\infty} n^{-2} < \infty$

By Borel–Cantelli lemma, we conclude that $\bar{Z}_{n2} = O(\frac{logn}{n})^{1/2}$ as $n\rightarrow\infty$. Therefore, as $\tau_{n2}(x)\rightarrow 0$ as $n \rightarrow \infty$,

$$(p_{n,2}(x) - p)\, d_{12}(x) = \bar{Z}_{n2} + \tau_{n2} \overline{\overline{a.s.}}\, O(\frac{logn}{n})^{1/2}$$

i.e. $p_{n,2}(x) - p = O(\frac{logn}{n})^{1/2}$ a.s. for each x as $n\rightarrow\infty$.

(II) is an immediate consequence of part(I). Hence the result follows.

Corollary 2.3: Under the conditions of Theorem 2.1 on $\{F_i(x)\}$,

I. $p_{n,1}(x) - p = O(\frac{\log n}{n})^{\frac{1}{2}}$ **a.s.**

II. $\tilde{\lambda}_n(x) = \tilde{H}_n(x) - H(x) = O(\frac{\log n}{n})^{\frac{1}{2}}$ **a.s. as n** $\rightarrow \infty$

Proof: The proof follows exactly the similar line of argument as for the proof of Theorem 2.3.

3. Optimal bandwidth $a_{n,opt}$

We select the optimal $a_{n,opt}$ as that a_n for which MSE ($p_{n,2}$(x)) is the minimum. Solving the equation $\frac{\partial \text{MSE}\,(p_{n,2}(\text{x}))}{\partial x}$ = 0 for a_n;

$$\text{i.e. M} = \text{MSE}(p_{n,2}(\text{x})) = \frac{(\bar{H}_n(x)[1-\bar{H}_n(x)]-v_{nF}(x))}{n} - \frac{a_n}{n}\,\xi_{n,1}(\text{x}) + a_n^4\,\xi_{n,2}^2(\text{x})$$

$$\frac{\partial \text{M}}{\partial a_n} = 0 = -\frac{1}{n}\,\xi_{n,1}(\text{x}) + 4a_n^3\,\xi_{n,2}^2(\text{x})$$

so that

$$a_{n,opt} = \left[\frac{\xi_{n,1}(x)}{4\xi_{n,2}^2(\text{x})}\right]^{1/3} \cdot n^{-1/3} \tag{3.1}$$

where $\xi_{n,1}(\text{x}) = \bar{H}_n^{(1)}(\text{x})\psi_1(\text{K})$, $\xi_{n,2}^2(\text{x}) = \frac{1}{4}\,\bar{H}_n^{(2)^2}(\text{x})\mu_2^2(\text{K})$, $\psi_1(\text{K}) = 2\int t\text{K}(t)d\text{K}(t)$

4. Comparisons between the estimators

We first compare the performance of the proposed smoothed estimator $p_{n,2}$(x) with Boes-James type estimator $p_{n,1}$(x), when H_1(x), H_2(x) are known based on the minimum mean square error (MSE) criterion under non-i.i.d. set-up. Note that from Theorem 2.1 and Corollary 2.1,

$$\text{MSE}\,(p_{n,2}(\text{x})) < \text{MSE}\,(p_{n,1}(\text{x}))$$

$$\text{If } \frac{a_n}{n}\,\bar{H}_n^{(1)}(\text{x})\psi_1(\text{K}) > \frac{a_n^4}{4}\,\bar{H}_n^{(2)^2}(\text{x})\mu_2^2(\text{K})$$

$$\text{If } a_n\bar{H}_n^{(1)}(\text{x})\,\psi_1(\text{K}) > \text{n}a_n^4\,[\frac{1}{4}\,\bar{H}_n^{(2)^2}(\text{x})\mu_2^2(\text{K})] \tag{4.1}$$

for finite values of n. Since both terms on the left side of the above inequality are always positive and in view of $\text{n}a_n^4 \to 0$ for moderate n, (4.1) holds. The gain in precision of $p_{n,2}$(x) over $p_{n,1}$(x) is defined as

$$\frac{\text{MSE}\,(\text{p}_{\text{n},1}(x))-\text{MSE}\,(p_{n,2}(\text{x}))}{\text{MSE}(p_{n,1}(\text{x}))} \times 100.$$

5. Monte Carlo Simulation

A simulation study is carried out in the estimation of p by $p_{n,j}$(x); j=1,2 when two component distributions are known and are estimated by using empirical distribution function and kernel distribution function for Normal and Exponential populations. The procedure is given in appendix A.

Table 5.1. Simulation results of $p_{n,j}(x_0)$ and $p_{n,j}(x_0)$ for different sets N of sample size n with p=0.5 and X_0 = -2, -1, -0.5, 0.5, 1, 2 and X_0 = 0.2, 0.3, 0.33, 0.4, 0.5, 0.6 for Exponential population

p=0.5	N	H_1(x)=N(βt_{1i},0.5^2), H_2(x)= N(βt_{2i},3^2), H(x)=N(βt_i,(2.151)^2)					p=0.5	H_1(x)=Exp(2), H_2(x)=Exp(3), H(x)=Weibull(1.25,k=0.5)				
		$p_{n,1}(x_0)$	$p_{n,2}(x_0)$	$\widehat{MSE}$		Efficiency		$p_{n,1}(x_0)$	$p_{n,2}(x_0)$	$\widehat{MSE}$		Efficiency
				$p_{n,1}(x_0)$	$p_{n,2}(x_0)$					$p_{n,1}(x_0)$	$p_{n,2}(x_0)$	
X_0=-2 n =12	10	0.538	0.119	0.026	0.006	78.03	X_0=0.2 n=12	0.69	0.76	0.11	0.02	80.11
	25	0.525	0.140	0.035	0.011	70.04		0.67	0.76	0.12	0.03	70.94
	50	0.525	0.144	0.039	0.013	66.12		0.58	0.74	0.12	0.04	61.31
	75	0.516	0.148	0.044	0.014	67.36		0.54	0.76	0.11	0.04	59.82
	100	0.531	0.154	0.039	0.014	64.12		0.50	0.75	0.10	0.05	48.81
X_0=-1 n=12	10	0.369	0.228	0.051	0.017	67.21	X_0=0.3 n=12	0.1889	0.865	0.055	0.011	79.59
	25	0.431	0.248	0.059	0.015	74.57		0.211	0.874	0.062	0.010	83.34
	50	0.383	0.249	0.054	0.024	55.66		0.211	0.868	0.062	0.012	80.19
	75	0.377	0.252	0.047	0.025	47.14		0.211	0.869	0.062	0.012	81.22
	100	0.376	0.249	0.048	0.024	49.31		0.211	0.867	0.062	0.012	80.19
X_0= -0.5 n=12	10	0.460	0.300	0.106	0.038	64.14	X_0=0.33 n=12	0.57	0.86	0.1810	0.0614	66.09
	25	0.362	0.265	0.081	0.032	60.68		0.32	0.87	0.0680	0.0127	81.37
	50	0.405	0.257	0.078	0.039	49.74		0.34	0.89	0.0734	0.0104	85.83
	75	0.402	0.255	0.074	0.039	45.84		0.33	0.89	0.0717	0.0082	88.62
	100	0.396	0.248	0.077	0.038	49.66		0.35	0.88	0.0757	0.0115	84.84
X_0=0.5 n=12	10	0.463	0.388	0.041	0.033	19.42	X_0=0.4 n=12	0.50	0.85	0.08	0.02	73.61
	25	0.446	0.360	0.042	0.028	32.43		0.40	0.89	0.07	0.01	83.50
	50	0.438	0.342	0.051	0.031	38.39		0.34	0.90	0.05	0.01	85.67
	75	0.475	0.335	0.066	0.032	51.79		0.30	0.90	0.04	0.01	75.67
	100	0.461	0.325	0.068	0.033	50.95		0.28	0.90	0.03	0.01	65.91
X_0=1 n=12	10	0.58	0.16	0.081	0.021	73.55	X_0=0.5 n=12	0.47	0.35	0.08	0.05	42.27
	25	0.52	0.16	0.067	0.017	74.28		0.49	0.47	0.09	0.07	19.86
	50	0.51	0.19	0.061	0.021	65.92		0.39	0.40	0.08	0.06	26.23
	75	0.47	0.17	0.061	0.022	64.30		0.44	0.43	0.08	0.07	21.28
	100	0.47	0.18	0.056	0.024	57.39		0.40	0.41	0.08	0.06	24.23
X_0=2 n=12	10	0.439	0.127	0.045	0.012	73.97	X_0=0.6 n=12	0.071	0.879	0.034	0.011	67.70
	25	0.512	0.135	0.036	0.012	67.50		0.083	0.836	0.039	0.025	36.48
	50	0.525	0.169	0.039	0.022	43.46		0.071	0.822	0.034	0.028	17.11
	75	0.494	0.163	0.053	0.024	53.91		0.075	0.827	0.036	0.027	24.21
	100	0.518	0.153	0.047	0.021	55.75		0.071	0.822	0.034	0.028	17.16

6. Comments

In the paper it is shown that when the component normal populations with parameters are $N(\beta t_{1i}, 0.25)$, $N(\beta t_{2i}, 9)$ and the mean value of estimate $p_{n,1}(x_0)$ and $p_{n,2}(x_0)$ is close to its actual value p**.** The simulation results show that $\widehat{MSE}$ for smoothed parametric estimator is less than that of unsmoothed estimator for different values of x, uniformly for all samples. Thus, the smoothed estimator is a better estimator in terms of minimum MSE when compared to BJ estimator. The average gain in efficiency due to smoothing lies between 19% to 88% for different sets N of size n.

REFERENCES

ACHESON, M. A., McELWEE, E. M., (1951). Concerning the reliability of electron tubes, Proceedings of the IRE, Vol. 40, pp. 1204–1206.

BERNSTEIN, S. N., (1946). The Theory of Probabilities, Gastehizdat Publishing House, Moscow.

BOES, D. C., (1966). On the estimation of mixing distributions, Ann. Math. Statist. 37, pp. 177–188.

BLISCHKE, W. R., (1965). Mixtures of discrete distributions, Collected papers presented at the Int. Simp. Ed. By G. P. Patil, Oxford, Pergamon.

BRUNI, C. et al., (1985). On the inverse problem in cytofluorometry-recovering DNA distribution from FMF date, Cell Bio. Physics 5, pp. 5–19.

CHOI, K., BULGREN, W. G., (1968). An estimation procedures for mixtures distributions, J.R.S.S. Series B (Methodological), Vol. 30, No. 3, pp. 444–460.

CLARK, M. V., (1976). Some methods for statistical analysis of multi-model distributions and their application to grain-size data, J. Math. Geol. 8, pp. 267–282.

CHRISTENSEN, P. R. et al., (1980). Gamma-ray multiplicity moments from 86 kr reactions on 144, 154 at 490, Mev. Nuclear Phys. A 349, pp. 217–257.

DAVIS, D. J., (1952). An analysis of some failure date, Jour. Amer.Stat.Assoc., 47, pp. 113–150.

EPSTEIN, B., (1953). Statistical problems in life testing, proceeding of The Seventh Annual Convention, American Society of Quality Control.

EVERITT, B., HAND, D., (1981). Finite mixture distributions. Chapman and Hall, London.

FU, K. S., (1968). Sequential methods in pattern recognition and machine learning, Academic Press, New York.

HARRIS, E. K., (1958). On the probability of survival of bacteria in sea water, Biometrics, 14, pp. 195–206.

HERD, G. R., (1953). Heterogeneous distributions. Unpublished technical note.

HOSMER W. DAVID, (1973). A comparison of iterative maximum likelihood estimates of the parameters of a mixture of two normal distributions under three different types of sample. Biometrics, 29, pp. 761–770.

JAMES, I. R., (1978). Estimation of the Proportion in a mixture of two normal distributions from simple, rapid measurements, Biometrics, pp. 265–275.

JAYALAKSHMI, C., SUDHAKAR RAO, M., (2002). Estimation of mixing proportion using smooth distribution function in two population mixture model, Aligarh Jl. of Statistics, Vol. 22, pp. 91–99.

JEWEL, N. P., (1982). Mixture of Exponential distributions, Annals of Statistics, Vol. 10, No. 2, pp. 479–484.

KARL PEARSON, (1894). Contributions to the mathematical theory of evolution, Phil. Trans. Roy Soc. 185 A, pp. 71–110.

LINDSAY, B. G., (1995). Mixture models: Theory geometry and applications. NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 5, Institute of Mathematical Statistics, Hayward.

MENDENHALL, W., HADER, R. J., (1958). Estimation of parameters of mixed exponentially distributed failure time distributions from censored life test data, Biometrika 45(3-4), pp. 504–520.

MACDONALD, P. D. M., PITCHER, T. J., (1979). Age groups from size frequency data, J. Fish. Res. Bd. Cano. 36, pp. 987–1008.

McLACHLAN, G. J., BASFORD, K. E., (1988). Mixture Models: Inference and Applications to Clustering. New York: Marcel Dekker.

McLACHLAN, G. J., PEEL, D., (2000). Finite mixture models, Wiley, New York.

MERZ, P. H., (1980). Determination of absorption energy distribution by regularization and a characterization of certain absorption isotherms, J. Comput. Phys. 34, pp. 64–85.

STEEN, J. R., (1952). Life testing of electronic tubes, Unpublished paper presented at a summer statistics conference, University of North Carolina.

TITTERINGTON, D. M. et al., (1985). Statistical analysis of finite mixture distributions. John Wiley & Sons.

WILDE, R. D., (1952). Nature of rate of failure curves, Unpublished paper presented at a summer statistics conference, University of North Carolina.

ODELL, P. L., BASU, J. P., (1976). Concerning several methods for estimations crop acreages using remote sensing data, Commun. Statist. A 5, pp. 1091–1114.

VAN HOUWELINGEN, J. C., DEVRIES, (1974). Minimax estimation of mixing proportion of two known distributions, J. American Stat. Asso., 397, pp. 300–304.

VARLI, Y. et al., (1975). A statistical model for position emission tomography, JASA, 80, pp. 8–20.

APPENDIX A

Random samples of sizes n_1=6 and n_2=6 are generated from the two component mixtures of the normal populations with parameters $(\mu_1,\mu_2)=(\beta t_{1i},\beta t_{2i})$ and $(\sigma_1^2,\sigma_2^2)=(0.5^2,3^2)$ and with parameters (θ_1,θ_2) = (2,3) for Exponential populations. The mixed sample of size n = n_1+n_2 = 12 is generated from the normal population with parameters ($\mu = p\mu_1+q\mu_2$) = ($\beta t_i=p\beta t_{1i}+q\beta t_{2i}$) and σ^2 = ($p\sigma_1^2+q\sigma_2^2$), and in the case of Exponential population, the mixed sample is drawn from Weibull population with shape parameter k less than 1. Since Weibull distribution with shape parameter k<1 arises as a mixture of Exponential distributions (Jewel 1982), the samples of sizes n, n_1, n_2 are independent. Taking p=q= 0.5, β=0.1 and $t_i = \mp \frac{i}{n^{\delta}}$; j = 1, 2, δ =1.5 are selected in such a way that $\sum t_i$ = 0 and $\sum t_i^2 \rightarrow 0$. The present simulation study is to estimate parametric estimators such as $p_{n,1}(x)$ and $p_{n,2}(x)$ for $x=x_0$ defined as follows.

$$\mathbf{p}_{n,1}(\boldsymbol{x_0}) = \frac{\widetilde{\mathbf{H}}_{\mathbf{n}}(\boldsymbol{x_0})-\mathbf{H_2}(\boldsymbol{x_0})}{\mathbf{H_1}(\boldsymbol{x_0})-\mathbf{H_2}(\boldsymbol{x_0})} \quad \text{and} \quad \mathbf{p}_{n,2}(\boldsymbol{x_0}) = \frac{\widehat{\mathbf{H}}_{\mathbf{n}}(\boldsymbol{x_0})-\mathbf{H_2}(\boldsymbol{x_0})}{\mathbf{H_1}(\boldsymbol{x_0})-\mathbf{H_2}(\boldsymbol{x_0})}$$

where $\widetilde{\mathrm{H}}_{\mathrm{n}}(x_0)$, $\widehat{\mathrm{H}}_{\mathrm{n}}(x_0)$ are estimated by the usual empirical and kernel-based distribution functions and $H_j(x)$, j=1,2 such as

$$\widetilde{\mathrm{H}}_{\mathrm{n}}(x) = \frac{1}{\mathrm{n}} \sum_{i=1}^{n} I(X_i \le x) \text{ , if } I(X_i \le x), \text{ assign 1 otherwise 0.}$$

$$\hat{H}_{\mathrm{n}}(x_0) = \mathrm{n}^{-1} \sum_{i=1}^{n} K\left(\frac{x_0 - X_i}{a_n}\right); \; H_j(x) = \frac{1}{n_j} \sum_{i=1}^{n_j} F_{ji}(x_0).$$

Here, we used the Epanechnikov kernel function as $k(u) = \frac{3}{4}(1 - u^2)$; $|u| \le 1$ for Normal distribution, and for Exponential distribution we used Gaussian kernel function as $k(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$. The distribution function of the Epanechnikov kernel function is

$$K\left(\frac{x_0 - X_i}{a_n}\right) = \frac{3}{4}\left[\frac{x_0 - X_i}{a_n} - \frac{1}{3}\left(\frac{x_0 - X_i}{a_n}\right)^3 + \frac{2}{3}\right]$$

Thus, the estimators becomes

$$p_{n,1}(x_0) = \frac{\frac{1}{n}\sum_{i=1}^{n} I(X_i \le x_0) - \frac{1}{n_2}\sum_{i=1}^{n_2} F_{2i}(x_0)}{\frac{1}{n_1}\sum_{i=1}^{n_1} F_{1i}(x_0) - \frac{1}{n_2}\sum_{i=1}^{n_2} F_{2i}(x_0)}$$

$$p_{n,2}(x_0) = \frac{n^{-1}\sum_{i=1}^{n} K\left(\frac{x_0 - X_i}{a_n}\right) - \frac{1}{n_2}\sum_{i=1}^{n_2} F_{2i}(x_0)}{\frac{1}{n_1}\sum_{i=1}^{n_1} F_{1i}(x_0) - \frac{1}{n_2}\sum_{i=1}^{n_2} F_{2i}(x_0)} \qquad (5.1)$$

where $F_{ji}(x)$ are the cumulative distribution functions of Normal distribution and Exponential distribution.

All computations are done by using MS Excel, and the procedure is as follows.

1. Generate n uniform random numbers between (0,1).
2. Generate the cumulative distribution function of Normal and Exponential distribution by taking different means βt_{ji} and variances σ_j^2; j=1,2 at x = x_0
3. Generate the mixed normal and Weibull observations by taking different means $\beta t_i = p\beta t_{1i} + q\beta t_{2i}$ and variance $\sigma^2 = p\sigma_1^2 + q\sigma_2^2$ and $\theta = p\theta_1 + q\theta_2$ respectively with k<1.
4. Calculate (5.1) by taking X_0 = -2, -1, -0.5, 0.5, 1, 2 related to Normal distribution and X_0= 0.2, 0.3, 0.33, 0.4, 0.5, 0.6 related to Exponential distribution.

Generate N=100 mixed and component independent sample sets with sample sizes n=12 and n_1= 6 and n_2=6, so that $n = n_1+n_2$ and calculate $p_{n,1}(x_0)$ and $p_{n,2}(x_0)$, their mean values $\bar{p}_{n,j}(x_0) = \frac{1}{N}\sum p_{n,j}(x_0)$ and their mean square errors $\overline{MSE}\ (p_{n,j}(x_0)) = \frac{1}{N}\sum_{i=1}^{N}(p_{n,j}(x_0) - \bar{p}_{n,j}(x_0))^2$; j=1,2. Sets are ignored when p ≥1. The results are presented in table 5.1.