Nonparametric Kernel Estimation of Hazard and Density Functions from Duration of Breastfeeding Data

Tapan Kumar Chakrabarty
Department of Statistics
North Eastern Hill University
Shillong- 793022

Abstract

A number of new methods of analyzing time to occurrence variables for events of interest in demography e.g., marriage, mortality, birth, leaving parental home, postpartum amenorrhoea, breastfeeding etc. have been developed in the recent years using the World Fertility Survey (WFS) and the Demographic and Health Surveys (DHS) data. These methods rely upon retrospective information from life or birth histories and recollections of past events. Retrospective information of the sort is known to be affected by recall errors which result in the omission of events, the misplacement of dates, and the distortion of reports of duration. For example, analysis of breastfeeding information using retrospectively reported ages of weaning for all births that occurred during the three or five years preceding the survey commonly display marked heaping at durations 6, 12, and 18 months. The present article proposes to use a nonparametric kernel estimation procedure to obtain a smooth estimate of hazard function based on retrospectively reported duration data that can address the problem of heaping due to recall errors. Following Ramlau-Hansen (1983), smooth estimate of hazard of weaning is obtained by smoothing the increments of Nelson-Aalen (NA) cumulative hazard function estimate and illustrated using duration of breastfeeding data for six North Eastern states of India from the last two National Family and Health Surveys. Approximate bias and 95% confidence interval for these estimates are also obtained using their respective asymptotic expressions. Further, under additive error model, a kernel-type deconvolving density estimator (Wand and Jones, 1995) of durations of breastfeeding is proposed by smoothing the increments of Kaplan-Meier (KM) cumulative distribution function. Using simulated data it has been shown that in small and moderately censored samples these estimators can reduce the bias substantially.

Keywords: Life Table, Survival Distribution, Smoothing, kernel.

1. INTRODUCTION

During the past two decades studies on health and fertility implications of breastfeeding pattern have received much attention. The diverse sources of evidences suggest that, at least in developing countries, breastfeeding has non-trivial beneficial effects on infant health and survival (Briend et al, 1988; Cabigon, 1997; Huffman et al 1984; Majumder, 1991; Nath et al 1994). Increased risks of mortality and morbidity for children who are followed by other births after very short interval and hence are breastfed for shorter durations have been supported by diverse data sources (Cabigon, 1997; Goldberg et al, 184; Palloni et al 1986; Ratherford et al, 1989).

The duration and intensity of breastfeeding have significant effects on both mothers and children. These effects vary by both the duration and intensity of breastfeeding. Previous studies show that the countries in which there has been a decline either in initiation or duration of breastfeeding include Brazil, Thailand, Taiwan, South Korea and Mexico(Sousa et al, 1975, Chayovan et al, 1990; Milman, 1986). Studies in India have also shown a decline in breast-feeding trends, especially in urban areas (Chhabra et al, 1998). National Family Health Surveys in India reveal that breastfeeding is nearly universal and there is an increase of one month in the median duration of breastfeeding over a period of five years (IIPS, 1995; IIPS and ORC Macro, 2000).

Prompted by the reported decline in the average durations and proportions of women initially breastfeeding, WHO (1991) recommended to the developing countries initiation of breastfeeding soon after birth and only breast milk to babies up to 4–6 months of age. Under the Reproductive and Child Health Programme, the Government of India recommends that infants should be exclusively breastfed from birth to age four months. Most babies do not require any other foods or liquids during this period. By age seven months, adequate and appropriate complementary foods should be added to the infant's diet in order to provide sufficient nutrients for optimal growth. It is recommended that breastfeeding should continue, along with complementary foods, through the second year of life or beyond.

There is a great deal of variability in the findings on breastfeeding differentials and trends in developing countries which are not always in agreement and consistent. Estimates of average duration of breastfeeding , for example, can be quite sensitive to the data and the method of analysis employed. Generally, the estimates referred to are based on retrospective information from birth histories and on mothers' recollection of past breastfeeding behaviour. Retrospective information of the sort is known to be affected by recall errors which result in the omission of events, the misplacement of dates, and the distortion of reports of duration. Yet another factor justifies caution is that these findings may be model specific, i.e., they could be sensitive to the choice of model for estimates of parameters.

The present article proposes a set of kernel-type estimators for hazard and density function of breastfeeding duration based on retrospectively reported data that can address the shortcomings of the existing estimators. In section 2, I discuss the sources of data used in the present work and some of the methodological issues in analyzing the information on breastfeeding from large scale health surveys. Kaplan-Meier and Nelson-Aalen estimates of survival functions are obtained in section 3 using data from the last two National Family Health Surveys for six north-eastern states in India. In section 4, I propose a kernel-type estimator of hazard of weaning and obtain the large sample bias corrected estimates and confidence intervals. The last section is devoted to obtain a kernel-type deconvolving density estimator of durations of breastfeeding under non-negligible additive type recall error.

2. Data and Methodological Issues

For the present analysis, the information on breastfeeding in six north-eastern states of India: Manipur, Meghalaya, Mizoram, Nagaland, Arunachal Pradesh and Tripura have been pooled that are available from two different National Family Health Surveys (IIPS, 1995, 2000). The first National Family Health Survey (NFHS-1) conducted during February to June 1993, gathered information on a representative sample of 6266 ever married women aged 13 – 49. Whereas the NFHS-2, conducted during May 1999 to June 2000, gathered information on 6467 ever-married women aged 15 – 49. Information of breastfeeding was collected for the children of interviewed women born in the four years preceding the survey of NFHS-1 and three years preceding NFHS-2. For any given woman, a maximum of three births were included in the analysis for NFHS-1 whereas a maximum of only two births were included in the analysis of NFHS-2. For a total of 3525 children information on breastfeeding duration were collected for NFHS-1 of whom 1914 cases were still breast feeding at the date of interview or were breastfed until died. For NFHS-2, information on duration of breastfeeding on a total of 2792 children were collected of whom 2052 cases were still breastfeeding at the date of interview or have breastfed their child until died. The rest of the children completed their breast feeding. The duration of breastfeeding for the births who are still breastfed is calculated as the difference between their birth dates and the date of the survey.

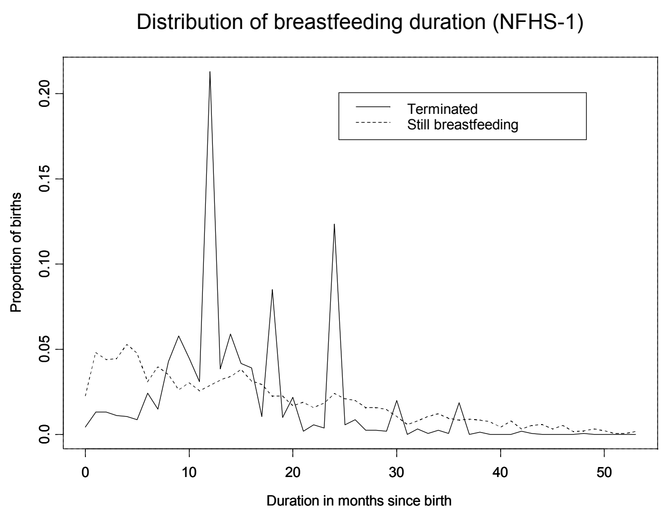
In an excellent review, Trussell et al (1992), argued that information on breastfeeding collected in both the World Fertility Survey (WFS) and the Demographic and Health Surveys (DHS) can be analysed in two ways. The *current-status* information (yes/no) on whether a woman was still breastfeeding her most recent child at the time of survey, can be used to compute the proportion of children who are still breastfed by single month of age at interview. Alternatively, retrospective reported ages of weaning for children who are no longer being breastfed can be used. The authors maintained that the current-status measures lead to unbiased estimates of the survival function for a sample of births that occur during a fixed period. The estimates of the survival function so obtained, however, often fail to decline monotonically and are subject to larger variations. On the other hand, retrospectively reported ages of weaning commonly display marked heaping at durations 6, 12, 18 and 24 months. Of course, such heaping may be genuine, and may reflect societal norms about appropriate weaning times. However, their comparison with the durations computed from the births, who are currently breastfeeding, show clearly that such heaping is less evident in the latter. We illustrate this by plotting the proportion of births for both the status category: children who are weaned and children who are still breastfed against their age in months in figure-1. Furthermore, data for the last closed interval represent, at least partially, a self selected group of short breast feeders and thus add selection bias to the existing reporting bias. In this study, I have chosen to analyse the breastfeeding information for all births that occurred during a fixed time period preceding the survey so that the results are not bias upward by excluding births from women with short birth intervals.

The information on breastfeeding duration in the current open interval is complete only for the children who were already weaned at the time of interview while it is not complete for those who were breastfeed until death and were breastfeeding at the time of interview. So for the analysis of breastfeeding duration, the durations which are not complete are regarded as right censored.

3. Estimation of Survival Function

We consider non-parametric procedure for estimating the probability S(x) of surviving to time x, using a random sample $X_1, X_2, \ldots, X_n$ of death times from a distribution F(x). The

Figure-1



X_i are censored on the right by random variables C_i, so that one observes only min(X_i, C_i) = Y_i, i=1,…..,n. The C_i's are are a random sample, drawn independently of the X_i, from a distribution G(c). We let $Y_{(1)} \leq \ldots \leq Y_{(n)}$ denote the ordered observations, and let $\delta_i = I\left(X_{(i)} \leq C_{(i)}\right)$ be an indicator for the event that $Y_{(i)}$ is uncensored.

Kaplan and Meier (1958) developed the nonparametric estimator of S(x) as

$$\hat{S}_{KM}(x) = \prod_{Y_{(i)} \leq x} \left(\frac{r_i - d_i}{r_i}\right)^{\delta_i} \qquad \ldots \quad \ldots(3.1)$$

where r_i = # alive at time $Y_{(i)}$- , d_i = # died at time $Y_{(i)}$. The Nelson-Aelen estimator of the survival function is

$$\hat{S}_{NA}(x) = e^{-\hat{\Lambda}(x)} \quad \text{with } \hat{\Lambda}(x) = \sum_{Y_{(i)} \leq x} \delta_i \frac{d_i}{r_i}. \qquad \ldots \quad \ldots(3.2)$$

Greenwood's formula for the variance of the survival function

$$\hat{V}(\hat{S}(x)) = (\hat{S}(x))^2 \sum_{Y_{(i)} \leq x} \frac{d_i}{r_i(r_i - d_i)}$$

$$\text{or} \qquad \hat{V}(\ln \hat{S}(x)) = \sum_{Y_{(i)} \leq t} \frac{d_i}{r_i(r_i - d_i)} \qquad \ldots \quad \ldots \quad (3.3)$$

The endpoints of a 100(1-α)% confidence interval for S(x) on the cumulative hazard or log-survival scale is given by

$$\exp(\ln \hat{S}(x) \pm z_{1-\alpha/2} \hat{se}(\ln \hat{S}(x))) \quad \ldots \quad \ldots \qquad (3.4)$$

Figure-3.1

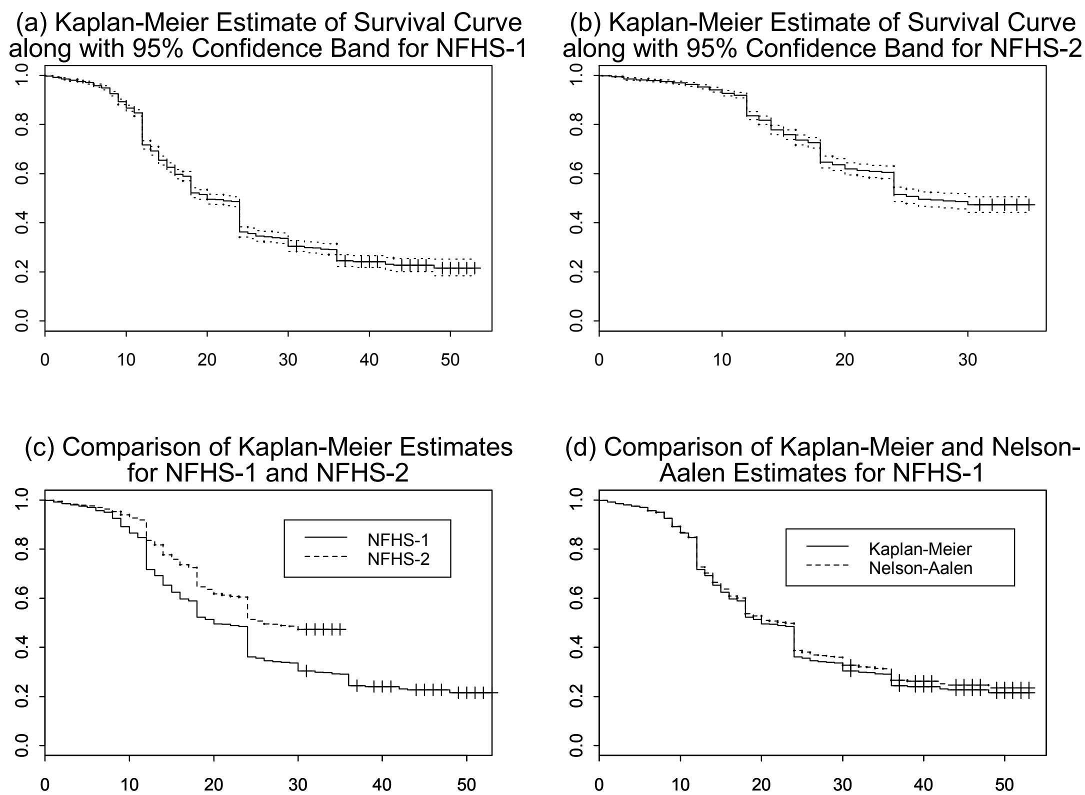


Figure 3.1 shows the plots of Kaplan-Meier and Nelson-Aalen estimates of survival function and the jumps in these estimates may be observed at times multiple of 6 months as discussed earlier. Figure 3.1 (a) and 3.1(b) provide the graphs of estimated Kaplan-Meier survival functions along with their 95% confidence band based on NFHS-1 and NFHS-2 data of breast feeding duration. Fig. 3.1(c) shows the comparison of the survival curves for NFHS-1 and NFHS-2 and a longer breast feeding experience is evident in the later survey from this comparison, the estimated survival curve for NFHS-2 lying completely above that of NFHS-1.

4. **The Kernel function Estimators of Hazard**

The estimates $\hat{S}_{KM}(t)$ and $\hat{S}_{NA}(t)$ clearly share some disadvantages in the present context. First, both of these depict jumps at durations of multiple of six months as the reported durations are coupled with recall error. Second, they do not provide a useful estimate of hazard rate $\lambda(t) = -(d/dt)\ln S(t)$, which is often of real interest.

It is natural to seek for an estimator of $\lambda(t)$ that can be motivated through the Nelson-Aalen cumulative hazard function $\Lambda(t)$. We shall first obtain a smoothed estimator of hazard function by smoothing the increments of the Nelson-Aalen cumulative hazard function $\hat{\Lambda}(t)$ defined in (2.2). Following Ramlau-Hansen (1983), and Anderson et al (1993), the kernel function estimator for the hazard function $\hat{\lambda}(t)$ may be defined as

$$\hat{\lambda}(t) = b^{-1} \int K\left(\frac{t-s}{b}\right) d\hat{\Lambda}(s)$$

where K is a bounded function vanishing outside [-1,1] and $\int_{-1}^{1} K(x)dx = 1$, which we call the kernel and b is a positive number called the bandwidth. As earlier, denoting $t_{(1)} < t_{(2)} < \ldots < t_{(n)}$ as successive jump times, $\hat{\lambda}(t)$ may equivalently be written as

$$\hat{\lambda}(t) = b^{-1} \sum_j K\left(\frac{t-t_{(j)}}{b}\right) l_j$$

where l_j is the size of the jump of $\hat{\Lambda}$ at t_j. Thus, $\hat{\lambda}(t)$ is a weighted mean of the increments l_j of the Nelson-Aalen estimator over [t-b, t+b]. The integral of $\hat{\lambda}$, which results in a smoothed estimate of cumulative hazard $\Lambda(t)$, is

$$\int_0^t \hat{\lambda}(u)du = b^{-1} \sum_j l_j \int_0^t K\left(\frac{u-t_{(j)}}{b}\right) du = b^{-1} \sum_j \alpha_j(t) l_j \,.$$

The coefficients $\alpha_j(t)$ are known functions which do not depend on l_j. The estimator $\hat{\lambda}$ is pointwise consistent and under stronger conditions it is uniformly consistent (Anderson et al, 1993).

Figure 4.1(a) contains estimates of the hazard function λ(t) for NFHS-I data on breastfeeding durations, using bandwidth b = 10 months and the following three kernels, all of which take the value 0 outside of [-1, 1]; the *triangle* kernel

$$K_T(x) = (1 - |x|), \quad -1 \le x \le 1,$$

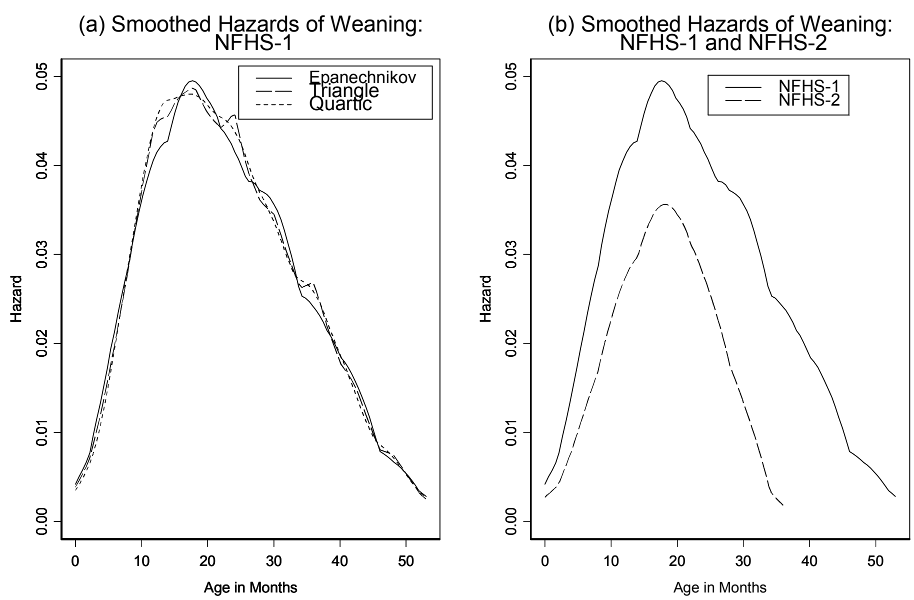
the *Epanechnikov* kernel

$$K_E(x) = \frac{3}{4}(1 - x^2), \quad -1 \le x \le 1,$$

and the *quartic* kernel

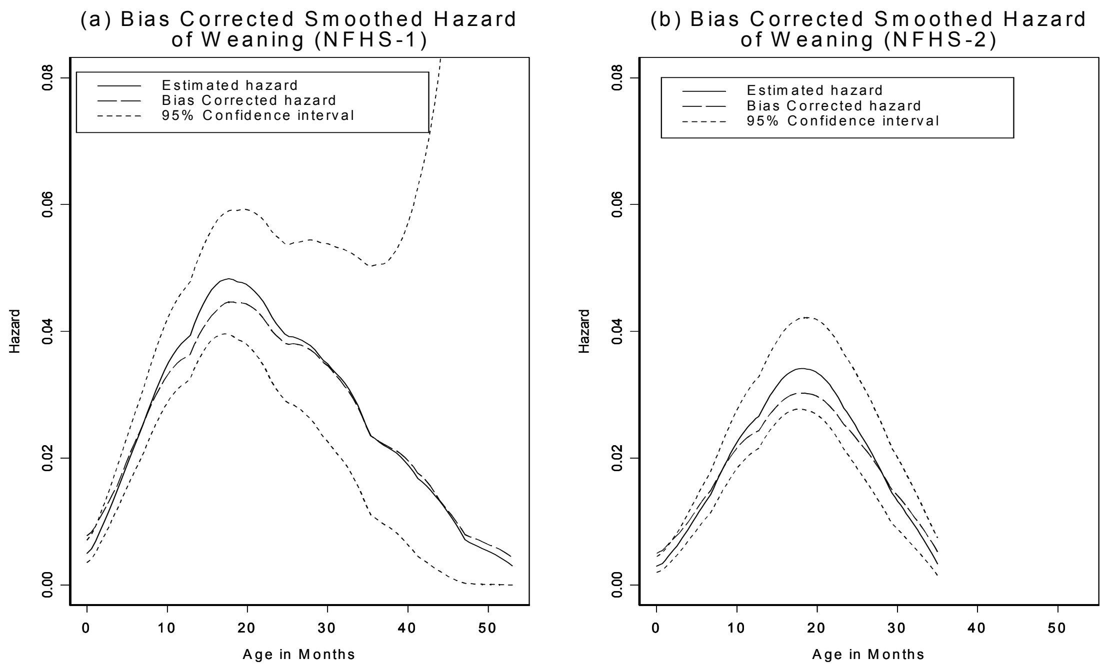
$$K_Q(x) = \frac{15}{16}(1 - x^2)^2, \quad -1 \le x \le 1.$$

Figure-4.1



It is seen that the triangle kernel gives slightly less smoothing than the other two kernel functions. But, broadly speaking, the three estimates are roughly identical, all displaying a clearly increasing weaning hazard with age up to the age of 18 months and thereafter the hazard declines sharply. Figure 4.1(b) compares the hazard rate estimated from NFHS-1 and NFHS-2 data. As it is expected, the estimated hazard is found to be substantially lower in NFHS-2. However, the overall pattern remains unchanged. The function attains its maximum at the age of 18 months and then sharply declines.

Figure 4.2



4.1 Large sample bias and confidence interval

Andersen et al (1983) has studied the asymptotic behaviou of the mean integrated squared error and optimal bandwidth for estimating $\lambda(t)$. Let $t \in \mathcal{T}$, where $\mathcal{T}$ is a fixed continuous time interval and let $0 < t_1 < t_2 < t$ be fixed numbers and that λ is twice continuously differentiable on $[t_1\text{-}c, t_2+c] \subset \mathcal{T}$ for some $c > 0$. Restricting our choice to the kernel functions satisfying

$$\int_{-1}^{1} K(t)dt = 1, \quad \int_{-1}^{1} tK(t)dt = 0, \quad \text{and} \quad \int_{-1}^{1} t^2 K(t)dt = k_2 > 0,$$

we have the following asymptotic expression for the bias of the kernel estimator

$$E\left[\hat{\lambda}(t)-\lambda(t)\right]\approx\frac{1}{2}b^2\lambda''(t)k_2$$

where $\lambda''(t)$ is the second derivative of λ(t). An estimate of the local bias using a particular kernel K and bandwidth b may, thus, be obtained as $\frac{1}{2}b^2\hat{\lambda}''(t)k_2$, where $\hat{\lambda}''(t)$ is some estimate of the second derivative of λ(t). To estimate $\lambda''(t)$, consider some kernel K_1 which is twice differentiable on the whole line. Then, we get

$$\hat{\lambda}''(t)=\frac{1}{b_1^{\,3}}\int K_1''\left(\frac{t-t_{(j)}}{b_1}\right)d\hat{\Lambda}(t).$$

Using the simplest symmetric twice differentiable kernel

$$K_1(x)=\frac{35}{32}\left(1-x^2\right)^3 I_{(-1,1)}(x)$$

we see that

$$K_1''(x)=-\frac{105}{16}\left(1-6x^2+5x^4\right)I_{(-1,1)}(x).$$

Further, the large sample variance of $\hat{\lambda}(t)$, may be calculated as

$$E\left[\hat{\lambda}(t)-\lambda(t)\right]^2\approx\hat{\tau}^2(t)=b^{-1}\sum_j\left\{\frac{K((t-t_{(j)})/b)}{r(t_{(j)})}\right\}^2 d(t_{(j)})\ .$$

Also, the 95% confidence limits for the smoothed hazard function $\hat{\lambda}(t)$ is obtained as

$$\hat{\lambda}(t)\exp\left\{\pm1.96\frac{\hat{\tau}(t)}{\hat{\lambda}(t)}\right\}.$$

Figure 4.2 provides bias corrected estimates $\hat{\lambda}(t)+\frac{1}{2}b^2\hat{\lambda}''(t)k_2$, and the 95% confidence limits for NFHS-1 and NFHS-2 data on reported age at weaning. In the present case, the influence of bias correction for the smoothed hazard estimate does not appear to be substantial. Although, 95% confidence limits are provided for these estimates, care has to be exercised in the interpretation of these limits in view of the inherent bias.

5. The Density function Estimators

A kernel density estimator of f_X can be motivated through the Kaplan-Meier estimator of the distribution function F_X, which is given by

$$\hat{F}_{KM}(x) = 1 - \hat{S}_{KM}(x) = \begin{cases} 0 & , \quad 0 \le x \le Y_{(1)} \\ 1 - \prod_{Y_{(i)} \le x} \left(\frac{r_i - d_i}{r_i} \right) & , \quad Y_{(i)} \le x \le Y_{(i+1)} \\ 1 & , \quad x > Y_{(n)} \end{cases}$$

The kernel estimator of $f_Y(x)$ induced by $\hat{F}_{KM}$ is then

$$\hat{f}_Y(x,h) = h^{-1} \int K\left(\frac{x-y}{h}\right) d\hat{F}_{KM}(y)$$

$$= h^{-1} \sum_j K\left(\frac{x - Y_{(j)}}{h}\right) s_j \qquad (5.1)$$

where s_j is the size of the jump of $\hat{F}_{KM}$ at $Y_{(j)}$.

We now assume that the observable times $Y_1, Y_2, \ldots\ldots, Y_n$ are contaminated with non-negligible recall error such that

$$W_i = Y_i + Z_i \,, \quad i = 1, \ldots\ldots n$$

and, for each i, Z_i is a random variable that is independent of Y_i and has known density f_Z, which we call the error density. If we apply the ordinary kernel estimate to the $W_1, \ldots, W_n$ then we will obtain a consistent estimate of convolution

$$f_W = f_Y * f_Z$$

rather than f_Y which we aim to estimate. Estimation of f_Y requires that we take into account the fact that it is convolved with f_Z to give the density of the error contaminated data. Thus the estimation of f_Y is a problem of deconvolution type. A kernel type solution is obtained by using Fourier transform (or characteristic function) properties and noting that

$$\phi_{f_W}(t) = \phi_{f_Y}(t)\phi_{f_Z}(t)$$

where ϕ_g is used to denote the c.f. of a density g. According to the Fourier inversion theorem, the target density can be written as

$$f_Y(y) = (2\pi)^{-1}\int e^{-ity}\phi_{f_Y}(t)dt$$
$$= (2\pi)^{-1}\int e^{-ity}\{\phi_{f_W}(t)/\phi_{f_Z}(t)\}dt$$

provided $\phi_{f_Z}(t) \neq 0 \quad \forall t$. An estimate of $f_Y(y)$ is obtained by replacing f_W by its kernel estimator $\hat{f}_W(y,h) = (nh)^{-1}\sum_{i=1}^{n} K\left(\frac{y - W_i}{h}\right)$ to obtain

$$\hat{f}_Y(y,h) = (2\pi)^{-1}\int e^{-ity}\{\phi_{\hat{f}_W(y,h)}(t)/\phi_{f_Z}(t)\}dt$$

which is the deconvolving kernel density estimator (Stefanski and Carroll, 1990). It can be shown (Wand and Jones, 1995) that the deconvolving kernel density estimator of the target density is

$$\hat{f}_Y(y,h) = (nh)^{-1}\sum_{i=1}^{n} K^Z\left(\frac{y - W_i}{h}, h\right) \quad \ldots\ldots.. \qquad (5.2)$$

where

$$K^Z(u,h) = (2\pi)^{-1}\int e^{-ity}\{\phi_K(t)/\phi_{f_Z}(t/h)\}dt \qquad (5.3)$$

ϕ_K being the characteristic function of the kernel K used in estimating $\hat{f}_W$. Thus, the kernel $K^Z\left(\cdot,\ h\right)$ is to be used for estimating f_Y instead of K. This effective kernel differs from K in that its shape depends on the bandwidth. We now use $K^Z(\cdot,h)$ of (5.3) to rewrite (5.1) as

$$\hat{f}_Y(x,h) = h^{-1}\sum_j K^Z\left(\frac{x - Y_{(j)}}{h}, h\right) s_j \qquad (5.4)$$

5.1 Simulation Study of Small sample Bias

We generate data through simulation to examine the small sample bias of the estimator $\hat{f}_Y(x,h)$ in (5.4). The effective kernel is obtained for two different error density functions. When the error density is Laplacian

$$f_Z(x) = (2\sigma)^{-1}\exp(-|x|/\sigma), \quad -\infty < x < \infty; \ \ \sigma > 0$$

and that $K(x) = \Phi(x)$, the standard normal kernel, the effective kernel for deconvolution of Laplacian error is

$$K^Z(x,h) = \Phi(x)\left\{1 + (\sigma/x)^2\left(x^2 - 1\right)\right\} . \tag{5.1.1}$$

Supposing instead that the error variable has a $N(0,\sigma^2)$ distribution, the effective kernel would be

$$K^Z(x,h) = \frac{1}{\sqrt{2\pi\left(1-\left(\sigma/h\right)^2\right)}}\exp\left(-\frac{x^2}{2\left(1-\left(\sigma/h\right)^2\right)}\right). \tag{5.1.2}$$

Further, we use $F_X(x) = 1-\exp(-x/\lambda_1)$ and $G_C(c) = 1-\exp(-c/\lambda_2)$. The choices of pair of values for (λ_1,λ_2) give rise to desired censoring proportion to indicate no censoring, moderate censoring and heavy censoring. Observations have been simulated from F_X and G_C for three different sample sizes n=30, 60, 100. The pulling mechanism to contaminate the simulated data is defined through the following function

$$f(y) = \begin{cases} 6, & 4 \le y < 9 \\ 12, & 9 \le y < 15 \\ 18, & 15 \le y \le 20 \\ 24, & 21 \le y \le 27 \\ 30, & 28 \le y \le 32 \end{cases}$$

Table 5.1 presents the results showing bias at selected time points. We denote by B_0, B_L and B_N the bias due to the density estimate $\hat{f}_Y(x)$ in (5.1) without accounting for recall error, the bias due to the Laplacian error corrected density estimate and the bias due to

Table 5.1. Small sample Bias B_0, B_L and B_N for three different sample sizes and three different censoring percentages at selected time points

n	λ_1	λ_2	π	Time	Bias		
					B_0	B_L	B_N
30	8	-	0		h = 1.8	h = 1.6, σ = 0.8	h = 2.2 ,σ = 0.9
				6	-0.009	-0.003	-0.006
				12	-0.002	0.000	-0.001
				18	-0.002	-0.001	-0.002
				24	0.000	0.000	-0.001
30	12	36	27		h = 2.0	h = 2.0, σ = 0.7	h = 2.5 ,σ = 1.0
				6	0.000	0.001	0.001
				12	-0.016	-0.009	-0.010
				18	0.009	0.013	0.012
				24	-0.013	-0.010	-0.010
				30	0.000	0.000	0.000
				36	0.000	0.000	0.000
30	14	12	53		h = 2.5	h = 2.5, σ = 1.0	h = 3.0 ,σ = 1.0
				6	0.015	0.014	0.014
				12	-0.025	-0.017	-0.020
				18	-0.003	-0.003	-0.003
				24	-0.001	-0.002	-0.002
60	18	-	0		h = 3.0	h = 3.0, σ = 1.0	h = 3.5 ,σ = 1.0
				6	0.006	0.005	0.005
				12	-0.009	-0.006	-0.006
				18	0.002	0.000	0.000
				24	-0.002	-0.001	-0.001
				30	0.002	0.001	0.001
				36	0.000	0.000	0.000
60	18	18	43		h = 2.5	h = 2.5, σ = 1.0	h = 3.5 ,σ = 1.0
				6	0.012	0.012	0.012
				12	-0.012	-0.006	-0.008
				18	0.001	-0.001	-0.001
				24	-0.005	-0.003	-0.004
				30	0.002	0.003	0.003
60	22	12	77		h = 2.0	h = 2.0, σ = 1.0	h = 2.5 ,σ = 1.0
				6	-0.002	0.005	0.001
				12	-0.005	0.001	-0.002
				18	0.001	-0.001	0.000
				24	0.002	-0.001	0.000
				30	0.000	0.000	0.000
100	8	-	0		h = 2.0	h = 2.0, σ = 0.8	h = 2.5 ,σ = 1.0
				6	-0.002	0.001	0.001
				12	0.000	0.000	0.000
				18	-0.004	-0.003	-0.003
				24	-0.001	-0.002	-0.002
				30	-0.003	-0.001	-0.001
				36	0.001	0.001	0.001
100	18	22	39		h = 2.0	h = 2.0, σ = 1.0	h = 2.5 ,σ = 1.0
				6	0.003	0.003	0.003
				12	-0.001	0.003	0.004
				18	0.001	0.001	0.000
				24	-0.005	-0.003	-0.004
				30	-0.001	0.000	0.000
100	22	18	60		h = 2.5	h = 2.5, σ = 1.0	h = 2.5 ,σ = 1.0
				6	-0.002	0.002	0.000
				12	0.005	0.004	0.004
				18	0.001	0.001	0.001
				24	0.003	0.004	0.003
				30	0.000	0.000	0.000

the normal error corrected density estimate respectively. Table 5.1 illustrates the following general findings: (i) All estimators are fairly unbiased. (ii) The kernel density estimator $\hat{f}_Y(x)$ by smoothing the Kaplan-Meier distribution function exhibits more bias than the other two estimators except at lower values of time (t=6) in small sample samples (n=30, 60) with moderate to heavy censoring percentage (π = 27, 77). (iii) For large sample size (n=100) with moderate to heavy censoring, the effect of error correction in the bias reduction is little.

5.2 Density Estimate from Duration of Breastfeeding Data

We now obtain smooth density estimates from duration of breastfeeding data of NFHS-1 and NFHS-2. Figure 5.2(a) provides three estimates of density of breastfeeding duration using NFHS-1 data: (i) smoothed estimate from Kaplan-Meier distribution function, (ii) estimate using Laplacian error correction, and (iii) estimate using normal error correction. Figure 5.2 (b) compares the smoothed estimates for NFHS-1 and NFHS-2. The right skewness of the distributions are evident from the plots. The curve for NFHS-1 is more skewed than the other. The plots in fig 5.1(a) show hardly any effect of error correction; smooth density sharply rises reaching its maximum at 14 months and then trails off gradually. The median duration of breastfeeding is 22 months for NFHS-1. It describes well the situation in the whole population. The relatively long right tail is a result of the few subjects who had long breastfeeding experience. Figure 5.2 (b), on the other hand, is featured by a relatively heavier right tail; it shows that the NFHS-2 density curve reaches its maximum at 15 months while the median duration is 30 months. A longer breastfeeding experience is evident in the later survey from this comparison.

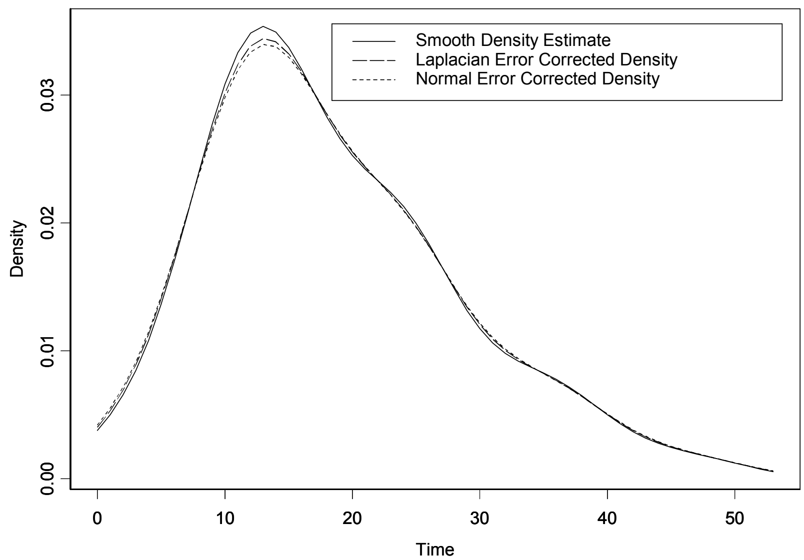
6. Concluding Remarks.

The kernel estimators of hazard and density functions for breastfeeding duration proposed in this article use the idea of convolving a kernel weight with the hazard and density estimates induced by the natural estimate of the cumulative distribution function. The estimators have attractive mean squared error properties (Wand and Jones, 1995) and are pointwise consistent (Anderson et al, 1993).

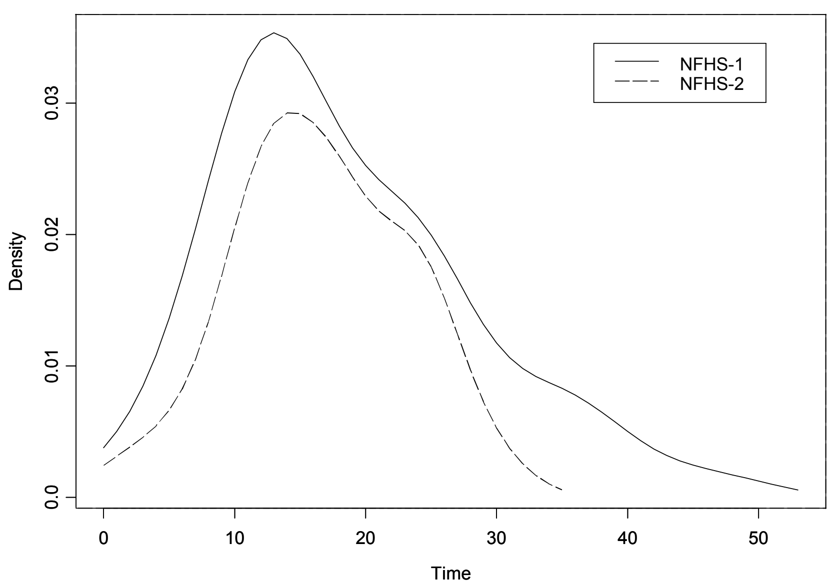
For large samples, the effect of bias correction is little, subject to the appropriate selection of bandwidth parameter. One can make judicious use of any of the several hi-tech bandwidth selectors to implement the proposed estimators. One can also get fairly precise confidence interval approximately except when only a few subjects are remaining in the risk set.

Figure-5.2

(a) Density of Breastfeeding Duration (NFHS-1)



(b) Comparison of Breastfeeding Density Estimates



REFERENCES

1. Akin, J.S.,Bilsborrow, R.E., Guillkey, D.K. and Popkin, B.M.(1986). Breastfeeding patterns and determinants in the Near East: An analysis for four countries. Pop. Stud. 40(2), 247-262.
2. Anderson, J.E., Roudrigues, W. and Thome, A.M.T.(1983). Analysis of breastfeeding in Northeastern Brazil: Methodological and policy considerations. Stud. Fam. Plann. 16(3), 154-163.
3. Anderson, P.K.,Borgan, Ø., Gill, R.D., and Keiding, N. (1993). Statistical Models Based on Counting Processes. Springer-Verlag New York, Inc.
4. Bongaarts, J. and Potter, R. G.(1983). Fertility, Biology and Behavior. New York, Academic Press.
5. Chakrabarty, T.K.& Singh, Ksh. A.(2005). Duration of Breastfeeding in Northeast India : Estimates of Survival Function from Large Scale Health Surveys, Health and Population, Perspective & Issues, National Institute of Health and Family welfare, New Delhi, (Forthcoming issue).
6. _______ (2005). Analysis of Duration of Breast-feeding with Missing Covariate Data: Contributions to Applied and Mathematical Statistics, vol.3, Department of Statistics, Gauhati University,
7. Chhabra, P., Grover, V.L., Aggrawal, O.P., and Dubey, K.K.(1998). Breast feeding pattern in urban resettlement colony of Delhi. Ind. J. Pediat., 65,872-875.
8. Ferry, B. (1981). Breastfeeding: World Fertility Survey Comparative Studies, No. 13, International Statistical Institute, Voorburg.
9. Haaga, J. (1988). Reliability of retrospective survey data on infant feeding, Demography, 25 (2), pp. 307-314.
10. Huffman, S.L.(1984). Determinants of breastfeeding in developing countries: Overview and policy implications. Stud. Fam. Plann., 15(4), 170-183.
11. International Institute for Population Sciences(1995). National Family Health Survey, 1992-93: Northeastern States(MCH and Family Planning),IIPS, Mumbai.
12. International Institute for Population Sciences and ORC Macro.(2002). National Family Health Survey(NFHS-2), 1998-99: Northeastern States, IIPS, Bombay.
13. Jain, A.K. and Bongaarts, J.(1981). Breastfeeding: Patterns, correlates and fertility effects. Stud. Fam. Plann., 12(3), 79-99.
14. Mott, S.H.(1984). A note on the determinants of breastfeeding duration in an African country. Soc. Biol., 31(3), 270-289.
15. Nath, D.C., Singh, K.K. and Land, K.C.(1993). Breastfeeding and postpartum amenorrhoea in a traditional society: A hazards model analysis. Soc. Biol., 40(1/2), 74-86.
16. Nath, D.C., Land, K. C. and Singh, K. K. (1994). The role of breast-feeding beyond post-partum amenorrhoea in the return of fertility in India : A life table and hazards model analysis, J. Biosoc. Sc., **26**, 191-206.

17. Nath, D.C. and Goswami, G.(1997). Determinants of Breast-feeding in an urban society of India. Hum. Bio.,60(4), 557-573.
18. Ramlau-Hansen, H. (1983). Smoothing counting process intensities by means of kernel functions. Ann. Statist. 11, 453-466.
19. Smith, D.J. and Ferry, B.(1984). Correlates of Breastfeeding: World Fertility Survey, Comparative Studies, 41, Int. Stat. Inst., Voorburg.
20. Smith, D.P.(1985). Breastfeeding, contraception, and birth intervals in developing countries. Stud. Fam. Plann., 16(3), 154-163.
21. Stefanski, L. and Carroll, R.J.(1990) Deconvolving kernel density estimators. Statistics 2, 169-184.
22. Trussel, J. et al(1992). Trends and Differentials in Breastfeeding Behaviour: Evidence from the WFS and DHS. Population Studies, 46(2), pp. 285-307.
23. Victoria, C.G., Smith, P.G. and Vaugham, I.P.(1987). Evidence for protection by breastfeeding against infant death from infectious diseases in Brazil. Lancet, 2, 319-322.
24. Wand, M.P., and Jones, M.C. (1995). Kernel Smoothing. Chapmann & Hall/CRC, www.crcpress.com.
25. World Health Organization(1989). Research on improving infant feeding practices to prevent diarrhoea or reduce its severity: Memorandum for a John Hopkins University/ WHO Meeting Bull., WHO,67(1),27-33.