NISS

On Nonparametric Regression for Current Status Data

Deborah Burr and Shanti Gomatam

Technical Report Number 127
July, 2002

National Institute of Statistical Sciences
19 T. W. Alexander Drive
PO Box 14006
Research Triangle Park, NC 27709-4006
www.niss.org

On Nonparametric Regression for Current Status Data

Deborah Burr
Epidemiology and Biometrics
Ohio State University
Columbus, Ohio 43210

Shanti Gomatam[†]
National Institute of
Statistical Sciences
RTP, North Carolina 27709

July 25, 2002

Abstract

We study the problem of nonparametric estimation of the conditional distribution function when we have current status data on the outcome variable and a single continuous-valued covariate. An estimator of the conditional distribution function $F(Y|X=x)$, called the local nonparametric maximum likelihood estimator (LNPMLE) is proposed. This estimator is a locally weighted version of the nonparametric maximum likelihood estimator (NPMLE) for current status data in the absence of covariates. The primary goal of this work is to obtain an expression for the optimal bandwidth used to pick neighborhood size. The asymptotic distribution of the LNPMLE of the conditional distribution function at a point, $F(t|X=x)$, is studied, and the asymptotically optimal bandwidth is shown to be of the order $n^{-1/7}$. The LNPMLE of the conditional distribution function can be obtained as a solution to a weighted isotonic regression problem. A plug-in estimate is suggested for the bandwidth, and the computation of the LNPMLE is illustrated on a simulated sample.

[*]The authors would like to thank Hani Doss, Brett Presnell, and Jon Wellner for their valuable comments and suggestions, and B. Narasimhan for the use of his random number generation code. The second author would like to acknowledge NSF grant DMS-9631278 for partial support of this work.

[†]S. Gomatam is also Assistant Professor in the Department of Mathematics, University of South Florida, Tampa, Florida.

1 Introduction

In many studies, the time to occurrence of some event is of interest. In some cases it is not possible to observe directly the time at which the event of interest occurs but one is able to observe a time interval in which the event has occurred. A particular case arises in those situations where the testing is destructive, or when observing the subject at more than one time is not possible. A specific practical situation of this sort is an experiment in which rats are injected with a dose of a carcinogenic substance, and are operated on at certain times to see if a cancerous tumor has developed in a particular organ. When the rat must be sacrificed, or the organ either must be extracted in order to do a complete pathological examination, or is damaged or altered, no second measurement is possible. So each rat is observed at a single time and it is known whether it shows the presence of the tumor at that time. Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables denoting the time to development of cancer, and let T_i be the time at which the i^{th} rat is operated upon, for $i = 1, 2, \ldots, n$. The data available in these cases, i.e., the combination of the examination/censoring time T, and the event indicator, $\delta = I(Y \leq T)$, are known as current status data. This censoring mechanism is also referred to as interval-censoring, case 1 (Groeneboom and Wellner, 1992).

One could also consider the Y's above as differences between an initiating event (time of injection of carcinogen in the example above) and a subsequent event (time of tumor formation). Implicit in the above description of current status observations is the assumption that the time of the initiating event is known. An alternative formulation for current status observations is to consider both the initiating time and the time to subsequent event as being current status observations — in such cases the data are referred to as doubly censored current status data (see Rabinowitz and Jewell (1996)).

Consider the situation where information on a covariate X, for example the dose of the carcinogen injected in the rat, is also available. The observed data are then the triplets (T_1, δ_1, X_1), (T_2, δ_2, X_2), $\ldots$, (T_n, δ_n, X_n), where $\delta_i = I(Y_i \leq T_i)$, and X_i is the dose administered to the i^{th} subject. The problem of nonparametric estimation of the conditional distribution function, when current status data on the outcome variable and a single continuous-valued covariate are available, is studied here. We assume that the time of the initiating event is known.

Estimation of the distribution function for current status observations in the absence of covariates has been studied by various authors. Ayer et al (1955) show that the nonparametric maximum likelihood estimate (NPMLE) can be easily computed and establish its consistency. Turnbull's (1976) work, applicable to a broader class of censored and truncated data than just current status, describes a "self-consistency algorithm" (a special case of the EM algorithm (Dempster, Laird, and Rubin, 1976)). Groeneboom and Wellner (1992) show that the NPMLE can be characterized as a solution to an isotonic regression problem, establish consistency and convergence in distribution, and show asymptotic normality of the mean. Huang and Wellner (1995) show that the self-consistency equation is satisfied by the NPMLE, and prove asymptotic normality of the NPMLEs of linear functionals. Van der Laan (1994) shows that the NPMLE is efficient. Jewell, Malani and Vittinghoff (1994) consider nonparametric estimation of the distribution function for doubly censored current status data.

Work on the regression case for current status data includes that of Murphy, van der Vaart, and Wellner (1999) who consider asymptotic properties of the MLE and penalized MLE of the slope parameter in linear regression; Rabinowitz, Tsiatis and Aragon (1995) considered estimation and inference using score statistics for the regression parameters in a linear model for current status data, assuming the error distribution is unknown; Rabinowitz and Jewell (1996) establish a correspondence between current status data and doubly censored current status data when the initiating event can be assumed to be uniformly distributed; Rossini and Tsiatis (1994) study estimation in the context of the proportional odds regression model; van der Laan, Bickel, and Jewell (1994) introduce a regularized MLE and obtain results for linear regression for both current status and doubly censored current status data. A useful summary of work on current status and interval censored data is given by Huang and Wellner (1996). Other related references are Diamond, McDonald, and Shah (1986), Jewell and Shiboski (1990), Diamond and McDonald (1991), and Keiding (1991).

To the best of our knowledge, existing regression estimators considered so far have been at most semi-parametric, i.e., either distributional assumptions are made on the error terms, or a parametric or semi-parametric form has been assumed for the model. In addition, most of the work on regression has centered on estimation of the mean. The focus in this work is on nonparametric estimation of the conditional distribution function for current status data. We propose an estimator called the "local nonparametric maximum likelihood estimator" (LNPMLE), which is a locally smoothed modification of the nonparametric maximum likelihood estimator (NPMLE) of the distribution function for the no-covariate case discussed by Ayer et al (1955) and Groeneboom and Wellner (1992).

The LNPMLE depends on the size of the neighborhood used to obtain the estimator; hence an appropriate choice of neighborhood is critical to the estimation process. Neighborhood size is typically expressed in terms of the bandwidth h. We will obtain an expression for the asymptotically optimal bandwidth using the bias and variance of the asymptotic distribution of the LNPMLE of $F(Y = t|X = x)$. This asymptotically optimal bandwidth is shown to be of order $n^{-1/7}$, whereas, under similar assumptions, the optimal bandwidth for estimating the distribution function with uncensored data is of order $n^{-1/5}$.

We also show that the LNPMLE can be obtained as the solution to a weighted isotonic regression problem. A method is suggested for estimating the optimal bandwidth, and a simulated sample is used to demonstrate the computation.

In Section 2 below we present the estimation method. Section 3 gives the distribution theory result and the derivation of the asymptotically optimal bandwidth. Section 4 discusses computation of the estimate, and illustrates this computation on a simulated sample.

2 The Estimation Method

Consider current status data, consisting of random observation times and event indicators (T_1, δ_1), (T_2, δ_2), $\dots$, (T_n, δ_n), corresponding to the occurrence times Y_1, Y_2, $\dots$, Y_n. The Y_i and T_i are non-negative random variables, assumed independent, with distribution functions F and G respectively. We will assume, without loss of generality, that the T_i's are arranged in increasing order, i.e. $T_1 \leq T_2 \leq \cdots \leq T_n$. The log likelihood, up to an additive term that does not involve F, is

$$\phi(F) = \sum_{i=1}^{n} \Big\{ \delta_i \log\Big(F(T_i)\Big) + (1 - \delta_i) \log\Big(1 - F(T_i)\Big) \Big\}. \tag{2.1}$$

The NPMLE of F is the $\hat{F}$ that maximizes ϕ subject to the constraint $F(T_1) \leq F(T_2) \leq \cdots \leq F(T_n)$. Obtaining such an estimate is equivalent to finding the F that maximizes

$$\psi(F) = \int_{R^2} \Big\{ I_{y \leq t} \log\Big(F(t)\Big) + I_{y > t} \log\Big(1 - F(t)\Big) \Big\} dP_n(y, t), \tag{2.2}$$

where P_n is the empirical probability measure of the pairs (Y_i, T_i), $1 \leq i \leq n$, subject to the constraint above.

As only the values of $\hat{F}$ at the observation points matter for this maximization problem, the NPMLE can be arbitrarily taken to be a distribution function which is piecewise constant with jumps only at the observations points T_i (Groeneboom and Wellner (1992)). It is possible that the function obtained by the maximization of $\phi(F)$ may be a sub-distribution function, i.e., $\hat{F}(t) < 1$ at each observation point t. In such cases, the location of the remaining mass is not specified. With this understanding, the NPMLE is uniquely determined.

Groeneboom and Wellner (1992) have shown that the values of $\{F(T_i)\}_{i=1,\ldots,n}$ that maximize $\phi(F)$ are the solutions to a certain isotonic regression problem, namely the minimization of $\sum_{i=1}^{n}(\delta_i - u_i)^2$ with respect to $(u_1, u_2, \ldots, u_n)$, subject to $u_1 \leq u_2 \leq \cdots \leq u_n$. Here the u_i's represent the values of the distribution function F at the n values of T_i.

Groeneboom and Wellner (1992) also show that when n current status observations are available, we have

$$\frac{n^{\frac{1}{3}}(\hat{F}(t) - F(t))}{\left\{\frac{1}{2}F(t)\Big(1 - F(t)\Big)f(t)/g(t)\right\}^{\frac{1}{3}}} \quad \xrightarrow{\mathcal{D}} \quad 2V, \tag{2.3}$$

where V is the last time at which standard two-sided Brownian motion minus the parabola $y = x^2$ reaches it maximum. Here $f(t)$ and $g(t)$ are the derivatives of F and G at t, where G is the distribution function of T. It is assumed that $f(t)$ and $g(t)$ are strictly positive, Y and T are independently distributed, and the probability measure induced by F is absolutely continuous with respect to the probability measure induced by G.

An unusual feature of the result above is that $\hat{F}(t)$ has an $n^{1/3}$ rate of convergence, unlike the usual $n^{1/2}$ rate which holds for the estimators of F for uncensored or right-censored data, for example. Also, the limit distribution is not the usual normal limit.

We will modify the above estimation method to obtain an estimator of the conditional distribution function. In the situation above, n independent current status observations drawn from the distribution F are used in order to estimate F. In our framework we assume that the conditional distribution function is smooth in the covariate values, i.e., $F_x(y)$ is smooth in x for fixed y. Thus, although multiple observations from F_{x_0} may not be possible, weighted observations from F_x, where x is in the neighborhood of x_0, serve as surrogates for a sample from F_{x_0}. The i^{th} point in the neighborhood is assigned a weight, w_i, generated by a smoothing method — points that are close to x_0 are given higher weights than those that are further away from x_0.

Let $(T_{(1)}, \delta_{(1)}, X_{(1)})$, $(T_{(2)}, \delta_{(2)}, X_{(2)})$, $\ldots$, $(T_{(n)}, \delta_{(n)}, X_{(n)})$ denote the observations, arranged in ascending order of the T values. The local nonparametric maximum likelihood estimator (LNPMLE) of the conditional distribution function F_x, at the target point x_0, is $\hat{F}_{x_0}$ which maximizes

$$\eta(F_{x_0}) = \sum_{i=1}^{n} w_i \Big\{ \delta_{(i)} \log\Big(F_{x_0}(T_{(i)})\Big) + (1 - \delta_{(i)}) \log\Big(1 - F_{x_0}(T_{(i)})\Big) \Big\}, \tag{2.4}$$

subject to

$$F_{x_0}(T_{(1)}) \leq F_{x_0}(T_{(2)}) \leq \cdots \leq F_{x_0}(T_{(n)}).$$

Here η is the local log likelihood (except for an additive constant that does not involve F_{x_0}). This local log likelihood, defined over points in the covariate neighborhood of the target point, is just a weighted version of the log likelihood in (2.1). As noted in the no-covariate case, the maximization above determines the LNPMLE only at observed T values in the covariate-neighborhood of x_0. We will follow the convention of using a linear interpolation for values of T that do not lie on this grid. Subject to the above conditions, and allowing for the possibility of obtaining a sub-distribution function as an estimate, the LNPMLE is unique. In Section 4 we show that an isotonic regression characterization can be obtained for the LNPMLE $\hat{F}_x$.

3 Asymptotic Distribution and Optimal Bandwidth

We will see that optimal bandwidths for current status data will be larger than those required in estimation for uncensored data, i.e., for current status observations one would expect a slower reduction in bandwidth as sample size increases. When uncensored observations are used, optimal bandwidths for nonparametric regression function estimation are of the order of $n^{-1/5}$ under the assumption that the regression function is twice continuously differentiable. We show heuristically that for current status data the order of the aysmptotically optimal bandwidth is $n^{-1/7}$ under similar assumptions on the covariate distribution.

We are interested in extending the result stated in Groeneboom and Wellner (1992), in Equation (2.3) above, to estimates of F_x based on kernel weights. The procedure proposed here amounts to estimating the conditional distribution function by imitating the NPML estimation on the "n_x" points in the covariate-neighborhood, instead of using all n points. When estimating at the target point x, points are assigned weights w_i according to their distance from the target covariate value. We will denote by $N_{x,h}$ those points for which w_i is non-zero. The following Nadaraya-Watson weights (Nadaraya (1964), Watson (1964))

$$w_i = w_x(X_i) = \frac{K(\frac{x-X_i}{h})}{\sum_{i=1}^{n} K(\frac{x-X_i}{h})} \tag{3.5}$$

are considered in (2.4), where K is a kernel, X_i are observed values of the covariate, and h is the bandwidth. When weights from a uniform kernel are considered, equally weighted points in a neighborhood of length $2h$ around the target point x are used. In this case, the estimation procedure is equivalent to that outlined in Groeneboom and Wellner (1992) for the no-covariate case operating only on the n_x points in the neighborhood.

In the case of the uniform kernel, each point in the support of the kernel is assigned the same weight, thus the effective sample size n_x is just the number of points that are assigned non-zero weights by the kernel. However, since a kernel function may assign unequal, non-zero weights to all points on the real line, for non-uniform kernels it is not appropriate to take n_x to be the number of points that are assigned non-zero weights. The inverse of the sum of the squared weights, $(\sum_i^n w_i^2)^{-1}$, provides a natural measure of the effective sample size in this case.

In the estimation scheme discussed by Groeneboom and Wellner (1992) current status observations from the distribution function F are available for its estimation. Our estimation scheme, on the other hand, uses neighboring observations to estimate the conditional distribution function F_x. Each (T, δ) point in the neighborhood corresponds to a Y that comes from a distribution F_z (*not* F_x), where z is the covariate value associated with Y. Assume that the covariates are randomly and independently distributed. Consider an (as of yet) unobserved covariate value z, that will be generated according to a probability distribution L, along with an additional $n-1$ points, to form a sample of size n. When estimating at the target point x, the weight assigned to the random observation with covariate value z is denoted $w_x(z)$. Denote the probability law of the covariate, conditional on the covariate being in the neighborhood $N_{x,h}$, by L_x. As we will effectively be sampling the covariate values from a weighted mixture of covariate densities in the neighborhood, $dL_x(z) = w_x(z)dL(z)/\int_{N_{x,h}} w_x(u)dL(u)$. The values of the unobservable random variable Y with covariate values in the neighborhood of x will arise from a mixture of distributions, which will be denoted by F_{K_x}. For simplicity we assume that T is independent of X. Let G denote the distribution of T, and g the corresponding density. This gives

$$F_{K_x}(t) = \int_{N_{x,h}} F_z(t)dL_x(z) = \int_{N_{x,h}} F_z(t)\frac{w_x(z)}{\int_{N_{x,h}} w_x(u)dL(u)}dL(z).$$

Letting

$$w_x^*(z) = \frac{w_x(z)}{\int w_x(u)dL(u)} = \frac{K(\frac{x-z}{h})}{\int K(\frac{x-u}{h})dL(u)},$$

we define

$$F_{K_x}(t) = \int_{N_{x,h}} w_x^*(z)F_z(t)dL(z). \tag{3.6}$$

3.1 Asymptotic Distribution of $\hat{F}_x(t)$

In order to study the asymptotic properties of $\hat{F}_x(t)$ we decompose $\hat{F}_x(t) - F_x(t)$ into two parts as follows

$$\hat{F}_x(t) - F_x(t) = \Big(\hat{F}_x(t) - F_{K_x}(t)\Big) + \Big(F_{K_x}(t) - F_x(t)\Big), \tag{3.7}$$

and study the asymptotic properties of each of the two terms.

We wish to extend the result in Groeneboom and Wellner (1992) on the asymptotic distribution of $\hat{F}$ to the first term on the right in (3.7). As we now have an effective sample size of n_x and are sampling from F_{K_x}, heuristically extending the result for fixed h as $n \to \infty$, we have

$$\frac{(n_x)^{\frac{1}{3}}(\hat{F}_x(t) - F_{K_x}(t))}{\left\{\frac{1}{2}F_{K_x}(t)(1 - F_{K_x}(t))f_{K_x}(t)/g(t)\right\}^{\frac{1}{3}}} \xrightarrow{\mathcal{D}} 2V, \tag{3.8}$$

where $f_{K_x}(t) = d(F_{K_x}(t))/dt$, and $g(t)$ is the density of T. We assume that $f_{K_x}(t)$, $g(t)$ are positive, $Y|X = x$ and T are independently distributed, and the probability measure induced by F_x is absolutely continuous with respect to the probability measure induced by G (the distribution function of T). Further, assuming that (3.8) holds for all h in a small neighborhood around 0[1], it also holds as $n \to \infty$ and $h \to 0$.

Let us represent the denominator of the left hand side of (3.8) by $D_x(h)$, i.e.,

$$D_x(h) = \left\{\frac{1}{2}F_{K_x}(t)(1 - F_{K_x}(t))f_{K_x}(t)/g(t)\right\}^{\frac{1}{3}}$$

and let

$$D_x = \left\{\frac{1}{2}F_x(t)(1 - F_x(t))f_x(t)/g(t)\right\}^{\frac{1}{3}}. \tag{3.9}$$

As $h \to 0$, $F_{K_x}(t) \to F_x(t)$, thus $D_x(h) \to D_x$. Combining this with (3.8) and using Slutsky's theorem gives

$$\frac{(n_x)^{\frac{1}{3}}(\hat{F}_x(t) - F_{K_x}(t))}{\left\{\frac{1}{2}F_x(t)(1 - F_x(t))f_x(t)/g(t)\right\}^{\frac{1}{3}}} \xrightarrow{\mathcal{D}} 2V. \tag{3.10}$$

3.1.1 Approximation to $F_{K_x}(t) - F_x(t)$

We now consider the asymptotics on

$$\frac{(n_x)^{\frac{1}{3}}(F_{K_x}(t) - F_x(t))}{\left\{\frac{1}{2}F_x(t)(1 - F_x(t))f_x(t)/g(t)\right\}^{\frac{1}{3}}},$$

the appropriately scaled version of the second term in (3.7). For simplicity we only consider those points x for which a full kernel neighborhood is possible, i.e., we will not deal with points at the edges with incomplete neighborhoods.

Consider the following assumptions:

[1]Showing that this holds requires an appropriate modification of results in Groeneboom and Wellner (1992), which is not shown here.

A_1**:** The distribution $L(z)$ has a density $l(z)dz$ in the neighborhood of x, and there exists $h_1 > 0$ such that, for $|\delta| < h_1$,

$$l(x+\delta) = l(x) + \delta l'(x) + \frac{\delta^2 l''(x)}{2!} + O(\delta^3) \tag{3.11}$$

A_2**:** There exists $h_2 > 0$ such that, for each δ with $|\delta| < h_2$, there exists a z_1 in the interval adjoining x and $x+\delta$, such that

$$\begin{aligned} F_{x+\delta}(t) &= F_x(t) + \delta \frac{\partial F_z(t)}{\partial z}\bigg|_x + \frac{\delta^2}{2!}\frac{\partial^2 F_z(t)}{\partial^2 z}\bigg|_x + \frac{\delta^3}{3!}\frac{\partial^3 F_z(t)}{\partial^3 z}\bigg|_{z_1} \\ &= F_x(t) + \delta g_1(t,x) + \frac{\delta^2}{2} g_2(t,x) + O(\delta^3), \end{aligned} \tag{3.12}$$

where $g_1(t,x) = \frac{\partial F_z(t)}{\partial z}|_x$ and $g_2(t,x) = \frac{\partial^2 F_z(t)}{\partial z}|_x$. We are assuming here that $\partial^3 F_z(t)/\partial^3 z$ is uniformly bounded in the neighborhood of x and can be ignored.

Let $h_0 = \min(h_1, h_2)$ so both (3.11) and (3.12) hold for $|z - x| < h_0$.

A_3**:** The kernel has compact support. Hence we can choose $N_{x,h}$ such that $N_{x,h} \subseteq N_{x,h_0}$ and the kernel "vanishes" outside $N_{x,h}$. That is,

$$K\left(\frac{x-z}{h}\right) = 0, \quad \text{for all } z \in N_{x,h}^c \tag{3.13}$$

For the derivations below, the case of the uniform kernel is considered first as the results and derivations for this case are more intuitive. Generalization of these results to the case of bounded kernels follows naturally and is outlined later.

When K represents the uniform kernel, using (3.12) and ignoring terms of order $(z-x)^3$ and higher we can write

$$\begin{aligned} F_{K_x}(t) - F_x(t) &= \int_{N_{x,h}} F_z(t) w_x^*(z) dL(z) - F_x(t) \\ &\approx \int_{N_{x,h}} \left\{(z-x)g_1(t,x) + \frac{(z-x)^2}{2} g_2(t,x)\right\} dL_x(z). \end{aligned} \tag{3.14}$$

For the uniform kernel $w_x(z) = 1/2h$ for $z \in (x-h, x+h)$. The conditional density of the covariate, given that the covariate value lies in the neighborhood $N_{x,h}$, is given by

$$dL_x(z) = \frac{w_x(z)dL(z)}{\int w_x(z)dL(z)} = \frac{\frac{1}{2h} l(z)dz}{\frac{1}{2h}\int_{-h}^{h} l(x+u)du} \approx \frac{l(z)dz}{2hl(x)}, \qquad \text{for } z \in (x-h, x+h)$$

Using the above and (3.11), and ignoring terms that are $O(h^3)$ and smaller, it is easy to show that

$$\int_{N_{x,h}} (z-x) dL_x(z) \approx \frac{h^2 l'(x)}{3l(x)}, \tag{3.15}$$

and

$$\int_{N_{x,h}} (z-x)^2 dL_x(z) = \frac{h^2}{3}. \tag{3.16}$$

Substituting from (3.15) and (3.16) into (3.14) we get

$$\begin{aligned} F_{K_x}(t) - F_x(t) &= \frac{h^2 l'(x)}{3l(x)} g_1(t,x) + \frac{h^2}{3}\frac{g_2(t,x)}{2} \\ &= \left(g_1(t,x)l'(x) + \frac{g_2(t,x)l(x)}{2}\right)\frac{h^2}{3l(x)}. \end{aligned} \tag{3.17}$$

Hence,

$$\frac{n_x^{\frac{1}{3}}(F_{K_x}(t) - F_x(t))}{\left\{\frac{1}{2}F_x(t)(1-F_x(t))f_x(t)/g(t)\right\}^{\frac{1}{3}}} \approx \frac{n_x^{\frac{1}{3}} h^2}{3l(x)D_x}\left(g_1(t,x)l'(x) + \frac{g_2(t,x)l(x)}{2}\right), \tag{3.18}$$

where D_x is as defined in (3.9).

As we are using the uniform kernel, for sufficiently small h the number of points n_x contained in the neighborhood $N_{x,h}$ can be approximated by $n_x = 2nhl(x)$, where n is the total number of observations. Thus, the right hand side of the approximation in (3.18) reduces to

$$\frac{(2nhl(x))^{\frac{1}{3}} h^2}{3l(x)D_x}\left(g_1(t,x)l'(x) + \frac{g_2(t,x)l(x)}{2}\right) = \frac{(2nh^7)^{\frac{1}{3}}}{3l(x)^{\frac{2}{3}} D_x}\left(g_1(t,x)l'(x) + \frac{g_2(t,x)l(x)}{2}\right)$$

i.e.,

$$\frac{n_x^{\frac{1}{3}}(F_{K_x}(t) - F_x(t))}{\left\{\frac{1}{2}F_x(t)(1-F_x(t))f_x(t)/g(t)\right\}^{\frac{1}{3}}} \approx \frac{(2nh^7)^{\frac{1}{3}}}{3l(x)^{\frac{2}{3}} D_x}\left(g_1(t,x)l'(x) + \frac{g_2(t,x)l(x)}{2}\right) \tag{3.19}$$

If $nh^7 \to M_1$ as $n \to \infty$, and $h \to 0$, then the above expression converges to a constant $C_x^1(t)$, defined as

$$C_x^1(t) = \frac{(2M_1)^{\frac{1}{3}}}{3l(x)^{\frac{2}{3}} D_x}\left(g_1(t,x)l'(x) + \frac{g_2(t,x)l(x)}{2}\right).$$

Combining the results in (3.10) and (3.19) we have, if $nh^7 \to M_1$ as $n \to \infty$, and $h \to 0$,

$$\frac{n_x^{\frac{1}{3}}(\hat{F}_x(t) - F_x(t))}{\{\frac{1}{2}F_x(t)(1-F_x(t))f_x(t)/g(t)\}^{\frac{1}{3}}} \xrightarrow{\mathcal{D}} 2V + C_x^1(t). \tag{3.20}$$

The above result also holds when we have non-uniform kernels satisfying assumption A_3. In this case we define $n_x = \left(\sum_i^n w_i^2\right)^{-1}$, and as shown in the Appendix, the following approximation to $F_{K_x}(t) - F_x(t)$ can be obtained when $\left(\sum_i^n w_i^2\right)^{-\frac{1}{3}} h^3/C_h \to M$

$$F_{K_x}(t) - F_x(t) \approx \left(g_1(t,x)l'(x) + \frac{g_2(t,x)l(x)}{2}\right)\frac{I_{K,2}h^3}{C_h}, \tag{3.21}$$

where

$$C_h = \int K(\frac{x-z}{h})dL(z), \tag{3.22}$$

$$I_{K,2} = \int u^2 K(u)du, \tag{3.23}$$

and D_x is as defined in (3.9). Thus,

$$\begin{aligned} \frac{n_x^{\frac{1}{3}}(F_{K_x}(t) - F_x(t))}{\{\frac{1}{2}F_x(t)(1-F_x(t))f_x(t)/g(t)\}^{\frac{1}{3}}} &= \frac{\left(\sum_i^n w_i^2\right)^{-\frac{1}{3}}(F_{K_x}(t) - F_x(t))}{D_x} \\ &\approx \frac{\left(\sum_i^n w_i^2\right)^{-\frac{1}{3}} h^3 I_{K,2}}{D_x C_h}\Big(g_1(t,x)l'(x) + \frac{g_2(t,x)l(x)}{2}\Big). \end{aligned} \tag{3.24}$$

Hence, for kernels satisfying the assumptions A_1, A_2 and A_3, if $\left(\sum_i^n w_i^2\right)^{-\frac{1}{3}} h^3/C_h \to M$, we have[2]

$$\frac{\left(\sum_i^n w_i^2\right)^{-\frac{1}{3}}(F_{K_x}(t) - F_x(t))}{\left\{\frac{1}{2}F_x(t)(1-F_x(t))f_x(t)/g(t)\right\}^{\frac{1}{3}}} \longrightarrow C_x(t), \tag{3.25}$$

where

$$C_x(t) = \frac{M I_{K,2}}{D_x}\Big(g_1(t,x)l'(x) + \frac{g_2(t,x)l(x)}{2}\Big). \tag{3.26}$$

Combining the results in (3.10) and (3.25) we have the following asymptotic distribution for the appropriately scaled version of $\hat{F}_x(t)$ for kernels satisfying assumptions A_1, A_2, and A_3 with $\left(\sum_i^n w_i^2\right)^{-\frac{1}{3}} h^3/C_h \to M$

$$\frac{\left(\sum_i^n w_i^2\right)^{-\frac{1}{3}}(\hat{F}_x(t) - F_x(t))}{\{\frac{1}{2}F_x(t)(1-F_x(t))f_x(t)/g(t)\}^{\frac{1}{3}}} \xrightarrow{\mathcal{D}} 2V + C_x(t) \tag{3.27}$$

3.2 Asymptotic Approximations to the Bias and Variance of $\hat{F}_x(t)$

In order to obtain an explicit expression for the asymptotically optimal bandwidth $h_{\text{opt},x}$, the bandwidth that minimizes the mean square error (MSE) in terms of the underlying distributions, we consider an approximation to the MSE of $\hat{F}_x(t)$. We use the bias and variance of the asymptotic distribution in order to approximate the MSE.

[2]For the uniform kernel $\left(\sum_i^n w_i^2\right)^{-\frac{1}{3}} h^3/C_h \approx (2nh^7)^{\frac{1}{3}}/l(x)^{\frac{2}{3}}$.

The random variable V in the limiting distribution above has 0 mean and finite variance (see, for example, Groeneboom and Wellner (2001)). Let us denote $\mathrm{Var}(2V)$ by v. Hence from (3.27) we have the following approximation to the bias of the asymptotic distribution of $\hat{F}_x$:

$$\mathrm{Bias}_{\mathrm{AD}}(\hat{F}_x(t)) \approx C_x(t) D_x n_x^{-\frac{1}{3}}, \tag{3.28}$$

where the subscript AD refers to the asymptotic distribution of $\hat{F}_x(t)$ obtained in (3.27).

Recall $C_h = \int K(\frac{x-z}{h}) dL(z)$. For standard (symmetric, bounded) kernels with compact support (i.e., the Uniform, Epanechnikov, Biweight and triangular kernels), ignoring terms that are $O(h^3)$ and smaller, $C_h \approx c_1 h$, where c_1 is a constant with respect to h and n. In particular, $c_1 = l(x)$ for the uniform kernel under assumption A_1. Reverting to the expression given by the approximation in (3.24) and using $C_h \approx c_1 h$, we have

$$\mathrm{Bias}_{\mathrm{AD}}(\hat{F}_x(t)) \approx \left(g_1(t,x) l'(x) + \frac{g_2(t,x) l(x)}{2} \right) \frac{I_{K,2}}{c_1} h^2 \tag{3.29}$$

Note that for the uniform kernel $n_x \approx 2nhl(x)$, for the other standard kernels $n_x = \left(\sum_i^n w_i^2 \right)^{-1} \approx c_2 nh$, where c_2 does not depend on h or n. Thus, using (3.27) and $\mathrm{Var}(2V) = v$, an approximation for the variance of the asymptotic distribution of $\hat{F}_x(t)$ can be given by

$$\begin{aligned} \mathrm{Var}_{\mathrm{AD}}(\hat{F}_x(t)) &\approx {D_x}^2 v (n_x)^{-\frac{2}{3}} \\ &\approx \frac{{D_x}^2 v}{(c_2)^{\frac{2}{3}}} (nh)^{-\frac{2}{3}}. \end{aligned} \tag{3.30}$$

Using (3.29) and (3.30) we can obtain an approximate expression for the asymptotic MSE (AMSE) of $\hat{F}_x(t)$.

$$\begin{aligned} \mathrm{AMSE}(\hat{F}_x(t)) &\approx \mathrm{Bias}^2_{\mathrm{AD}}(\hat{F}_x(t)) + \mathrm{Var}_{\mathrm{AD}}(\hat{F}_x(t)) \\ &\approx K_1 h^4 + K_2 (nh)^{-\frac{2}{3}}, \end{aligned} \tag{3.31}$$

where $K_1 = \left((g_1(t,x) l'(x) + g_2(t,x) l(x)/2) \frac{I_{K,2}}{c_2} \right)^2$ and $K_2 = {D_x}^2 v / (c_2)^{\frac{2}{3}}$.

Differentiating the expression for AMSE with respect to h and setting the derivative to 0, we obtain the following value for the asymptotically optimal bandwidth $h_{\mathrm{opt},x}$ at t

$$\begin{aligned} h_{\mathrm{opt},x}(t) &= \left(\frac{K_2}{6K_1} \right)^{\frac{3}{14}} n^{-\frac{1}{7}} \\ &= \left(\frac{c_1^2 {D_x}^2 v}{6 c_2^{\frac{2}{3}} I_{K,2}^2 \left(g_1(t,x) l'(x) + \frac{g_2(t,x) l(x)}{2} \right)^2} \right)^{\frac{3}{14}} n^{-\frac{1}{7}}. \end{aligned} \tag{3.32}$$

Thus, the asymptotically optimal bandwidth in this case is of the order of $n^{-\frac{1}{7}}$, and the above expression allows its computation when some quantities based on the underlying distributions and the kernel are available.

4 Computation of the Estimate

In the previous sections an estimator for the conditional distribution function was introduced and a heuristic approximation for the optimal bandwidth was established. However, it is not apparent how one would carry out the maximization in (2.4) and compute the estimate of F_x. Recall that, in the absence of covariates, Groeneboom (1992) established that the NPMLE of F can be characterized as the solution to an isotonic regression problem. Applying Theorem 1.5.1 of Robertson et al (1988) to the convex function with $\Phi(s) = s\log(s)+(1-s)\log(1-s)$ above, we see that the maximizer of $\eta(F_x)$ defined in (2.4) is given by the solution to the weighted isotonic regression problem, i.e., of minimizing $\sum_{i=1}^{n} w_i(x)(\delta_i - u_i)^2$ with respect to $(u_1, u_2, \ldots, u_n)$, subject to $u_1 \leq u_2 \leq \cdots \leq u_n$. (An alternative proof of the above is also possible from modifications of results in Brunk (1955).)

Thus the Pool Adjacent Violators (PAV) Algorithm, a widely used method for the computation of the isotonic regression, can be used to compute $\hat{F}_x$. The PAV algorithm is explained in greater detail in Barlow et al (1972, Ch. 1, pp. 13–18) and Robertson et al (1988, Ch. 1, pp. 8–10).

Clearly, before the PAV algorithm can be implemented the neighborhood size must be chosen and the weights for each point in the neighborhood must be computed. Obtaining a value for $h_{\text{opt},x}(t)$ is a non-trivial problem in itself. One way of obtaining a value for $h_{\text{opt},x}(t)$ is by fitting a simple parametric model to the data, and using the distribution-based constants from this fit in (3.32). This strategy is implemented in the example given in the following subsection.

4.1 Illustration on a Simulated Sample

In this subsection we describe the computation of the plug-in bandwidth, and its use to estimate $F_x(t)$, for a randomly generated sample. Besides kernel-based constants and sample size, the expression for $h_{\text{opt},x}(t)$ in (3.32) involves quantities based on the distributions of T, X, and that of Y given $X = x$. The distribution of T can be specified by the investigator, and when a designed experiment is possible, so is the distribution of X. In any case, the X and T are completely observable, and their densities may be (parametrically) estimated. However, the form of the conditional distribution of Y given X cannot be assumed to be known. We overcome this problem by (mis-)specifying a parametric form for the conditional distribution of Y (we use an exponential, the true distribution is a gamma), using the sample values to obtain estimates of its parameters, and using quantities obtained from this approximation to obtain plug-in values for $h_{\text{opt},x}(t)$. We have found that the use of the plug-in value gives good results, in spite of the model

Table 4.1: *Values of* $\hat{F}_x(t)$ *and* $|\hat{F}_x(t) - F_x(t)|$ *(in parentheses) for simulated data set.* $h_{\mathrm{opt},x}$ *was obtained using the exponential distribution approximation.*

$X\backslash T$	0.106	0.347	0.660	1.154	2.203
0.105	0	0.042	0.252	0.364	0.870
	(0.006)	(0.015)	(0.086)	(0.001)	(0.171)
0.310	0	0.042	0.293	0.459	0.889
	(0.009)	(0.035)	(0.079)	(0.013)	(0.106)
0.484	0	0.042	0.293	0.475	0.867
	(0.011)	(0.053)	(0.036)	(0.036)	(0.029)
0.701	0	0.042	0.293	0.500	0.971
	(0.014)	(0.077)	(0.016)	(0.084)	(0.083)
0.893	0	0.042	0.293	0.514	0.967
	(0.018)	(0.099)	(0.062)	(0.128)	(0.047)

mis-specification (see below). For comparison, the estimate $\hat{F}_x(t)$ is also obtained when the expression for the optimal bandwidth is obtained using the constants calculated from the true distribution of $Y|X$.

In the simulated data set of size 1000 used in the illustration below, weights are obtained from a uniform kernel, the covariate $X \sim \mathrm{U}(0,1)$, $T \sim \mathrm{Exp}(\lambda = 1.0)$ (mean=$1/\lambda$), and $Y|X = x \sim \mathrm{Gamma}(\alpha = 2, \beta = 1 + x)$ (mean=α/β). We note here that the choice of distribution for T is a critical component of the estimation procedure. It is obvious, for instance, that for a choice that gives $\delta = 0$ (or $\delta = 1$) for all observations, there is very little information on Y. An appropriate choice would produce a scatter of 0's and 1's for values of δ over the range of values of X in the sample.

The conditional distribution of Y given $X = x$ is approximated by an exponential distribution with a parameter that is linear in the covariate. The linear predictor $\beta_0 + \beta_1 x$ is related to the conditional mean of Y given $X = x$ by the reciprocal function. Thus β_0 and β_1 in the distribution function $F_x(y) = 1 - \exp(-(\beta_0 + \beta_1 x)y)$ must be estimated. The sample observations are used to find the optimal estimates of the parameters β_0 and β_1 by maximizing the current status likelihood for the exponential distribution, i.e., by maximizing

$$\xi(\beta_0, \beta_1) = \prod_{i=1}^{n} (1 - e^{-(\beta_0+\beta_1 x_i)t_i})^{\delta_i} (e^{-(\beta_0+\beta_1 x_i)t_i})^{1-\delta_i}. \tag{4.33}$$

with respect to β_0 and β_1. This maximization is carried out by using the Splus function `ms()`. The `ms()` function generated $\hat{\beta}_0 = 0.3784$, $\hat{\beta}_1 = 0.5635$ for the sample used.

Table 4.2: *Values of* $\hat{F}_x(t)$ *and* $|\hat{F}_x(t) - F_x(t)|$ *(in parentheses) for simulated data set.* $h_{\text{opt},x}$ *was obtained using the gamma distribution constants.*

$X \backslash T$	0.106	0.347	0.660	1.154	2.203
0.105	0	0.043	0.257	0.378	0.833
	(0.006)	(0.014)	(0.091)	(0.013)	(0.134)
0.310	0	0.033	0.255	0.387	0.842
	(0.009)	(0.044)	(0.041)	(0.059)	(0.059)
0.484	0	0.045	0.271	0.471	0.870
	(0.011)	(0.049)	(0.014)	(0.040)	(0.032)
0.701	0	0.061	0.322	0.521	0.969
	(0.014)	(0.058)	(0.013)	(0.063)	(0.081)
0.893	0	0.080	0.317	0.682	0.964
	(0.018)	(0.061)	(0.038)	(0.040)	(0.044)

These estimates of β_0 and β_1 were used in conjuction with other information to compute optimal bandwidths. Under the simulation conditions, $l(x)$ is the uniform density on [0,1], so $l'(x) = 0$, $D_x = (\beta_0 + \beta_1 x)(1 - e^{-(\beta_0+\beta_1 x)t}e^{-2(\beta_0+\beta_1 x)t})/(2g(t))$, and $g_2(t,x) = -(\beta_1 t)^2 e^{-(\beta_0+\beta_1 x)t}$. The uniform kernel was used to generate the neighborhood weights. For the uniform kernel, $c_1 = 1, I_{K,2} = \frac{2}{3}$, and $c_2 = 2$. The value $v = 1.05423856$ was used for Var$(2V)$ (see Groeneboom and Wellner (2001)).

The estimate $\hat{F}_x$ was computed for all combinations of x and t values corresponding to the $(10^{th}, 20^{th}, \ldots, 90^{th})$ sample percentiles of the observed X and T values. For any value x the PAV algorithm actually computes estimates of F_x only for values of t which correspond to observed T_i associated with the X_i values in the neighborhood of x. Some of the t values may not belong to the neighborhood grid, hence no estimate is computed for these values. In such cases we either use linear interpolation (if the t value for which the estimation is required is between two values on the T_i grid), set $\hat{F}_x$ to be 0 (if the corresponding t value is less than the smallest T_i value in the neighborhood grid), or set $\hat{F}_x$ to be 1 (if the corresponding t value is greater than the largest T_i value in the neighborhood grid).

Table 4.1 gives the results for the above simulation when the asymptotically optimal bandwidth was estimated using the exponential model. The table presents the value of $\hat{F}_x$, and the absolute difference between $\hat{F}_x$ and F_x, for five of the X and five of the T percentiles, i.e., the $(10^{th}, 30^{th}, 50^{th}, 70^{th}$, and $90^{th})$. For the same sample used above, the asymptotically optimal bandwidths were also calculated using quantities computed from the true gamma distribution — all other conditions remained the same. Table 4.2 presents these results. The estimates of $\hat{F}_x(t)$ using each of these bandwidths and true values of $F_x(t)$ are graphed in Figure 4.1.

From the tables and graphs we note that the estimates appear to do reasonably well overall. There is no obvious systematic difference between the estimates obtained using the plug-in bandwidth via the exponential approximation, and those obtained using the true underlying distribution.

5 Concluding Remarks

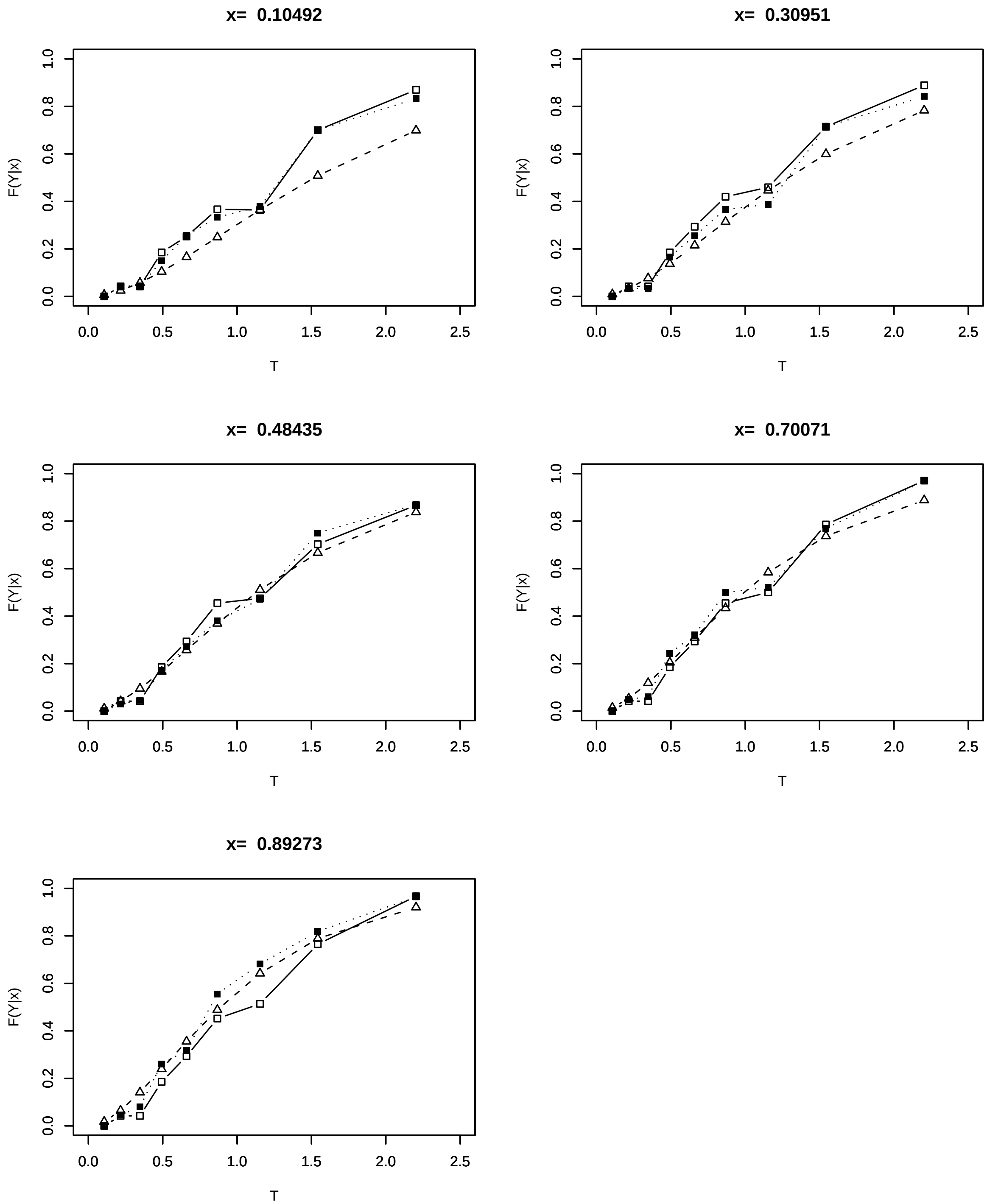
We have proposed a completely nonparametric estimator, $\hat{F}_x$, for the conditional distribution function for current status data, in the presence of a single continuous-valued covariate. The asymptotically optimal bandwidth, derived using heuristic weak convergence results for $\hat{F}_x(t)$, depends on values of both x and t, and is shown to be proportional to $n^{-1/7}$. We showed that the estimator $\hat{F}_x(t)$ can be obtained as a solution to an isotonic regression problem, and is easily computed. A method for obtaining a plug-in estimate of the bandwidth was suggested, and an example where this plug-in choice was contrasted with the optimal bandwidth obtained by using constants from the true distribution was presented. The plug-in choice is seen to perform favorably on the sample used.

Only Nadaraya-Watson weights based on kernel methods have been considered here. The estimation scheme can be extended to incorporate other weighting methods. In principle, one could also extend the methodology to situations with multiple covariates. However, the "curse of dimensionality" is likely to cause a problem for large dimensions of the covariate vector. Study of the multiple covariate situation is planned for future work.

The unusual $n^{-1/7}$ rate of the optimal bandwidth is particularly notable. This rate has also been obtained by Groeneboom (1998) for the bandwidth in the deconvolution problem, and used by Roy Choudhury (1998) in the problem of deblurring images that are blurred by Poisson noise.

For current status data in the absence of covariates, Groeneboom and Wellner (1992) have shown that $\hat{F}(t)$ converges at an $n^{1/3}$ rate, whereas the mean of F converges at the $n^{1/2}$ rate. This is unlike uncensored data, in that the same $n^{1/2}$ rate is attained for estimation of the distribution function and the mean under similar assumptions. A conjecture that warrants additional investigation is that a better rate can be obtained for the nonparametric estimation of the (conditional) mean in the presence of a covariate too.

Figure 1: *Graph of F_x (triangle), $\hat{F}_x$ when h_{opt} was obtained from estimated distribution of $Y|X$ (unfilled square), and $\hat{F}_x$ when h_{opt} was obtained from true distribution $Y|X$ (filled square).*



Appendix: Approximation for $F_{K_x}(t) - F_x(t)$

We show here that (3.25) holds for a kernel that satisfies assumptions A_1, A_2 and A_3 for all points x for which a full kernel neighborhood is possible. D_x, C_h, $I_{K,2}$, and $C_x(t)$ are as defined in (3.9), (3.22), (3.23), and (3.26) respectively, $g_1(t,x) = \frac{\partial F_z(t)}{\partial z}|_x$, and $g_2(t,x) = \frac{\partial^2 F_z(t)}{\partial z}|_x$.

Using (3.12) and the three assumptions in Section 3, when we have symmetric probability kernels, we have the following

$$\begin{aligned} F_{K_x}(t) &= \int_z w_x^*(z) F_z(t) dL(z) \\ &= \int_{z\in N_{x,h}} w_x^*(z) F_z(t) dL(z) + \int_{z\in N_{x,h}^c} w_x^*(z) F_z(t) dL(z) \\ &\approx \int_{z\in N_{x,h}} w_x^*(z) F_x(t) dL(z) \\ &\quad + \int_{z\in N_{x,h}} w_x^*(z)(z-x) g_1(t,x) dL(z) \\ &\quad + \int_{z\in N_{x,h}} w_x^*(z)(z-x)^2 g_2(t,x) dL(z). \end{aligned} \tag{A.1}$$

The first term in (A.1) can be simplified as follows using A_3:

$$\int_{N_{x,h}} w_x^*(z) F_x(t) dL(z) = F_x(t) \int_{N_{x,h}} w_x^*(z) dL(z) = F_x(t). \tag{A.2}$$

Using (3.11) we can simplify the second term in (A.1) as follows:

$$\begin{aligned} \int_{N_{x,h}} w_x^*(z)(z-x) g_1(t,x) dL(z) &= g_1(t,x) \int_{N_{x,h}} w_x^*(z)(z-x) dL(z) \\ &= g_1(t,x) \int_{N_{x,h}} \frac{(z-x)K(\frac{x-z}{h})}{C_h} dL(z) \\ &= \frac{g_1(t,x)}{C_h} \int_{N_{x,h}} (z-x)K(\frac{x-z}{h})\Big(l(x) + (z-x)l'(x) + O(z-x)^2\Big) dz \end{aligned}$$

Transforming $u = \frac{x-z}{h}$, and denoting the support of the kernel by U, we get

$$\begin{aligned} \int_{N_{x,h}} w_x^*(z)(z-x) g_1(t,x) dL(z) &\approx \frac{g_1(t,x)h^2}{C_h}\Big\{-\int_U uK(u)l(x)du + \int_U hu^2 l'(x)K(u)du\Big\} \\ &= g_1(t,x) l'(x) I_{K,2} \frac{h^3}{C_h}. \end{aligned} \tag{A.3}$$

Finally,

$$\begin{aligned} \int_{N_{x,h}} w_x^*(z) \frac{(z-x)^2}{2} g_2(t,x) dL(z) &= \frac{g_2(t,x)}{2} \int_{N_{(x,h)}} w_x^*(z)(z-x)^2 dL(z) \\ &= \frac{g_2(t,x)}{2C_h} \int_{N_{(x,h)}} K(\frac{x-z}{h})(z-x)^2 dL(z) \end{aligned} \tag{A.4}$$

Transforming $u = \frac{x-z}{h}$, and denoting the support of the kernel by U, we get

$$\begin{aligned}
\int_{N_{x,h}} w_x^*(z) \frac{(z-x)^2}{2} g_2(t,x) dL(z) &\approx \frac{g_2(t,x)h^3}{2C_h} \Big\{ \int_U u^2 K(u) l(x) du - \int_U h u^3 K(u) l'(x) du \Big\} \\
&= \frac{g_2(t,x) l(x) I_{K,2}}{2} \frac{h^3}{C_h} \qquad \text{(A.5)}
\end{aligned}$$

Substituting for (A.2), (A.3) and (A.5) in (A.1) we get

$$F_{K_x}(t) - F_x(t) \approx \left(g_1(t,x) l'(x) + \tfrac{g_2(t,x) l(x)}{2} \right) \tfrac{I_{K,2} h^3}{C_h}, \qquad \text{(A.6)}$$

and this establishes (3.25).

References

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., and Silverman E. (1955), "An Empirical Distribution Function for Sampling with Incomplete Information," *The Annals of Mathematical Statistics*, 26, 641–647.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., Brunk, H. D. (1972), *Statistical Inference under Order Restrictions,* Wiley.

Brunk, H. D. (1955), "Maximum Likelihood Estimates of Monotone Parameters," *The Annals of Mathematical Statistics*, 26, 607–616.

Dempster, A.P., Laird, N. M., Rubin, D. B. (1977), "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, Series B, 39, 1–38.

Diamond, I. D., and McDonald, J. W. (1991), "Analysis of Current Status Data," in *Demographic Applications of Event History Analysis,* eds. J. Trussell, R. Hankinson, and J. Tilton, Oxford, U.K.: Oxford University Press, pp. 231–252

Diamond, I. D., McDonald, J. W., and Shah, I. H. (1986), "Proportional Hazards Models for Current Status Data: application to the Study of Differentials in Age at Weaning in Pakistan," *Demography,* 23, 607–620.

Groeneboom, P. (1998), "Nonparametric Estimation for Inverse Problems: Algorithms and Asymptotics," *Technical report No. 344, December 1998*, Department of Statistics, University of Washington.

Groeneboom, P. and Wellner, J. (1992), *Information Bounds and Nonparametric Maximum Likelihood Estimation*, Birkhauser.

Groeneboom, P. and Wellner, J. (2001), "Computing Chernoff's distribution," *Journal of Computational and Graphical Statistics,* 10, 388–400.

Huang, J., and Wellner, J. (1995), "Asymptotic Normality of the NPMLE of Linear Functionals for Interval Censored Data, Case I," *Statistical Neerlandica*, 49, 153–163.

Huang, J., and Wellner, J. (1996), "Interval Censored Survival Data: A Review of Recent Progress," *Proceedings of the First Seattle Symposium in Biostatistics*, D. Y. Lin and T. R. Fleming, editors.

Jewell, N. P., and Shiboski, S. C. (1990), "Statistical Analysis of HIV Infectivity Based on Partner Studies," *Biometrics*, 46, 1133–1150.

Jewell, N. P., Malani, H. M. and Vittinghoff, E. (1994), "Nonparametric Estimation for a Form of Doubly Censored Data, with Application to Two Problems in AIDS," *Journal* *of the American Statistical Association*, 89, No. 425, 7–18.

Keiding, N. (1991), "Age-Specific Incidence and Prevalence," (with discussion), *Journal of the Royal Statistical Society,* Series A, 154, 371–412.

Murphy, S. A., van der Vaart, A. W., and Wellner, J. A. (1999), "Current Status Regression," *Mathematical Methods in Statistics*, 8, 407–425.

Nadaraya, E. A. (1964), "On Nonparametric Estimates of Density Functions and Regression Curves," *Theory of Probability and its Applications*, 15, 134–137.

Rabinowitz, D., and Jewell, N. P. (1996), "Regression with Doubly Censored Current Status Data," *Journal of the Royal Statistical Society*, Series B, 58, 541–550.

Rabinowitz, D., Tsiatis, A., and Aragon, J. (1995), "Regression with Interval-Censored Data," *Biometrika*, 82, 501–513.

Rossini, A., and Tsiatis, A. (1994), "A Semiparametric Proportional Odds Regression Model for the Analysis of Current Status Data," *Journal of the American Statistical Association*, 91, No. 434, 713–721.

Robertson, T., Wright, F. T., and Dykstra, R. L. (1988), *Order Restricted Statistical Inference*, New York: Wiley.

Roy Choudhury, K. (1998), "Additive Mixture Models for Multichannel Image Data", *Ph. D. dissertation,* Department of Statistics, University of Washington.

Turnbull, B. W. (1976), "The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data," *Journal of the Royal Statistical Society*, Series B, 38, 290–295.

van der Laan, M. (1994), "Proving Efficiency of NPMLE and Identities," *Technical report, April 1994*, Group in Biostatistics, University of California, Berkeley.

van der Laan, M., Bickel, P., and Jewell, N. P. (1994), "Singly and Doubly Censored Current Status Data: Estimation, Asymptotics and Regression," *Technical report, November 1994*, Group in Biostatistics, University of California, Berkeley.

Watson, G. S. (1964), "Smooth Regression Analysis," *Sankhya Series A*, 26, 359–372.