Nonparametric Statistics, Vol. 4, pp. 203–209
Reprints available directly from the publisher
Photocopying permitted by license only

Printed in the United States of America

SOME HEURISTICS OF KERNEL BASED ESTIMATORS OF RATIO FUNCTIONS

P. N. PATIL,[1] M. T. WELLS[2] and J. S. MARRON[1,2]

[1] University of North Carolina at Chapel Hill and [2] Cornell University

(Received: January 3, 1991; Revised: December 1, 1993; Accepted December 15, 1993)

Ratio functions for which nonparametric estimators have been considered include the hazard rate and density under random censoring. One estimation method involves individual estimates of the numerator and denominator. An alternative targets the entire function not the separate pieces. The two estimators are not comparable in terms of Mean Integrated Squared Error. However, the second type is seen to be more natural because it admits an elegant and useful martingale representation, and also is pictorially more attractive.

KEYWORDS: Censored and uncensored samples, cumulative hazard function, hazard rate, kernel estimators, martingale.

1. INTRODUCTION

Kernel type estimators of ratio functions, such as (a) the density under random censoring (b) the hazard rate and (c) the hazard rate under random censoring, have been studied by several authors (eg. Watson and Leadbetter [11, 12], Rice and Rosenblatt [4], Singpurwalla and Wong [7], Tanner and Wong [8, 9] Marron and Padgett [3] and Lo, Mack and Wang [2]). The kernel estimators of such ratio functions involve evaluation of a cumulative distribution function (c.d.f.) estimate. There are usually two choices of how this may be done, as shown in Section 2, resulting in two related but different estimators. Both are based on delta sequence smoothing, introduced by Watson and Leadbetter [12], Földes and Révész [1], Rice and Rosenblatt [4], and Walter and Blum [10]. The aim of this paper is the comparison of these two estimators. Asymptotic analysis of their mean integrated square errors, as done in Section 3, shows that the variances are the same but with regard to bias neither estimator is uniformly superior to the other. In particular for some choices of the underlying density function bias of one estimator is smaller, while for other choices bias of the other estimator is smaller.

In this note we provide new insight into the choice between these two estimators. We argue that one of the estimators is more natural. The basis of our comparison of these estimators is their structure. One estimator may be viewed as separately estimating the numerator and denominator of the target function, while the other is motivated only in terms of the entire function. We consider the latter type of estimator to be more natural in two senses. First, it admits an important and simple type of martingale representation which the former does not seem to have and second, the "effective kernel" of the latter type of estimator is smooth while the "effective kernel" of the former has undesirable discontinuities.

2. THE ESTIMATORS AND NOTATION;

Let $X_1^0, X_2^0, \ldots, X_n^0$ denote the independent identically distributed (i.i.d.) survival times of n items or individuals that are censored on the right by i.i.d. random variables $U_1, U_2, \ldots, U_n$ which are independent of the X_i's. Denote the common distribution function of the X_i^0's by F^0 and that of the U_i's by H. It is assumed that F^0 is absolutely continuous with density f^0 and that H is continuous. The observed randomly right censored data are denoted by the pairs (X_i, Δ_i) $i=1, 2, \ldots, n$ where

$$X_i = \min\{X_i^0, U_i\} \quad \text{and} \quad \Delta_i = I_{[X_i^0 \leq U_i]}.$$

The X_i's form an i.i.d. sample from a distribution F where $1-F=(1-F^0)(1-H)$. A general formulation of the target function in all of the cases discussed in Section 1 is $\eta(x)$ where

$$\eta(x) = \frac{(1-H(x))f^0(x)}{Q(x)}, \tag{2.1}$$

for $Q(x)>0$, where $Q(x)$ is a non-increasing function such that $0 \leq Q(x) \leq 1$, $z \in \mathcal{R}$.

REMARK 2.1. (i) If $Q(x)=1-F(x)$ then we have the case of hazard rate estimation in the censored data setting. (ii) If $Q(x)=1-H(x)$ then we have the case of density estimation in the censored setting.

REMARK 2.2. If the censoring random variable has all its mass at ∞ then $H(x)=0$ for all $x \in \mathcal{R}$ and $F^0(x) \equiv F(x)$. (i) If $Q(x)=1-F(x)$ we have the case of hazard rate estimation in the uncensored data setting. Also note that (ii) if we take $Q(x) \equiv 1$ we get the usual probability density.

To motivate estimators of $\eta(x)$, define the sub-distribution functions

$$F^+(x) = P[X_i \leq x, \Delta_i = 0] = \int_0^x (1-F^0(u))\, \mathrm{d}H(u)$$

$$F^-(x) = P[X_i \leq x, \Delta_i = 1] = \int_0^x (1-H(u))\, \mathrm{d}F^0(u).$$

Clearly $F(x)=F^+(x)+F^-(x)$. Define the empirical distribution functions

$$F_n(x) = F_n^+(x) + F_n^-(x),$$

where

$$F_n^+(x) = n^{-1} \sum_{i=1}^n I_{[X_i \leq x]}(1-\Delta_i), \qquad F_n^-(x) = n^{-1} \sum_{i=1}^n I_{[X_i \leq x]}\Delta_i.$$

Let $F_x^0(x)$ be the Kaplan-Meier estimate of F^0, and let H_n be such that $(1-F_n(x)) = (1-F_n^0(x))(1-H_n(x))$. Observe that $n^{-1}\sum_{i=1}^n K_h(x-X_i)\Delta_i$

provides an estimate of $f^0(x)(1-H(x))$, see Marron and Padgett [3]. We assume $Q(x)$ to be one of the functions considered in the Remarks 1 and 2. Define $Q_n(x)$ to be the empirical version of $Q(x)$. Then the straightforward ratio estimator $\eta_{n,1}(x)$ of $\eta(x)$, obtained by plugging estimators into the numerator and denominator, is

$$\eta_{n,1}(x) = n^{-1} \sum_{i=1}^{n} \frac{\Delta_i K_h(x-X_i)}{Q_n(x)} \tag{2.2}$$

which is one of the estimators considered by Marron and Padgett [3] and Watson and Leadbetter [11].

Define the cumulative target function

$$\Gamma(x) = \int_0^x \eta(u)\,du = \int_0^x \frac{(1-H(u))}{Q(u)}\,\mathrm{d}F^0(u) = \int_0^x \frac{\mathrm{d}F^-(u)}{Q(u)}.$$

An obvious estimate of $\Gamma(x)$ is its empirical version,

$$\Gamma_n(x) = \int_0^x \frac{(1-H_n(u))}{Q_n(u)}\,\mathrm{d}F_n^0(u) = \int_0^x \frac{\mathrm{d}F_n^-(u)}{Q_n(u)}.$$

Now we consider $\eta(x)$ as a function on its own, rather than a ratio and approximate it by $\eta^*(x) = \int K_h(x-u)\eta(u)\,\mathrm{d}u$ where $K_h(x-u)$ is a kernel. Then find the estimator of $\eta^*(x)$. That is,

$$\begin{aligned}\eta(x) \approx \eta^*(x) &= \int K_h(x-u)\eta(u)\,\mathrm{d}u = \int K_h(x-u)\,\mathrm{d}\Gamma(u)\\ &= \int K_h(x-u)\,\mathrm{d}\Gamma_n(u) + \int K_h(x-u)\,\mathrm{d}[\Gamma(u)-\Gamma_n(u)].\end{aligned} \tag{2.3}$$

Such heuristic motivation gives the estimator

$$\eta_{n,2}(x) = \int K_h(x-u)\,\mathrm{d}\Gamma_n(u) = n^{-1} \sum_{i=1}^{n} \frac{K_h(x-X_i)\Delta_i}{Q_n(X_i)} \tag{2.4}$$

This is the estimator Tanner and Wong [9] have used and is one of the estimators Watson and Leadbetter [11] and Marron and Padgett [3] have considered in their settings, although no motivation of this type is given. The representation in (2.3) shows that the approximation of the function $\eta(x)$ by $\eta^*(x)$ is the bias part of the error, while the random part of the error is $\int K_h(x-u)\,\mathrm{d}[\Gamma(u)-\Gamma_n(u)]$. In the next section it is shown that the random part of the error admits a useful martingale representation.

Note that the only difference between (2.2) and (2.4) is the argument of the denominator, x for the former and X_i for the latter. The main objective of this note is to understand how these two possibilities compare.

3. COMPARISONS OF THE ESTIMATORS

3.1. Mean and Variance

First note that as Q_n converges to Q at the fast rate of $n^{-1/2}$, (see for eg. Serfling [5]), asymptotically $\eta_{n,1}(x)$ and $\eta_{n,2}(x)$ are equivalent to

$$\bar{\eta}_{n,1}(x) = n^{-1} \sum_{i=1}^{n} \frac{\Delta_i K_h(x-X_i)}{Q(x)} \quad \text{and} \quad \bar{\eta}_{n,2}(x) = n^{-1} \sum_{i=1}^{n} \frac{K_h(x-X_i)\Delta_i}{Q(X_i)}.$$

THEOREM 1. Assume that the conditions of Theorem 2 of Tanner and Wong [9] hold. Then,

$$E[\tilde{\eta}_{n,1}(x)-\eta(x)]=\int K(u)\left[\eta(x-hu)\frac{Q(x-hu)}{Q(x)}-\eta(x)\right]\mathrm{d}u,$$

$$\operatorname{Var}(\tilde{\eta}_{n,1}(x))=(nh)^{-1}\left[\int K^2(u)\,\mathrm{d}u\right]\eta(x)[Q(x)]^{-1}+o(n^{-1}h^{-1}).$$

And

$$E[\tilde{\eta}_{n,2}(x)-\eta(x)]=\int K(u)[\eta(x-hu)-\eta(x)]\,\mathrm{d}u,$$

$$\operatorname{Var}(\tilde{\eta}_{n,2}(x))=(nh)^{-1}\left[\int K^2(u)\,du\right]\eta(x)[Q(x)]^{-1}+o(n^{-1}h^{-1}).$$

Proof. See Watson and Leadbetter [11] and Tanner and Wong [9].

By Theorem 1, $\mathrm{MISE}(\tilde{\eta}_{n,1}(x))=an^{-1}h^{-1}+b_1+o(n^{-1}h^{-1})$ and $\mathrm{MISE}(\tilde{\eta}_{n,2}(x))=an^{-1}h^{-1}+b_2+o(n^{-1}h^{-1})$, with

$$a=\left[\int K^2(u)\,\mathrm{d}u\right]\int \eta(x)[Q(x)]^{-1}w(x)\,\mathrm{d}x,$$

$$b_i=\int B_i^2(x,h)w(x)\,\mathrm{d}x,\quad i=1,2,$$

where

$$B_1(x,h)=\int K(u)\left[\eta(x-hu)\frac{Q(x-hu)}{Q(x)}-\eta(x)\right]\mathrm{d}u$$

and $B_2(x,h)=\int K(u)[\eta(x-hu)-\eta(x)]\,\mathrm{d}u$, and $w(x)$ is a weight function which is chosen so that the integrals are finite.

The difference in the asymptotic analysis of these MISE's shows up in the bias part. Insight into how these quantities compare is easily obtained by Taylor series expansion. In particular, if η and Q are twice differentiable, then as $h\to 0$,

$$\begin{aligned}B_1^2(x,h)&=h^4k^2\eta''^2(x)+h^4k^2[\eta(x)Q''(x)+2\eta'(x)Q'(x)]^2Q^{-2}(x)\\&\quad+2h^4k^2[\eta(x)\eta''(x)Q''(x)+2\eta'(x)\eta''(x)Q'(x)]Q^{-1}(x)+o(h^4),\\B_2^2(x,h)&=h^4k^2\eta''^2(x)+o(h^4),\end{aligned}$$

where $k=\int(u^2/2)K(u)\,\mathrm{d}u$. Note that at an inflection point of $\eta(x)$, $B_1^2(x,h)=h^4k^2[\eta(x)Q''(x)+2\eta'(x)Q'(x)]^2Q^{-2}(x)+o(h^4)$ and $B_2^2(x,h)=o(h^4)$, so that $\eta_{n,2}$ is better. On the other hand, at an inflection point of $Q(x)$, squared bias in $\eta_{n,1}(x)$ will be smaller than $\eta_{n,2}(x)$ when $(2\eta'Q')^2Q^{-2}+4\eta'\eta''Q'Q^{-1}<0$, but close to 0. But this inequality is equivalent to $[\eta'Q'][\mathrm{d}/\mathrm{d}x(n'Q)]<0$. Since this is sometimes true and sometimes false, the two estimators are not comparable in terms of MSE. Which is better depends on the underlying setting, and the point where the estimation is being done.

3.2. *Martingale Representation*

To see the martingale structure of the random error term, $\int K_h(x-u)\,\mathrm{d}[\Gamma(u)-\Gamma_n(u)]$, note that

$$n^{1/2}\{\Gamma_n(x)-\Gamma(x)\}=\int_0^x \frac{1}{Q_n(u)}\,\mathrm{d}M_n(u)$$

which is a stochastic integral of the predictable process $Q_n(u)^{-1}$ with respect to the martingale

$$M_n(x)=n^{1/2}\left[F_n^-(x)-\int_0^x \frac{Q_n(u)}{Q(u)}\,\mathrm{d}F(u)\right],$$

with associated filtration $F_x=\sigma(I_{[X_i^0\le u]}, I_{[X_i^0\le u]}\Delta_i, I_{[U_i\le u]}\Delta_o\colon 1\le i\le n,\ u\le x)$. The weak convergence of the process $n^{1/2}\{\Gamma_n(x)-\Gamma(x)\}$ is discussed in Chapter 7 of Shorack and Wellner [6].

The elegance and power of this representation is seen in Wells [13] who used the representation in (2.3) to show that $(nh)^{1/2}\{\eta_{n,2}(x)-\eta(x)\}$ converges weakly to a Gaussian process on $D[0,\infty)$. The main step of the proof is the application of Rebolledo's central limit theorem to

$$(nh)^{1/2}\{\eta_{n,2}(x)-\eta^*(x)\}=\int \frac{h^{-1/2}K\left(\dfrac{x-u}{h}\right)}{Q_n(u)}\,\mathrm{d}M_n(u)$$

On the other hand, $\eta_{n,1}$ has no such simple representation that the authors have found. This makes $\eta_{n,2}$ not only more technically tractable, but also provides an intuitive sense in which it is the more natural of the two. The martingale approach is more natural as well since it reduces the estimation problem to the classical White Noise problem. The White Noise problem is perhaps one of the most studied problems in nonparametric function estimation.

3.3. *Practical Considerations*

Another difference in the above discussed estimators, which seems to have been overlooked so far, is the form of their "effective kernel" which we define as follows. Effectively one has to put the kernel function $K_h(x-X_i)[Q_n(x)]^{-1}$ and $K_h(x-X_i)[Q_n(X_i)]^{-1}$ at each sample point respectively for the estimators $\eta_{n,1}(x)$ and $\eta_{n,2}(x)$ and then average them together. So we call $K_h(x-X_i)[1-Q_n(x)]^{-1}$ and $K_h(x-X_i)[1-Q_n(X_i)]^{-1}$ the effective kernels of the estimators $\eta_{n,1}(x)$ and $\eta_{n,2}(x)$ respectively. To bring out the difference between these two types of effective kernels we consider the special case where $H(x)=0$ for all $x\in\mathfrak{R}$ and $Q_n(x)=1-F_n(x)$. This difference is made evident in Figure 1. The effective kernels are shown as functions of x, the location of estimation. They are shown for observations X_1, X_2, X_3 in the locations represented by the vertical lines. Effective kernels are shown, centered at both X_1 and at X_2. The dotted lines are the $\eta_{n,1}$ kernels, while $\eta_{n,2}$ kernels are shown as dashed lines. Note that both types of kernels get larger with increasing x, which reflects the fact that $1-F_n(\cdot)$ is a decreasing function. Disturbing features of the effective kernel of $\eta_{n,1}$ are

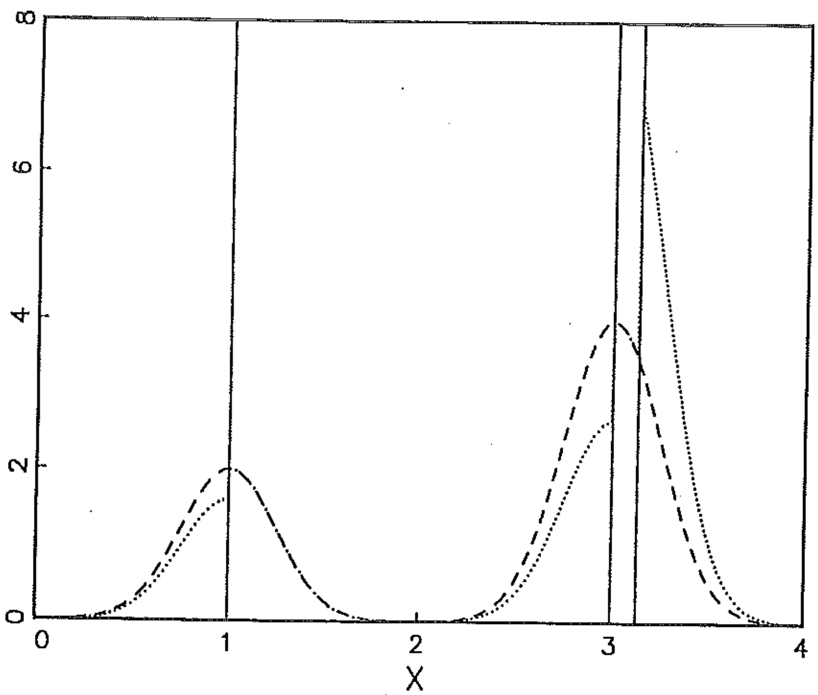

Figure 1. Effective kernels for $\eta_{n,1}$ – dotted lines; $\eta_{n,2}$ – dashed lines, vertical lines are location of the data points.

the jump discontinuities at each of the sample observations caught in the window width of the kernel (caused by the discontinuities in $1-F_n(x)$). On the other hand the effective kernel of $\eta_{n,2}$ is a smooth curve. Thus $\eta_{n,2}$ will give a smooth estimate of η, while the estimator $\eta_{n,1}$ always results in a discontinuous estimate of η. This is an undesirable property for a smoothing method, which we view as another considerable disadvantage of $\eta_{n,1}$ with respect to $\eta_{n,2}$.

Acknowledgement

This research was supported by the Mathematical Sciences Institute at Cornell University, and by the National Science Foundation Grant 8902973.

References

1. A. Földes and P. Révész, A general method for density estimation. *Studia Sci. Math. Hungar,* **9,** (1974), 81–82.
2. S. H. Lo, Y. P. Mack and J. L. Wang, Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator. *Probab. Theory Rel. Fields* **80,** (1989), 461–473.
3. J. S. Marron and W. J. Padgett, Asymptotically optimal bandwidth selection for kernel density estimators from randomly right censored samples. *Ann. Statist.* **15,** (1987), 1520–1535.
4. J. Rice and M. Rosenblatt, Estimation of the log survivor function and hazard function. *Sankhya Ser. A* **38,** (1976), 60–78.
5. R. J. Serfling, *Approximation Theorems of Mathematical Statistics.* Wiley, New York (1980).
6. G. R. Shorack and J. A. Wellner, *Empirical Processes with Applications to Statistics,* Wiley, New York, (1986).
7. N. D. Singpurwalla and M. Y. Wong, Kernel estimators of the failure rate function and density function: An analogy. *J. Amer. Statist. Assoc.* **78,** (1983), 478–481.
8. M. A. Tanner and W. H. Wong, Data-based nonparametric estimation of the hazard function with applications to model diagnostics and exploratory analysis. *J. Amer. Statist. Assoc.* **79,** (1984), 174–182.

9. M. A. Tanner and W. H. Wong, The estimation of the hazard function from randomly censored data by kernel method. *Ann. Statist.* **11,** (1983), 989–993.
10. G. Walter and J. R. Blum, Probability density estimation using delta sequences. *Ann. Statist.* **7,** (1979), 328–340.
11. G. S. Watson and M. R. Leadbetter, Hazard analysis, I. *Biometrika* **51,** (1964), 175–184.
12. G. S. Watson and M. R. Leadbetter, Hazard analysis II. *Sankhya Ser. A* **26,** (1964), 101–116.
13. M. T. Wells, On the estimation of hazard rates and their extreme from general randomly censored data. To be published, (1989).