A kernel estimator of cumulative hazard function and its application to reliability problem.

González Henríquez J. J. Saavedra Santana, P.

Abstract

A kernel estimator of integrated hazard is considered and the bandwidth optimal is investigated. A problem of reliability such that the optimun replacement policy depends on survival funtion is also considered. By means of simulation study, the optimun stopping time obtained through of the proposed kernel estimator of the survival function is compared whit the obtained by Kaplan-Meier estimator

A kernel estimator of cumulative hazard function

In this section we will study the large sample properties of a kernel estimator of a cumulative hazard function from randomly censored data. To this end, let $T_1, T_2, ..., T_n$ be i.i.d. random variables with a continuous distribution F_T, which represents the life of the items (or individuals) under observation. Associated with each T_i is an independent censoring variable C_i. Further $C_1, C_2, ..., C_n$ are assumed to be i.i.d. from a distribution F_C. Due to random censoring, only $X_i = \min(T_i, C_i)$ and $\delta_i = I_{\{T_i \le C_i\}}$ are observable. Formally, we define the estimator of $\Lambda_T(x) = -\log(1 - F_T(x))$ by

$$\widehat{\Lambda}(x;h) = \sum_{j=1}^{n} \frac{\delta_{(j)} W\left(\frac{X_{(j)}-x}{h}\right)}{(n-j+1)}. \tag{1}$$

Here, $X_{(1)}, X_{(2)}, ..., X_{(n)}$ and $\delta_{(1)}$, $\delta_{(2)}, ...,$ $\delta_{(n)}$ represents the ordered X's and δ's, respectively and W is a distribution funtion, which probability density w is bounded, symmetric and compactly supported. The parameter h controls the degree of smoothing.

The following theorems shows small and large sample expressions for the mean and variance of the estimator $\widehat{\Lambda}(x;h)$.

Theorem 1 $E[\widehat{\Lambda}(x;h)] = \int (1 - F^n(y))\lambda_T(y) W(\frac{x-y}{h})dy,$

$$Var[\widehat{\Lambda}(x;h)] = \int I_n(F(y)\lambda_T(y)W^2(\tfrac{x-y}{h})dy + 2\iint_{y\le z} \{F^n(y) - F^n(y)F^n(z)$$
$$-\tfrac{1-F(y)}{F(z)-F(y)}[F^n(y) - F^n(z)]\Big\} \lambda_T(y)\lambda_T(z)W(\tfrac{x-y}{h})W(\tfrac{x-z}{h})dydz$$

Theorem 2 *Assume that $\Lambda_T(x)$ has two continuous derivatives and the bandwidth $h \to 0$ and $nh \to \infty$ as $n \to \infty$,. Then,*

i) $E\left[\widehat{\Lambda}(x;h)\right] = \Lambda(x;h) + \frac{\lambda_T'(x)\mu_2(K)}{2}h^2 + o(h^2)$

ii) $Var\left[\widehat{\Lambda}(x;h)\right] = \int_{-\infty}^{x} \frac{\lambda_T(y)}{n(1-F(y))}dy - \frac{\lambda_T(x)\mu_1(K^*)}{n(1-F(x))}h + o(n^{-1}) + o(h)$

where K^* is a probability density.

As an inmediate consequence of the above theorem, $\widehat{\Lambda}(x;h)$ is a pointwise consistent estimator of $\Lambda_T(x)$ in quadratic mean.

Optimal Bandwidth

The estimator $\widehat{\Lambda}(x;h)$ is consistent if the bandwidth h tends to zero sufficiently slowly as n increases. To be able to study more closely how the bandwich should tend to zero, and even be able to pick an optimal bandwidth, we first have to select a measure for the global perfomance the estimator. Like most of the literature on nonparametric estimation, we will choice as optimal smoothing parameter h that mimimises $MISE(h) = E\int\left\{\widehat{\Lambda}(x;h) - \Lambda(x;h)\right\}^2 dx$. So the global criterion $MISE(h)$ can be deriver easily from theorem 2.

$$MISE(h) = \frac{\lambda_T'(x)^2\mu_2(K)^2}{4}h^4 + \int_{-\infty}^{x} \frac{\lambda_T(y)}{n(1-F(y))}dy - \frac{\lambda_T(x)\mu_1(K^*)}{n(1-F(x))}h + o(h^2) + o(n^{-1}) + o(h) \quad (2)$$

The minimization of the leading term in (2) results an asymptotically bandwidth which is give by

$$h_o = \left(\frac{\mu_1(K^*)\int \lambda_T(x)(1-F(x))^{-1}dx}{n\left\|\lambda_T'(y)\right\|_2^2 \mu_2(K)^2}\right)^{\frac{1}{3}} \quad (3)$$

A application in reliability

Consider a functioning system with specified life distribution F, and probability of survival to age x, $S(x) = 1 - F(x)$. If the sytem fails prior t units of time after installaction, it is replaced at that failure with cost C_1. Otherwise, the system is replaced t units of time after its installation with cost C_2. The objetive is to minimize the expected long-run average cost,

$$C(t) = \frac{C_1 F(t) + C_2 S(t)}{\int_0^t S(u)du} \tag{4}$$

It is assumed that there is a unique and finite time, ϕ^*, where $C(t)$ attains a global minimun. There are several conditions which guarantee this, see Bergman (1979). An example of such a condition, which we will use here, is that the failure rate $f(x)/S(x)$ be strictly increasing to infinity with x.

In 1991 Aras and whitacker developed a sequential nonparametric scheme to estimate the optimal age replacement policy ϕ^*. In that paper, an estimator of the cost function $C(t)$ was the plug-in estimator

$$\widehat{C}_n(t) = \frac{C_1 - (C_1 - C_2)S_n(t-)}{\int_0^t S_n(u)du} \tag{5}$$

where $S_n(t)$ was the well known product-limit estimator. Thus, an estimator of ϕ^* is ϕ_n where a minimun is reached. Here, we present a variation on the Aras and Whitaker scheme. The sequential estimators $\{\phi_n^*\}$ are the same but $S_n(t)$ will be estimated as $e^{-\widehat{\Lambda}(t;h)}$.

A Monte-Carlo study was undertaken to demostrate the usefulness of the variation proposed. For the model of the problem, we took $C_1 = 1000$ and $C_2 = 100$. The Weibull distribution was used with density

$$f(t) = \alpha\lambda(\lambda t)^{\alpha-1}e^{-(\lambda t)^{\alpha}}$$

where $\alpha = 0.01$ and $\lambda = 2$. This produces a mean of 88.6227 and standard deviation of 46.3251 for the life times of the units. It also ensures a unique, finite $\phi^* = 33.6451$ and an optimal cost $C(\phi^*) = 6.0561$.

Figure 1 describes the performance of the estimator ϕ_{200} in the different model proposed. The alternative method presents lowest standard deviation and one ligth bias that the sheme of Aras and Withaher.

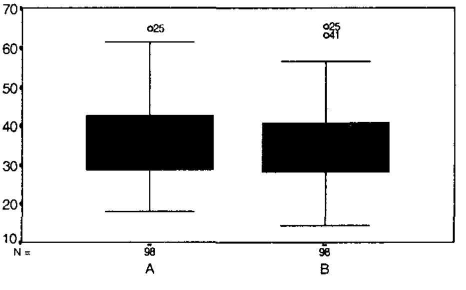

Figure 1:

References

[1] G. Aras, L. R. Whitacker, "A sequential scheme to estimate the optimal age replacement policy", *Conmun. Statist.-Theory Meth.*, **21**, 8, (1992), 2325–2338.

[2] M.A. Tunner, Wing Hung Wong "The estimation of the hazard function from randomly censored data by the kernel method", *The annals of Statistics*, **11**, 3, (1983), 989–993.

1. Departamento de Matemáticas. Universidad de Las Palmas de Gran Canaria. Campus Universitarior de Tafira. 35017. Las Palmas.