RMIT
UNIVERSITY

On nonparametric estimation of a reliability function

Zardasht, Vali; Zeephongsekul, Panlop; Asadi, M

ON NONPARAMETRIC ESTIMATION OF A RELIABILITY FUNCTION

V. Zardasht [1], P. Zeephongsekul [2] and M. Asadi [1]

[1] Department of Statistics

  University of Isfahan

  Isfahan, Iran.

[2] School of Mathematical and Geospatial Sciences

  RMIT University

  GPO Box 2476V, Melbourne 3001, Australia.

  panlopz@rmit.edu.au

ABSTRACT This paper considers the properties of a nonparametric estimator developed for a reliability function which is used in many reliability problems. Properties such as asymptotic unbiasedness and consistency are proven for the estimator and using U-statistics, weak convergence of the estimator to a normal distribution is shown. Finally, numerical examples based on an extensive simulation study are presented to illustrate the theory and compare the estimator developed in this paper with another based directly on the ratio of two empirical distributions studied in Zardasht and Asadi (2010).

1. INTRODUCTION

In reliability theory and related areas, methods of comparison between different reliability estimators have received a great deal of attention in the research literature. This is because these estimators allow reliability engineers and other decision makers to assess different products and their designs, different inspection and maintenance policies, different resource allocation policies and other similar processes. These comparisons are usually implemented by applying some well known stochastic orderings to pairs of random variables, usually interpreted as the lifetimes of two systems of interest. Depending on the applications, these orderings can be partial orderings or total orderings (refer to Shaked and Shantikhumar (1993) and Shaked and Shantikhumar (2007) for a comprehensive discussion on various types of stochastic orderings).

Let X and Y be two random variables with distribution functions F and G and survival functions $\bar{F} = 1 - F$ and $\bar{G} = 1 - G$, respectively. Assume that X and Y represent the lifetimes of two systems of interest, e.g. the products produced by two different companies, then the quantity $P(X > Y)$ gives the reliability of X relative to Y. Earlier on, it was pointed out by Brown and Rutemiller (1973) that in order to ensure that a designed product is long lasting, one should consider the quantity $P(X > Y)$ and then choose system with lifetime X or with lifetime Y depending on whether this probability is greater or less than 0.5 respectively. Indeed, based on whether $P(X \geq Y)$ is ≤ 0.5 or > 0.5, there is an ordering between the two random variables which is called stochastic precedence ordering. We say $X \leq_{sp} Y$ if $P(X \geq Y) \leq 0.5$. (For the significance of this precedence ordering and its

application, we refer the reader to Samanieo and Hollander (2008), Boland et al. (2004) and Singh and Misra (1994).)

Dynamic formulations of reliability theory are of considerable interest in the analysis and comparison of working systems in real time. In fact, when assessing and inspecting working systems, one is interested in the prospective performance of operating systems and comparing their residual lifetimes and related quantities in a dynamic setting. Motivated by this and stochastic precedence ordering, Zardasht and Asadi (2010) have introduced the reliability function $R(t) = P(X_t > Y_t)$ to compare two residual lifetimes $X_t = X - t | X > t$ and $Y_t = Y - t | Y > t$ where X and Y are two independent non-negative continuous random variables and the symbol $X|A$ represents the random variable X conditioned on the event A. They have exhibited several properties of $R(t)$ and plotted them for various underlying distributions.

We note that for $t \geq 0$,

$$
\begin{aligned}
R(t) &= P(X_t > Y_t) \\
&= P(X - t > Y - t \mid X > t, Y > t) \\
&= \frac{P(X > Y > t)}{P(X > t, Y > t)} \\
&= \frac{\int_t^\infty \bar{F}(x)\, dG(x)}{\bar{F}(t)\bar{G}(t)},
\end{aligned}
\tag{1}
$$

provided that $\bar{F}(t) > 0$ and $\bar{G}(t) > 0$.

The function $R(t)$ can be fruitfully applied in many reliability problems and we give here a couple of examples. Firstly, in burn-in procedures (refer to Block and Savits (1997)), components or systems are subjected to a period of intensive use or accelerated testing for

a period of time, say b, prior to being released into general usage. Therefore, the lifetime of components that survive the burn-in procedure is actually X_b. Thus, $R(b)$ can be used for comparison between successfully burned-in products as a stochastic precedence ordering. This has its own advantage as it is also a total ordering. Secondly, from an inspection point of view, $R(t)$ can be a useful reliability function to consider in general failure models (see Cha and Mi (2007) and the references therein). Consider a system composed of two components or parts, say A and B, which are connected to each other in a series format. We assume that the failure of A causes a minor system failure, Type I failure, whereas that of B causes a catastrophic failure, Type II failure. If X and Y denote the lifetime of A and B respectively, then because the lifetime of the system is $T = min\{X, Y\}$, it follows that the reliability of the whole system is predicated on the occurrence of Type II failure and therefore a reasonable measure of reliability is $R(t) = P(Y = T | T > t)$, i.e. the probability that the component with lifetime Y causes the system failure given that the system has survived up to time t. This function is in contrast with a similar failure probability introduced by Cha and Mi (2007) which is defined as $p(t) = P(Y = T | T = t)$, i.e. the probability that component B causes the system failure given that the system fails at time t. From an inspection and maintenance point of view, high values of $R(t)$ suggests that more attention and inspection should be devoted to maintaining and probable replacement of component B.

Given the potential applications of $R(t)$, it is incumbent on practitioners to estimate $R(t)$ when no prior information are available on the underlying distributions of X and Y. In this paper, our objective is to provide a nonparametric estimator of $R(t)$. Let $X_1, \ldots, X_n$ and

$Y_1, \ldots, Y_m$ be two independent, random samples from the populations of X and Y respectively. Zardasht and Asadi (2010) have considered the following nonparametric estimator of $R(t)$ based on empirical distributions and investigated its property by using simulated data:

$$R_{nm}(t) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} I(X_i > Y_j > t)}{\sum_{i=1}^{n} \sum_{j=1}^{m} I(X_i > t, Y_j > t)}. \tag{2}$$

(In the sequel, the symbol $I(A)$ will refer to the indicator function of the set A.)

As is well known, a notable deficiency of empirical estimators is that they do not take into account the smoothness of the corresponding functions being estimated. Also, the empirical estimators are not able to estimate beyond the largest observation which could lead to large bias close to the boundaries. If it is known that $R(t)$ is a smooth function of t, then we would expect to produce a better estimator by applying a kernel density smoother to the empirical estimator. In the case where $t = 0$ which corresponds to the static case, Baklizi and Eidous (2006) have used kernel method to give a nonparametric estimation of $P(X > Y)$.

Note that $R(t)$ can also be written as

$$R(t) = \frac{\int_t^\infty \bar{F}(x) \, dG(x)}{\int_t^\infty \bar{F}(x) \, dG(x) + \int_t^\infty \bar{G}(x) \, dF(x)} = \frac{R_1(t)}{R_1(t) + R_2(t)}, \tag{3}$$

where $R_1(t) = \int_t^\infty \bar{F}(x) \, dG(x)$ and $R_2(t) = \int_t^\infty \bar{G}(x) \, dF(x)$. In this paper, we replace F and G in (3) with their corresponding kernel estimators (see e.g. Azzalini (1981), Wand and Jones (1995)). Let $k_i(x)$ be a continuous kernel density function and $K_i(x) = \int_{-\infty}^{x} k_i(u) du$, $\bar{K}_i(x) = 1 - K_i(x)$, $i = 1, 2$. The kernel estimators of F and G are

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} K_1\left(\frac{x - X_i}{h_1}\right) \quad \text{and} \quad \hat{G}(y) = \frac{1}{m} \sum_{j=1}^{m} K_2\left(\frac{y - Y_j}{h_2}\right)$$

5

respectively and their corresponding survival functions estimators are

$$\hat{\bar{F}}(x) = \frac{1}{n} \sum_{i=1}^{n} \bar{K}_1(\frac{x - X_i}{h_1}) \quad \text{and} \quad \hat{\bar{G}}(y) = \frac{1}{m} \sum_{j=1}^{m} \bar{K}_2(\frac{y - Y_j}{h_2}).$$

Also, the kernel estimators of their density functions f and g are

$$\hat{f}(x) = \frac{1}{nh_1} \sum_{i=1}^{n} k_1(\frac{x - X_i}{h_1}) \quad \text{and} \quad \hat{g}(y) = \frac{1}{mh_2} \sum_{j=1}^{m} k_2(\frac{y - Y_j}{h_2})$$

respectively. Thus, analogous to (3) we propose the following smooth kernel estimator of $R(t)$:

$$\hat{R}(t) = \frac{\hat{R}_1(t)}{\hat{R}_1(t) + \hat{R}_2(t)} \tag{4}$$

where

$$\hat{R}_1(t) = \int_t^\infty \hat{\bar{F}}(x)\hat{g}(x)dx = \frac{1}{nmh_2} \sum_{i=1}^{n} \sum_{j=1}^{m} \int_t^\infty \bar{K}_1(\frac{x - X_i}{h_1})k_2(\frac{x - Y_j}{h_2})dx$$

and

$$\hat{R}_2(t) = \int_t^\infty \hat{\bar{G}}(x)\hat{f}(x)dx = \frac{1}{nmh_1} \sum_{i=1}^{n} \sum_{j=1}^{m} \int_t^\infty \bar{K}_2(\frac{x - Y_j}{h_2})k_1(\frac{x - X_i}{h_1})dx.$$

The main objective of this paper is to investigate some key properties of the proposed estimator defined in (4). The paper is organized as follows. In Section 2, we propose and provide proofs of some asymptotic properties of $\hat{R}(t)$. Since a key component of kernel smoothing technique is the selection of an optimal bandwidth, Section 3 is devoted to a discussion of this important aspect of kernel smoothing. Section 4 looks at a numerical example and finally, some concluding remarks are given in Section 5.

2. ASYMPTOTIC PROPERTIES OF $\hat{R}(t)$

The main purpose of this section is to investigate the asymptotic unbiasedness and consistency of $\hat{R}(t)$ and to find its asymptotic distribution. We first give a list of regularity conditions on the distribution functions F and G and kernel densities $k_i(x)$ with $K_i(x) = \int_{-\infty}^{x} k_i(u)du$, $\bar{K}_i(x) = 1 - K_i(x)$, $i = 1, 2$, that will be referenced in the theorems to follow. In the sequel, we also let $n' = \min\{n, m\}$ and $N = n + m$.

A1: $k_i(u)$ is bounded;

A2: $\int u k_i(u)du = 0$;

A3: $\int u^2 k_i(u)du = \tau_i^2 < \infty$;

A4: $\int k_i^2(u)du < \infty$;

 B: $h_1, h_2 \to 0$ and $nh_1, mh_2 \to \infty$, and $Nh_1^4, Nh_2^4 \to 0$, as $n' \to \infty$;

C1: F and G are absolutely continuous with densities f and g, respectively;

C2: F has finite mean, is bounded, continuously differentiable up to the second order and has bounded second derivative;

C3: g, g' and g'' exist, are bounded and continuous.

Under the assumptions A1–A4, B and C2–C3, it follows using standard Taylor series expansion of the kernel density function that (see, e.g., Wand and Jones (1995))

$$E[\hat{f}(x)] = f(x) + \frac{1}{2}h_1^2\tau_1^2 f''(x) + o(h_1^2), \tag{5}$$

$$\text{and } Var[\hat{f}(x)] = \frac{1}{nh_1}f(x)\int k_1^2(u)du + o\left(\frac{1}{nh_1}\right). \tag{6}$$

7

A similar approach (refer to Azzalini (1981) or Lejeune and Sarda (1992)) also gives

$$E[\hat{\bar{F}}(x)] = \bar{F}(x) - \frac{1}{2}h_1^2 f'(x)\tau_1^2 + o(h_1^2), \tag{7}$$

$$\text{and } Var[\hat{\bar{F}}(x)] = \frac{\bar{F}(x)F(x)}{n} + \frac{2h_1}{n}f(x)\int w\bar{K}_1(w)K_1(w)\,dw + o\left(\frac{h_1}{n}\right). \tag{8}$$

Analogous equations to (5)-(8) hold true for the expectation and the variance of $\hat{g}(y)$ and $\hat{\bar{G}}(y)$ with obvious changes and will not be displayed here.

The first theorem gives the asymptotic bias and variance of $\hat{R}(t)$.

Theorem 1. Assume conditions A1–A4, B and C1–C3 hold. Then, for any t

$$E[\hat{R}(t)] \rightarrow R(t)$$

$$\text{and } Var[\hat{R}(t)] \rightarrow 0$$

as $n' \rightarrow \infty$.

The next theorem gives the asymptotic distribution of $\hat{R}(t)$.

Theorem 2. Assume conditions A1–A4, B and C1-C3 hold and in addition, $\frac{n}{N} \rightarrow p, \frac{m}{N} \rightarrow 1-p$ where $0 < p < 1$ as $n' \rightarrow \infty$. Then

$$\sqrt{N}(\hat{R}(t) - R(t)) \overset{d}{\rightarrow} \mathbf{N}(0, \sigma^2(t)), \tag{9}$$

($\overset{d}{\rightarrow}$ represents convergence in distribution). $\mathbf{N}(0, \sigma^2(t))$ represents the normal random variable with mean 0 and variance $\sigma^2(t)$ where

$$\sigma^2(t) = \frac{1}{p}[a^2(t)\sigma_{11}(t) + b^2(t)\sigma_{21}(t)] + \frac{1}{1-p}[a^2(t)\sigma_{22}(t) + b^2(t)\sigma_{12}(t)],$$

8

$$\sigma_{11}(t) = \int_t^\infty [\bar{G}(t) - \bar{G}(x)]^2 f(x)dx - R_1^2(t) \quad , \quad \sigma_{12}(t) = \int_t^\infty \bar{F}^2(x)g(x)dx - R_1^2(t),$$

$$\sigma_{22}(t) = \int_t^\infty [\bar{F}(t) - \bar{F}(x)]^2 g(x)dx - R_2^2(t) \quad , \quad \sigma_{21}(t) = \int_t^\infty \bar{G}^2(x)f(x)dx - R_2^2(t),$$

$$a(t) = \frac{R_2(t)}{[R_1(t) + R_2(t)]^2} \quad \text{and} \quad b(t) = \frac{-R_1(t)}{[R_1(t) + R_2(t)]^2}.$$

Due to their length and being of a somewhat technical nature, we have postponed the proofs of both theorems to the Appendix.

3. CHOICE OF BANDWIDTHS

The bandwidth h of a kernel density estimator plays an important role in determining the accuracy of the kernel estimation. It represents a window on each observation and determines the degree of smoothness of the resulting estimator, with increasing smoothness being achieved as the bandwidth increases. Since the performance of an estimator is commonly assessed by the Mean Squared Error (MSE), this criterion is often used as a means of determining an optimal bandwidth. In our case, the MSE of $\hat{R}(t)$ does not present itself in a closed form since $\hat{R}(t)$ is a ratio estimator. Therefore, the minimization of the MSE using first–order condition through differentiation with respect to the bandwidths h_1 and h_2 is not a straightforward task. However, there are some classical bandwidth selection methods (cf. Wand and Jones (1995)) that can be chosen to select appropriate bandwidths for this problem. In the kernel density estimation context, a well-known and perhaps the most popular method for estimating bandwidth is Least Squared Cross–Validation (LSCV) introduced by Rudemo (1982) and Bowman (1984). This method is based on the minimization of the Mean Integrated Squared Error (MISE) of the estimators, which, when applied to our problem,

select values h_1 and h_2 which minimize the following MISE:

$$MISE\{\hat{R}(t)\} \quad = \quad E[\int [\hat{R}(t) - R(t)]^2 dt]. \tag{10}$$

This is equivalent to minimizing

$$MISE\{\hat{R}(t)\} - \int R^2(t)dt \quad = \quad E[\int \hat{R}^2(t)dt - 2\int \hat{R}(t)R(t)dt], \tag{11}$$

since $\int R^2(t)dt$ does not depend on h_1 or h_2. The last term in (11) depends on the unknown $R(t)$. We therefore estimate this term by the usual method of "leave–one–out" averaging implicit in cross–validation which is equivalent to choosing h_1 and h_2 to minimize

$$C(h_1, h_2) \quad = \quad \int \hat{R}^2(t)dt - \frac{2}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\int \hat{R}^{-(i,j)}(t)I(X_i > Y_j > t)dt$$

where $\hat{R}^{-(i,j)}(t)$ is $\hat{R}(t)$ defined by (4) constructed using all the data points except x_i and y_j.

However, extensive preliminary simulation trials revealed that the selected bandwidths obtained through this minimization procedure produce estimators with large bias. This is conjectured to have arisen due to the functional form of $\hat{R}(t)$, which is a combination of $\hat{R}_1(t)$ and $\hat{R}_2(t)$ where h_1 and h_2 play different role (i.e. in $\hat{R}_1(t), h_1$ is used for estimating of the survival function of X whilst it is used for estimating of the density function of Y in $\hat{R}_2(t)$). We have therefore adopted the approach of choosing h_1 and h_2 for $\hat{R}_1(t)$ and $\hat{R}_2(t)$ *separately* by minimizing their corresponding MISE.

Thus, the selection of h_1 and h_2 in $\hat{R}_1(t)$ for estimating $R_1(t)$ is reduced to choosing h_1 and h_2 to minimize

$$C_1(h_1, h_2) \quad = \quad \int \hat{R}_1^2(t)dt - \frac{2}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\int \hat{R}_1^{-(i,j)}(t)I(X_i > Y_j > t)dt \tag{12}$$

10

where

$$\hat{R}_1^{-(i,j)}(t) = \int_t^\infty \bar{\hat{F}}^{(-i)}(x)\hat{g}^{(-j)}(x)dx,$$

$\bar{F}^{(-i)}(x)$ and $\hat{g}^{(-j)}(x)$ are the kernel estimators of $\bar{F}(x)$ and $g(x)$ constructed using all data

points except x_i and y_j respectively.

Similarly, selection of h_1 and h_2 in $\hat{R}_2(t)$ reduces to minimizing

$$C_2(h_1, h_2) = \int \hat{R}_2^2(t)dt - \frac{2}{nm}\sum_{i=1}^n\sum_{j=1}^m\int \hat{R}_2^{-(i,j)}(t)I(Y_j > X_i > t)dt \tag{13}$$

where

$$\hat{R}_2^{-(i,j)}(t) = \int_t^\infty \bar{\hat{G}}^{(-j)}(x)\hat{f}^{(-i)}(x)dx,$$

$\bar{G}^{(-j)}(x)$ and $\hat{f}^{(-i)}(x)$ are the kernel estimators of $\bar{G}(x)$ and $f(x)$ constructed from all the

data points except y_j and x_i respectively. We note here that

$$E[C_i(h_1, h_2)] = MISE\{\hat{R}_i(t)\} - \int R_i^2(t)dt$$

$i = 1, 2$.

4. NUMERICAL RESULTS

Simulation exercises were undertaken to study the performances of our estimator and

assess the performances of this estimator, comparing it with the empirical estimator given

in (2). In the simulation, we considered the following two cases for the distribution of X and

Y respectively:

(i) X and Y are exponentially distributed with means 4 and 5 respectively, i.e.

$$\bar{F}(t) = e^{-t/4}, \quad \bar{G}(t) = e^{-t/5}.$$

(ii) X is exponentially distributed with mean 1 and Y has a Weibull distribution with shape

parameter 2 and scale parameters 1, i.e.

$$\bar{F}(t) = e^{-t}, \quad \bar{G}(t) = e^{-t^2}.$$

One can readily show using (1), that in case (i), $R(t) = 4/9$, and in case (ii)

$$R(t) = 1 - \frac{1}{\sqrt{2}r_Z(\sqrt{2}t + \frac{\sqrt{2}}{2})}$$

where $r_Z(t)$ is the hazard rate of the standard normal random variable, i.e. $r_Z(t) = \frac{\Phi'(t)}{1-\Phi(t)}$

and $\Phi(x)$ is the standard normal cumulative distribution function. The Epanechnikov kernel,

i.e $k_1(x) = k_2(x) = 0.75(1 - x^2)$, $|x| < 1$, is used in both examples.

In each case, we ran 1000 simulation trials of different sizes $n = m$ where:

$$n = m = 5, \quad n = m = 10, \quad n = m = 30 \text{ and } n = m = 50.$$

For each sample and at different values of t, we obtained the optimal bandwidths h_1 and h_2

through minimization of $C_1(h_1, h_2)$ and $C_2(h_1, h_2)$ and use each set to estimate $\hat{R}_1(t)$ and

$\hat{R}_2(t)$ separately as discussed in the previous section. Unlike the empirical estimator, the

estimators that appear in the numerator and denominator of out smoothed estimator are

biased. Therefore in our simulations, we have used bootstrapping to reduce their bias (refer

to Hall (1990)). Altogether, 300 replications of bootstrapping were applied. In addition, to

avoid the drawback of the usual kernel smoothing for positive valued random variables (see,

for example, Bagai and Rao (1995)), we applied log transformation to the data noting that

the final results are invariant under this monotonic transformation. For each sample, the

values of our smooth estimator (4) and empirical estimator (2) were obtained. This is then repeated for 1000 samples and their corresponding bias (i.e. the average of the estimates minus $R(t)$), mean–squared errors (i.e. the average of the sum of the squared difference between the estimate and $R(t)$) and relative efficiencies (defined by (14) below) calculated. Through simulation, we also obtained confidence intervals with nominal confidence level of 0.95 and estimate the total coverage rate, i.e. the proportion of intervals which cover the true value $R(t)$.

Table 1–Table 2 summarize the results of the 1000 simulation trials for both examples. The notations $T_s(T_e), B_s(B_e)$ and $MSE_s(MSE_e)$ refer to the total coverage rate, bias and mean–squared error of the smooth estimator (the empirical estimator) respectively. We also displayed the relative efficiency between the two estimators defined by

$$eff \quad = \quad \frac{MSE_s}{MSE_e}. \tag{14}$$

(**Table 1–Table 2 TO BE PLACED HERE.**)

As is evident from the above tables, the performance of our smooth estimator is generally better than that of the empirical estimator. Although both estimator underestimate $R(t)$ for large values of t, the smooth estimator produces better results when considering bias, mean–squared error and efficiency. The total coverage rates are also slightly better although they both deteriorate for large values of t. From a practical point of view, the main objective of the simulation is to show that the performance of the smoothed estimator is comparable

13

or exceeds that of the empirical estimator. From the results obtained, we may conclude that this is indeed the case here. It is anticipated that with larger samples, the improvement of the smooth estimator over empirical estimator would be further enhanced.

5. CONCLUSION

In this paper, we have considered some properties of a smoothed nonparametric estimator of a reliability function based on two independent random samples. A nonparametric approach provides a viable alternative to the conventional method based on knowing the exact distributions of the survival functions of the underlying variables, which is often not possible nor realistic in practical applications. This estimator was shown to be asymptotically unbiased and consistent, and also to converge in distribution to a normal random variable. An extensive simulation exercise was undertaken to compare between the performance of this estimator and one which uses directly the empirical distributions of the true underlying distributions and the results argue favorably for the smoothed estimator.

There are several quantitative measures of reliability, e.g. Mean Residual Life (MRL) (c.f. Abdous and Berred (2005)), which are used by practitioners in disciplines such as engineering, medical statistics and finance, to assess the reliability of products and systems. The method discussed in this paper could be used to estimate these measures when the underlying distributions are unknown or imprecise.

Finally, some issues have surfaced in the course of conducting this research and they have to do with the selection of optimal bandwidths when the function to be estimated is in the form of a ratio, such as $R(t)$. In this paper, the optimal bandwidths were obtained by

14

considering the components $R_1(t)$ and $R_2(t)$ of the ratio separately as discussed in Section 3. Since both estimators are biased, it is expected that the nonparametric estimator of $R(t)$, will require a much larger sample size than the ones used here in order to achieve a satisfactory level of accuracy. Thus, it will be necessary to seek an alternative method of selecting bandwidths in order to reduce bias when estimating a reliability function of the type considered in this paper.

ACKNOWLEDGEMENTS

APPENDIX

Proof of Theorem 1.

Under the given assumptions and using the equations (5) and (7) we obtain

$$
\begin{aligned}
|E[\hat{R}_1(t)] - R_1(t)| \ &\leq \ \int_t^\infty |E[\hat{\bar{F}}(y)] - \bar{F}(y)|E[\hat{g}(y)]dy + \int_t^\infty \bar{F}(y)|E[\hat{g}(y)] - g(y)|dy \\
&\leq \ O(h_1^2)[\int_t^\infty g(y)dy + \frac{1}{2}h_2^2\tau_2^2 \int_t^\infty g''(y)dy] + O(h_2^2) = O(h_1^2 + h_2^2),
\end{aligned}
$$

where the last equality follows by omitting the smaller order terms. Thus, it follows that

$$
E[\hat{R}_1(t)] \ = \ R_1(t) + O(h_1^2 + h_2^2). \tag{15}
$$

Similarly, it can be shown that

$$
E[\hat{R}_2(t)] \ = \ R_2(t) + O(h_1^2 + h_2^2). \tag{16}
$$

15

Under the given conditions, one can easily show that $E[\hat{\bar{F}}(x)\hat{\bar{F}}(y)] \to \bar{F}(x)\bar{F}(y)$ and $E[\hat{g}(x)\hat{g}(y)] \to g(x)g(y)$ as $n' \to \infty$. Since

$$E[\hat{R}_1^2(t)] = \int_t^\infty \int_t^\infty E[\hat{\bar{F}}(x)\hat{\bar{F}}(y)]E[\hat{g}(x)\hat{g}(y)]dx\,dy,$$

applying dominated convergence theorem and (15) imply that $Var[\hat{R}_1(t)] \to 0$ as $n' \to \infty$. Likewise, $Var[\hat{R}_2(t)] \to 0$ as $n' \to \infty$. Also, using Holder's inequality in conjunction with (15) and (16), one can show that $E[\hat{R}_1(t) - R_1(t)][\hat{R}_2(t) - R_2(t)]$ goes to zero as $n' \to \infty$.

Expanding $\frac{\hat{R}_1(t)}{\hat{R}_1(t)+\hat{R}_2(t)}$ in a Taylor series about $(R_1(t), R_2(t))$, it follows that

$$\hat{R}(t) - R(t) = \frac{R_2(t)}{[R_1(t) + R_2(t)]^2}[\hat{R}_1(t) - R_1(t)] - \frac{R_1(t)}{[R_1(t) + R_2(t)]^2}[\hat{R}_2(t) - R_2(t)]$$
$$+ o\left(\sqrt{[\hat{R}_1(t) - R_1(t)]^2 + [\hat{R}_2(t) - R_2(t)]^2}\right). \qquad (17)$$

Since $Var[\hat{R}_1(t)] \to 0$ and $Var[\hat{R}_2(t)] \to 0$ as $n' \to \infty$, then $[\hat{R}_1(t) - R_1(t)]^2 + [\hat{R}_2(t) - R_2(t)]^2 = o_p(1)$ by Chebyshev's inequality. Therefore, the last term in (17) is $o_p(1)$ and this, along with (15) and (16), proves asymptotic unbiasedness. Finally, squaring both sides of (17) and using the second moment properties of $\hat{R}_1(t)$ and $\hat{R}_2(t)$ stated earlier, consistency will follow.

Proof of Theorem 2.

Let

$$\psi_{1t}(X_i, Y_j) = \frac{1}{h_2}\int_t^\infty \bar{K}_1(\frac{y - X_i}{h_1})k_2(\frac{y - Y_j}{h_2})dy,$$

$$\psi_{2t}(X_i, Y_j) = \frac{1}{h_1}\int_t^\infty \bar{K}_2(\frac{y - Y_j}{h_2})k_1(\frac{y - X_i}{h_1})dy.$$

16

Then, we have

$$\hat{R}_1(t) = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \psi_{1t}(X_i, Y_j),$$

$$\hat{R}_2(t) = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \psi_{2t}(X_i, Y_j).$$

$\hat{R}_1(t)$ and $\hat{R}_2(t)$ are in the form of a two–sample U–Statistics and we can use the projection method of Hajek (1968) to approximate $\hat{R}_1(t)$ and $\hat{R}_2(t)$ by their projections $\hat{R}_1^*(t)$ and $\hat{R}_2^*(t)$ defined by

$$\hat{R}_1^*(t) - E[\hat{R}_1(t)] = \frac{1}{n} \sum_{i=1}^{n} \varphi_{11}(X_i) + \frac{1}{m} \sum_{j=1}^{m} \varphi_{12}(Y_j),$$

$$\hat{R}_2^*(t) - E[\hat{R}_2(t)] = \frac{1}{n} \sum_{i=1}^{n} \varphi_{21}(X_i) + \frac{1}{m} \sum_{j=1}^{m} \varphi_{22}(Y_j),$$

and where

$$\varphi_{11}(x) = E[\psi_{1t}(X_i, Y_j)|X_i = x] - E[\hat{R}_1(t)], \qquad \varphi_{12}(x) = E[\psi_{1t}(X_i, Y_j)|Y_j = x] - E[\hat{R}_1(t)],$$

$$\varphi_{21}(x) = E[\psi_{2t}(X_i, Y_j)|X_i = x] - E[\hat{R}_2(t)], \qquad \varphi_{22}(x) = E[\psi_{2t}(X_i, Y_j)|Y_j = x] - E[\hat{R}_2(t)].$$

Assuming A1–A4 hold, then Theorem 2.1 in Grams and Serfling (1973) shows that both $\hat{R}_1(t) - \hat{R}_1^*(t)$ and $\hat{R}_2(t) - \hat{R}_2^*(t)$ are asymptotically negligible and therefore $\hat{R}_1(t)$ and $\hat{R}_2(t)$ have the same asymptotic distributions as $\hat{R}_1^*(t)$ and $\hat{R}_2^*(t)$ respectively. Let

$$Z_{nm} = \sqrt{N}(\hat{R}_1^*(t) - E[\hat{R}_1(t)], \hat{R}_2^*(t) - E[\hat{R}_2(t)]) = \sqrt{N}[\frac{1}{n} \sum_{i=1}^{n} Z_{1i} + \frac{1}{m} \sum_{j=1}^{m} Z_{2j}]$$

where

$$Z_{1i} = (\varphi_{11}(X_i), \varphi_{21}(X_i)) \quad \text{and} \quad Z_{2j} = (\varphi_{12}(Y_j), \varphi_{22}(Y_j)).$$

17

Since the summands of $\hat{R}_1^*(t)$ and $\hat{R}_2^*(t)$ are independent, they may be dealt with by standard central limit theorem. We will show that Z_{nm} converges in distribution to a bivariate normal distribution. To accomplish this, it is sufficient, using the Cramér-Wold device, to prove that $\mathbf{c}'Z_{nm}$ converges to a normal distribution for any $\mathbf{c} = (c1, c2)$ in R^2. For any $i = 1, 2, \ldots, n$, we have

$$\varphi_{11}(X_i) + E[\hat{R}_1(t)] = \int_t^\infty \bar{K}_1(\frac{y - X_i}{h_1})E[\frac{1}{h_2}k_2(\frac{y - Y_1}{h_2})]dy, \tag{18}$$

$$\text{and} \quad Var[\varphi_{11}(X_i)] = \int_{-\infty}^\infty \left(\int_t^\infty \bar{K}_1(\frac{y - x}{h_1})E[\frac{1}{h_2}k_2(\frac{y - Y_1}{h_2})]dy\right)^2 f(x)dx$$

$$- E^2[\hat{R}_1(t)]. \tag{19}$$

Let $\sigma_{11}(t)$ be the limit of $Var[\varphi_{11}(X_i)]$ on letting $n' \to \infty$. Firstly, under the given conditions and Theorem 2.1, as $m \to \infty$, applying the bounded convergence gives

$$Var[\varphi_{11}(X_i)] \to \int_{-\infty}^\infty \left(\int_t^\infty \bar{K}_1(\frac{y - x}{h_1})g(y)dy\right)^2 f(x)dx - \left(\int_t^\infty E[\hat{\bar{F}}(x)]g(x)dx\right)^2$$

$$= \int_{-\infty}^\infty [\bar{G}(t)\bar{K}_1(\frac{t - x}{h_1}) - \int_{\frac{t-x}{h_1}}^\infty k_1(v)\bar{G}(x + h_1 v)dv]^2 f(x)dx \tag{20}$$

$$- \left(\int_t^\infty E[\hat{\bar{F}}(x)]g(x)dx\right)^2$$

after applying an integration by parts and a change of variables. We next note that when $x < t$, the squared terms in the integrand of (20) converges to 0 as $h_1 \to 0$ and when $x > t$ it converges to

$$(\bar{G}(t)\bar{K}_1(-\infty) - \bar{G}(x)\int_{-\infty}^\infty k_1(v)dv)^2 = (\bar{G}(t) - \bar{G}(x))^2.$$

Therefore, it follows that

$$\sigma_{11}(t) = \int_t^\infty [\bar{G}(t) - \bar{G}(x)]^2 f(x)dx - R_1^2(t). \tag{21}$$

18

Similarly,

$$\varphi_{12}(Y_j) + E[\hat{R}_1(t)] \;=\; \frac{1}{h_2}\int_t^\infty E[\bar{K}_1(\frac{y-X_1}{h_1})]k_2(\frac{y-Y_j}{h_2})dy,$$

$$\text{and}\;\; Var[\varphi_{12}(Y_j)] \;=\; \int\left(\frac{1}{h_2}\int_t^\infty E[\bar{K}_1(\frac{y-X_1}{h_1})]k_2(\frac{y-z}{h_2})dy\right)^2 g(z)dz - E^2[\hat{R}_1(t)].$$

Let $\sigma_{12}(t)$ be the limit of $Var[\varphi_{12}(Y_j)]$ as $n' \to \infty$. As $n \to \infty$,

$$\begin{aligned}
Var[\varphi_{12}(X_i)] \;\to\; & \int_{-\infty}^\infty \left(\frac{1}{h_2}\int_t^\infty \bar{F}(y)k_2(\frac{y-z}{h_2})dy\right)^2 g(z)dz - \left(\int_t^\infty \bar{F}(x)E[\hat{g}(x)]dx\right)^2 \\
= & \int_{-\infty}^\infty \left(\int_{\frac{t-z}{h_2}}^\infty \bar{F}(z+h_2 v)k_2(v)dv\right)^2 g(z)dz - \left(\int_t^\infty \bar{F}(x)E[\hat{g}(x)]dx\right)^2.
\end{aligned}$$

$$(22)$$

Then, letting $h_2 \to 0$ in (22), we obtain

$$\sigma_{12}(t) \;=\; \int_t^\infty \bar{F}^2(z)g(z)dz - R_1^2(t) \qquad\qquad (23)$$

using the same reasonings as in the previous case.

Likewise, we can show that as $n' \to \infty$

$$Var[\varphi_{21}(X_i)] \;\to\; \sigma_{21}(t) = \int_t^\infty \bar{G}^2(z)f(z)dz - R_2^2(t) \qquad\qquad (24)$$

and

$$Var[\varphi_{22}(Y_j)] \;\to\; \sigma_{22}(t) = \int_t^\infty [\bar{F}(t)-\bar{F}(x)]^2 g(x)dx - R_2^2(t). \qquad\qquad (25)$$

Let

$$\tilde{\varphi}_{11}(X_i) = \int_t^\infty \bar{K}_1(\frac{y-X_i}{h_1})E[\frac{1}{h_2}k_2(\frac{y-Y_1}{h_2})]dy,$$

and

$$\tilde{\varphi}_{21}(X_i) = \frac{1}{h_1}\int_t^\infty E[\bar{K}_2(\frac{y-Y_1}{h_2})]k_1(\frac{y-X_i}{h_1})dy.$$

19

Notice that $E[\tilde{\varphi}_{11}(X_i)] = E[\hat{R}_1(t)]$ and $E[\tilde{\varphi}_{21}(X_i)] = E[\hat{R}_2(t)]$. Then,

$$
\begin{aligned}
Cov[\varphi_{11}(X_i), \varphi_{21}(X_i)] &= E\left[\tilde{\varphi}_{11}(X_i)\tilde{\varphi}_{21}(X_i)\right] - E[\hat{R}_1(t)]E[\hat{R}_2(t)] \\
&= E\left[\tilde{\varphi}_{11}(X_i)\{\tilde{\varphi}_{21}(X_i) - E[\hat{R}_2(t)]\}\right] + E[\hat{R}_2(t)]\{E[\tilde{\varphi}_{11}(X_i)] - E[\hat{R}_1(t)]\} \\
&= E\left[\tilde{\varphi}_{11}(X_i)\{\tilde{\varphi}_{21}(X_i) - E[\hat{R}_2(t)]\}\right].
\end{aligned}
$$

Under assumption C3,

$$
\begin{aligned}
|\tilde{\varphi}_{11}(X_i)| &\leq \int_t^\infty \bar{K}_1(\frac{y - X_i}{h_1})[g(y) + O(h_2)]dy \\
&\leq \sup_y g(y)h_1 \int_{\frac{t-X_i}{h_1}}^\infty \bar{K}_1(w)dw + O(h_2)h_1 \int_{\frac{t-X_i}{h_1}}^\infty \bar{K}_1(w)dw \\
&= O(h_1) + O(h_1 h_2)
\end{aligned}
$$

almost surely, hence

$$
\begin{aligned}
|Cov[\varphi_{11}(X_i), \varphi_{21}(X_i)]| &\leq \int \tilde{\varphi}_{11}(x)|\tilde{\varphi}_{21}(x) - E[\hat{R}_2(t)]|f(x)dx \\
&\leq 2E[\hat{R}_2(t)](O(h_1) + O(h_1 h_2)).
\end{aligned}
$$

Therefore, it follows that

$$
\lim_{n' \to \infty} Cov[\varphi_{11}(X_i), \varphi_{21}(X_i)] = 0.
$$

Similarly, one can show that

$$
\lim_{n' \to \infty} Cov[\varphi_{12}(Y_j), \varphi_{22}(Y_j)] = 0.
$$

Let $\sigma_{nm}^2(t) = \lim_{n' \to \infty} Var(\mathbf{c}'Z_{nm})$; using the assumptions $\frac{n}{N} \to p$ and $\frac{m}{N} \to 1 - p$ where $0 < p < 1$, it follows that

$$
\sigma_{nm}^2(t) = \frac{1}{p}(c_1^2\sigma_{11}(t) + c_2^2\sigma_{21}(t)) + \frac{1}{1-p}(c_1^2\sigma_{22}(t) + c_2^2\sigma_{12}(t)) > 0.
$$

Under assumption A1, $\varphi_{ij}(.), i, j = 1, 2$ are bounded and the conditions stated in Corollary 3.1, Grams and Serfling (1973) hold, giving rise to the following Berry-Esséen type bound:

$$\sup_x \left| Pr\left[(\mathbf{c}'Z_{nm}/\sqrt{Var(\mathbf{c}'Z_{nm})} \leq x \right] - \Phi(x) \right| = O(N^{-\frac{1}{2}+\varepsilon})$$

for every $\varepsilon > 0$ and where $\Phi(x)$ refers to the standard normal cumulative distribution function. Therefore, as $N \to \infty$, we have

$$\mathbf{c}'Z_{nm} \xrightarrow{d} \mathbf{N}(0, \sigma_{nm}^2(t))$$

and hence

$$Z_{nm} \xrightarrow{d} \mathbf{N}(0, \frac{1}{p}\Sigma_1 + \frac{1}{1-p}\Sigma_2)$$

where

$$\Sigma_1 = \begin{pmatrix} \sigma_{11}(t) & 0 \\ 0 & \sigma_{21}(t) \end{pmatrix} \quad \text{and} \quad \Sigma_2 = \begin{pmatrix} \sigma_{12}(t) & 0 \\ 0 & \sigma_{22}(t) \end{pmatrix}.$$

Finally, let $\tilde{Z}_{nm} = \sqrt{N}(\hat{R}_1(t) - R_1(t), \hat{R}_2(t) - R_2(t))$, then we obtain

$$
\begin{aligned}
\tilde{Z}_{nm} &= \sqrt{N}(\hat{R}_1(t) - E[\hat{R}_1(t)], \hat{R}_2(t) - E[\hat{R}_2(t)]) + \sqrt{N}(E[\hat{R}_1(t)] - R_1(t), E[\hat{R}_2(t)] - R_2(t)) \\
&= \sqrt{N}(\hat{R}_1(t) - E[\hat{R}_1(t)], \hat{R}_2(t) - E[\hat{R}_2(t)]) + \sqrt{N}(O(h_1^2 + h_2^2), O(h_1^2 + h_2^2))
\end{aligned}
$$

where the last equation follows from equations (15) and (16). Therefore, under assumption B, Slutsky's theorem implies

$$\tilde{Z}_{nm} \xrightarrow{d} \mathbf{N}(0, \frac{1}{p}\Sigma_1 + \frac{1}{1-p}\Sigma_2)$$

and a simple application of the *delta method* (Lehmann 1999), p. 85) completes the proof of the theorem.

21

BIBLIOGRAPHY

Abdous, B. and Berred, A. (2005). Mean residual life estimation. *Journal of Statistical Planning and Inference*, **132**, pp. 3–19.

Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika*, **68**, 326 –328.

Bagai, I., and Rao, B.L.S., P, (1995). Kernel Type Density Estimators for Positive Valued Random Variables *Sankhya, Series A,* **57**, 56-67.

Baklizi, A., and Eidous, O. (2006). Nonparametric Estimation of $P(X < Y)$ Using Kernel Methods. *Metron,* vol. LXIV,**1**, 47-60.

Block, H. and Savits, T. (1997). Burn–in. *Statist. Sci.*, **12**, 1–13.

Boland, P.J., Singh, H., Cukic, B. (2004). The stochastic precedence ordering with applications in sampling and testing. *J. Appl. Prob.*, **41**, 73-82.

Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, **71**, 353–360.

Brown, G. G., and Rutemiller, H. C. (1973). Evaluation of $P(X > Y)$ when both X and Y are from Three-Parameter Weibull distributions. *IEEE Trans. on Reliab.*, **22**, 78–82.

Cha, J. H. and Mi, J. (2007). Some Probability Functions in Reliability and Their Applications. *Naval Research Logistics*, **54**, 128–135.

Grams, W.F. and Serfling, R.J. (1973). Convergence rates for U–Statistics and related statistics. *The Annals of Statistics*, **1**, 153–160.

Hajek, J. (1968). Asymptotic normality of simple linear statistics under alternatives. *Ann. Math. Statist.*, **39**, 325–346.

Hall, P. (1990). Using the Bootstrap to Estimate Mean Squared Error and Select Smoothing Parameter in Nonparametric Problems. *Journal of Multivariate Analysis*, **32**, 177-203.

Jones, M.C. and Foster, P.J. (1996). A simple nonnegative boundary correction method for kernel density estimation. *Statistica Sinica*, **6**, 1005–1013.

Lejeune, M. and Sarda, P. (1992). Smooth estimators of distribution and density functions. *Comput. Statist. Data Anal.*, **14**, 451–471.

Lehmann, E. L. (1999). *Elements of Large-Sample Theory.* New York: Springer–Verlag.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scan. J. Stat.*, **9**, 65–78.

Samanieo, F. J., and Hollander, M. (2008). On Comparing the Reliability of Arbitrary Systems via Stochastic Precedence. *Advances in Mathematical Modeling for Reliability*, Bedford, T., Quigley, J., Walls, L., Alkali, B., Daneshkhah, A. and G. Hardman (Editors), 129–137.

Singh, H., Misra, N. (1994). On redundancy allocations in systems. *J. Appl. Prob.*, **31**, 1004–1014.

Shaked, M., and Shanthikumar, J. G. (1994). *Stochastic Orders and Their Applications.*

Academic Press.

Shaked, M., and Shanthikumar, J. G. (2007). *Stochastic Orders.* Springer.

Wand, M. and Jones, M.C. (1995). *Kernel Smoothing.* London: Chapman and Hall.

Zardasht, V. and Asadi, M. (2010). Evaluation of $P(X_t > Y_t)$ when both X_t and Y_t are residual lifetimes of two systems. *Statistica Neerlandica,* doi:10.1111/j.1467-9574.2010.00464.x