International Journal of Data Science and Analysis
2018; 4(6): 106-111
http://www.sciencepublishinggroup.com/j/ijdsa
doi: 10.11648/j.ijdsa.20180406.11
ISSN: 2575-1883 (Print); ISSN: 2575-1891 (Online)

Estimation of the Survivorship Function Using the Cox-Proportional Hazard Model with Relaxed Tsiatis Assumptions

Valerie Atieno Odhiambo, George Otieno Orwa, Romanus Odhiambo

Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

Email address:
Valerieatieno254@gmail.com (V. A. Odhiambo)

To cite this article:
Valerie Atieno Odhiambo, George Otieno Orwa. Estimation of the Survivorship Function Using the Cox-Proportional Hazard Model with Relaxed Tsiatis Assumptions. *International Journal of Data Science and Analysis.* Vol. 4, No. 6, 2018, pp. 106-111.
doi: 10.11648/j.ijdsa.20180406.11

Received: October 10, 2018; **Accepted**: October 22, 2018; **Published**: January 25, 2019

Abstract: Survival analysis is the primary statistical method of analysing time to event data. The most popular method for estimating the survivor function is the Cox-Proportional Hazard model. It assumes that the effect on the hazard function of a particular factor of interest remains unchanged throughout the observation. This is known as Proportional Hazards. Tsiatis assumed that the underlying hazard function is constant over distinct intervals. In the current study, no shape assumption is imposed other than that the hazard function is a smooth function with an arbitrary choice of a smoother. Such an approach involves the implementation of kernel-smoothing of the initial hazard estimate which have proved in studies to provide a trade-off between bias and variance. The cross-validation and plug-in bandwidth selectors are considered to determine the optimal bandwidth, h to be used as a smoothing parameter. Consequently, the survivorship function is estimated using the Cox-Proportional Hazards model. Proper application of the smoothing procedure is seen to improve the statistical performance of the resulting hazard rate estimator. No constraints are implored on the form of the underlying hazard proving to be less bias than Tsiatis' method. This implies that the kernel smoothed survivorship function is more appropriate than the common standard techniques in survival analysis as it provides piecewise smooth estimates. Coverage probabilities of the estimate are then obtained which are found to be more accurate and closer to the nominal level compared to those estimated by Tsiatis.

Keywords: Survivorship Function, Cox-Proportional Hazard, Kernel Smoothing, Bandwidth Selection

1. Introduction

Survival times are data that measure follow-up time from a well-defined starting point to the occurrence of a given event. Standard statistical techniques cannot be easily applied because the underlying distribution is not normal and the data are often censored. In survival analysis, data is called censored when there is a follow up time and the event has not yet occurred or is not known to have occurred.

Survivorship function is defined as the probability of surviving at least to time t. There are various methods that have been proposed to estimate the survivorship function of any survival analysis data. One among them is the Kaplan Meier method proposed by Kaplan and Meier [1]. The most popular method for estimating the survivorship function is by using the Cox-Proportional Hazard model which has a variety of advantages. One of them is that covariates must not be changing over time. It is a semi-parametric model that makes fewer assumptions than a typical parametric method.

A hazard function is defined as the conditional probability of dying at time t having survived to that time. Though parametric models provide convenient ways to analyse lifetime data, the necessary assumptions when violated can lead to erroneous analysis. In estimating the hazard function under Cox-PH model, no assumption is made about the probability distribution of the hazard except for smoothing. Kernel smoothing specifically provides a powerful methodology for gaining insights into data. Effective use of these methods requires the choice of a smoothing parameter known as the bandwidth. A non-parametric smoothed estimate of the hazard function can be obtained by implementing the Nadaraya [2] estimator which proved to be asymptotically unbiased.

Breslow [3] came up with a proposal to estimate the underlying hazard function and assumed that the hazard rate is a constant between death times. This assumption was adopted by Tsiatis [4] and used to estimate the survivor function under Cox-PH model. The current study seeks to estimate the underlying hazard function by smoothing through it first. The exponential relationship between the smoothed hazard estimate and the survivorship function is implemented to obtain the parameter estimates. An asymptotic variance of the estimated survivorship function is further be derived to determine if the confidence intervals obtained will give accurate coverage probabilities.

2. Some Related Works

Regression models for survival analysis with censored data have been used quite extensively in the past few years. One very popular model is known as the Cox-Proportional Hazard model which is a method for investigating the effect of several variables upon the time a specified event takes to happen. Kernel based methods for the smooth, non-parametric estimation of the hazard function have received considerable attention in statistical literature. The results obtained can be used to estimate the survivorship function.

Rosenblatt and Murray [5] carried out a detailed analysis of the bias and covariance properties of a number of estimates of the log survival function and hazard functions. Estimates were also considered when one has dependence. The study was quite useful in studies of mortality. However, the actual estimator of the survivor function was not done and the usage of the said estimator not clearly discussed.

For each individual in the analysis of censored death times, Cox [6] assumed that they were affected by one or more explanatory variables. He wanted to explore the consequences of allowing the underlying hazard to be arbitrary. His interests were mainly in the regression parameters. He determined that if the function has sensible properties, then the parameters would only be slightly affected. An adoption of the assumption that some smoothness in the distribution function would therefore be reasonable enough which the current study considers.

A method which gives a piece-wise smooth estimate for the hazard function given by Cox's model was described by Anderson and Senthilselvan [7]. The penalized likelihood estimation introduced by Goodd and Gaskins [8] was used. This gave a quadratic spline with discontinuities in the slope at death times. The hazard function and corresponding survival curves were estimated and proved to be less bias compared to rougher estimates by Kalbfleisch and Prentice [9]. However, the hazard function could take negative values of certain values in the smoothing parameter which was a drawback. An alternative smoothing method is therefore determined in the current study.

The bandwidth selection procedure to be considered is of great importance. Three bandwidth selection procedures were compared in detail by Gijbels [10] where the bootstrap outperformed the plug-in and cross-validation method. This leaves a leeway for the comparison of cross-validation and plug-in which is investigated in the study.

On coverage probabilities, Lin and Fleming [11] constructed confidence bands for survival curves under the Cox-PH model. The distribution of the normalized cumulative hazard estimator was approximated by a zero-mean Gaussian process whose distribution was easily generated through simulation. Since the choice of the weight function affected the widths of the band at different time points, two weight functions were considered namely the equal-precision band by Nair [12] and Hall-Wellner band by Bie, Borgan and Liestol [13]. The proposed bands maintained their coverage probabilities near the nominal level even for small sample sizes with heavy censoring.

3. Methodology

3.1. Cox-Proportional Hazard Model

Cox-Proportional Hazards regression model is a tool that is used for studying the dependency of survival time on predictor variables. It's given by:

$$\lambda(t) = \lambda_0(t)e^{(\alpha+\beta_1 x_{i1}+\beta_2 x_{i2}+\cdots+\beta_k x_{ik})}$$

where $\lambda(t)$ is the expected hazard at time t, $\lambda_0(t)$ is the baseline hazard and represents the hazard when all of the predictors; $x_1, x_2, \ldots, x_p$ are equal to zero. A prominent feature of the Cox's model is that the baseline hazard function is measured non-parametrically, and so unlike most other statistical models, the survival times are not assumed to follow a particular statistical distribution and the t in $\lambda(t)$ indicates that the hazard function may vary over time.

3.2. Estimation of Cox- Proportional Hazard Model Parameters

Parameter estimates in the Cox-PH model are obtained by maximizing the partial likelihood as opposed to the full likelihood itself. Assuming that n is the number of individuals observed and k is the number of failure times which are assumed to be distinct, the partial likelihood is given by:

$$L(\beta) = \prod_{i=1}^{k} \frac{e^{x_i\beta}}{\sum_{j=1}^{n} e^{x_j\beta}}$$

$$l(\beta) = logL(\beta) = \sum_{i=1}^{k}\left\{x_i\beta - log\left[\sum_{j=1}^{n} e^{x_j\beta}\right]\right\}$$

Maximization is accomplished by setting the first partial derivatives equal to zero. It can then be solved by the Newton-Raphson iteration method and the matrix of second partial derivatives.

3.3. Estimating the Survivorship Function

3.3.1. A review of the Cox-PH Model Approach

The survivorship function of the Cox-PH model is given by:

$$S(t) = S_0(t)^{e^{\beta' x}}$$

where

$$S_0(t) = e^{-\Lambda_0(t)}$$

is the baseline survival function and

$$\Lambda_0(t) = \int_0^t \lambda_0(u)du$$

is the cumulative baseline hazard function.

In the estimation of $S(t)$, $\hat{\beta}$ is substituted for β from the model parameters. Estimation of $\Lambda_0(t)$ and $S_0(t)$ can be done using two different methodologies that are discussed in this section. One of them is that of Tsiatis which is compared to a proposed methodology of the current study.

3.3.2. A Review of Tsiatis Approach

The survival probability in Cox's model is estimated by using likelihood techniques similar to Breslow [14]. Estimates are derived for the underlying cumulative hazard function and the underlying survival probability. Tsiatis computed the estimates for the asymptotic variance of the cumulative hazard function and the survival probability for a given set of covariates x.

Tsiatis uses:

$$S_0(t) = e^{-\Lambda_0(t)}$$

where

$$\Lambda_0(t) = \sum_{t_i \le t} \frac{\delta_i}{\sum_{j \in R_i} e^{\beta' x_j}}$$

An assumption is made that the underlying hazard function is constant over distinct intervals. That is, let $0 < t_1 < \cdots < t_{k+1}$, then

$$\lambda_0(t) = \lambda_i$$

for $t_i < t \le t_{i+1}$, $i = 1, \dots., k$

Under this assumption, standard likelihood techniques are used to estimate the survivorship function as well as its variance. The cumulative hazard function for an individual with covariates x evaluated at t_{i+1} is:

$$\Lambda(t_{t+i}|x) = \sum_{\alpha=1}^{i} \lambda_\alpha e^{(\beta' x)} \Delta t_\alpha$$

where $\Delta t_\alpha = \Delta t_{\alpha+1} - t_\alpha$ denoted by Λ^i. The maximum likelihood estimate of Λ^i is equal to:

$$\Lambda^i = \sum_{\alpha=1}^{i} \lambda_\alpha e^{(\beta' x)}$$

The survivorship function will therefore be given by:

$$S(t) = exp - \left(\sum_{\alpha=1}^{i} \lambda_\alpha e^{(\beta' x)} \right)$$

3.3.3. The Proposed Approach with Relaxed Tsiatis Assumptions

In the current study, the interest is to determine the unknown parameters of the underlying hazard and survivorship functions. Let $t_1 < t_2 < \cdots < t_m$ be the distinct time deaths and censoring, n_i be the number of deaths and censoring at time t_i and m_i be the number of deaths at time t_i. The unknown parameter β in $\lambda(t, x) = \lambda_0(t) e^{\beta' x}$ is estimated using the arguments of the conditional likelihood.

To estimate $\Lambda_0(t)$, Breslow assumed $\lambda_0(t)$ is constant between uncensored observations and estimated the hazard rate between t_i and t_{i-1} for $i = 1, \dots, k$ as:

$$\lambda_i = \frac{m_i}{(t_i - t_{i-1}) \sum_{j=1}^{n} e^{\beta' x_j}}$$

The survival function was therefore estimate by:

$$S(t) = \left\{ \prod_{i=1}^{l} 1 - \frac{m_i}{\sum_{j \in R_i}^{n} e^{\beta' x_j}} \right\}^{e^{\beta' x}}$$

It is noted that $\widehat{\Lambda}(t)$ and Breslow's $\hat{S}(t)$ are inconsistent with the assumption of a continuous underlying hazard function. Furthermore, Breslow's estimate can take negative values especially when the number in the i^{th} risk set is small. This motivates the current study to use a smooth estimate for Λ_0 instead to determine any significant differences.

The hazard smoothing technique to be used is based on that studied by Ramlau-Hansen [15] and Tanner and Wong [16]. Having estimated β, the log likelihood function for $\lambda_0(t)$ can be written as:

$$l[\lambda_0(t)] = \sum_{i=1}^{m} \left\{ m_i \log[\lambda_0(t_i)] + \beta\, s_i - \sum_{j=1}^{m} e^{\beta z_{ij}} \int_0^{t_i} \lambda_0(u)\, du \right\}$$

where s_i is the sum of the covariates of the m_i individuals failing at time t_i and z_{ij} is the vector of being censored at time t_i. Since $l[\cdot]$ is a functional of $\lambda_0(t)$, it cannot be maximized with respect to $\lambda_0(t)$, because it's unbounded. Ramlau showed how to impose a smoothness constraint on $l[\cdot]$ to obtain a bounded estimate of the function $\lambda_0(t)$. He estimated the cumulative hazard function and terminated his work.

In the current study, the survivorship function is further estimated after smoothing is done. A kernel function is used to estimate the smoothed underlying hazard function which is given by:

$$\lambda_0(t) = \frac{1}{h(t)} \sum_{i=1}^{n} K_t \left(\frac{t - t_{(i)}}{h(t)} \right) \frac{\delta_i}{n - i + 1}$$

where $K(\cdot)$ is a kernel function, h is the bandwidth and

$0 < t_{(1)} < \cdots < t_{(n)}$ are distinct ordered times.

The underlying cumulative hazard function can then be estimated as:

$$\Lambda_0(t) = \int_0^t \lambda_0(t)dt$$

and consequently the survivorship function is estimated by:

$$S(t) = S_0(t)^{e^{\beta' x}}$$

where $S_0(t)$ is the estimated baseline survival function. The survivorship function obtained is a smoothed version of the step function estimate by Breslow and Tsiatis. It is therefore consistent and desirable with the assumption of a continuous underlying hazard function. It also gives an estimate of the survival function for any time $t \leq t_k$ rather than just the observed time deaths.

3.4. Asymptotic Variance of the Survivorship Function

A simple application of the delta method is used to estimate the asymptotic variance of the survival function. This is given by:

$$Var[S(t)] = S(t)^2\, Var[\Lambda(t)]$$

The asymptotic variance of the cumulative hazard function is first derived. The likelihood for n individuals with covariates can be expressed as:

$$\prod_{j=1}^{n}\prod_{i=1}^{k}\left[\left\{\lambda_0(t)e^{\beta' x_j}\right\}^{d_{ij}} exp\left\{-\lambda_0(t)e^{\beta' x_j}\right\}D_{ij}\right]$$

where d_{ij}= 1 if individual j dies in the interval i and d_{ij}= 0 otherwise. D_{ij} is the time spent in the i^{th} interval by individual j. The log likelihood would be equal to:

$$L = \sum_{i=1}^{k} d_i \log \lambda_0(t) + \sum_{j\in D}\beta' x_j - \sum_{i=1}^{k}\lambda_i \sum_{i=1}^{k} e^{\beta' x_j} D_{ij}$$

where $d_i = \sum_{j=1}^{n} d_{ij}$ is the number of deaths in the i^{th} interval and D denotes the index set for individuals who died. Maximization is accomplished by setting the first partial derivatives equal to zero. The log-likelihood however cannot be maximized with respect to $\lambda_0(t)$ as the smoothed function is restricted to the class of continuous functions which have piecewise continuous first derivatives. The asymptotic variance of $\Lambda(t)$ using the δ- method will be equal to:

$$Var\,[\Lambda(t)] = f'\, Cov\big(\hat{\lambda}_0(t)\ \hat{\beta}\big)\, f$$

where f is the vector of first partial derivatives of the cumulative hazard function. The evaluation of these values yields:

$$Var\,[\Lambda(t)] = \left\{e^{\beta' x}\right\}^2 \lambda_0(t) + z'\, Var(\beta)\, z$$

where

$$z' = \left(\sum_{\alpha=1}^{i}\frac{d_\alpha \Delta t_\alpha}{\sum_{j=1}^{n} e^{\beta' x_j} D_{\alpha j}}\left[\frac{\sum_{j=1}^{n} x_{jl} e^{\beta' x_j} D_{\alpha j}}{\sum_{j=1}^{n} e^{\beta' x_j} D_{\alpha j}} - x_l\right], l = 1, \dots, p\right)$$

and $Var(\beta) = \left(\frac{\partial^2 L(\beta)}{\partial \beta_l \partial \beta_l}\right)^{-1}$

The asymptotic variance of the survivorship function is then estimated as:

$$Var[S(t)] = S(t)^2\, Var[\Lambda(t)]$$

For a case where $\beta = 0$, the variance of the survivor function reduces to the asymptotic variance of the classic Kaplan-Meier estimate.

3.5. Coverage Probabilities

The coverage probability for an interval estimate is the proportions of instances in which the sample statistic obtained from infinite independent and identical replications of the experiment is contained. They are important in that they can be used as a measure of accuracy done by checking these values against the selected confidence level if they are equivalent. To estimate the coverage probabilities, the following steps are taken:

i. Simulate many samples of size n from the population.

ii. Estimate the survivorship function using Cox-PH model while relaxing the Tsiatis assumption.

iii. Estimate the asymptotic variance for the survivorship function.

iv. Construct the point-wise confidence interval for the estimate as $S(t) \pm constat * [SE(S(t))]$ where the constant is given by $z\alpha/_2$

v. Compute the proportion of samples for which the survivor function parameter is contained in the confidence interval. The proportion is an estimate for the empirical coverage probability for the confidence interval.

4. Simulation Study

A simulation study was carried out whereby n independent subjects were involved in a survival study and only right censored data were obtained. For a subject i, T_i denotes the survival time of interest and C_i denotes the censoring times. We also assume that C_i and T_i are independent for $i = 1, 2, \dots, n$. The actual observations consist of (X_i, δ_i) where $X_i = \min(T_i, C_i)$ and $\delta_i = I[T_i \leq C_i]$ is an indicator of the censoring status of X_i. The following quantities are used to simulate the data set:

i. Sample size: 100

ii. Covariates: One categorical and one continuous covariates were generated, that is $X1 \sim bin(n, 0.45)$ and $X2 \sim N(44, 2.915)$

ii. Hazard: Followed a Weibull distribution, $Wei(0.5, 4)$

iv. Censoring: Exponential censoring times were generated, $C \sim \exp(0.01)$

v. Regression parameters -0.5 and -0.6 were also used

Simulations were replicated 1000 times for the three sample sizes 100, 300 and 500 respectively to determine if the confidence intervals obtained from the proposed method provide accurate coverage probabilities after estimating the survival function. Comparisons are made to that of Tsiatis' estimates and the choice of bandwidth selected (plug-in and cross-validation) for kernel smoothing are also observed.

5. Results and Discussions

An assessment of the Proportional Hazards assumption confirmed its violation. From literature, two solutions of dealing with the violation of the assumption are considered where the covariates can either be stratified or extending the PH model appropriately. The extension of the model could be viewed in the classical context of considering more covariates after stratification, including interactions between the covariates or considering a different family for the error term. In an even more robust way, the assumptions can be relaxed as in Tsiatis [17]. The covariates are therefore assumed to be zero and onto which the current study is built up.

For visual comparison, the corresponding survival functions are given together with the estimates computed by Tsiatis in Figure 1. It's evident that the Tsiatis survivor curve is a step-function which is common with most survivorship functions. The proposed curves smooth out the discontinuities found on Tsiatis' curve and are consistent with the assumption of a continuous hazard. For a large sample size, the smoothed survival probabilities are higher at most time points. At say time $t = 80$, the probability of survival for all three curves is approximately 87% as they coincide at this time point.

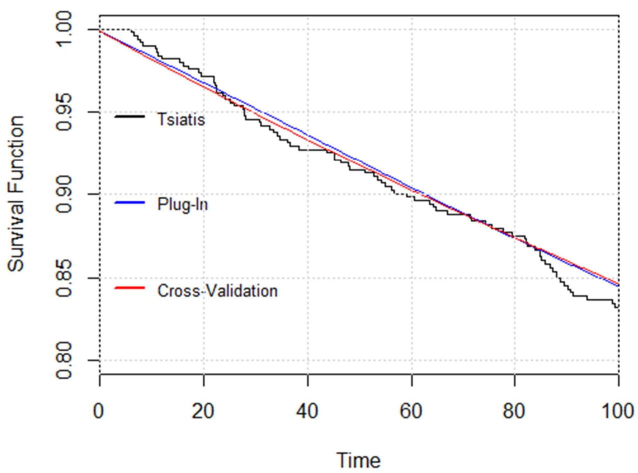

Figure 1. *Tsiatis, Cross-Validation and Plug-In survivorship functions.*

This goes to show that the target point provides crucial information about the dataset proving that smoothing often considers the important peaks and leaves out white noise. Both the plug-in and cross-validation bandwidth selectors are competitive in the sense that neither can be claimed to be the best in estimating the survivorship function from a visual perspective.

Figure 2 presents the asymptotic variance estimates of the survivor function. The variances in the respective techniques increases as time increases attributed to fewer observations as more subjects experience the event of interest. The variances also decrease as the sample size increases which is an indication that the asymptotic normality assumption is appropriate. The smoothed survivor functions variances record lower variances compared to those of Tsiatis' estimates. Estimators with low variances are generally preferable hence the proposed models provide better estimates due to discontinuities smoothed out during estimation.

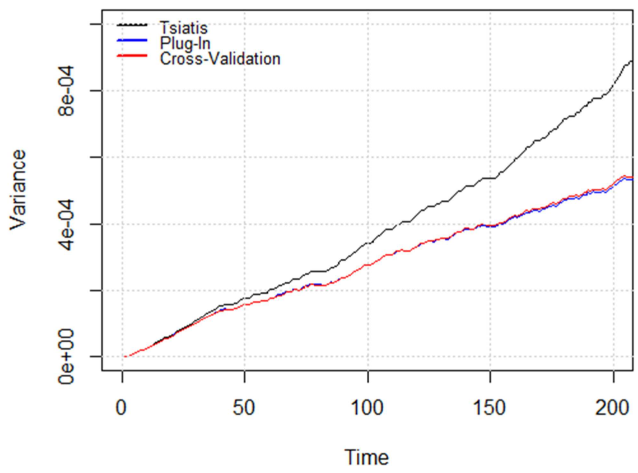

Figure 2. *Point-wise asymptotic variance estimates for the survivorship function.*

Simulation enabled the computation of the proportion of samples for which the survivor function parameter is contained in the confidence interval. Table 1 shows the coverage probabilities repeated for three sample sizes and 90%, 95% and 99% confidence levels. It's noted that the coverage remains about constant and close to their respective nominal levels which is good. The proposed smoothing techniques coverages are however seen to be more accurate than Tsiatis'. More accurate coverages implies narrower or more precise confidence intervals which is always preferable. No trend is observed with increase in the sample size hence the coverage probabilities are independent of the sample size. The coverage probability of 90%, 95% and 99% is only true when the population is normally distributed (which the simulated data violates) or when the sample sizes are large enough to invoke the Central Limit Theorem.

Table 1. *Coverage Probabilities of the Estimated Survivorship Function.*

Confidence Level	n	Coverages		
		Tsiatis	**Plug-In**	**Cross-Validation**
90%	100	0.889	0.903	0.911
	300	0.892	0.890	0.902
	500	0.894	0.889	0.898
95%	100	0.939	0.942	0.953
	300	0.937	0.939	0.947
	500	0.941	0.946	0.941
99%	100	0.979	0.980	0.981
	300	0.980	0.983	0.987
	500	0.976	0.982	0.986

6. Conclusions

In the estimation of the survivorship function using the Cox-PH model, kernel smoothing is more appropriate than the common standard techniques in survival analysis. Kernel smoothing which is a non-parametric estimator method uncovers features in the data which parametric approaches such as Tsiatis' might not reveal. For instance, the smoothed hazard rates provide more information such as the immediate risks attached to a given subject. They also eliminate white noise by smoothing out any discontinuities. An advantage of the proposed methodology is that it gives piecewise smooth estimates and does not have negative values. Furthermore, no constraints are implored on the form of the underlying hazard proving to be less bias than Tsiatis' method. On the choice of bandwidth, the cross-validation and plug-in selectors are very competitive such that neither can be said to outdo the other when estimating the survivor function. Generally, empirical estimators are reduced on application of kernel smoothing which the study concurs as observed by the lower variances and coverage probabilities that are accurate and closer to the nominal level.

References

[1] Kaplan E. L. and Meier P., "Nonparametric Estimation from Incomplete Observations," Journal of the American Statistical Association, 1958.

[2] Nadaraya E. A., "Some New Estimates for Distribution Functions," Theory of Probability and Its Applications, vol. 9, no. 3, pp. 497-500, 1964.

[3] Breslow N., "Covariance Analysis of Censored Survival Data," Biometrics, vol. 30, pp. 89-99, 1974.

[4] Tsiatis Anastasios A., "A Heuristic Estimate of the Asymptotic Variance of the Survival Probability in Cox's Regression Model," University of Wisconsin, 1978.

[5] Rosenblatt John Rice and Murray, "Estimation of the Log Survivor Function and Hazard Function," Sankhya: The Indian Journal of Statistics, Series A (1961-2002), vol. 38, pp. 60-78, 1976.

[6] Cox D. R., "Regression Models and Life-Tables," Journal of the Royal Statistical Society. Series D (The Statistician), 1972.

[7] Anderson J. A. and Senthilselvan A., "Smooth Estimates for the Hazard Function," Journal of the Royal Statistical Society. Series B (Methodological), vol. 42, pp. 322--327, 1980.

[8] Goodd I. J. and Gaskins R. A., "Nonparametric roughness penalties for probability densities," Biometrika, vol. 58, pp. 255-277, 1971.

[9] Kalbfleisch J. D. and Prentice R. L., "Marginal likelihoods based on Cox's regression and life model," Biometrika, vol. 60, no. 2, pp. 267-278, 1973

[10] Gijbels A. D., "Practical bandwidth selection in deconvolution kernel density estimation," Computational Statistics and Data Analysis, vol. 45, pp. 249 - 267, 2004

[11] Lin D. Y., Fleming T. R. and Wei L. J., "Confidence Bands for Survival Curves Under the Proportional Hazards Model," Biometrika, vol. 81, no. 1, pp. 73-81, 1994.

[12] Nair V. N., "Confidence Bands for Survival Functions With Censored Data: A Comparative Study," Technometrics, vol. 26, no. 3, pp. 265-275, 1984.

[13] Bie O., Borgan O. and Liestol K., "Confidence Intervals and Confidence Bands for the Cumulative Hazard Rate Function and Their Small Sample Properties," Scandinavian Journal of Statistics, vol. 14, no. 3, pp. 221-233, 1987.

[14] Breslow N. E., "Discussion of Professor Cox's paper," J Royal Stat Soc B, vol. 34, pp. 216-217, 1972.

[15] Ramlau-Hansen H., "Smoothing Counting Process Intensities by Means of Kernel Functions," The Annals of Statistics, vol. 11, no. 2, pp. 453-466, 1983.

[16] Tanner M. A. and Wong W. H., "The Estimation of the Hazard Function from Randomly Censored Data by the Kernel Method," The Annals of Statistics, vol. 11, pp. 989-993, 1983.

[17] Tsiatis Anastasios A., "A Large Sample Study of Cox's Regression Model," The Annals of Statistics, vol. 9, pp. 93-108, 1981.