

The Cause Specific Hazard Quantile Function

P. G. Sankaran
Cochin University of
Science and Technology
Cochin, 682022

Isha Dewan
Indian Statistical Institute
New Delhi, 110016

Dileep Kumar M
Cochin University of
Science and Technology
Cochin, 682022

Abstract

In this paper, we discuss modeling and analysis of competing risks data using the quantile function. We introduce and study the cause specific hazard quantile function. We present competing risks models using various functional forms for the cause specific hazard quantile functions. A non-parametric estimator of the cause specific hazard quantile function is derived. Asymptotic properties of the estimator are studied. Simulation studies are carried out to assess the performance of the estimator. Finally, we apply the proposed procedure to real life data sets.

Keywords: competing risks, quantile density function, non-parametric estimators.

1. Introduction

In survival studies, it is common that the failure of subjects may be attributed to more than one cause. Competing risks models are usually employed to analyze such type of data. In the competing risks set up, for each subject under study we observe a random vector (T, J) where T represents lifetime (possibly censored) and $J = \{1, 2, \dots, k\}$ is a set of possible causes of failure. Assume that the causes of failure are mutually exclusive. Two frameworks are often employed to deal with standard competing risks data viz. cumulative incidence function formulations and cause specific hazard formulations.

The cumulative incidence function $F_j(t)$ is the probability of failure before time t due to cause j given by,

$$F_j(t) = P[T \leq t, J = j], \quad j = 1, 2, \dots, k. \quad (1)$$

Note that $F(t) = \sum_{j=1}^k F_j(t)$ is the distribution function of T .

The cause specific hazard function $h_j(t)$ of T is defined as,

$$h_j(t) = \lim_{\Delta t \rightarrow 0} \frac{P[T < t + \Delta t, J = j | T \geq t]}{\Delta t}, \quad j = 1, 2, \dots, k. \quad (2)$$

The $h_j(t)$ is the instantaneous rate of failure due to the cause j at time t given the subject has survived up to time t .

Let $f_j(t) = \frac{d}{dt}F_j(t)$ be the cause specific density of T . If the density $f_j(t)$ exists, (2) can be written as,

$$h_j(t) = \frac{f_j(t)}{S(t)}, \quad (3)$$

where $S(t)$ is the survival function of T .

Another important function of interest used for the analysis of competing risks data is the sub-survival function $S_j(t)$, defined by,

$$S_j(t) = \exp\left(-\int_0^t h_j(x)dx\right). \quad (4)$$

The function (4) does not represent a proper survival function of an observable random variable (Lawless (2003)). Further $S_j(t) \neq 1 - F_j(t)$.

When the causes of failure are mutually exclusive and exhaustive, then the hazard rate of T , $h(t)$ can be written as,

$$h(t) = \sum_{j=1}^k h_j(t).$$

Thus, $S(t)$ is uniquely determined by the following identity,

$$S(t) = \exp\left(-\sum_{j=1}^k \int_0^t h_j(u)du\right) = \prod_{j=1}^k S_j(t).$$

From (3), we get the cumulative incidence function $F_j(t)$ as,

$$F_j(t) = \int_0^t S(u)h_j(u)du.$$

For properties and applications of (1), (2) and (4), see Carriere and Kocher (2000), Lawless (2003) and Crowder (2012).

An alternative approach for modeling and analysis of statistical data is to use the quantile function. Both the distribution function and the quantile function convey the same information about the random mechanism of the subject with different implications. The concepts and methodologies based on the distribution function are traditionally employed in statistical theory. However, the quantile function has several interesting properties that are not shared by the distribution function. For example, the sum of two quantile functions is again a quantile function. Parzen (1979) discussed non-parametric statistical modeling of data using the quantile function. Recently, Peng and Fine (2007) developed non-parametric inference procedures for competing risks data using the quantile function. Nair and Sankaran (2009) presented basic reliability concepts viz, hazard rate, mean residual life function etc. in terms of the quantile function. Sankaran, Nair, and Sreedevi (2010) derived a test procedure for comparing various risks using sub-quantile functions and Soni, Dewan, and Jain (2015) proposed tests for successive comparison of quantiles using the quantile functions. Soni, Dewan, and Jain (2012) developed a non-parametric estimator of the quantile density function. Various properties and applications of the quantile function are available in Gilchrist (2000) and Nair, Sankaran, and Balakrishnan (2013).

The objective of the present work is to supplement the work of Peng and Fine (2007) by introducing quantile based concepts in the competing risks set up. We define the cause specific hazard quantile function which is the quantile version of (2). The proposed study has several advantages. In many practical situations, the well known parametric models are not appropriate for the analysis of lifetime data. The quantile approach provides new quantile function models, as shown in Section 2, which are useful for the modeling and analysis of lifetime data. In survival

studies, censoring is common. In such contexts, quantile based analysis is more appropriate as quantiles are more robust (Nair *et al.* (2013)). Finally, the quantile approach gives an alternative methodology for the statistical analysis of competing risks data.

The rest of the article is organized as follows. In Section 2, we present definitions of quantile based reliability concepts useful in competing risks theory. Section 3 discusses non-parametric estimation of the cause specific hazard quantile functions and study asymptotic properties of the estimators. A simulation study is carried out in Section 4 to assess finite sample properties of the estimators. The proposed estimation procedure is illustrated on two real data sets in Section 5. Finally, Section 6 provides major conclusions of the study.

2. Cause specific hazard quantile functions

Let T be a non-negative continuous random variable representing the lifetime of a subject with distribution function $F(t)$ and density function $f(t)$. Assume that $F(t)$ is strictly increasing. Denote $Q(u) = \inf\{t : F(t) \geq u\}$ as the quantile function of T . Since $F(t)$ is strictly increasing, we have $Q(u) = F^{-1}(u)$. Let $Q_j(u)$ be the sub-quantile function defined by,

$$Q_j(u) = \inf\{t : F_j(t) \geq u\}. \quad (5)$$

Since $F_j(\infty) < 1$, $Q_j(1) = v_j < \infty$. Let $q(u) = \frac{d}{du}Q(u)$ and $q_j(u) = \frac{d}{du}Q_j(u)$ be the quantile density and the sub-quantile density functions, respectively (see Peng and Fine (2007)). Taking derivative on both sides of the identity $F(Q(u)) = u$, we get $f(Q(u)) = \frac{1}{q(u)}$. Nair and Sankaran (2009) defined the hazard quantile function of T which is the quantile version of $h(t)$ as,

$$\Lambda(u) = h(Q(u)) = \frac{f(Q(u))}{1 - F(Q(u))} = \frac{1}{(1 - u)q(u)}. \quad (6)$$

It is shown that $\Lambda(u)$ uniquely determines the quantile function, $Q(u)$, by

$$Q(u) = \int_0^u \frac{dp}{(1 - p)\Lambda(p)}. \quad (7)$$

We now define the cause specific hazard quantile function as,

$$\Lambda_j(u) = h_j(Q(u)) = \frac{f_j(Q(u))}{1 - F(Q(u))} = \frac{f_j(Q(u))}{(1 - u)}. \quad (8)$$

The quantity $\Lambda_j(u)$ is interpreted as the conditional probability of the failure of the subject in the next small interval of time due to cause j given the survival of the subject at $100(1 - u)\%$ point of the distribution.

Note that,

$$\frac{d}{du}F_j(Q(u)) = q(u)f_j(Q(u)) = \frac{f_j(Q(u))}{f(Q(u))}. \quad (9)$$

It is easy to see from (6) and (8), that $\sum_{j=1}^k \Lambda_j(u) = \Lambda(u)$.

Thus the hazard quantile function is the sum of the cause specific hazard quantile functions. Further, note that,

$$\frac{d}{du}(F_j(Q(u))) = \frac{\Lambda_j(u)}{\Lambda(u)}.$$

Therefore,

$$F_j(Q(u)) = \int_0^u \frac{\Lambda_j(p)}{\Lambda(p)} dp,$$

or

$$Q(u) = Q_j \left(\int_0^u \frac{\Lambda_j(p)}{\Lambda(p)} dp \right). \quad (10)$$

The identity (10) enables us to determine $Q(u)$ or $Q_j(u)$ from $\Lambda_j(u)$.

Example 2.1 (Constant cause specific hazard quantile function). Suppose that the cause specific hazard quantile function corresponding to j th risk is constant. That is,

$$\Lambda_j(u) = a_j, \quad a_j > 0 \quad \text{for } 0 < u < 1.$$

Hence the hazard function $\Lambda(u) = \sum_{j=1}^k a_j$ is a constant. This leads to the fact that the lifetime T has an exponential distribution with $F(t) = 1 - e^{-(\sum_{j=1}^k a_j)t}$. Then the quantile function is,

$$Q(u) = \frac{-\log(1-u)}{\sum_{j=1}^k a_j}.$$

Example 2.2 (Proportional hazards model). When

$$h_j(t) = \pi_j h(t), \quad \pi_j > 0,$$

we obtain,

$$\Lambda_j(u) = \pi_j \Lambda(u),$$

and $Q(u) = Q_j(\pi_j u)$. Thus the cause specific hazard functions are proportional.

3. Non-parametric estimation of the cause specific hazard quantile function

We develop a non-parametric estimator of $\Lambda_j(u)$ under right censoring using the kernel density estimation approach. Suppose that the lifetime T is randomly right censored by a variable Z . Then we observe a random vector $(Y, \delta, \delta J)$ where $Y = \min(T, Z)$ and $\delta = I(T \leq Z)$. Note that δJ is 0 for a censored observation, otherwise it is the cause of failure. Denote $G(t)$ and $H(t)$ as the distribution functions of Z and Y , respectively. When Z and T are independent, we have,

$$1 - H(t) = (1 - F(t))(1 - G(t)).$$

The tuples $(Y_i, \delta_i, \delta_i J_i)$ are assumed to be realizations of random variables $(Y, \delta, \delta J)$, for subjects $1, 2, \dots, n$. Thus δJ equals zero for a censored observation, otherwise it is the cause of failure. If censoring is assumed, the Kaplan-Meier estimator of $S(t)$ for the ordered failure times $Y_{(1)} < Y_{(2)} < \dots < Y_{(n)}$, corresponding to $Y_i, i = 1, 2, 3, \dots, n$, is given by,

$$\hat{S}(t) = \prod_{k: Y_{(k)} < t} \left(1 - \frac{d_k}{n_k}\right), \quad (11)$$

where d_k is the number of failures at $Y_{(k)}$ and n_k is the number of subjects at risk in $Y_{(k)}$; $k = 1, 2, \dots, n$. Then the non-parametric estimator of $F(t)$ is $\hat{F}(t) = 1 - \hat{S}(t)$.

Let $Y_{j(1)} < Y_{j(2)} < \dots < Y_{j(n_j)}$ be ordered failure times due to risk at j . The Kaplan-Meier estimator of $S_j(t)$ is obtained as

$$\hat{S}_j(t) = \prod_{k: Y_{j(k)} < t} \left(1 - \frac{d_{jk}}{n_{jk}}\right), \quad (12)$$

where d_{jk} is the number of failures at $Y_{j(k)}$ and n_{jk} is the number of subjects at risk in $Y_{j(k)}$.

Let

$$\hat{S}_{ji} = \begin{cases} \hat{S}_j(Y_{j(i-1)}) - \hat{S}_j(Y_{j(i)}) & i = 2, \dots, n_j - 1 \\ \hat{S}_j(Y_{j(n_j)}) & i = n_j \end{cases}$$

and

$$S^*(i) = \begin{cases} 0 & \text{if } i = 0 \\ \hat{F}(Y_{(i)}) & \text{if } i = 1, 2, \dots, n-1 \\ 1 & \text{if } i = n \end{cases}$$

A simple non-parametric estimator of $\Lambda_j(u)$ is given by,

$$\hat{\Lambda}_j(u) = \frac{\hat{f}_j(\hat{Q}(u))}{1-u}, \quad (13)$$

where $\hat{Q}(u) = \inf\{x : \hat{F}(x) > u\}$ is the non-parametric estimator of $Q(u)$ and

$$\hat{f}_j(\hat{Q}(u)) = \frac{1}{h(n)} \int_0^1 K\left(\frac{p-u}{h(n)}\right) d(\hat{F}_j(\hat{Q}(p))). \quad (14)$$

Function $K(x)$ is a kernel function satisfying following conditions:

- (a) $K(x) \geq 0$ for all x and $\int_{-\infty}^{\infty} K(x)dx = 1$;
- (b) $K(x)$ is symmetric about zero;
- (c) $K(x)$ has finite support and
- (d) $K(x)$ satisfies the Lipschitz condition.

Denote $\delta_{(i)}$ as the indicator function corresponding to $Y_{(i)}$. Then a non-parametric estimator of $\hat{f}_j(\hat{Q}(u))$ given in (14) becomes,

$$\hat{f}_j(\hat{Q}(u)) = \frac{1}{h(n)} \sum_{i=1}^n \hat{S}_{ji} K\left(\frac{S^*(i)-u}{h(n)}\right) I(\delta_{(i)} = j), \quad j = 1, 2, 3, \dots, k. \quad (15)$$

Substituting (15) in (13), we get an estimator of $\Lambda_j(u)$.

We now establish asymptotic properties of $\Lambda_j(u)$. We first prove strong consistency of $\Lambda_j(u)$, $j = 1, 2, \dots, k$.

Theorem 1. Suppose that $K(x)$ satisfies conditions (a) to (d). Assume that both $F(x)$ and $K(x)$ are differentiable. Then $\sup_u |\hat{\Lambda}_j(u) - \Lambda_j(u)| \rightarrow 0$ as $n \rightarrow \infty$ for $j = 1, \dots, k$.

The proof is given in Appendix A.

In the following theorem, we prove the limiting distribution of $\sqrt{n}(\hat{\Lambda}_j(u) - \Lambda_j(u))$.

Theorem 2. As $n \rightarrow \infty$, for fixed u ($0 < u < 1$), $\sqrt{n}(\hat{\Lambda}_j(u) - \Lambda_j(u))$, $j = 1, \dots, k$ follows a normal distribution with mean 0 and variance $\sigma_j^2(u)$, where,

$$\sigma_j^2(u) = \mathbb{E} \left[\frac{1}{h(n)(1-u)} \int_0^1 K^*(u, p) Z(p) dp \right]^2, \quad \text{with } Z(p) = \sqrt{n}[\hat{F}_j(\hat{Q}(p)) - F_j(Q(p))].$$

The proof is given in Appendix B.

Remark 1. Since the analytical expressions of $\sigma_j^2(u)$ is complex, we have to use the bootstrap procedure for estimating the variance of $\hat{\Lambda}_j(u)$, $j = 1, 2, \dots, k$. The bootstrap method is based on the resampling method from the original data. We take B samples of size n from the original data using random sampling with replacement. The bootstrap samples are $(Y_i^{(k)}, \delta_i^{(k)}, J_i^{(k)})$, $k = 1, 2, \dots, B$; $i = 1, 2, \dots, n$. We then compute $\hat{\Lambda}_j(u)$, using original data set and the estimate of $\Lambda_j(u)$ using the bootstrap sample k is $\hat{\Lambda}_j^{(k)}(u)$, $k = 1, 2, \dots, B$. We then compute the bias by taking differences $\hat{\Lambda}_j^{(k)}(u) - \hat{\Lambda}_j(u)$, $j = 1, 2$; $k = 1, 2, \dots, B$. Then using these differences, the average bias and MSE are calculated.

4. Simulations

We now carry out extensive simulation studies to find out mean square error (MSE) and bias of the estimator $\hat{\Lambda}_j(u)$ for the uncensored as well as the censored case. We consider two causes of failure. We take different samples of size 50, 100 and 200. We generated 5000 data sets for each scenario. The order of sub-quantiles considered are $u = 0.2 (0.2) 0.8$. Simulations are carried out for uncensored and censored cases to find the average bias and MSE of the estimators. We have employed the triangular, uniform and Epanechnikov kernel functions in simulation studies. However, results are being reported for the Epanechnikov kernel as this provides the smallest MSE. The Epanechnikov kernel is defined by,

$$K(u) = 0.75(1 - u^2)I(|u| \leq 1).$$

To generate random numbers, we consider the following two quantile function models.

(1) Linear cause specific hazard quantile function (Midhu, Sankaran, and Nair (2014))

Suppose that the cause specific hazard quantile function for the cause j is given by the function,

$$\Lambda_j(u) = a_j + b_j u, \quad a_j > 0, \quad a_j + b_j > 0, \quad 0 < u < 1.$$

Then we obtain,

$$\begin{aligned} \Lambda(u) &= A + Bu, \text{ where } A = \sum_{j=1}^k a_j \text{ and } B = \sum_{j=1}^k b_j, \\ \text{and } Q(u) &= \log \left(\frac{A + Bu}{A(1 - u)} \right)^{\frac{1}{A+B}}. \end{aligned}$$

(2) Weibull cause specific hazard model (Crowder (2012)).

The cause specific hazard quantile function for the risk j is,

$$\Lambda_j(u) = \phi \xi_j^{-\phi} \left(-\frac{1}{\beta} \log(1 - u) \right)^{1 - \frac{1}{\phi}}. \quad (16)$$

The relation $\Lambda(u) = \sum_{j=1}^k \Lambda_j(u)$ and identity (7) provide the hazard quantile function and quantile function as,

$$\Lambda(u) = \phi \beta \left(-\frac{1}{\beta} \log(1 - u) \right)^{1 - \frac{1}{\phi}} \text{ and } Q(u) = \left(-\frac{1}{\beta} \log(1 - u) \right)^{\frac{1}{\phi}}, \quad (17)$$

where $\beta = \sum_{j=1}^k \xi_j^{-\phi}$.

Since the proposed estimator of the cause specific hazard quantile function is based on the kernel function, the choice of bandwidth is an important issue. For the construction of kernel type estimator of a quantile function, Padgett (1986) has considered separate bandwidths for different regions of $u \in (0, 1)$ in such a way that the mean squared error (MSE) is minimum. In our study, we calculate the optimum bandwidths corresponding to different values of u such as 0.2, 0.4, 0.6, and 0.8. The average of the optimal bandwidths obtained for different values of u is employed for the construction of the proposed estimators.

To perform the simulation study, we use the same parameter combinations for the linear cause specific hazard quantile function model in both censored as well as uncensored cases. The same procedure is adopted for the Weibull cause specific hazard model. The parameter values chosen for the linear cause specific hazard quantile function model are $a_1 = \frac{1}{2}$, $b_1 = 3$, $a_2 = \frac{1}{3}$, and $b_2 = 2$. For the Weibull model, we take $\phi = 3$, $\xi_1 = 1$ and $\xi_2 = 2$.

4.1. Results for the uncensored case

We first consider the linear cause specific hazard quantile function for different sample sizes $n = 50, 100$, and 200 . The estimators $\hat{\Lambda}_j(u)$, $j = 1, 2$ are calculated for all values of u ($0 < u < 1$), which provides the smooth curves. Then the average bias and MSE of the estimators are computed. The bandwidths for $\hat{\Lambda}_1(u)$ and $\hat{\Lambda}_2(u)$ are obtained as 0.52 and 0.64 respectively. Figures 1(a) and 1(b) show mean of the estimators and true values of $\Lambda_j(u)$, $j = 1, 2$ for $n = 200$. The results for $n = 50$ and 100 are similar. Table 1 presents the average bias and MSE

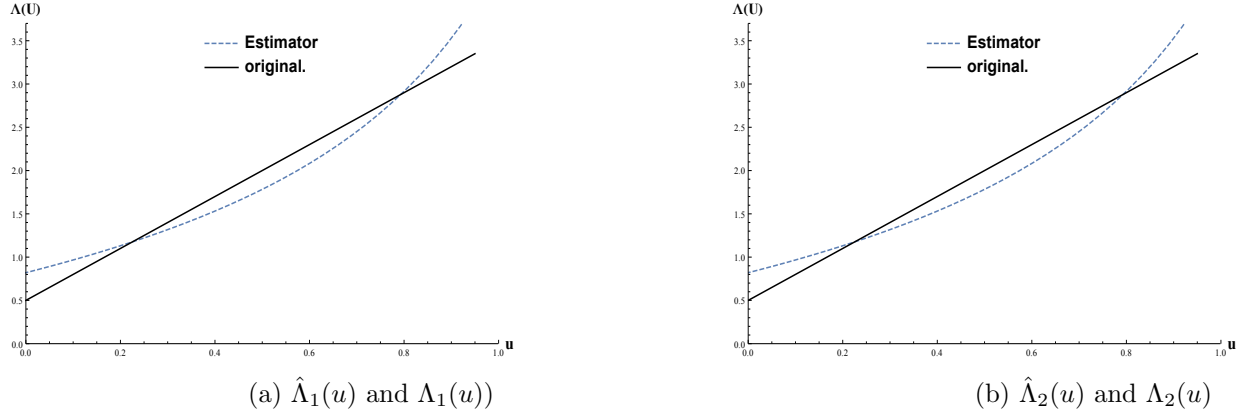


Figure 1: Mean of the estimators and true values of $\Lambda_j(u)$, $j = 1, 2$ for the linear cause specific hazard model with optimal bandwidths for $n = 200$ (uncensored).

of $\Lambda_j(u)$, $j = 1, 2$, for $n = 50, 100$ and 200 . Both average bias and MSE decrease as sample size increases.

Table 1: Average bias and MSE of $\hat{\Lambda}_1(u)$ and $\hat{\Lambda}_2(u)$ for the linear cause specific hazard model (uncensored) for the optimal bandwidths.

n			u			
			0.2	0.4	0.6	0.8
50	$\hat{\Lambda}_1(u)$	MSE	0.0562	0.2758	0.3318	0.1729
		BIAS	-0.1445	-0.4912	-0.4314	0.5814
	$\hat{\Lambda}_2(u)$	MSE	0.0384	0.1477	0.1686	0.3029
		BIAS	-0.1225	-0.3077	-0.2773	0.2433
100	$\hat{\Lambda}_1(u)$	MSE	0.0453	0.2400	0.2766	0.1525
		BIAS	-0.1312	-0.4732	-0.4040	0.3315
	$\hat{\Lambda}_2(u)$	MSE	0.0328	0.1303	0.1417	0.2478
		BIAS	-0.1221	-0.3008	-0.2727	0.6292
200	$\hat{\Lambda}_1(u)$	MSE	0.0406	0.2238	0.2523	0.1380
		BIAS	-0.1239	-0.4638	-0.3921	0.3307
	$\hat{\Lambda}_2(u)$	MSE	0.0303	0.1221	0.1293	0.1734
		BIAS	-0.1203	-0.2961	-0.2566	0.1922

We then consider the Weibull cause specific hazard model (15). The estimators $\hat{\Lambda}_j(u)$, $j = 1, 2$ are calculated. The bandwidths which give minimum MSE for $\hat{\Lambda}_1(u)$ and $\hat{\Lambda}_2(u)$ are 0.72 and 0.44 respectively. Figures 2(a) and 2(b) show mean of the estimators of $\Lambda_j(u)$, $j = 1, 2$ for $n = 200$. Table 2 gives average bias and MSE of the estimators of the cause specific hazard quantile functions. Note that both average bias and MSE decrease as sample size increases.

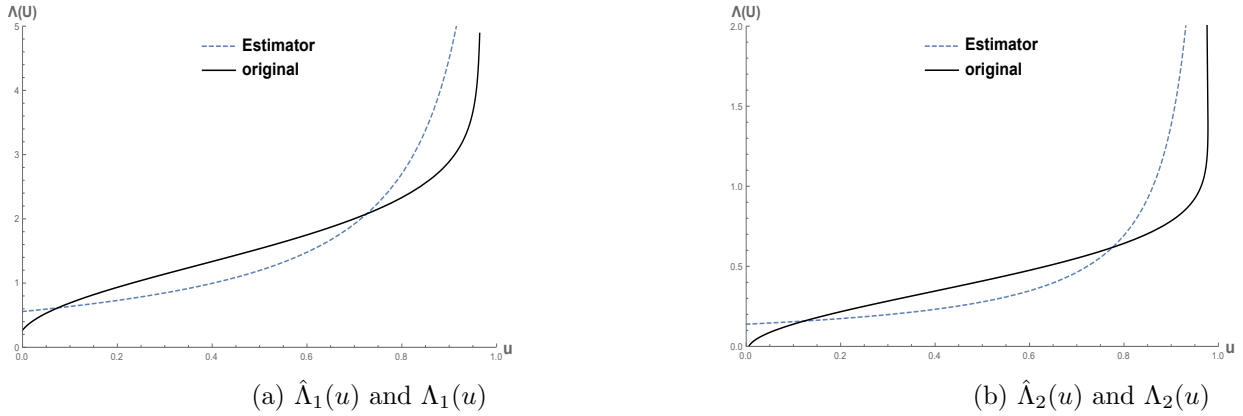


Figure 2: Mean of the estimators and true values of $\Lambda_j(u)$, $j = 1, 2$ for the Weibull cause specific hazard model with optimal bandwidths for $n = 200$ (uncensored).

Table 2: Average bias and MSE for $\hat{\Lambda}_1(u)$ and $\hat{\Lambda}_2(u)$ for the Weibull cause specific hazard model (uncensored) for the optimal bandwidths.

n			u			
			0.2	0.4	0.6	0.8
50	$\hat{\Lambda}_1(u)$	MSE	0.0395	0.1511	0.1546	0.1776
		BIAS	-0.1670	-0.3421	-0.2833	0.4186
	$\hat{\Lambda}_2(u)$	MSE	0.0160	0.0253	0.0301	0.0296
		BIAS	-0.0560	-0.1248	-0.1296	0.1410
100	$\hat{\Lambda}_1(u)$	MSE	0.0354	0.1387	0.1368	0.1594
		BIAS	-0.1555	-0.3319	-0.2822	0.3956
	$\hat{\Lambda}_2(u)$	MSE	0.0142	0.0207	0.0231	0.0232
		BIAS	-0.0554	-0.1009	-0.0932	0.1125
200	$\hat{\Lambda}_1(u)$	MSE	0.0335	0.1328	0.1285	0.1262
		BIAS	-0.1533	-0.3231	-0.1541	0.3492
	$\hat{\Lambda}_2(u)$	MSE	0.0136	0.0189	0.0207	0.0165
		BIAS	-0.0447	-0.0804	-0.0748	0.0168

4.2. Results for the censored case

The censored observations are generated using uniform distribution $U(0, C)$, where C is chosen such that 20% observations are censored. We first consider the linear cause specific hazard quantile function model. We compute the average bias and MSE of the estimators $\hat{\Lambda}_j(u)$, $j = 1, 2$. The bandwidths which give minimum MSE for $\hat{\Lambda}_1(u)$ and $\hat{\Lambda}_2(u)$ are 0.67 and 0.31 respectively. Figure 3 shows the mean of the estimators and the original values of $\Lambda_j(u)$, $j = 1, 2$ for $n = 200$. Table 3 presents the average bias and MSE under censoring. Both average bias and MSE decrease as sample size increases.

We generate observations from the Weibull cause specific hazard model with the censoring scheme given above. The $\hat{\Lambda}_j(u)$, $j = 1, 2$ are calculated and the average bias and MSE of the estimators are computed. The bandwidths which give minimum MSE for $\hat{\Lambda}_1(u)$ and $\hat{\Lambda}_2(u)$ are 0.59 and 0.38 respectively. Figure 4 shows the mean of the estimators and true values of $\Lambda_j(u)$, $j = 1, 2$ for $n = 200$. Table 4 presents average bias and MSE of the estimators of $\Lambda_j(u)$, $j = 1, 2$. It follows that the average bias and MSE of $\hat{\Lambda}_j(u)$, $j = 1, 2$ are small and both decrease as sample size increases.

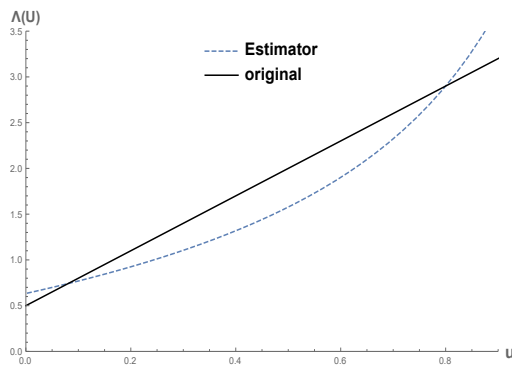
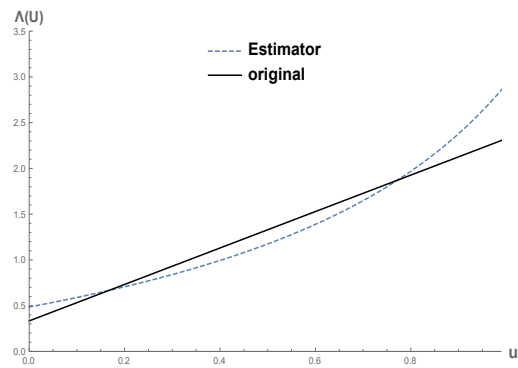
(a) $\hat{\Lambda}_1(u)$ and $\Lambda_1(u)$ (b) $\hat{\Lambda}_2(u)$ and $\Lambda_2(u)$

Figure 3: Mean of the estimators and true values of $\Lambda_j(u)$, $j = 1, 2$ for the linear cause specific hazard model with optimal bandwidths for $n = 200$ (censored).

Table 3: Average bias and MSE for $\hat{\Lambda}_1(u)$ and $\hat{\Lambda}_2(u)$ for the linear cause specific hazard model (censored) for the optimal bandwidths.

n			u			
			0.2	0.4	0.6	0.8
50	$\hat{\Lambda}_1(u)$	MSE	0.0680	0.2764	0.3901	0.7679
		BIAS	-0.2179	-0.5230	-0.5746	0.6104
	$\hat{\Lambda}_2(u)$	MSE	0.0548	0.1622	0.1862	0.4400
		BIAS	-0.1691	-0.3778	-0.4054	0.6832
100	$\hat{\Lambda}_1(u)$	MSE	0.0618	0.2401	0.1875	0.7560
		BIAS	-0.1898	-0.4870	-0.5239	0.6020
	$\hat{\Lambda}_2(u)$	MSE	0.0337	0.1991	0.1856	0.4110
		BIAS	-0.1506	-0.3533	-0.3699	0.2796
200	$\hat{\Lambda}_1(u)$	MSE	0.0587	0.2284	0.1582	0.7032
		BIAS	-0.1760	-0.4697	-0.4997	0.5468
	$\hat{\Lambda}_2(u)$	MSE	0.0323	0.1978	0.1816	0.3591
		BIAS	-0.1415	-0.3412	-0.3523	0.2642

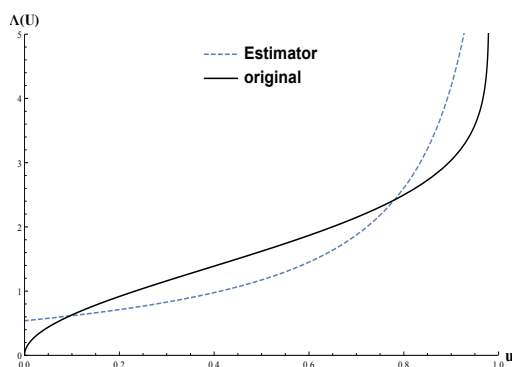
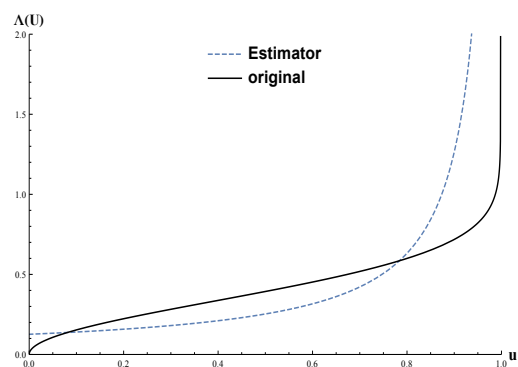
(a) $\hat{\Lambda}_1(u)$ and $\Lambda_1(u)$ (b) $\hat{\Lambda}_2(u)$ and $\Lambda_2(u)$

Figure 4: Mean of the estimators and true values of $\Lambda_j(u)$, $j = 1, 2$ for the Weibull cause specific hazard model with optimal bandwidths for $n = 200$ (censored).

5. Real data illustration

In this section, we apply the proposed procedure to two real life data sets. The first one is

uncensored data and the second one is censored data.

Hoel Data (Hoel (1972)) The data were obtained from a laboratory experiment on two groups of RFM strain male mice which had received a radiation dose of 300r at an age of 5-6 weeks. The first group of mice lived in a conventional laboratory environment while the second group was in a germ-free environment.

Table 4: Average bias and MSE of $\hat{\Lambda}_1(u)$ and $\hat{\Lambda}_2(u)$ for the Weibull cause specific hazard model (censored) for the optimal bandwidths.

n			u			
			0.2	0.4	0.6	0.8
50	$\hat{\Lambda}_1(u)$	MSE	0.0406	0.1512	0.1885	0.4045
		BIAS	-0.1750	-0.3827	-0.3864	0.6372
	$\hat{\Lambda}_2(u)$	MSE	0.0168	0.0275	0.0339	0.0315
		BIAS	-0.0697	-0.1264	-0.1360	0.1430
100	$\hat{\Lambda}_1(u)$	MSE	0.0363	0.1415	0.1879	0.4041
		BIAS	-0.1619	-0.3658	-0.3619	0.6267
	$\hat{\Lambda}_2(u)$	MSE	0.0149	0.0276	0.0341	0.0260
		BIAS	-0.0562	-0.1059	-0.1059	0.1344
200	$\hat{\Lambda}_1(u)$	MSE	0.0349	0.1501	0.1730	0.3757
		BIAS	-0.1554	-0.3574	-0.3497	0.6232
	$\hat{\Lambda}_2(u)$	MSE	0.0157	0.0247	0.0295	0.0157
		BIAS	-0.0468	-0.0962	-0.0917	0.0177

There are three major causes for death viz. thymic lymphoma, reticulum cell sarcoma and other cause. All mice died at the end of the study so that there is no censoring. We considered data from first group of 99 mice for analysis. We combine the last two causes since the number of deaths due to reticulum cell sarcoma is small. Thus two causes for the analysis are thymic lymphoma (J_1) and other causes (J_2) which includes reticulum cell sarcoma. The interest is to compare the mortality from these two modes of death. The estimators of $\Lambda_j(u)$, $j = 1, 2$ are computed as described in Section 3. The bandwidth which minimizes the bootstrap MSE has been chosen. Bandwidths thus obtained for $\hat{\Lambda}_j(u)$, $j = 1, 2$, are 0.71 and 0.29 respectively. Figure 5 shows the cause specific hazard quantile functions. From Figure 5, it is clear that the

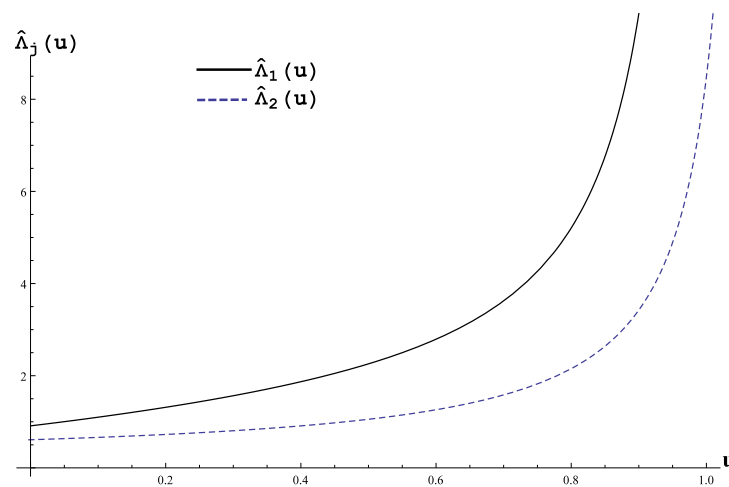


Figure 5: Estimates of cause specific hazard quantile functions for Hoel data.

cause specific hazard quantile function due to thymic lymphoma is uniformly smaller than that due to other causes. We also observe that the two cause specific hazard functions are closer to

each other at the tails. The major cause of failure is not thymic lymphoma J_1 , but other causes J_2 .

Davis and Lawrance Data (Davis and Lawrance (1989)). They considered the tyre-testing data, which measure the failure times at hourly intervals of 171 tyres with 12% right censoring. The major causes of failures are,

- (i) an open joint on the inner lines,
- (ii) rubber chunking on the shoulder
- (iii) loose chunking, low on the shoulder,
- (iv) cracking of tread rubber,
- (v) cracking on the side wall,

and,

- (vi) all other causes of failures.

Since there are few failures due to certain causes, we grouped the causes into three major categories as,

cause 1 (J_1) - for causes (iii) and (v) - 34 failures,

cause 2 (J_2) - for cause (iv) - 69 failures,

and, cause 3 (J_3) - for causes (i), (ii), and (vi) - 48 failures.

The optimal bandwidths for $\hat{\Lambda}_1(u)$, $\hat{\Lambda}_2(u)$ and $\hat{\Lambda}_3(u)$ are 0.47, 0.61 and 0.51 respectively. Figure 6 shows the cause specific hazard quantile function due to three different causes. From Figure 6, it follows that the cause specific hazard quantile function due to cause 1 is larger than that of cause 2. Further the cause specific hazard quantile function due to cause 3 lies between the other two causes of failure. The major cause of failure is due to loose chunking, low on side wall and cracking on side wall (J_1).

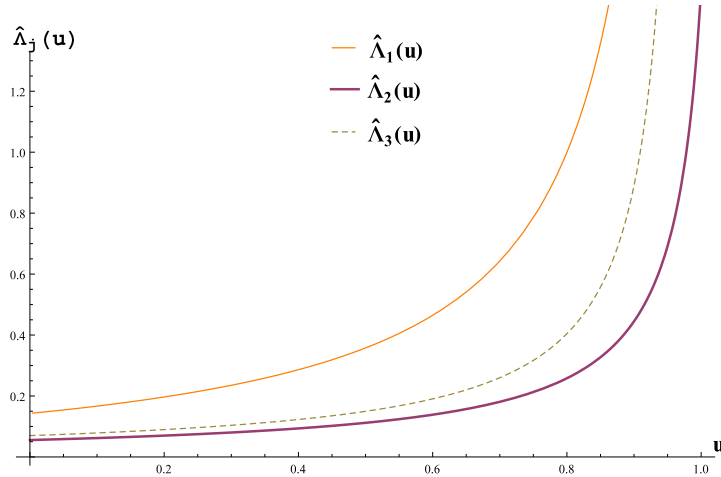


Figure 6: Estimates of cause specific hazard quantile functions for Davis and Lawrence data.

6. Conclusion

The present work introduced the concept of cause specific hazard quantile function, which is the quantile version of the cause specific hazard rates. The proposed methodology provided

new lifetime models useful for the analysis of competing risks data. The smooth kernel type estimator of cause specific hazard quantile function has been developed for uncensored as well as censored data. Asymptotic properties of the proposed estimator were studied. The estimator performs well in terms of average bias and MSE for linear cause specific hazard model as well as for Weibull cause specific hazard model. The procedure has been applied to two real lifetime data sets.

The proposed work based on the cause specific hazard quantile functions is an alternative method of modeling and analysis of competing risks data. This technique has the ability to pick up differences at extreme values of the data. The quantile models presented here will enable the practitioner to differentiate between effects of various risks. In survival studies, it is often interesting to compare various risks. The comparison of various risks can be done by developing non-parametric tests using $\hat{\Lambda}_j(u)$. The work in this direction will be reported elsewhere.

Acknowledgment

We thank the referee and the editor for their constructive comments. The first author is thankful for DST, government of India, for the financial support. The third author is thankful to Kerala State Council for Science Technology and Environment(KSCSTE) for the financial support.

References

- Andersen PK, Borgan O, Gill RD, Keiding N (1993). *Statistical Models Based on Counting Processes*. Springer, New York.
- Carriere KC, Kochar SC (2000). "Comparing Sub-survival Functions in a Competing Risks Model." *Lifetime Data Analysis*, **6**(1), 85–97.
- Crowder MJ (2012). *Multivariate Survival Analysis and Competing Risks*. CRC Press, Washington, D.C.
- Davis TP, Lawrance AJ (1989). "The Likelihood for Competing Risk Survival Analysis." *Scandinavian Journal of Statistics*, pp. 23–28.
- Gilchrist W (2000). *Statistical Modelling with Quantile Functions*. Chapman and Hall, London.
- Hoel DG (1972). "A Representation of Mortality Data by Competing Risks." *Biometrics*, **2**, 475–488.
- Lawless JF (2003). *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons, New Jersey.
- Midhu NN, Sankaran PG, Nair NU (2014). "A Class of Distributions with Linear Hazard Quantile Function." *Communications in Statistics-Theory and Methods*, **43**, 3674–3689.
- Nair NU, Sankaran PG (2009). "Quantile-based Reliability Analysis." *Communications in Statistics- Theory and Methods*, **38**, 222–232.
- Nair NU, Sankaran PG, Balakrishnan N (2013). *Quantile-Based Reliability Analysis*. Birkhauser Verlag, New York.
- Padgett WJ (1986). "A Kernel-type Estimator of a Quantile Function from Right-censored Data." *Journal of the American Statistical Association*, **81**(393), 215–222.

- Parzen E (1979). “Nonparametric Statistical Data Modeling.” *Journal of the American Statistical Association*, **74**, 105–131.
- Peng L, Fine JP (2007). “Nonparametric Quantile Inference with Competing-risks Data.” *Biometrika*, **94**, 735–744.
- Sankaran PG, Nair NU, Sreedevi EP (2010). “A Quantile Based Test for Comparing Cumulative Incidence Functions of Competing Risks Models.” *Statistics and Probability Letters*, **80**, 886–891.
- Serfling RJ (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.
- Soni P, Dewan I, Jain K (2012). “Nonparametric Estimation of Quantile Density Function.” *Computational Statistics and Data Analysis*, **56**, 3876–3886.
- Soni P, Dewan I, Jain K (2015). “Tests for Successive Differences of Quantiles.” *Statistics and Probability Letters*, **97**, 1–8.

Appendix A: Proof of Theorem 1.

From equation (9), we have $dF_j(Q(u)) = f_j(Q(u))q(u)du$. Then,

$$\hat{\Lambda}_j(u) - \Lambda_j(u) = \frac{1}{(1-u)h(n)} \int_0^1 K\left(\frac{p-u}{h(n)}\right) d\hat{F}_j(\hat{Q}(p)) - \frac{f_j(Q(u))}{1-u}. \quad (18)$$

We can write (18) as,

$$\begin{aligned} \hat{\Lambda}_j(u) - \Lambda_j(u) &= \frac{1}{(1-u)h(n)} \int_0^1 K\left(\frac{p-u}{h(n)}\right) d[\hat{F}_j(\hat{Q}(p)) - \hat{F}_j(Q(p)) + \hat{F}_j(Q(p)) - F_j(Q(p))] \\ &\quad + \frac{1}{(1-u)h(n)} \int_0^1 K\left(\frac{p-u}{h(n)}\right) dF_j(Q(p)) - \frac{f_j(Q(u))}{1-u}. \end{aligned} \quad (19)$$

Since, $\sup_u |\hat{Q}(u) - Q(u)| \rightarrow 0$ as $n \rightarrow \infty$ (Andersen, Borgan, Gill, and Keiding (1993)) and $\sup_x |\hat{F}_j(x) - F_j(x)| \rightarrow 0$ as $n \rightarrow \infty$ (Lawless (2003)), the first term on the right side of (19) tends to zero when n is large. Now consider,

$$\begin{aligned} \frac{1}{h(n)} \int_0^1 K\left(\frac{p-u}{h(n)}\right) d(F_j(Q(p))) &= \frac{1}{h(n)} \int_{-\infty}^{\infty} K\left(\frac{t-x}{h(n)}\right) d(F_j(t)) \\ &= \frac{1}{h(n)} \int_{-\infty}^{\infty} K\left(\frac{t-x}{h(n)}\right) f_j(t) dt. \end{aligned} \quad (20)$$

Let $\frac{t-x}{h(n)} = z$. Then (20) becomes,

$$\int_{-\infty}^{\infty} K(z) f_j(x + zh(n)) dz. \quad (21)$$

By Taylor series expansion of $f_j(x + zh(n))$, we obtain (21) as,

$$\int_{-\infty}^{\infty} K(z) [f_j(x) + zh(n)f_j'(x) + \dots] dz, \quad (22)$$

where prime denote derivative with respect to x .

As $n \rightarrow \infty$, $h(n) \rightarrow 0$ and hence (22) tends to,

$$\int_{-\infty}^{\infty} K(z) f_j(x) dz = f_j(x). \quad (23)$$

Using (20) and substituting $x = Q(u)$ in (23), the equation (19) becomes,
 $\sup_u |\hat{\Lambda}_j(u) - \Lambda_j(u)| \rightarrow 0$ as $n \rightarrow \infty$.

Appendix B: Proof of Theorem 2.

From Theorem 1 as $n \rightarrow \infty$ and $h(n) \rightarrow 0$, the expression (19) asymptotically reduces to,

$$\sqrt{n}(\hat{\Lambda}_j(u) - \Lambda_j(u)) = \frac{\sqrt{n}}{h(n)(1-u)} \int_0^1 K\left(\frac{p-u}{h(n)}\right) d[\hat{F}_j(\hat{Q}(p)) - F_j(\hat{Q}(p)) + F_j(\hat{Q}(p)) - F_j(Q(p))]. \quad (24)$$

Using integration by parts, (24) becomes,

$$\begin{aligned} \sqrt{n}(\hat{\Lambda}_j(u) - \Lambda_j(u)) = \frac{\sqrt{n}}{h(n)(1-u)} & \left(\int_0^1 K^*(u, p) [\hat{F}_j(\hat{Q}(p)) - F_j(\hat{Q}(p))] dp \right. \\ & \left. + \int_0^1 K^*(u, p) [F_j(\hat{Q}(p)) - F_j(Q(p))] dp \right), \quad (25) \end{aligned}$$

where $K^*(u, p) = \frac{dK}{dp}\left(\frac{p-u}{h(n)}\right)$.

From Andersen *et al.* (1993), for $0 < u < 1$, we have $\sqrt{n}(\hat{Q}(u) - Q(u))$ is asymptotically normal

with mean zero and variance $\sigma_1^{2*}(u) = (S(u))^2 \int_0^u \frac{(-dS(t))}{S(t)S^*(t)}$, where $S^*(t)$ is the probability that a unit is alive and uncensored at time t .

It follows from Lawless (2003) that for $0 < x < \infty$, $\sqrt{n}(\hat{F}_j(x) - F_j(x))$ is asymptotically normal with mean zero and variance $\sigma_2^{2*}(x)$, which can be estimated as given in Section 9.2 of Lawless (2003).

Using functional delta method and Slutsky's theorem (Serfling (1980)), we get that for $0 < u < 1$, $\sqrt{n}(\hat{\Lambda}_j(u) - \Lambda_j(u))$ follows normal distribution with mean as zero and variance $\sigma_j^2(u)$, where $\sigma_j^2(u) = \mathbb{E}\left[\frac{1}{h(n)(1-u)} \int_0^1 K^*(u, p) Z(p) dp\right]^2$, with $Z(p) = \sqrt{n}(\hat{F}_j(\hat{Q}(p)) - F_j(Q(p)))$. This completes the proof.

Affiliation:

P. G. Sankaran
 Department of Statistics
 Cochin University of
 Science and Technology
 Cochin-682022
 E-mail: sankaran.pg@gmail.com

Austrian Journal of Statistics
 published by the Austrian Society of Statistics

<http://www.ajs.or.at/>
<http://www.osg.or.at/>

Volume 48
 January 2019

Submitted: 2017-04-04
 Accepted: 2018-11-20