한국통계학회논문집 제3권 2호
1996년 8월 pp. 119-127

# Confidence Bands for Survival Function Based on Hjort Estimator[1)]

**Byung-Gu Park[2)], Kil-Ho Cho[3)], Woo-Dong Lee[4)] and Young-Joon Cha[5)]**

## Abstract

In this paper, we derive the Hall-Wellner band and the equal precision band for survival function based on Hjort when the data are randomly right censored. The bands are illustrated and compared by applying them to data from a preoperative radiation therapy.

## 1. Introduction

In industrial life testing and medical follow-up studies, censoring is common because of time limits and other restrictions on data collection. It is important in these situations to propose a nonparametric estimator and to obtain confidence bands for the survival function. For the randomly right censored data, Kaplan and Meier (1958) provided a nonparametric estimator of survival function. This Kaplan-Meier estimator(KME) reduces to the usual empirical survival function in the absence of censoring.

Gillespie and Fisher (1979) constructed asymptotic confidence bands for the survival function using the weak convergence of the KME for censored survival data. Hall and Wellner (1980) developed asymptotic confidence bands that reduce to the Kolmogorov band in the absence of censoring. Nair (1984) showed that the usual confidence intervals based on Greenwood's variance formula can be adapted to yield a large-sample confidence band. It is compared with the censored versions of the Kolmogorov band and Rényi band. Recently, in Bayesian set-up, Hjort (1990) proposed an estimator of the cumulative hazard function. This nonparametric Bayes estimator smooths the estimator at censored observations but KME does not. And Hjort estimator has merit to construct the confidence band. Some simulation results shows

1) The Present Studies were Supported (in part) by the Basic Research Institute Program, Ministry of Education, 1995, Project No. BSRI-95-1403.
2) Professor, Department of Statistics, Kyungpook National University, Taegu, Korea.
3) Associate Professor, Department of Statistics, Kyungpook National University, Taegu, Korea.
4) Full-time Lecturer, Department of Statistics, Kyungsan University, Kyungsan, Kyungpook, Korea.
5) Associate Professor, Department of Statistics, Andong National University, Andong, Kyungpook, Korea.

that Hjort estimator has smaller mean squared error(MSE) than that of KME.

In this paper, through the relationship between cumulative hazard function and survival function, we derive an estimator of survival function(H-estimator) and asymptotic confidence bands for the survival function based on Hjort. In Section 2, the equal precision(EP) band and the Hall-Wellner(HW) band based on Hjort are derived. In Section 3, two bands are compared by the real data set and some concluding remarks are given.

## 2. Confidence Bands for Survival Function

Let $T_1, T_2, \cdots, T_n$ be independent and identically distributed (i.i.d.) random times with a common continuous distribution function $F(t)$ on $[0,\infty)$ with $F(0)=0$. Let $S(t)=\Pr(T_i>t)=1-F(t)$ be the survival function. Consider the hazard function of $F(t)$ as $\lambda(t)=f(t)/(1-F(t))$ and the cumulative hazard function $\Lambda(t)$ of $F(t)$ as

$$\begin{aligned}\Lambda(t) &= \int_0^t \frac{1}{S(t)}\,dF(t)\\ &= -\log(S(t)).\end{aligned} \tag{2.1}$$

Let $C_1, C_2, \cdots, C_n$ be i.i.d. random censoring times with a continuous survival function $G(t)$. Assume $T_i$ are independent of $C_i$ for each $i$. Let

$$X_i=(T_i\wedge C_i)=\min(T_i, C_i),\quad i=1,2,\cdots,n,$$

and let the distribution function of $X_i$ be $H(t)$. Then $1-H(t)=S(t)G(t)$. Under the random censorship model, $T$ and $C$ are not directly observable, but we can observe only $(X_1,\delta_1),(X_2,\delta_2),\cdots,(X_n,\delta_n)$, where $\delta_i=1$, if $T_i<C_i$ or $\delta_i=0$, if $T_i\geq C_i$.

Let $X_{(1)}<X_{(2)}<\cdots<X_{(n)}$ be the ordered times of $X_1,X_2,\cdots,X_n$, and define $\delta_{(i)}$ to be the value of $\delta_i$ associated with $X_{(i)}$.

Recently, the Bayes estimator of the cumulative hazard function $\Lambda(t)$ under squared error loss was proposed by Hjort as

$$\hat{\Lambda}(t)=\int_0^t \frac{c(s)d\Lambda_0(s)+dN(s)}{c(s)+Y(s)},$$

where $N(t)=\sum_{i=1}^n I_{\{X_i\leq t,\ \delta_i=1\}}$ and $Y(t)=\sum_{i=1}^n I_{\{X_i\geq t\}}$, and $c(s)$ and $\Lambda_0(s)$ are parameters of prior knowledge of risk and cumulative hazard function at time $s$, respectively.

Now consider the estimator of survival function based on $\widehat{\Lambda}(t)$. By the relationship (2.1), we can propose an estimator $\widehat{S}(t) = \exp(-\widehat{\Lambda}(t))$, and we call $\widehat{S}(t)$ as H-estimator. Then H-estimator is also a comparable estimator to Kaplan-Meier estimator.

In the following, $D[0, T]$ denotes the space of all real valued functions on $[0, T]$, which are right continuous with left hand limit, and equipped with Skorohod metric(see Billingsley (1968)).

Now, we prove the asymptotic properties of $\widehat{S}(t)$ by following Theorem 2.1.

**Theorem 2.1.** Let $T$ be such that $H(T)<1$. Assume that for any $s\in[0, T]$, as $n\to\infty$

(i) $Y(s) \xrightarrow{p} \infty$;

(ii) $\frac{c(s)}{Y(s)} \xrightarrow{p} 0$,

(iii) $\inf_{0\le s\le T} c(s)\neq 0$.

(iv) $c(s)$ is predictable and bounded.

Then,

(1) $\sup_{s\in[0,T]} | \widehat{S}(s)-S(s) | \xrightarrow{p} 0$.

(2) $\sqrt{n}(\widehat{S}-S) \xrightarrow{d} Z^*$ in $D[0, T]$, where $Z^*$ is mean zero Gaussian process and the covariance function is given by

$$Cov(Z^*(s), Z^*(t)) = S(s)S(t)\int_0^{s\wedge t}\frac{dF(u)}{(1-H(u))S(u)},\quad 0\le s,\ t\le T. \qquad (2.2)$$

(3) $\sqrt{n}\left(\frac{\widehat{S}(t)-S(t)}{S(t)/(1-K(t))}\right) \xrightarrow{d} B^0(K(t)),\ 0\le t\le T.$ where $B^0$ is the Brownian Bridge on $(0,1)$, $K(t)=\frac{\sigma^2(t)}{1+\sigma^2(t)}$ and $\sigma^2(t)=\int_0^t\frac{1}{1-H(s)}\,d\Lambda(s)=\int_0^t\frac{dF(s)}{(1-H(s))S(s)}$.

**Proof.** (1) Let $\Lambda^*(t)=\int_0^t\frac{c(s)}{c(s)+Y(s)}\,d\Lambda_0(s)$. Then

$$\sup_{s\in[0,T]} | \widehat{\Lambda}(s)-\Lambda(s) | \le \sup_{s\in[0,T]} | \widehat{\Lambda}(s)-\Lambda^*(s) | + \sup_{s\in[0,T]} | \Lambda^*(s)-\Lambda(s) | .$$

Since

$$\Lambda^*(t)-\Lambda(t)=\int_0^t\frac{\frac{c(s)}{Y(s)}}{\frac{c(s)}{Y(s)}+1}\,(d\Lambda_0(s)-d\Lambda(s)),$$

by assumption (ii), $\sup_{t\in[0,T]} |\Lambda^*(t)-\Lambda(t)| \xrightarrow{p} 0$. Now, it is sufficient to show that $\sup_{t\in[0,T]} |\hat{\Lambda}(t)-\Lambda^*(t)|$ converges to 0 in probability. Since

$$\sup_{t\in[0,T]} |\hat{\Lambda}(t)-\Lambda^*(t)| = \sup_{t\in[0,T]} \left| \int_0^t \frac{1}{c(s)+Y(s)}\, dM(s) \right|$$

where $M$ is a square integrable martingale, it suffices to show that

$$\sup_{t\in[0,T]} \left\{ \int_0^t \frac{1}{c(s)+Y(s)}\, dM(s)\right\}^2 \xrightarrow{p} 0.$$

By Lenglart's inequality,

$$P\left\{\sup_{t\in[0,T]} \left\{ \int_0^t \frac{1}{c(s)+Y(s)}\, dM(s)\right\}^2 \geq \varepsilon\right\} \leq \frac{\eta}{\varepsilon} + P\left\{ \int_0^T \left( \frac{1}{c(s)+Y(s)}\right)^2 Y(s)d\Lambda(s) \geq \eta\right\}.$$

By assumption (i) and (ii), the second term in righthand side converges 0 in probability. Hence the result (1) follows by Taylor expansion of $\exp(-\hat{\Lambda}(t))$.

(2) Let

$$\begin{aligned} Z_n(t) &= \sqrt{n}(\hat{\Lambda}(t)-\Lambda(t)) \\ &= \sqrt{n}(\hat{\Lambda}(t)-\Lambda^*(t)+\Lambda^*(t)-\Lambda(t)) \\ &= U_n(t)+V_n(t) \end{aligned}$$

where $U_n(t)=\sqrt{n}(\hat{\Lambda}(t)-\Lambda^*(t))$ and $V_n(t)=\sqrt{n}(\Lambda^*(t)-\Lambda(t))$. By assumptions (ii), (iv) and $\frac{n}{Y(s)} \xrightarrow{p} (1-H(s))^{-1}$,

$$\sup_{t\in[0,T]} |V_n(t)| = \sup_{t\in[0,T]} \left| \int_0^t \frac{\frac{c(s)}{\sqrt{n}} \frac{n}{Y(s)}}{\frac{c(s)}{Y(s)}+1} (d\Lambda_0(s)-d\Lambda(s)) \right| \xrightarrow{p} 0.$$

By assumption (ii), (iii) and (iv), $U_n$ is a square integrable martingale with covariation process

$$\begin{aligned} \langle U_n, U_n\rangle(t) &= \int_0^t n\left(\frac{Y(s)}{c(s)+Y(s)}\right)\left(\frac{1}{c(s)+Y(s)}\right)d\Lambda(s) \\ &= \int_0^t \left(\frac{c(s)}{Y(s)}+1\right)^{-1}\left(\frac{c(s)}{n}+ Y(s)\, n\right)^{-1} d\Lambda(s) \\ &\xrightarrow{p} \int_0^t (1-H(s))^{-1} d\Lambda(s), \end{aligned}$$

by Glivenko-Cantelli theorem. Define the $\varepsilon$-decomposition of $U_n$ as

$$\overline{U}_n^{\,\varepsilon}(t)=\sqrt{n}\int_0^t \frac{1}{c(s)+Y(s)} I\left(\frac{\sqrt{n}}{c(s)+Y(s)} \geq \varepsilon\right) dM(s)$$

and

$$\underline{U}_n^{\,\varepsilon}(t)=\sqrt{n}\int_0^t \frac{1}{c(s)+Y(s)} I\left(\frac{\sqrt{n}}{c(s)+Y(s)} < \varepsilon\right) dM(s).$$

Then $\overline{U}_n^{\,\varepsilon}$ and $\underline{U}_n^{\,\varepsilon}(t)$ are square integrable martingale and hold the assumptions of martingale central limit theorem(MCLT) in Gill (1980) . Therefore, by MCLT,

$$\sqrt{n}(\hat{\Lambda}-\Lambda) \xrightarrow{d} Z \text{ in } D[0, T],$$

where $Z$ is a mean 0 Gaussian process with covariance function

$$Cov(Z(s), Z(t)) = \int_0^{s\wedge t} \frac{dF(u)}{(1-H(u))S(u)}, \quad 0 \leq s, t \leq T.$$

By using the Taylor expansion of $\exp(-\hat{\Lambda}(t)), \sqrt{n}(\hat{S}-S) \xrightarrow{d} Z^* \text{ in } D[0, T]$ .

(3) In law, for each $0 \leq t \leq T$,

$$Z^*(t)_{\{0\leq t\leq T\}} = \left\{ \frac{B^0(K(t))S(t)}{1-K(t)} \right\}_{\{0\leq t\leq T\}}$$

since they are both mean zero Gaussian process with covariance function (2.2). Hence the result follows.

Using Theorem 2.1, we can construct the $100(1-\alpha)\%$ asymptotic confidence bands for $S(t), t\in[0, T]$ , based on $\hat{S}(t)$.

Let

$$\hat{\sigma}^2(t) = n\int_0^t \frac{dN(s)}{(Y(s)-1)Y(s)} = n \sum_{i:X_i\leq t} \frac{\delta_i}{(n-i+1)(n-i)}$$

and

$$\hat{K}(t) = \frac{\hat{\sigma}^2(t)}{1+\hat{\sigma}^2(t)}.$$

Then, by the result (1) of Theorem 2.1, $\hat{\sigma}^2(t)$ and $\hat{K}(t)$ are consistent estimators of $\sigma^2(t)$ and $K(t)$, respectively. So, a consistent estimator of the asymptotic variance of $\hat{S}(t)$, is

$$\hat{S}^2(t)\,\hat{\sigma}^2(t)/n.$$

By the above fact and Theorem 2.1, the $100(1-\alpha)\%$ pointwise confidence interval for $S(t)$ is constructed by

$$\hat{S}(t) \pm z_{\frac{\alpha}{2}} n^{-\frac{1}{2}} \hat{S}(t)\, \hat{\sigma}(t) \tag{2.3}$$

where $z_{\frac{\alpha}{2}}$ is the $(1-\frac{\alpha}{2})$ *th* standard normal quantile.

Now, we consider the simultaneous confidence bands for $S(t)$, $0 \le t \le T$, where $T$ is such that $H(T) < 1$. In practice we can choose $T < T_n$, the largest uncensored observation to ensure $H(T) < 1$. From Theorem 2.1, we first derive the equal precision confidence band for $S(t)$. For fixed $a$, $b$ such that $0 < a < b < 1$,

$$\begin{aligned} \sup_{a \le K(t) \le b} \left| \frac{\sqrt{n}(\hat{S}(t) - S(t))}{S(t)\sigma(t)} \right| &= \sup_{a \le K(t) \le b} \sqrt{n} \left| \frac{(\hat{S}(t) - S(t))}{S(t)/(1-K(t))} \right| [K(t)(1-K(t))]^{-\frac{1}{2}} \\ &\xrightarrow{d} \sup_{a \le K(t) \le b} | B^0(K(t)) | [K(t)(1-K(t))]^{-\frac{1}{2}} \\ &= \sup_{a \le u \le b} | B^0(u) | [u(1-u)]^{-\frac{1}{2}}, \end{aligned} \tag{2.4}$$

since the weak convergency holds by Theorem 5.1 in Billingsley (1968) and the last equality follows by the fact that $K(t)$ is continuous on $[0, T]$.

Using the equation (2.4), we can obtain the next corollary which deals with the EP band.

**Corollary 2.1** Under the assumptions of Theorem 2.1, as $n \to \infty$, for $0 < a < b < 1$,

$$\sup_{a \le K(t) \le b} \sqrt{n} \left| \frac{(\hat{S}(t) - S(t))}{\hat{S}(t)\, \hat{\sigma}(t)} \right| \xrightarrow{d} \sup_{a \le u \le b} | B^0(u) | [u(1-u)]^{-\frac{1}{2}}$$

**Proof.** Since $\hat{K}(t) \xrightarrow{p} K(t)$ and $\hat{S}(t) \xrightarrow{p} S(t)$, the $100(1-\alpha)\%$ EP band for $S(t)$ is determined as

$$\hat{S}(t) \pm n^{-\frac{1}{2}} C^1_{\frac{\alpha}{2}} \hat{S}(t)\, \hat{\sigma}(t), \tag{2.5}$$

where $C^1_{\frac{\alpha}{2}} = C^1_{\frac{\alpha}{2}}(a, b)$ satisfies

$$P\left\{ \sup_{a \le u \le b} \frac{| B^0(u) |}{\sqrt{[u(1-u)]}} \le C^1_{\frac{\alpha}{2}} \right\} = 1 - \alpha.$$

The quantile point of $\sup_{a \le u \le b} \frac{| B^0(u) |}{\sqrt{[u(1-u)]}}$ can be obtained by following approximation.

$$P\left\{\sup_{a\le u\le b}\frac{|B^0(u)|}{\sqrt{u(1-u)}}\ge d\right\}\approx\frac{1}{2}\,d\phi(d)\log\left\{\frac{b(1-a)}{a(1-b)}\right\}$$

where $\phi(\cdot)$ is the standard normal density function.

From now on, HW band will be obtained. Using Theorem 2.1, for $0<b=K(T)<1$, we can directly obtain

$$\begin{aligned}\sup_{0\le t\le T}\left|\frac{\sqrt{n}(\hat{S}(t)-S(t))}{S(t)/(1-K(t))}\right| &\xrightarrow{d}\sup_{0\le t\le T}|B^0(K(t))|\\ &=\sup_{0\le u\le b}|B^0(u)|\end{aligned}.$$

As the same reason as Corollary 2.1, we can conclude the following.

**Corollary 2.2** Under the assumptions of Theorem 2.1, as $n\to\infty$,

$$\sup_{0\le t\le T}\left|\frac{\sqrt{n}(\hat{S}(t)-S(t))}{\hat{S}(t)/(1-\hat{K}(t))}\right|\xrightarrow{d}\sup_{0\le u\le b}|B^0(u)|.$$

By Corollary 2.2, the $100(1-\alpha)\%$ HW band for $S(t)$ is obtained as

$$\hat{S}(t)\pm n^{-\frac{1}{2}}\,C^2_{\frac{\alpha}{2}}\,\hat{S}(t)/(1-\hat{K}(t)) \tag{2.6}$$

where $C^2_{\frac{\alpha}{2}}$ satisfies

$$P\left\{\sup_{0\le u\le b}|B^0(u)|<C^2_{\frac{\alpha}{2}}\right\}=1-\alpha.$$

Hall and Wellner (1980) gave selected quantile points of $\sup_{0\le u\le b}|B^0(u)|$ for a given values of $b$.

## 3. Illustrative Example and Conclusion

In a recent published retrospective study, Fortier et al (1986) have investigated the effects of varying dosage of preoperative radiation therapy in recent cancer. Their interest was in the likelihood of local recurrence as well as survival time. From 1972 to 1983, 56 patients with proven adenocarcinoma of the rectum but no evidence of distinct metastases were treated preoperatively with radiation daily over a five-week period. Total dosage ranged from 4000 to 6000 rad. Surgery to excise the tumor followed four to seven weeks later. Survival times was measured from the first day of radiation treatment. All times were allocated to the monthly interval in which they fell.

We calculate EP and HW bands for the true survival function with 21 patients who were received the dosage of radiation less than 5000 rad. Figure 1 shows $\hat{S}(t)$ and 95% EP and HW bands for survival function. The largest uncensored observation is 78 ( $T_n=78$) and the second largest uncensored observation is 54. So we calculate the bands at time 54 ( $T=54$).

The EP band was derived with $a=0.05$, $b=\hat{K}(T_n)$and hence the corresponding critical value $C^1_{0.025}=3.0102$. The HW band was computed with $b=0.715$ and using the linear interpolation and Table 1 of Hall and Wellner(1980), we obtain the critical value $C^2_{0.025}=1.3463$. From Figure 1, we can observe that the EP band is narrower than HW band in a region from time 7 to 24. But the HW band is narrower except that region.

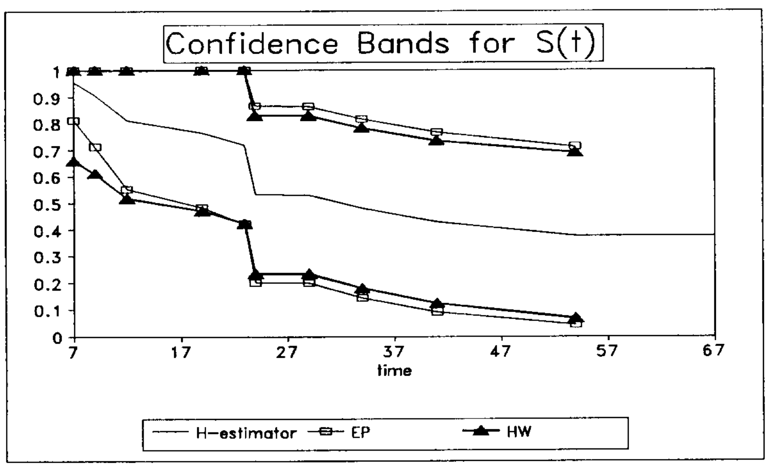


Figure 1. Confidence Bands for $S(t)$

## References

[1] Billingsley, P. (1968). *Convergence of Probability Measures*, John Wiley & Sons, Inc., New York.

[2] Breslow, N. E. and Crowley, J. (1974). A Large Sample Study of the Life Table and Product Limit Estimators under Random Censorship, *The Annals of Statistics*, Vol. 2, 435-453.

[3] Fortier, G. A., Constable, W. C., Meyers, H., and Wanebo, H. J. (1986). Preoperative radiation therapy for rectal cancer. *Arch. Surg.* 121, 1380-1385.

[4] Gill, R. D. (1980). *Censoring and Stochastic Integrals*. Mathematical Centre Tracts 124, Mathematisch Centrum, Amsterdam.

[5] Gillespie, M. J. and Fisher, L. (1979). Confidence Bands for the Kaplan-Meier Survival Curve Estimate, *The Annals of Statistics*, Vol. 7, 920-924.

[6] Hall, W. J. and Wellner, J. A. (1980). Confidence Bands for a Survival Curve From Censored Data, *Biometrika*, Vol. 67, 1, 133-143.

[7] Hjort, N. L. (1990). Nonparametric Bayes Estimators based on Beta Processes in Models for Life History Data, *The Annals of Statistics*, Vol. 3, 1259-1294.

[8] Kaplan, E. L. and Meier, P. (1958). Nonparametric Estimation from Incomplete Observations, *Journal of the American Statistical Association*, Vol. 53, 457-481.

[9] Nair, V. N. (1984). Confidence Bands for Survival Functions with Censored Data: A Comparative Study, *Technometrics*, Vol. 26, 265-275.