The Pure and Applied Mathematics **2** (1995), No 2, pp. 83–89
J. Korea Soc. of Math. Edu. (Series B)

ESTIMATING MOMENTS OF THE SURVIVAL TIME FROM CENSORED OBSERVATIONS

INHA JUNG, KANG SUP LEE

ABSTRACT. A Bayes estimator of the survival distribution function due to Susarla and Van Ryzin(1976) is used to estimate the mth moment of a survival time on the basis of censored observations in a random censorship model. Asymptotic normality of the estimator is proved using the functional version of the delta method.

1. Introduction

Consider two independent sets of nonnegative random variables

$$T_1, ..., T_n \sim iid\ F$$

$$C_1, ..., C_n \sim iid\ G$$

, where F and G are right-tailed continuous distribution functions, i.e., for each $u \geq 0$,

$$F(u) = P(T_1 > u), \quad G(u) = P(C_1 > u).$$

Assume that $F(0) = G(0) = 1$. In a random censorship model, one observes only $(\tilde{T}_1, \triangle_1), ..., (\tilde{T}_n, \triangle_n)$, where for each $i = 1, ..., n$,

$$\triangle_i = 1_{\{T_i \leq C_i\}}, \quad \tilde{T}_i = \min\{T_i, C_i\}. \tag{1.1}$$

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

Suppose that the survival times have a finite mth moment

$$\mu_m = ET_1{}^m = m\int_0^\infty u^{m-1}F(u)du$$

for a positive integer $m \geq 1$.

In this paper we consider the problem of estimating μ_m based only on data (1.1). A natural estimator of μ_m is given by $\hat{\mu}_m = \int_0^\infty mu^{m-1}\hat{F}(u)du$, whenever the random function $\hat{F}$ is a good estimator of F. The product-limit estimator of Kaplan and Meier is among the most well-known estimators of F. Susarla and Van Ryzin(1976) obtained a Bayes estimator of F using Ferguson's Dirichlet process prior which we shall now introduce. Let $\alpha(\cdot)$ be a positive finite measure over $\mathcal{B}(\mathbb{R}^+)$. Define a random function $\hat{F}_\alpha$ on $[0,\infty)$ by

$$\hat{F}_\alpha(t) = \frac{\alpha(t)+N^+(t)}{\alpha(\mathbb{R}^+)+n}\prod_{i=1}^n\left(\frac{\alpha(\tilde{T}_i^-)+N^+(\tilde{T}_i^-)+\lambda_i}{\alpha(\tilde{T}_i^-)+N^+(\tilde{T}_i^-)}\right)^{1_{\{\Delta_i=0,\tilde{T}_i\leq t\}}} \tag{1.2}$$

, where

$$N^+(\cdot) = \#\text{ observations } > \cdot,$$
$$\alpha(u) = \alpha(u,\infty),\ u \geq 0$$
$$\lambda_i = \#\text{ observations at } \tilde{T}_i,\ i = 1,\dots,n.$$

The estimator $\hat{F}_\alpha$ is Bayes estimator of F with respect to the Ferguson's Dirichlet process prior with parameter $\alpha(\cdot)$ and under the squarred error loss

$$L(\hat{F},F) = \int_0^\infty (\hat{F}(u)-F(u))^2dW(u).$$

The product-limit estimator corresponds to the limiting case of $\hat{F}_\alpha$ as $\alpha(\mathbb{R}^+)\to 0$. Among a number of advantages of $\hat{F}_\alpha$ over product-limit estimator is that the Bayes estimator $\hat{F}_\alpha$ is well-defined over the whole interval $[0, \infty)$ while the product-limit estimator is not. We consider the estimators of μ_m of the form

$$\hat{\mu}_m = \int_0^{K_n} m u^{m-1} \hat{F}_\alpha(u) du \tag{1.3}$$

, where $K_n \to \infty$ as $n \to \infty$. Susarla and Van Ryzin(1980) discusses the reason for considering estimators of the form (1.3) instead of $\int_0^\infty m u^{m-1} \hat{F}_\alpha(u) du$ in their estimating μ_1, the mean survival time. We prove in section 2 the asymptotic normality of the estimator $\hat{\mu}_m$ given by (1.3) when the following conditions hold:

$$\begin{aligned} \tau &= \sup\{t : F(t) > 0\} \\ &\leq \sup\{t : G(t) > 0\} \end{aligned} \tag{c1}$$

$$K_n = o(n^{\frac{1}{2m}}) \ as \ n \to \infty \tag{c2}$$

In section 3, we discuss some points in which our result in an improvement of Susarla and Van Ryzin(1980). Throughout the paper we use the following notations.

$$H = FG \tag{1.4}$$

$$\tilde{H}(s) = P(\Delta_1 = 1, \tilde{T}_1 \leq s) = -\int_0^s G dF, \ s \geq 0. \tag{1.5}$$

2. Asymptotic normality

We state the main result about the asymptotic normality.

Theorem 2.1. *Let $\hat{\mu}_m$ be given by (1.2) and (1.3). Assume (c1) and (c2). If*

$$\sigma_m^2 = \int_0^\infty H^{-2}\left(\int_0^\infty m u^{m-1} F(u)du\right)^2 d\tilde{H} < \infty \tag{2.1}$$

then

$$\sqrt{n}(\hat{\mu}_m - \int_0^{K_n} m u^{m-1} F(u)du) \xrightarrow{\mathcal{D}} N(0,\sigma_m^2) \text{ as } n \to \infty. \tag{2.2}$$

For the Proof of Theorem 2.1 we need the following results.

Theorem 2.2. *(Susarla and Van Ryzin(1978)) Let $T < \infty$ and $H(T) > 0$. Let F and G be continuous. Then on $(0,T]$*

$$\sqrt{n}(\hat{F}_\alpha - F) \xrightarrow{\mathcal{D}} Z \tag{2.3}$$

, where Z is a mean 0 Gaussian process with covariance structure given for $s \leq t$ by

$$cov(Z(s),Z(t)) = -F(s)F(t)\int_0^s H^{-1}F^{-1}dF. \tag{2.4}$$

Theorem 2.3. *(Billingsley(1968)) Let $X_{tn} \xrightarrow{\mathcal{D}} X_t$ for all $t < \infty$ as $n \to \infty$ and $X_t \xrightarrow{\mathcal{D}} X$ as $t \to \infty$. Suppose*

$$\lim_{t\to\infty} \overline{\lim_{n}}\, P\{\rho(X_{tn}, Y_n) > \epsilon\} = 0 \tag{2.5}$$

for each $\epsilon > 0$. Then, $Y_n \xrightarrow{\mathcal{D}} X$ as $n \to \infty$.

For each $t > 0$, let

$$\mu_m(t) = \int_0^t m u^{m-1} F(u)du \tag{2.6}$$

$$\hat{\mu}_m(t) = \int_0^t m u^{m-1} \hat{F}_\alpha(u)du.$$

Theorem 2.4. *Let $t > 0$ be fixed and $H(t) > 0$. Then as $n \to \infty$,*

$$\sqrt{n}(\hat{\mu}_m(t) - \mu_m(t)) \xrightarrow{\mathcal{D}} N(0, \sigma_m^2(t)) \tag{2.7}$$

, where

$$\sigma_m^2(t) = \int_0^t H^{-2}(s) \left(\int_s^t m u^{m-1} F(u) du \right)^2 d\tilde{H}(s). \tag{2.8}$$

Proof. Let $\mathcal{F}_t$ be the set of all bounded, real-valued functions on $(0, t]$ which are continuous from the right. Define ϕ on $\mathcal{F}_t$ by

$$\phi(A) = \int_0^t m u^{m-1} A(u) du, \ A \in \mathcal{F}_t.$$

Then ϕ is a bounded linear functional on $\mathcal{F}_t$ which is compactly differentiable with derivatives $d\phi(A) = \phi$ at all $A \in \mathcal{F}_t$. The functional version of the delta method applied to the weak convergence (2.3) yields

$$\sqrt{n}(\phi(\hat{F}_\alpha) - \phi(F)) \xrightarrow{\mathcal{D}} d\phi(F) \cdot Z = \phi(Z) \tag{2.9}$$

It follows from the covariance structure of the Gaussian process Z given by (2.4) that the random variable $\phi(Z) = \int_0^t m u^{m-1} Z(u) du$ has distribution $N(0, \sigma_m^2(t))$. □

Proof of Theorem 2.1. *For an application of Theorem 2.3., let*

$$X_{tn} = \sqrt{n}(\hat{\mu}_m(t) - \mu_m(t)),$$

$$Y_n = \sqrt{n} \left(\int_t^{K_n} m u^{m-1} \hat{F}_\alpha(u) du - \int_t^{K_n} m u^{m-1} F(u) du \right),$$

$$X_t \sim N(0, \sigma_m^2(t)),$$

$$X \sim N(0, \sigma_m^2)$$

, where $\hat{\mu}_m(t)$, $\mu_m(t)$, $\sigma_m^2(t)$, σ_m^2 are given by (2.6), (2.1), (2.8), respectively. Then for each $t > 0$, $X_{tn} \xrightarrow{\mathcal{D}} X_t$ as $n \to \infty$ by Theorem 2.4. Also $X_t \xrightarrow{\mathcal{D}} X$ as $t \to \infty$ since $\sigma_m^2(t) \to \sigma_m^2$ as $t \to \infty$. Therefore, by Theorem 2.3., proof will be completed if we showed that (2.5) holds : for each $\varepsilon > 0$,

$$\lim_{t\to\infty} \limsup_n P\{|\sqrt{n} \int_t^{K_n} mu^{m-1}(\hat{F}_\alpha(u) - F(u))du| > \varepsilon\} = 0. \tag{2.10}$$

Viewing (1.2) and $EN^+(u) = nH(u) = 0$ one sees from the Markov inequality that for each $t > \tau$ and $\varepsilon > 0$,

$$\begin{aligned} & P\{|\sqrt{n} \int_t^{K_n} mu^{m-1}(\hat{F}_\alpha(u) - F(u))du| > \varepsilon\} \qquad (2.11) \\ & \leq P\{\frac{m}{t^{m-1}}\sqrt{n}K_n^m \hat{F}_\alpha(t) > \varepsilon\} \\ & \leq P\{\frac{m}{t^{m-1}}\sqrt{n}K_n^m \frac{\alpha(t) + N^+(t)}{\alpha(\mathbb{R}^+) + n} > \varepsilon\} \\ & \leq \frac{m}{t^{m-1}}\sqrt{n}K_n^m \frac{\alpha(\mathbb{R}^+)}{\alpha(\mathbb{R}^+) + n} \cdot \varepsilon^{-1}. \end{aligned}$$

Since (c2) is assumed, the last term of (2.11) converges to 0 as $n \to \infty$ proving (2.10).□

3. Conclusion

The functional version of the delta method presented in Gill and Johansen(1990) provides us with a very powerful tool in proving convergence in distribution of various statistics. We used this method to prove Theorem 2.1., an obvious extension of Susarla and Van Ryzin(1980) in which the asymptotic normality of $\hat{\mu}_1$, the estimate of the mean survival time, is proved using the method of U-statistics in several steps and under some complicated rate conditions on $\{K_n\}$ combined with F and G. We achived greater simplification both in the proof and rate conditions on $\{K_n\}$ compared to Susarla and Van Ryzin(1980).

REFERENCES

1. Billingsley, P., *Convergence of probability measures.*, John Wiley & Sons, New York (1968).
2. Gill, R. D. and Johansen, S., *A survey of product-integration with a view towards applications in survival analysis.*, Ann. Statist. **18** (1990), 1501-1294.
3. Susarla, V. and Van Ryzin, J., *Nonparametric Bayesian estimation of survival curves from incomplete observations.*, J. Amer. Statist. Assoc. **61** (1976), 897-902.
4. Susarla, V. and Van Ryzin, J., *Large sample theory for a Bayesian nonparametric survival curve estimator based on censored samples.*, Ann. Statist. **6** (1978), 755-768.
5. Susarla, V. and Van Ryzin, J., *Large sample theory for an estimator of the mean survival time from censored samples.*, Ann. Statist. **8** (1980), 1002-1016.

Inha Jung
Dept. of Computer Science and Statistics,
Ajou University,
Suwon, 441-749, KOREA.

Kang Sup Lee
Dept. of Computer Science and Statistics,
Dankook University,
Seoul, 140-714, KOREA.