Confidence bands for the quantile function under random censoring

Einmahl, J.H.J.

Published in:
Journal of Statistical Planning and Inference

Publication date:
1993

Journal of Statistical Planning and Inference 36 (1993) 69-75
North-Holland

Confidence bands for the quantile function under random censoring

John H.J. Einmahl

Eindhoven University of Technology, Eindhoven, The Netherlands

Received 8 October 1990; revised manuscript received 18 May 1992

Abstract: We derive asymptotic confidence bands for the quantile function F^{-1} in the random censorship model. As a key tool we use the Chibisov–O'Reilly theorem for the product-limit process instead of a central limit theorem for the product-limit quantile process. This reduces the assumptions on F to only continuity.

AMS Subject Classification: Primary 62G15, 62G30; secondary 60F05, 60F17.

Key words and phrases: Confidence bands; central limit theorem; quantile function; random censoring; weight function.

1. Introduction and main result

The classical random censorship model deals with i.i.d. random vectors (Z_1, δ_1), $(Z_2, \delta_2), \ldots$, obtained from two independent sequences of i.i.d. random variables $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ sitting on a probability space (Ω, A, P) and taking values in $(0, \infty)$, in the following way:

$$Z_i = X_i \wedge Y_i, \quad \delta_i = 1_{\{X_i \leqslant Y_i\}}. \tag{1.1}$$

Denote the cumulative distribution function of X_i, Y_i and Z_i by F, G and $H = 1 - (1-F)(1-G)$ respectively and let $\theta \in (0,1)$ be an arbitrary but fixed number such that $\theta < F(T_H)$, where $T_H = \sup\{x\colon H(x) < 1\}$. Throughout assume F and G to be continuous on $[0, Q(\theta)]$, where Q is the quantile function pertaining to F, defined by

$$Q(t) = \inf\{x\colon F(x) \geqslant t\}, \quad 0 \leqslant t \leqslant 1. \tag{1.2}$$

The product-limit estimator $\hat{F}_n$ of F is given by

$$\hat{F}_n(x) = 1 - \prod_{Z_{(i)} \leqslant x} (1 - \delta_{(i)}/(n-i+1)), \quad x \geqslant 0,$$

Correspondence to: Dr. J.H.J. Einmahl, Department of Mathematics and Computing Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.

where $0<Z_{(1)}\leqslant\cdots\leqslant Z_{(n)}$ are the order statistics of the Z_i, $1\leqslant i\leqslant n$, and the $\delta_{(i)}$ are the corresponding δ's. The associated product-limit process is defined by

$$\alpha_n(x)=n^{1/2}(\hat{F}_n(x)-F(x)),\quad x\geqslant 0. \tag{1.3}$$

The quantile function Q is naturally estimated by

$$\hat{Q}_n(t)=\inf\{x:\ \hat{F}_n(x)\geqslant t\},\quad 0<t\leqslant 1. \tag{1.4}$$

Set $\hat{Q}_n(t)=0$ for $t\leqslant 0$ and $\hat{Q}_n(t)=\infty$ for $t>1$.

It is the purpose of this note to give asymptotic confidence bands for the quantile function on $(0,\theta)$. Our main tool will be the Chibisov–O'Reilly theorem for the product-limit process α_n (Csörgö, Csörgö and Horváth, 1987; Einmahl and Koning, 1992), instead of a central limit theorem for the product-limit quantile process, and therefore our result can be presented without any additional condition on F.

For this presentation we need some more notation. First we introduce the class of weight functions

$$\mathscr{Q}=\{q:[0,1]\to[0,\infty):\ q \text{ non-decreasing and continuous and } q>0 \text{ on } (0,1]\}.$$

Define the following condition:

(a) There exist $M>0$, $K\geqslant 1$ and $\delta>0$ such that for all $0<x<y\leqslant\delta$, $q(y)/q(x)\leqslant K(y/x)^M$.

Note that this condition is fulfilled if $q/I^{\gamma}\downarrow, \gamma\geqslant 0$, ($I$ the identity function) with $M=\gamma$, $K=1$ and any $0<\delta\leqslant 1$. Write

$$\mathscr{Q}_{\mathrm{a}}=\{q\in\mathscr{Q}:\ q \text{ satisfies condition (a)}\}.$$

Define empirical (sub-)distribution functions by

$$H_n(x)=n^{-1}\sum_{i=1}^{n}1_{[0,x)}(Z_i),\quad x\geqslant 0,$$

$$H_n^1(x)=n^{-1}\sum_{i=1}^{n}\delta_i 1_{[0,x]}(Z_i),\quad x\geqslant 0,$$

and write

$$D_n(x)=\int_0^x(1-H_n(u))^{-2}\,\mathrm{d}H_n^1(u),\quad 0\leqslant x<Z_{(n)}.$$

Now define $h_n=D_n\circ\hat{Q}_n$. Let W be a standard Wiener process and assume $q\in\mathscr{Q}$ is such that $\sup_{0<t\leqslant 1}|W(t)|/q(t)<\infty$ a.s. For $\alpha\in(0,1)$ define $c:=c(\alpha,q)$ by

$$\mathrm{P}\left(\sup_{0<t\leqslant 1}|W(t)|/q(t)\geqslant c\right)=\alpha.$$

Theorem. *Let* $0<\theta<F(T_H)$ *and assume* $q \in \mathcal{Q}_a$ *satisfies*

$$\int_0^1 \frac{1}{t}\exp\left(\frac{-\lambda q^2(t)}{t}\right)\mathrm{d}t<\infty \quad \textit{for all } \lambda>0, \tag{1.5}$$

then for all $\alpha \in (0,1)$

$$P[\hat{Q}_n\{t-cn^{-1/2}[h_n(\theta)]^{1/2}(1-t)q[h_n(t)/h_n(\theta)]\} \leqslant Q(t)$$
$$<\hat{Q}_n\{t+cn^{-1/2}[h_n(\theta)]^{1/2}(1-t)q[h_n(t)/h_n(\theta)]\},\ 0<t\leqslant\theta]$$
$$\to 1-\alpha \quad \text{as } n\to\infty.$$

Corollary. *Let* $0<\theta<F(T_H)$, $0\leqslant\gamma<\frac{1}{2}$ *and* $c:=c(\alpha, I^{\gamma})$. *Then for all* $\alpha\in(0,1)$

$$P[\hat{Q}_n\{t-cn^{-1/2}[h_n(\theta)]^{1/2-\gamma}(1-t)[h_n(t)]^{\gamma}\} \leqslant Q(t)$$
$$<\hat{Q}_n\{t+cn^{-1/2}[h_n(\theta)]^{1/2-\gamma}(1-t)[h_n(t)]^{\gamma}\},\ 0<t\leqslant\theta]$$
$$\to 1-\alpha \quad \textit{as } n\to\infty.$$

Remark 1. The Theorem or Corollary, specialized to the case $q\equiv 1$, can be compared to Theorem 5.3 in Aly, Csörgö and Horváth (1985). Observe that even in this unweighted case their conditions on F are much more restrictive. Note also the difference of the left endpoints of the intervals on which the bands are valid.

Remark 2. It is also possible to obtain, by a similar but more straightforward technique, asymptotic confidence bands for F itself. Then, however, the left endpoint of the interval has to be strictly positive. Therefore Corollary 1 on page 324 in Shorack and Wellner (1986) is incorrect. (Take, e.g., $G\equiv 0$ and note that the bands can not be valid on $(0, Z_{(1)})$ if $\psi(0)=\infty$; $\psi=1/q$). Note that in there, a transformation to Brownian bridge is used instead of one to Brownian motion (Wiener process). For other related results the reader is referred to Hall and Wellner (1980) and Nair (1981).

2. Proof of the main result

We first state the sufficiency part of the Chibisov–O'Reilly theorem for the process $\alpha_n/(1-F)$. Throughout set $0/0:=0$ and define

$$D(x)=\int_0^x (1-H(u))^{-2}\,\mathrm{d}H^1(u), \quad 0\leqslant x\leqslant Q(\theta)$$

with $H^1(u)=P\,(Z_i\leqslant u;\delta_i=1)$.

Fact (see Einmahl and Koning, 1992). Under the conditions of the theorem there exists a sequence $W_1, W_2, \ldots$ of standard Wiener processes such that

$$\sup_{0\leqslant x\leqslant Q(\theta)}\left|\frac{\alpha_n(x)}{1-F(x)}-W_n(D(x))\right|\Big/ q\left(\frac{D(x)}{D(Q(\theta))}\right)\to_{\mathrm{P}} 0 \quad (n\to\infty). \tag{2.1}$$

Writing $h = D \circ Q$, we will also need the following

Lemma. *Let* $0 < \theta < F(T_H)$. *Then for* $q \in \mathcal{Q}_a$ *we have*

$$\sup_{0 \leqslant t \leqslant \theta} \left| \left(\frac{h(\theta)}{h_n(\theta)} \right)^{1/2} \frac{q(h(t)/h(\theta))}{q(h_n(t)/h_n(\theta))} \right| = O_P(1) \quad (n \to \infty) \tag{2.2}$$

and, for any $0 < \eta \leqslant \theta$,

$$\sup_{\eta \leqslant t \leqslant \theta} \left| \left(\frac{h(\theta)}{h_n(\theta)} \right)^{1/2} \frac{q(h(t)/h(\theta))}{q(h_n(t)/h_n(\theta))} - 1 \right| \to_P 0 \quad (n \to \infty). \tag{2.3}$$

Proof of the lemma. Note that there exists a constant c_1 such that for all $0 \leqslant a \leqslant b \leqslant \theta$

$$h(b) - h(a) \leqslant c_1 (b - a) \tag{2.4}$$

and hence that h is continuous. Moreover, it is easily seen that h is strictly increasing. We first prove

$$\sup_{0 \leqslant t \leqslant \theta} |h_n(t) - h(t)| \to_P 0 \quad (n \to \infty). \tag{2.5}$$

By the triangle inequality we have

$$\sup_{0 \leqslant t \leqslant \theta} |h_n(t) - h(t)| \leqslant \sup_{0 \leqslant t \leqslant \theta} |h_n(t) - D(\hat{Q}_n(t))| + \sup_{0 \leqslant t \leqslant \theta} |D(\hat{Q}_n(t)) - h(t)|. \tag{2.6}$$

Note that for any $0 < \varepsilon < 1 - \theta$

$$\limsup_{n \to \infty} \sup_{0 \leqslant t \leqslant \theta} \hat{Q}_n(t) = \limsup_{n \to \infty} \hat{Q}_n(\theta) \leqslant Q(\theta + \varepsilon) \quad \text{a.s.}$$

So

$$\limsup_{n \to \infty} \sup_{0 \leqslant t \leqslant \theta} |h_n(t) - D(\hat{Q}_n(t))| \leqslant \limsup_{n \to \infty} \sup_{0 \leqslant x \leqslant Q(\theta + \varepsilon)} |D_n(x) - D(x)|. \tag{2.7}$$

It is well known that, for $\varepsilon > 0$ small enough, the right side of (2.7) tends to zero almost surely.

It is also well known that $\hat{Q}_n = Q \circ \hat{\Gamma}_n^{-1}$, where $\hat{\Gamma}_n^{-1}$ is the generalized inverse of $\hat{\Gamma}_n$ (cf. (1.2) and (1.4)) and $\hat{\Gamma}_n$ is the product-limit estimator pertaining to the $F(X_i)$ and $F(Y_i)$, $1 \leqslant i \leqslant n$. (See for details, e.g., Beirlant and Einmahl, 1990.) Hence, from (2.4),

$$\sup_{0 \leqslant t \leqslant \theta} |D(\hat{Q}_n(t)) - h(t)| = \sup_{0 \leqslant t \leqslant \theta} |h(\hat{\Gamma}_n^{-1}(t)) - h(t)| \leqslant c_1 \sup_{0 \leqslant t \leqslant \theta} |\hat{\Gamma}_n^{-1}(t) - t|. \tag{2.8}$$

The right side of (2.8) is known to tend to zero in probability (e.g., by using

$$\sup_{0\leqslant s\leqslant\theta} |\hat{\Gamma}_n(s)-s| \to_{\mathrm{P}} 0 \quad (n\to\infty)$$

and a Bahadur-Kiefer type result.) Combination of the results obtained so far yields (2.5), which in turn implies

$$\frac{h_n(\theta)}{h(\theta)}=\frac{h_n(\theta)-h(\theta)}{h(\theta)}+1\to_{\mathrm{P}} 1 \quad (n\to\infty). \tag{2.9}$$

We have for any $0<\eta\leqslant\theta$

$$\begin{aligned}
&\sup_{\eta\leqslant t\leqslant\theta}\left|\left(\frac{h(\theta)}{h_n(\theta)}\right)^{1/2}\frac{q(h(t)/h(\theta))}{q(h_n(t)/h_n(\theta)}-1\right| \\
&\quad\leqslant \sup_{\eta\leqslant t\leqslant\theta}|q[h(t)/h(\theta)]/q[h_n(t)/h_n(\theta)]|\,|\{[h(\theta)/h_n(\theta)]^{1/2}-1\} \\
&\qquad+\sup_{\eta\leqslant t\leqslant\theta}|q[h(t)/h(\theta)]/q[h_n(t)/h_n(\theta)]-1|.
\end{aligned} \tag{2.10}$$

From (2.9) and (2.10) we see that (2.3) is proved once we show

$$\sup_{\eta\leqslant t\leqslant\theta}|q[h(t)/h(\theta)]/q[h_n(t)/h_n(\theta)]-1|\to_{\mathrm{P}} 0 \quad (n\to\infty). \tag{2.11}$$

We have trivially by (2.5) and (2.9)

$$\sup_{\eta\leqslant t\leqslant\theta}\left|\frac{h(t)}{h(\theta)}-\frac{h_n(t)}{h_n(\theta)}\right|\leqslant\sup_{\eta\leqslant t\leqslant\theta}\left|\frac{h(t)-h_n(t)}{h(\theta)}\right|+\left|\frac{h_n(\theta)}{h(\theta)}-1\right|\to_{\mathrm{P}} 0 \quad (n\to\infty).$$

Now uniform continuity of q yields (2.11) and hence (2.3).

Now consider (2.2). Because of (2.9) it is sufficient to prove that

$$\sup_{0\leqslant t\leqslant\theta} q\left(\frac{h(t)}{h(\theta)}\right)\Big/q\left(\frac{h_n(t)}{h_n(\theta)}\right)=\mathrm{O_P}(1) \quad (n\to\infty).$$

Note that condition (a), possibly with another K, holds for $\delta=1$. Hence it is enough to show that

$$\sup_{0\leqslant t\leqslant\theta} h(t)/h_n(t)=\mathrm{O_P}(1) \quad (n\to\infty). \tag{2.12}$$

We have trivially

$$\begin{aligned}
\sup_{0\leqslant t\leqslant\theta}\frac{h(t)}{h_n(t)}&=\sup_{0\leqslant t\leqslant\theta}\frac{\int_0^{Q(t)}[1-H(x)]^{-2}\,\mathrm{d}H^1(x)}{\int_0^{\hat{Q}_n(t)}[1-H_n(x)]^{-2}\,\mathrm{d}H_n^1(x)} \\
&\leqslant\{1-H[Q(\theta)]\}^{-2}\sup_{0\leqslant t\leqslant\theta}H^1[Q(t)]/H_n^1[\hat{Q}_n(t)] \\
&\leqslant\{1-H[Q(\theta)]\}^{-2}\sup_{0\leqslant t\leqslant\theta}t/H_n^1[\hat{Q}_n(t)].
\end{aligned} \tag{2.13}$$

Observe that for small enough $\varepsilon>0$

$$\sup_{0\leqslant x\leqslant Q(\theta+\varepsilon)}\frac{\hat{F}_n(x)}{H_n^1(x)}\leqslant\sup_{0\leqslant x\leqslant Q(\theta+\varepsilon)}\frac{\sum_{Z_{(i)}\leqslant x}\delta_{(i)}/(n-i+1)}{H_n^1(x)}$$

$$=\mathrm{O_P}(1)\quad(n\to\infty).$$

Hence

$$\sup_{0\leqslant t\leqslant\theta}\frac{t}{H_n^1(\hat{Q}_n(t))}=\sup_{0\leqslant t\leqslant\theta}\frac{t}{\hat{F}_n(\hat{Q}_n(t))}\,\frac{\hat{F}_n(\hat{Q}_n(t))}{H_n^1(\hat{Q}_n(t))}$$

$$\leqslant\sup_{0\leqslant t\leqslant\theta}\frac{t}{t}\cdot\sup_{0\leqslant t\leqslant\theta}\frac{\hat{F}_n(\hat{Q}_n(t))}{H_n^1(\hat{Q}_n(t))}=\mathrm{O_P}(1)\quad(n\to\infty).$$

Combining this with (2.13) yields (2.12) and hence (2.2). □

Proof of the theorem. From the Fact above we easily deduce that

$$\sup_{0\leqslant t\leqslant\theta}\left|\frac{n^{1/2}(\hat{F}_n[Q(t)]-t)}{1-t}-W_n(h(t))\right|/q\left(\frac{h(t)}{h(\theta)}\right)\to_{\mathrm{P}}0\quad(n\to\infty).\tag{2.14}$$

Note that

$$\sup_{0\leqslant t\leqslant\theta}\left|\frac{n^{1/2}(\hat{F}_n[Q(t)]-t)}{[h_n(\theta)]^{1/2}(1-t)q[h_n(t)/h_n(\theta)]}-\frac{W_n[h(t)]}{[h(\theta)]^{1/2}q[h(t)/h(\theta)]}\right|$$

$$\leqslant\sup_{0\leqslant t\leqslant\theta}\left(\frac{h(\theta)}{h_n(\theta)}\right)^{1/2}\frac{q[h(t)/h(\theta)]}{q[h_n(t)/h_n(\theta)]}$$

$$\cdot\sup_{0\leqslant t\leqslant\theta}\left|\frac{n^{1/2}(\hat{F}_n[Q(t)]-t)}{1-t}-W_n[h(t)]\right|\Big/\left([h(\theta)]^{1/2}q\left(\frac{h(t)}{h(\theta)}\right)\right)$$

$$+\sup_{0\leqslant t\leqslant\theta}\left|\left(\frac{h(\theta)}{h_n(\theta)}\right)^{1/2}\frac{q[h(t)/h(\theta)]}{q[h_n(t)/h_n(\theta)]}-1\right|$$

$$\cdot W_n[h(t)]\Big/\left([h(\theta)]^{1/2}q\left(\frac{h(t)}{h(\theta)}\right)\right).\tag{2.15}$$

Now it follows from a routine argument based on (2.14), (2.2) and (2.3) that the left side of (2.15) is $\mathrm{o_P}(1)$ $(n\to\infty)$, which in turn yields

$$\sup_{0\leqslant t\leqslant\theta}\frac{|n^{1/2}[\hat{F}_n(Q(t))-t]|}{(h_n(\theta))^{1/2}(1-t)q[h_n(t)/h_n(\theta)]}$$

$$\xrightarrow[d]{}\sup_{0\leqslant t\leqslant\theta}\frac{|W(h(t))|}{(h(\theta))^{1/2}q[h(t)/h(\theta)]}\quad(n\to\infty).\tag{2.16}$$

It is also easily seen that the right side of (2,16) is equal in distribution to

$$\sup_{0\leqslant t\leqslant 1}|W(t)|/q(t).$$

So for the c as in the theorem we have from (2.16) with probability tending to $1-\alpha$ $(n\to\infty)$

$$-c\leqslant\frac{n^{1/2}(\hat{F}_n[Q(t)]-t)}{(h_n(\theta))^{1/2}(1-t)q[h_n(t)/h_n(\theta)]}<c,\quad 0<t\leqslant\theta,$$

or equivalently

$$t - cn^{-1/2}(h_n(\theta))^{1/2}(1-t)q[h_n(t)/h_n(\theta)] \leqslant \hat{F}_n(Q(t))$$
$$< t + cn^{-1/2}(h_n(\theta))^{1/2}(1-t)q[h_n(t)/h_n(\theta)], \quad 0<t\leqslant\theta.$$

Now noting that for all $t\in(0,1)$ and $s\in(0,Q(\theta)]$

$$\{\hat{F}_n(s)<t\} = \{\hat{Q}_n(t)>s\},$$

the proof is complete. □

Proof of the corollary. Take $q=I^{\gamma}$ in the Theorem and note that (1.5) is fulfilled for $0\leqslant\gamma<\frac{1}{2}$. □

References

Aly, E.-E.A.A., M. Csörgö and L. Horváth (1985). Strong approximations of the quantile process of the product-limit estimator. *J. Multivariate Anal.* **16**, 185–210.

Beirlant, J. and J.H.J. Einmahl (1990). Bahadur–Kiefer theorems for the product-limit process. *J. Multivariate Anal.* 35, 276–294.

Csörgö, M., S. Csörgö and L. Horváth (1987). Estimation of total time on test transforms and Lorenz curves under random censorship. *Statistics* **18**, 77–97.

Einmahl, J.H.J. and A.J. Koning (1992). Limit theorems for a general weighted process under random censoring. *Canadian J. Statist.*, **20**, 77–89.

Hall, W.J. and J.A. Wellner (1980). Confidence bands for a survival curve from censored data. *Biometrika* **67**, 133–143.

Nair, V.N. (1981). Plots and tests for goodness of fit with randomly censored data. *Biometrika* **68**, 99–103.

Shorack, G.R. and J.A. Wellner (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.