Ann. Inst. Statist. Math.
Vol. 42, No. 1, 21-36 (1990)

CONFIDENCE BANDS FOR QUANTILE FUNCTION UNDER RANDOM CENSORSHIP

CHANG-JO F. CHUNG[1]*, MIKLÓS CSÖRGŐ[2]** AND LAJOS HORVÁTH[3]***

[1]*Geological Survey of Canada, 601 Booth Street, Ottawa, Ontario, Canada K1A 0E8*
[2]*Department of Mathematics and Statistics, Carleton University, Ottawa, Ontario, Canada K1S 5B6*
[3]*Department of Mathematics, The University of Utah, Salt Lake City, UT 84112, U.S.A.*

(Received October 31, 1987; revised February 20, 1989)

Abstract. Some new confidence bands are established for the quantile function from randomly censored data. The method does not require estimation of the density function. As an application, we construct bands for the quantile function of the length of fractures in the granitic plutons near Lac du Bonnet, Manitoba, where an Underground Research Laboratory is being built for the nuclear waste disposal program in Canada.

Key words and phrases: Censored data, quantile function, confidence band, Wiener process, granitic pluton.

1. Introduction

Many statistical experiments result in incomplete samples, even under well-controlled conditions. For example, clinical data for surviving most types of disease are usually censored by other competing risks to life which result in death. Those who survive a disease up to a certain time usually constitute a censored sample of those who were observed to also have had the same illness, but died of other reasons. Hence, if we are to infer something about the stochastic nature of the survival time of clinical trials of interest, frequently we can only base our work on the leftover (censored) sample of those who are still alive, the so-called uncensored observations. The operating time of machines, and that of all kinds of equipment, are subject to the same types of censorship. In particular, the fracture data in Section 3, which are of importance to Canada's nuclear waste disposal program, are also based on incomplete observations. Many fractures in rock masses are covered by vegetation and soil and cannot be completely

*Geological Survey of Canada Contribution Number 10386.
**Research partially supported by an NSERC Canada Grant.
***Research done while at Carleton University, supported by NSERC Canada and EMR Canada Grants of M. Csörgő.

measured. These types of situations are usually described by the random censorship model.

Let $X_1^0, \ldots, X_n^0$ be independent identically distributed random variables with continuous distribution function F, and let $Y_1, \ldots, Y_n$ be independent identically distributed random variables with continuous distribution function H. Suppose that the two sequences $\{X_i^0\}$ and $\{Y_i\}$ are independent. We can observe only the pairs (X_i, δ_i), $i = 1, 2, \ldots, n$, where X_i is the minimum of X_i^0 and Y_i, and δ_i is the indicator function of the event $X_i^0 \leq Y_i$; i.e.,

$$X_i = X_i^0 \wedge Y_i \quad \text{and} \quad \delta_i = \begin{cases} 1 & \text{if } X_i^0 \leq Y_i, \\ 0 & \text{if } X_i^0 > Y_i, \end{cases} \tag{1.1}$$

$i = 1, 2, \ldots, n$. Hence the $\{X_i\}$ are independent identically distributed random variables with distribution function G given by

$$1 - G(t) = (1 - F(t))(1 - H(t)), \qquad -\infty < t < \infty, \tag{1.2}$$

and the sub-distribution function of the uncensored observations is

$$\tilde{F}(t) = P\{X_i < t \text{ and } \delta_i = 1\} = \int_{-\infty}^{t} (1 - H(s))\,dF(s).$$

This model is called the random censorship model from the right.

Kaplan and Meier (1958) introduced the product-limit (PL) estimator F_n^0 for F, defined by

$$1 - F_n^0(t) = \begin{cases} \prod_{[1 \leq i \leq n: X_i \leq t]} \left[\dfrac{n - R_i}{n - R_i + 1} \right]^{\delta_i} & \text{if } t < X_{n:n}, \\ 0 & \text{if } t \geq X_{n:n}, \end{cases}$$

where $X_{n:n} = \max(X_1, \ldots, X_n)$ and R_i is the rank of $(X_i, 1 - \delta_i)$ in the lexicographic ordering of $\{(X_i, 1 - \delta_i)\}$, $i = 1, 2, \ldots, n$.

The weak convergence of the PL-process

$$\beta_n(t) = n^{1/2}(F_n^0(t) - F(t)), \qquad -\infty < t < \infty,$$

to a Gaussian process was proved by Breslow and Crowley (1974) and Aalen (1976). Burke *et al.* (1981, 1988) established strong approximations of β_n in terms of appropriate Gaussian processes. Using time transformations of the approximating processes, S. Csörgő and Horváth (1986) developed a family of confidence bands for F. For further results on the

PL-process, we refer to Gill (1980, 1983) and Lo and Singh (1986).

A parallel problem to estimating F is that of estimating the quantile function

$$Q(y) = \inf\{t\colon F(t) \geq y\}, \qquad 0 < y < 1 .$$

Computing the quantile function at specific values of y, we get frequently used parameters of a distribution such as, for example, the median, the quartiles and the interquartile distance. From a statistical point of view, the distribution function F and its quantile function Q represent two natural, complementary views of a distribution. In this paper we are interested in the quantile function, because we have to estimate the length of those fractures which occur at a given proportion.

A natural estimator of Q is the PL-quantile function

$$Q_n(y) = \inf\{t\colon F_n^0(t) \geq y\}, \qquad 0 < y < 1 .$$

A more direct way of describing Q_n can be given in terms of those observations $U_n^{(1)},\ldots, U_n^{(\nu_n)}$, arranged in increasing order, whose corresponding indicator variables δ equal 1. This means that $U_n^{(1)},\ldots, U_n^{(\nu_n)}$ are the uncensored observations in increasing order.

$$Q_n(y) = \begin{cases} U_n^{(i)} & \text{if} \quad F_n^0(U_n^{(i)}-) < y \leq F_n^0(U_n^{(i)}), \quad i = 1,\ldots,\nu_n , \\ X_{n:n} & \text{if} \quad F_n^0(U_n^{(\nu_n)}) < y < 1 , \end{cases}$$

i.e., the values of Q_n are the uncensored observations and $X_{n:n}$ ($X_{n:n} = U_n^{(\nu_n)}$ if $X_{n:n}$ is uncensored). The PL-quantile process is defined as

$$\rho_n(y) = \sqrt{n}\, f(Q(y))(Q(y) - Q_n(y)), \qquad 0 < y < 1 ,$$

where f is the derivative of F.

The weak convergence of ρ_n was proved by Sander (1975). Aly *et al.* (1985) proved strong approximation theorems for ρ_n. For further use in the sequel, we quote a result from Aly *et al.* (1985). Let

$$d(t) = \int_{-\infty}^{t} (1 - G(s))^{-2} d\tilde{F}(s) .$$

THEOREM 1.1. *Let* $t_F = \sup\{x\colon F(x) = 0\}$ *and* $T_F = \inf\{x\colon F(x) = 1\}$. *We assume*:

(i) *F is twice differentiable on* (t_F, T_F),

(ii) $f(t) = (d/dt)F(t) > 0$ *on* (t_F, T_F).

Then there exists a sequence of Wiener processes $\{W_n(t)\colon 0 \leq t < \infty\}$ *such that*

$$\sup_{\gamma \le t \le p_0} |\rho_n(t) - (1-t)W_n(d(Q(t)))| = o_P(1)\,, \tag{1.3}$$

for all $0 < \gamma \le p_0$ *if* $G(Q(p_0)) < 1$.

If, in addition to (i) *and* (ii), *for some* $r > 0$ *and* $p^* \in (0, 1]$, *we have*

$$\sup_{0 \le t \le p^*} t\,\frac{|f'(Q(t))|}{f(Q(t))^2} \le r\,, \tag{1.4}$$

then there is a positive constant C such that, with $\gamma(n) = Cn^{-1}$ loglog *n*,

$$\sup_{\gamma(n) \le t \le p_0} |\rho_n(t) - (1-t)W_n(d(Q(t)))| \overset{\text{a.s.}}{=} O[n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4}]\,, \tag{1.5}$$

if $p_0 < p^*$ *and* $G(Q(p_0)) < 1$.

The first part of Theorem 1.1 states a weak convergence of ρ_n to a transformed Wiener process. The second part gives information on the rate of convergence in this invariance principle for ρ_n.

Though the quantile function is the inverse of the distribution function and the PL-quantile function is the inverse of the product-limit estimator, one cannot, in general, obtain an asymptotically correct confidence band for Q by simply inverting the corresponding one for F. Indeed, unless F has finite support, we have

$$P\left\{\lim_{n\to\infty} \sup_{0<y<1} |Q_n(y) - Q(y)| = \infty\right\} = 1$$

in the uncensored and censored cases alike. While it is true that the PL-process $\beta_n(t)$ converges weakly to a mean zero time transformed Wiener process over $t \in (-\infty, T]$ if $T < T_G = T_F \wedge T_H$, where T_H is defined like T_F in Theorem 1.1, the covariance function of this mean zero Gaussian process is equal to $(1 - F(s))(1 - F(t))d(s \wedge t)$ with $d(\cdot)$ as in Theorem 1.1. Thus, already for the PL-process β_n, we have a completely non-distribution-free weak convergence. Hence, in order to obtain asymptotically valid, and hopefully distribution-free, confidence bands for F, the PL-process β_n will have to be transformed so that the limiting Gaussian process should become free of F and H. The Efron (1967) transform of β_n converges weakly, as $n \to \infty$, to a standard Wiener process. This leads to confidence bands for F which still depend on the usually unknown quantity $d(T)$. Gillespie and Fisher (1979) generalized this approach of obtaining $d(T)$-dependent confidence bands for F, while Hall and Wellner (1980) transformed the PL-process β_n differently and obtained confidence bands, which again depend on $d(T)$. To overcome the latter common difficulty of the mentioned bands so far, Nair (1981, 1982 and 1984), following up an earlier proposal of Aalen (1976), considered the process $\beta_n(t)/(1 - F(t))$ between the boundaries $\pm \lambda d_n^{1/2}(T)$, where λ is a positive constant and $d_n(\cdot)$ is an appropriate empirical version of $d(\cdot)$ (cf. Section 2 for the definition of d_n in this paper). The resulting band for F is asymptotically distribution and censor-free. The already mentioned family of confidence bands for F developed by S. Csörgő and Horváth (1986) incorporates the above bands and their relationship to one another within a comprehensive theory of bands. Their thorough analysis yields narrower bands and modifications which are asymptotically distribution and censor-free. A look at the form of these bands (cf. pp. 133–134 of S. Csörgő and Horváth (1986)) convinces one immediately that inverting these type of bands, so that this inversion should result in asymptotically correct bands for Q, is not at all immediate.

The approach taken by us here is similar to the above-sketched procedure for constructing confidence bands for F in terms of the weak convergence of β_n. Accordingly, it is based on Theorem 1.1, which provides us with an appropriate weak convergence setting for the PL-quantile process ρ_n. Adapting, then, the above-mentioned appropriate transformations of β_n for now transforming the PL-quantile process ρ_n, we obtain our asymptotically correct confidence bands for Q, as summarized by Theorems 2.1 and 2.2. The fact that the density-quantile function $f(Q(\cdot))$ has become a part of the PL-quantile process $\rho_n(\cdot)$ for the sake of an appropriate weak convergence of the latter, makes this just-mentioned procedure of transforming the process ρ_n somewhat more cumbersome, though possible under our conditions. Of course, if we estimate $f(Q)$ in the definition of ρ_n, then Theorem 1.1 can be used to construct an asymptotically correct confidence band for Q. This method, however, requires an extra density estimation with all its inherent problems which we would like to avoid (cf. Theorem 5.2 in Aly *et al.* (1985), for an example). Hence the inversion approach taken here to construct asymptotically correct confidence bands for Q.

In the uncensored case, Alexander (1980) proposed confidence bands for the quantile function by simply inverting the corresponding one for F. This inversion, however, cannot work in general, without any restrictions, as recognized also by Alexander (1982). M. Csörgő and Révész (1984) gave sufficient conditions for the validity of this procedure. Aly *et al.* (1985) constructed similar confidence bands for Q under the random censorship model. Here in Section 2, we establish a larger family of bands for Q which contain those of Aly *et al.* (1985). The properties of the obtained confidence bands will also be discussed there. The methodology developed is applied to estimate the quantile function of the length of fractures in the granitic plutons near Lac du Bonnet, Manitoba, Canada.

2. Confidence bands

Let

$$G_n(t) := n^{-1}\#\{1 \le i \le n \colon X_i \le t\}$$

and

$$\tilde{F}_n(t) := n^{-1}\#\{1 \le i \le n \colon X_i \le t \text{ and } \delta_i = 1\}$$

be the empirical distribution functions corresponding to G and $\tilde{F}$. An estimator of d then is

$$d_n(t) = \int_{-\infty}^{t} (1 - G_n(s))^{-2} d\tilde{F}_n(s) ,$$

for which Burke *et al.* (1981) proved

$$\sup_{-\infty < t \le T} |d_n(t) - d(t)| \overset{\text{a.s.}}{=} O(n^{-1/2}(\log n)^{1/2}) , \tag{2.1}$$

provided $F(T) < 1$. Let

$$b_n(t) = \frac{1-t}{\sqrt{n}}\left[c_1 \frac{d_n(Q_n(t))}{(d_n(Q_n(p_0)))^{1/2}} + c_2(d_n(Q_n(p_0)))^{1/2} \right], \tag{2.2}$$

and let $\{W(t) \colon 0 \le t < \infty\}$ stand for a Wiener process.

THEOREM 2.1. *If* (i), (ii) *of Theorem* 1.1 *hold and* $0 < \varepsilon \le p_0$, $G(Q(p_0)) < 1$, *then*

$$\lim_{n \to \infty} P\{Q_n(t - b_n(t)) \le Q(t) \le Q_n(t + b_n(t)) \colon \varepsilon \le t \le p_0\}$$
$$= P\left\{ |W(s)| \le c_1 s + c_2 \colon \frac{d(Q(\varepsilon))}{d(Q(p_0))} \le s \le 1 \right\}$$

for all $c_2 > 0$ *and* $c_1 + c_2 \ge 0$.

An obvious question is whether ε of this theorem could be replaced by zero. It is shown in M. Csörgő and Révész (1984) that this cannot be done, in general, in the uncensored case (cf. also pp. 35–36 in M. Csörgő (1983)), and their counter-example also works under random censorship. Here too, ε can be replaced by a sequence of constants converging to zero slowly enough.

THEOREM 2.2. *If* (i), (ii), (1.4) *hold and* $p_0 < p^*$, $G(Q(p_0)) < 1$, $n^{1/2}\varepsilon(n) \to \infty$, $\varepsilon(n) \to 0$ $(n \to \infty)$, *then*

$$\lim_{n\to\infty} P\{Q_n(t - b_n(t)) \le Q(t) \le Q_n(t + b_n(t)) \colon \varepsilon(n) \le t \le p_0\}$$
$$= P\{|W(s)| \le c_1 s + c_2 \colon 0 \le s \le 1\} ,$$

for all $c_2 > 0$ *and* $c_1 + c_2 \ge 0$.

The proofs of these two theorems are similar. The proof of Theorem 2.2 is somewhat more involved, and this is the one we detail here.

PROOF OF THEOREM 2.2. We outline this proof only for the one-sided case of $Q(t) \le Q_n(t + b_n(t))$. The complete proof goes along the same lines. By the mean value theorem

$$P\{Q(t) \le Q_n(t + b_n(t)) \colon \varepsilon(n) \le t \le p_0\}$$
$$= P\left\{ \rho_n(t) \le \rho_n(t) - \frac{f(Q(t))}{f(Q(t + b_n(t)))}\rho_n(t + b_n(t)) \right.$$
$$\left. + \sqrt{n}\, f(Q(t))(Q(t + b_n(t)) - Q(t)) \colon \varepsilon(n) \le t \le p_0 \right\}$$
$$= P\left\{ \rho_n(t) \le \left[\frac{f(Q(t))}{f(Q(t + b_n(t)))} - 1 \right] \rho_n(t + b_n(t)) \right.$$
$$+ \rho_n(t) - \rho_n(t + b_n(t)) + \sqrt{n}\, b_n(t)$$
$$\left. + \sqrt{n}\, b_n(t)\left[\frac{f(Q(t))}{f(Q(\Theta_n(t)))} - 1 \right] \colon \varepsilon(n) \le t \le p_0 \right\} ,$$

where

$$t \le \Theta_n(t) \le t + b_n(t) . \tag{2.3}$$

Next we show

$$\sup_{\varepsilon(n) \le t \le p_0} |\rho_n(t + b_n(t)) - \rho_n(t)| = o_P(1) , \tag{2.4}$$

$$\sup_{\varepsilon(n) \le t \le p_0} \left| \frac{f(Q(t))}{f(Q(t + b_n(t)))} - 1 \right| |\rho_n(t + b_n(t))| = o_P(1) \tag{2.5}$$

and

(2.6) $$\sup_{\varepsilon(n)\le t\le p_0} |\sqrt{n}\, b_n(t)| \left| \frac{f(Q(t))}{f(Q(\Theta_n(t)))} - 1 \right| = o_P(1)\,.$$

Using Step 3 in Aly *et al.* (1985) and (2.1), we get

(2.7) $$\sup_{\varepsilon(n)\le t\le p_0} |d_n(Q_n(t)) - d(Q(t))| = o_P(1)$$

and hence also

(2.8) $$\sup_{\varepsilon(n)\le t\le p_0} \sqrt{n}\, |b_n(t)| = O_P(n^{-1/2})\,.$$

Since for each $n \ge 1$, we have

$$\{(1-t)W_n(d(Q(t))) \colon 0 \le t \le p_0\} \overset{D}{=} \{(1-t)W(d(Q(t))) \colon 0 \le t \le p_0\}\,,$$

and the indicated Gaussian process has almost surely continuous sample paths, by (1.5) and (2.8), we obtain (2.4). It follows from (1.4) that

$$\left| \frac{d}{dt} \log f(Q(t)) \right| = \frac{|f'(Q(t))|}{[f(Q(t))]^2} \le rt^{-1}\,,$$

hence, we have

(2.9) $$\left[\frac{t_1 \wedge t_2}{t_1 \vee t_2} \right]^r \le \frac{f(Q(t_1))}{f(Q(t_2))} \le \left[\frac{t_1 \vee t_2}{t_1 \wedge t_2} \right]^r,$$

where $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$. By (2.8) and $n^{1/2}\varepsilon(n) \to \infty$ $(n \to \infty)$, we get

$$\sup_{\varepsilon(n)\le t\le p_0} \sqrt{n}\, |b_n(t)/t| = o_P(1)\,,$$

and therefore by (2.9)

$$\sup_{\varepsilon(n)\le t\le p_0} \left| \frac{f(Q(t))}{f(Q(t+b_n(t)))} - 1 \right|$$
$$\le \sup_{\varepsilon(n)\le t\le p_0} \left| \left[\frac{t+b_n(t)}{t} \right]^r - 1 \right| + \sup_{\varepsilon(n)\le t\le p_0} \left| \left[\frac{t}{t+b_n(t)} \right]^r - 1 \right| = o_P(1)\,.$$

Hence (1.5) and (2.8) now imply (2.5). Proof of (2.6) is the same. Combining (2.4)–(2.7) and (1.5), we have

$$P\{Q(t) \le Q_n(t + b_n(t)) \colon \varepsilon(n) \le t \le p_0\}$$

$$= P\left\{ (1-t)W(d(Q(t))) \le (1-t)\left[c_1 \frac{d(Q(t))}{(d(Q(p_0)))^{1/2}} + c_2 (d(Q(p_0)))^{1/2} \right] \colon \varepsilon(n) \le t \le p_0 \right\}$$

$$+ o(1) .$$

Using now the scale transformation of Observation of p. 29 in M. Csörgő and Révész (1981), we obtain

$$P\left\{ W(d(Q(t))) \le c_1 \frac{d(Q(t))}{(d(Q(p_0)))^{1/2}} + c_2 (d(Q(p_0)))^{1/2} \colon \varepsilon(n) \le t \le p_0 \right\}$$

$$= P\{W(s) \le c_1 s + c_2 \colon d(Q(\varepsilon(n)))/d(Q(p_0)) \le s \le 1\} .$$

Since $d(Q(\varepsilon(n))) \to 0$ $(n \to \infty)$, we have

$$\sup_{0 \le s \le d(Q(\varepsilon(n)))/d(Q(p_0))} |W(s)| = o_P(1) ,$$

and this also completes the proof of Theorem 2.2.

Theorem 5.3 in Aly *et al.* (1985) is a special case of Theorem 2.2. Namely, $c_1 = 0$ gives the former result. Our confidence bands for Q are of the form $[Q_n(t - b_n(t)), Q_n(t + b_n(t))]$, where the random function $b_n(t)$ depends on two parameters c_1 and c_2. These can be determined from the distribution of the weighted Wiener process. Indeed, in order to apply Theorems 2.1 and 2.2, we need to know how to compute the following two probabilities:

(2.10) $$\psi_a(c_1, c_2) = P\{|W(t)| \le c_1 t + c_2 \colon a \le t \le 1\}, \qquad 0 \le a \le 1 ,$$

and

(2.11) $$\psi(c_1, c_2) = \psi_0(c_1, c_2) = P\{|W(t)| \le c_1 t + c_2 \colon 0 \le t \le 1\} .$$

Anderson (1960) gave a formula for

(2.12) $$\Gamma_b(u', v', u, v) = P\{u' + v't \le W(t) \le u + vt \colon 0 \le t \le b\} .$$

Using now the fact that the Wiener process has stationary independent increments, we obtain

$$(2.13)\qquad \begin{aligned}&\psi_a(c_1, c_2)\\ &= P\{-c_2 - c_1(t+a) - W(a) \le W(t+a) - W(a)\\ &\qquad \le c_2 + c_1(t+a) - W(a) \colon 0 \le t \le 1-a\}\\ &= \frac{1}{\sqrt{2\pi a}} \int_{-q}^{q} \Gamma_{1-a}(-q-h, -c_1, q-h, c_1) e^{-h^2/2a} dh,\end{aligned}$$

where $q = c_1 a + c_2$.

For the case of $a = 0$, we have closed forms. Using (2.12), Gillespie and Fisher (1979) arrived at

$$(2.14)\qquad \begin{aligned}\psi(c_1, c_2) &= 2\Phi(c_1 + c_2) - 1\\ &+ 2\sum_{k=1}^{\infty}(-1)^k e^{-2k^2 c_1 c_2}\\ &\cdot[\Phi(2kc_2 + c_1 + c_2) - \Phi(2kc_2 - c_1 - c_2)].\end{aligned}$$

Here and throughout, Φ is the standard normal distribution function. Taking now $c_1 = 0$, we get

$$(2.15)\qquad \psi(0, c_2) = 2\Phi(c_2) - 1 + 2\sum_{k=1}^{\infty}(-1)^k[\Phi((2k+1)c_2) - \Phi((2k-1)c_2)].$$

Feller (1966) gave an equivalent form for the latter as follows:

$$(2.16)\qquad \psi(0, c_2) = \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k}{2k+1}\exp\left[\frac{-(2k+1)^2\pi^2 b}{8c_2^2}\right].$$

Since

$$\left\{\frac{W(t)}{1+t} : 0 \le t < \infty\right\} \stackrel{D}{=} \left\{B\left(\frac{t}{1+t}\right) : 0 \le t < \infty\right\},$$

for any Brownian bridge $\{B(s)\colon 0 \le s \le 1\}$, from (2.14) with $c_1 = c_2 = c$, we obtain

$$(2.17)\qquad \begin{aligned}\psi(c, c) &= P\left\{\sup_{0 \le t \le 1/2} |B(t)| \le c\right\}\\ &= 2\Phi(2c) - 1\\ &+ 2\sum_{k=1}^{\infty}(-1)^k e^{-2k^2c^2}[\Phi(2(k+1)c) - \Phi(2(k-1)c)].\end{aligned}$$

S. Csörgő and Horváth (1983) tabulated the distribution function $\psi(0, c_2)$ using the formula in (2.16). Extensive tabulation of $\psi(c, c)$ are given by Koziol and Byar (1975), and those of some selected percentile points of $\psi(c, c)$ can also be found in Hall and Wellner (1980). A more detailed discussion of formulae like those of (2.13)–(2.17) can be found in Chung (1986). Chung (1987) provided a computer package which computes the probabilities in (2.13) and (2.14).

Since the limiting distribution in Theorem 2.1 depends on the unknown $a = d(Q(\varepsilon))/d(Q(p_0))$, it cannot be used directly for constructing confidence bands for Q. However, on estimating a by $\hat{a}(n) = d_n(Q_n(\varepsilon))/d_n(Q_n(p_0))$, we have $\hat{a}(n) \to a$ a.s. $(n \to \infty)$. If we now fix c_1, there is only one $\hat{c}_2(n)$ so that

$$\psi_{\hat{a}(n)}(c_1, \hat{c}_2(n)) = 1 - \alpha .$$

Using the monotonicity of ψ, we obtain immediately that $\hat{c}_2(n) \to c_2$ a.s. $(n \to \infty)$ and c_2 satisfies

$$\psi_a(c_1, c_2) = 1 - \alpha .$$

We define $\hat{b}_n(t)$ so that we replace c_2 by $\hat{c}_2(n)$ in the definition of $b_n(t)$ in (2.2). Then

$$\lim_{n \to \infty} P\{Q_n(t - \hat{b}_n(t)) \le Q(t) \le Q_n(t + \hat{b}_n(t)) \colon \delta \le t \le p_0\} = 1 - \alpha . \tag{2.18}$$

If we wished to use Theorem 2.1 with $c_1 = c_2$, then there is only one $\hat{c}(n)$ so that

$$\psi_{\hat{a}(n)}(\hat{c}(n), \hat{c}(n)) = 1 - \alpha .$$

Then defining $\hat{b}_n(t)$ via replacing $c_1 = c_2$ by $\hat{c}(n)$, we have (2.18) again.

The width of the confidence bands of Theorems 2.1 and 2.2 is

$$\Delta_n(t) = Q_n(t + b_n(t)) - Q_n(t - b_n(t))$$

which can be easily seen to converge as

$$\lim_{n \to \infty} n^{1/2} \Delta_n(t) = 2 \frac{1 - t}{f(Q(t))} \left[c_1 \frac{d(Q(t))}{(d(Q(p_0)))^{1/2}} + c_2 (d(Q(p_0)))^{1/2} \right] \quad \text{a.s.}$$

for each $t \in (0, p_0]$. Consequently, in general, these band widths depend on t as well as on F and H, also in the limit for whatever choice of c_1 and c_2. Hence it is very difficult to give general advice on how to choose c_1 and c_2.

If the censoring is very heavy (we have only a few large uncensored observations), then $d(Q(p_0))$, the variance at the endpoint of the confidence band, is very large. In this case we prefer to have c_1 large and c_2 small. If $c_1 = 0$, then the confidence band is usually very wide. If the censoring is not heavy, then we can try $c_1 = c_2$ or take c_1 a little bit larger than c_2. We may always select a few c_1 and c_2 values, and then choose the best available confidence bands.

3. Applications

The nuclear waste disposal program in Canada involves emplacement of a vault containing waste in a stable geological formation such as the granitic plutons in the Canadian Shield. The rock mass surrounding the vault containing the waste acts as a natural barrier between the waste and the biosphere. It is assumed, however, that the ground water system contaminated with radionuclides from the waste eventually migrates through the rock mass and reaches the biosphere. The fracture system in the host rock forms the main migration pathways.

The fractures are approximately linear planes in nature, and we can only see linear lines on the surface of the rock mass. A preliminary analysis of the length of these fractures shows that they are spatially uncorrelated. This suggests that measurements may be viewed as independent observations. What we observe are linear lines on the surface of the rock mass, many of them partially covered by soil and vegetation. Coverage by soil and vegetation is not related to the nature of the fracture system. Thus we may assume that we are observing independent measurements which are randomly censored. Hence, studying the distribution and the quantile functions of the length of the partially covered fracture lines on the surface under the random censorship model is a first step toward understanding the fracture system in the host rock.

The granitic pluton near Lac du Bonnet, Manitoba, where the underground Research Laboratory is being built for the nuclear waste disposal program in Canada, is selected for our investigations here. Stone *et al.* (1984) mapped the fractures on the surface in the granite. We chose a small area from the map by Stone *et al.* (1984) for this study.

About 67% of the granite pluton is covered by soil and vegetation, and thus we can only observe the fracture lines on the exposed areas (about 33%) of the granite. Consequently, most of the fractures are extended beyond the exposed areas. We have observed 1567 fractures. Of these both ends are shown only for 256 fractures in the exposed areas whose lengths can be completely measured. The rest, namely 1311 fracture lines, are censored. The lengths of the longest uncensored and censored fracture lines are 0.991 and 2.361 (1 unit = 2.54 m), respectively. The length of the shortest fracture, which happens to be uncensored, is 0.001.

The product limit estimator F^0_{1567} and the corresponding PL-quantile function Q_{1567}, based on the above discussed observations, are shown in Fig. 1. Due to the heavy censoring, $F^0_{1567}(t)$ cannot be used for estimating $F(t)$ when $t > 0.9910$, since there are only a few uncensored observations which are larger than 0.9910. Similarly, and for the same reason, $Q_{1567}(y)$ cannot be used to estimate $Q(y)$ if $y > 0.3403$.

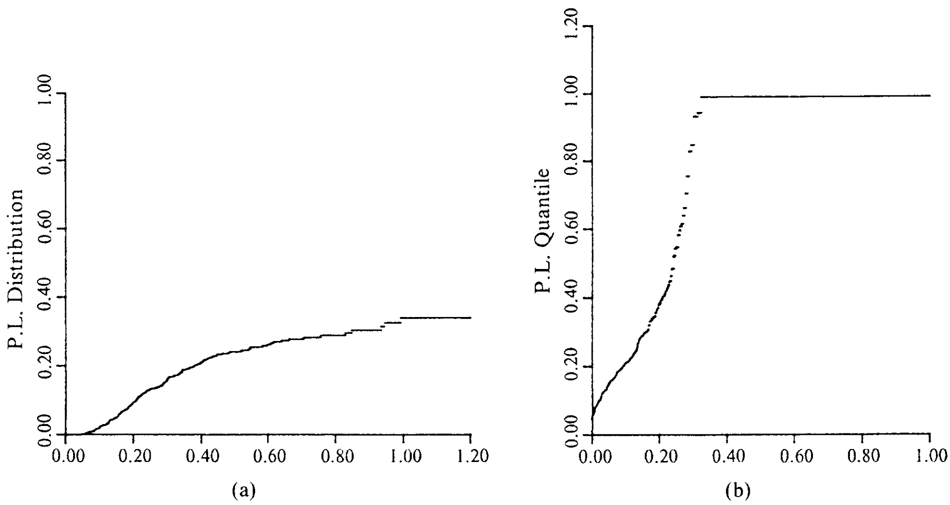

Fig. 1. (a) PL-estimator F^0_{1567} for F; (b) PL-quantile function Q_{1567} for Q. F and Q are the distribution function and quantile functions of the length of fractures in a part of Stone *et al.* (1984) map of a granitic pluton near Lac du Bonnet, Manitoba, Canada.

As an application of Theorem 2.1, we construct a 95% confidence band for the quantile function on the fixed interval $[0.1, 0.25]$. We computed $\hat{a}(1567) = 0.197$. Using the computer package of Chung (1987), we obtained Fig. 2 with $\hat{c}(1567) = 1.27297$. Now we use Theorem 2.2 to construct a 95% confidence band for Q. Let $\varepsilon(n) = n^{-1}$ logloglog n and $p_0 = 0.3$. Here, $\varepsilon(1567) = 0.0175$. Using again the computer package of Chung (1987), we obtain Fig. 3 with $(c_1, c_2) = (0, 2.241)$, $(1.273, 1.273)$ and $(3.689, 0.5)$.

The obtained results were quite distinct depending on the choice of the parameters c_1 and c_2. When a large value is chosen for c_1 as in Fig. 3(c), the left tail $(\varepsilon(1567) \leq t \leq 0.15)$ of the quantile function has a narrow band, whereas the band for $t \geq 0.25$ has a much wider width than the first two bands in Figs. 3(a) and 3(b).

The graphs of F^0_{1567} and Q_{1567} in Fig. 1 show that statistical inference can be made only for the lower 0.3-quantile of the distribution at hand. Also, for example, 90%, 85%, 80% and 75% of the fractures must be larger than 0.53, 0.73, 0.97 and 1.37 m, respectively. From the confidence band in

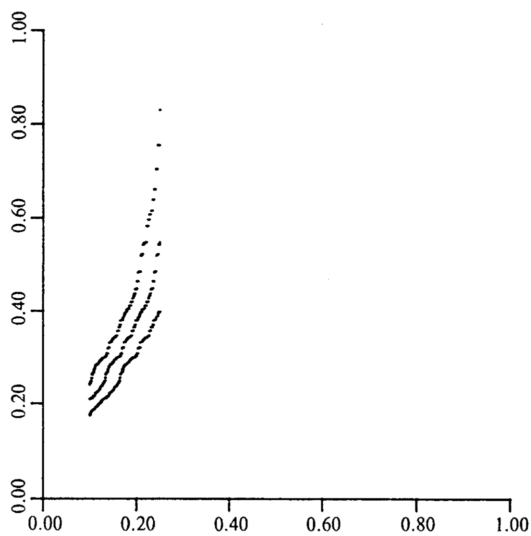

Fig. 2. 95% confidence band for Q on [0.1, 0.25] with $c(1567) = 1.27297$.

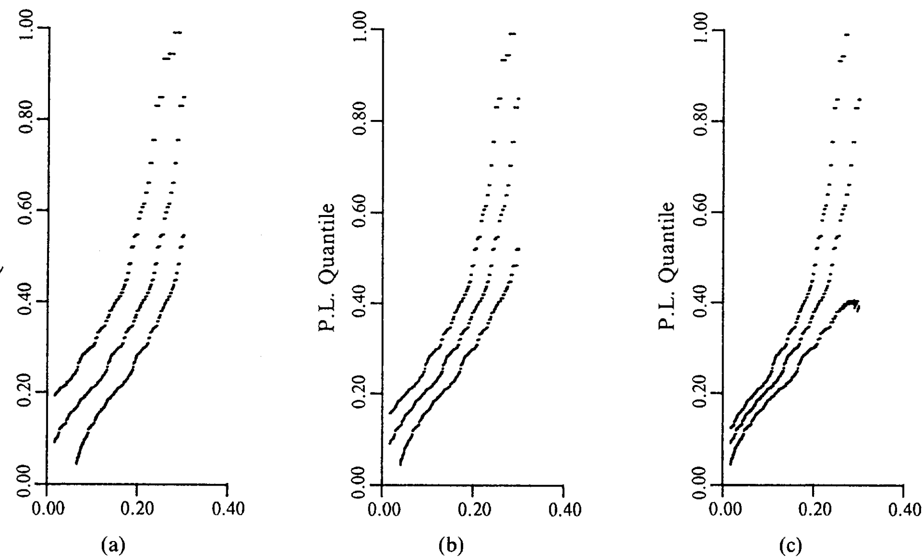

Fig. 3. 95% confidence band for Q on [0.0175, 0.3]. (a) $c_1 = 0$, $c_2 = 2.241$; (b) $c_1 = c_2 = 1.273$ and (c) $c_1 = 3.689$, $c_2 = 0.5$.

Fig. 3(b) we can, in addition, conclude also that simultaneously with probability 0.95 the errors of these four estimators are less than 0.26, 0.33, 0.47 and 1.03 m, respectively. Naturally, the width of this confidence band increases as we approach 0.3, due to having progressively less uncensored observations on the upper tail.

REFERENCES

Aalen, O. (1976). Nonparametric inference in connection with multiple decrement models, *Scand. J. Statist.*, **3**, 15–27.

Alexander, C. H. (1980). Simultaneous confidence bounds for the tail of an inverse distribution, *Ann. Statist.*, **8**, 1391–1394.

Alexander, C. H. (1982). Simultaneous confidence bounds, *Ann. Statist.*, **10**, 321.

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

Anderson, T. W. (1960). A modification of the sequential probability ratio test to reduce the sample size, *Ann. Math. Statist.*, **31**, 165–197.

Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship, *Ann. Statist.*, **2**, 437–453.

Burke, M. D., Csörgő, S. and Horváth, L. (1981). Strong approximation of some biometric estimates under random censorship, *Z. Wahrsch. Verw. Gebiete*, **56**, 87–112.

Burke, M. D., Csörgő, S. and Horváth, L. (1988). A correction to and improvement of "Strong approximations of some geometric estimates under random censorship", *Probab. Theory Related Fields*, **79**, 51–57.

Chung, C-J. F. (1986). Formulae for probabilities associated with Wiener and Brownian bridge processes, *Tech. Rep. Ser. Lab. Res. Stat. Probab.*, No. 79, Carleton University, Ottawa.

Chung, C-J. F. (1987). Wiener pack—A subroutine package for computing probabilities associated with Wiener and Brownian bridge processes, *Geol. Surv.*, Canada Paper, 87–12.

Csörgő, M. (1983). *Quantile Processes with Statistical Applications*, CBMS-NSF Regional Conference Series in Applied Mathematics, 42, SIAM, Philadelphia, Pennsylvania.

Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*, Academic Press, New York.

Csörgő, M. and Révész, P. (1984). Two approaches to constructing simultaneous confidence bounds for quantiles, *Probab. Math. Statist.* (Warsaw/Wroclaw), **4**, 221–236.

Csörgő, S. and Horváth, L. (1983). The baboons come down from the trees quite normally, *Probability and Statistical Inference*, *Proc. 4th Pannonian Symp. Math. Statist.*, (eds. W. Grossman, F. Konecny, G. Pflug and W. Wertz), 95–106, Reidel, Dordrecht.

Csörgő, S. and Horváth, L. (1986). Confidence bands from censored samples, *Canad. J. Statist.*, **14**, 131–144.

Efron, B. (1967). The two-sample problem with censored data, *Proc. fifth Berkeley Symp. on Math. Statist. Prob.*, Vol. 4, 831–853.

Feller, W. (1966). *An Introduction to Probability Theory and Its Applications*, Wiley, New York.

Gill, R. D. (1980). Censoring and stochastic integrals, Mathematical Centre Tracts, 124, Amsterdam.

Gill, R. D. (1983). Large sample behaviour of the product limit estimator on the whole line, *Ann. Statist.*, **11**, 49–58.

Gillespie, M. J. and Fisher, L. (1979). Confidence bands for the Kaplan-Meier survival curve estimate, *Ann. Statist.*, **7**, 920–924.

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

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations, *J. Amer. Statist. Assoc.*, **53**, 457–481.

Koziol, J. A. and Byar, D. P. (1975). Percentage points of the asymptotic distributions of one and two sample K-S statistics for truncated or censored data, *Technometrics*, **17**, 507–510.

Lo, S. H. and Singh, K. (1986). The product limit estimator and the bootstrap: Some asymptotic representations, *Probab. Theory Related Fields*, **71**, 455–465.

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

Nair, V. N. (1982). Goodness of fit tests for multiply right censored data, *Nonparametric Statistical Inference*, (eds. B. V. Gnedenko, M. L. Puri and I. Vincze), 653–666, North-Holland, Amsterdam.

Nair, V. N. (1984). Confidence bands for survival functions with censored data, A comparative study, *Technometrics*, **26**, 265–275.

Sander, J. M. (1975). The weak convergence of quantiles of the product-limit estimator, Tech. Report No. 5, Stanford University.

Stone, D., Kamineni, D. C. and Brown, A. (1984). Geology and fracture characteristics of the Underground Research Laboratory lease near Lac du Bonnet, Manitoba, Tech. Report-243, Atomic Energy of Canada Limited Research Company.