The Annals of Statistics
1988, Vol. 16, No. 3, 1113–1132

STRONG EMBEDDING OF THE ESTIMATOR OF THE DISTRIBUTION FUNCTION UNDER RANDOM CENSORSHIP

BY P. MAJOR AND L. REJTŐ[1]

Mathematical Institute of the Hungarian Academy of Sciences

In this paper the asymptotic behaviour of the product limit estimator F_n of an unknown distribution is investigated. We give an approximation of the difference $F_n(x) - F(x)$ by a Gaussian process and also by the average of i.i.d. processes. We get almost as good an approximation of the stochastic process $F_n(x) - F(x)$ as one can get for the analogous problem when the maximum likelihood estimator is approximated by a Gaussian random variable or by the average of i.i.d. random variables in the parametric case.

1. Introduction. The problem of estimating the distribution function under random censorship has been the subject of intense study over the last 10 years. Let $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$ be two sequences of i.i.d. random variables with the corresponding unknown distributions F and G. Let the sequences X_i and Y_i be independent of each other. Set $\delta_i = I(X_i \le Y_i)$ and $Z_i = \min(X_i, Y_i)$, $i = 1, \dots, n$, where $I(A)$ denotes the indicator function of the set A. If one considers the situation when only the censored sample (Z_i, δ_i), $i = 1, \dots, n$, is observed (and not the original data X_i, Y_i), the problem is to give a good estimator $F_n(x)$ of the distribution function $F(x)$ based on (Z_i, δ_i), $i = 1, \dots, n$. Then one would like to investigate the speed of convergence of $F_n(x)$ to $F(x)$. Kaplan and Meier (1958) proposed the following procedure to determine the product limit (PL) estimator. Put

$$N(u) = N(u, n) = \#\{Z_j: Z_j > u\} = \sum_{j=1}^{n} I(Z_j > u). \tag{1.1}$$

They gave an estimator for the survival function $1 - F(x)$ determined for continuous F and G by the formula

$$1 - F_n(u) = \bar{F}_n(u)$$
$$= \begin{cases} \prod_{j=1}^{n} \left[\dfrac{N(Z_j)}{1 + N(Z_j)} \right]^{I(Z_j \le u,\, \delta_j = 1)}, & \text{if } u \le \max(Z_1, \dots, Z_n), \\ 0, & \text{if } u \ge \max(Z_1, \dots, Z_n),\ \delta_n = 1, \\ \text{undefined}, & \text{if } u > \max(Z_1, \dots, Z_n),\ \delta_n = 0. \end{cases} \tag{1.2}$$

Breslow and Crowley (1974) proved that the sequence of PL estimator processes

Received March 1986; revised November 1987.

[1]Research supported by Hungarian National Foundation Research Grant No. 1803. Work partially done while visiting the Department of Mathematical Sciences at the University of Wisconsin-Milwaukee.

AMS 1980 *subject classifications.* 60F15, 60F17, 62G05.

Key words and phrases. Censored sample, product limit estimator.

$\sqrt{n}(F_n(u) - F(u))$ converges weakly to a Gaussian process with zero mean and a certain covariance function. Several papers have dealt with the problem of strong consistency of the PL estimator. Refer in particular to Földes and Rejtő (1981), Susarla and Van Ryzin (1978) and Csörgő and Horváth (1983). Burke, Csörgő and Horváth (1981) proved the first strong approximation result for the PL estimator process, obtaining $O((\log n)^{1/2}/n^{1/3})$ as their rate. This construction was based on a result of Komlós, Major and Tusnády (1975), where the exact rate of $O(\log n/n^{1/2})$ for the strong approximation of the empirical process by a sequence of Brownian bridges is proved. The question remained open as to whether one can get as good a rate of strong approximation for the PL estimator process by a sequence of Gaussian processes as can be obtained in the special case when there is no censoring. The aim of this paper is to show that the answer is in the affirmative.

During our investigation it became clear that to solve this problem it was first necessary to approximate $\sqrt{n}(F_n(u) - F(u))$ by a certain linear functional of an empirical process at an appropriate rate. We believe that the result that we obtained along this line is of separate interest; cf. Theorem 1. A very similar result is to be found in Lo and Singh (1986); however, their rate of approximation was not sharp enough for our purposes. In order to formulate our result we first must introduce some notation.

Put

$$\tilde{H}(u) = P(Z_i \le u, \delta_i = 1), \qquad \tilde{\tilde{H}}(u) = P(Z_i \le u, \delta_i = 0), \qquad H(u) = \tilde{H}(u) + \tilde{\tilde{H}}(u), \tag{1.3}$$

$$\tilde{H}_n(u) = \frac{1}{n}\sum_{j=1}^{n} I(Z_j \le u, \delta_j = 1), \qquad \tilde{\tilde{H}}_n(u) = \frac{1}{n}\sum_{j=1}^{n} I(Z_j \le u, \delta_j = 0). \tag{1.4}$$

Clearly,

$$\tilde{H}(u) = E\tilde{H}_n(u) = \int_{-\infty}^{u} (1 - G(t))\, dF(t), \qquad \tilde{\tilde{H}}(u) = E\tilde{\tilde{H}}_n(u) = \int_{-\infty}^{u} (1 - F(t))\, dG(t), \tag{1.5}$$

$$1 - H(u) = (1 - F(u))(1 - G(u)). \tag{1.6}$$

Let the functions $\overline{F}, \overline{G}, \overline{H}, \overline{F}_n, \overline{G}_n, \overline{H}_n$ denote $1 - F, 1 - G, \dots, 1 - H_n$. For the sake of simpler notation we assume that the distribution functions F and G are continuous although this restriction is not essential, as is shown in Remark 3.

Theorem 1. *Let T be such that $1 - H(T) > \delta$ with some $\delta > 0$. Then the process $F_n(u) - F(u)$, $-\infty < u < \infty$, $1 - H(u) > 0$, can be represented as*

$$F_n(u) - F(u) = (1 - F(u))(A(n,u) + B(n,u)) + R(n,u),$$

where

$$(1.7)\qquad \begin{aligned} A(n,u) &= \frac{\sqrt{n}\,(\tilde{H}_n(u) - \tilde{H}(u))}{\sqrt{n}\,(1 - H(u))} \\ &\quad - \frac{1}{\sqrt{n}} \int_{-\infty}^{u} \frac{\sqrt{n}\,(\tilde{H}_n(y) - \tilde{H}(y))}{(1 - H(y))^2}\, dH(y), \end{aligned}$$

$$(1.8)\qquad B(n,u) = \frac{1}{\sqrt{n}} \int_{-\infty}^{u} \frac{\sqrt{n}\,(H_n(y) - H(y))}{(1 - H(y))^2}\, d\tilde{H}(y)$$

are linear functionals of the empirical processes $\sqrt{n}\,(H_n(u) - H(u))$ *and* $\sqrt{n}\,(\tilde{H}_n(u) - \tilde{H}(u))$, *and the error term* $R(n,u)$ *can be bounded for* $u \le T$ *as*

$$(1.9)\qquad P\Big(\sup_{u \le T} n |R(n,u)| > x + \frac{C}{\delta}\Big) \le K e^{-\lambda x \delta^2}$$

for all $x > 0$, *where* $C > 0$, $K > 0$ *and* $\lambda > 0$ *are some universal constants.*

The following proposition is a relatively simple consequence of Komlós, Major and Tusnády (1975).

PROPOSITION 1. *The stochastic process* $A(n,u) + B(n,u)$ *defined in* (1.7) *and* (1.8) *can be approximated by an appropriate Gaussian process* $W(u)$, $-\infty < u < \infty$, $EW(u) = 0$, *with covariance function*

$$(1.10)\qquad EW(s)W(t) = EW(s)^2 = \int_{-\infty}^{s} \frac{dF(u)}{(1 - G(u))(1 - F(u))^2} \qquad \text{for } -\infty < s \le t < \infty,$$

in such a way that

$$(1.11)\qquad P\Big(\sup_{-\infty < u \le T} \big|\sqrt{n}\,(A(n,u) + B(n,u)) - W(u)\big| > \frac{C}{\delta}\log n + x\Big) < K e^{-\lambda \delta^2 x}.$$

A sequence of independent identically distributed Gaussian processes $W_1(u), W_2(u), \dots, EW_n(u) = 0$ *with covariance function* (1.10) *can be constructed in such a way that for all* $n = 1, 2, \dots,$

$$(1.12)\qquad P\Big(\sup_{k \le n} \sup_{-\infty < u \le T} \Big|k(A(k,u) + B(k,u)) - \sum_{j=1}^{k} W_j(u)\Big| > \frac{C}{\delta}\log^2 n + x \log n\Big) < K e^{-\lambda \delta^2 x}.$$

Here $C > 0$, $K > 0$ and $\lambda > 0$ are some universal constants; T and δ are the same as in Theorem 1.

Theorem 1 and Proposition 1 have the following consequence.

THEOREM 2. *We have*

$$P\left(\sup_{-\infty < u \le T} \left|\sqrt{n}\,(F_n(u) - F(u)) - \overline{W}(u)\right| > \frac{2C}{\delta}\log n + x\right) < 2Ke^{-\lambda\delta^2 x}, \tag{1.13}$$

where $\overline{W}(u) = (1 - F(u))W(u)$ and $W(u)$ is the same as in (1.10), *and*

$$P\left(\sup_{k \le n}\ \sup_{-\infty < u \le T} \left|k(F_k(u) - F(u)) - \sum_{j=1}^{k} \overline{W}_j(u)\right| > \frac{2C}{\delta}\log^2 n + x\log n\right) \le 2Ke^{-\lambda\delta^2 x}, \tag{1.14}$$

where $\overline{W}_j(u) = (1 - F(u))W_j(u)$ and the processes $W_j(u)$ are the same as in (1.12).

REMARK 1. In Theorem 1, $F_n(u) - F(u)$ is approximated by the average of bounded independent identically distributed stochastic processes [$A(n, u) + B(n, u)$ can be written in that way] and in Theorem 2 by a Gaussian process. The first type of approximation was given earlier by Lo and Singh (1986) and the second type by Burke, Horváth and Csörgő (1981), but with weaker results. Let us remark that as Proposition 1 shows, one of these approximation results implies the other with the same rate if this rate is worse than $O(\log^2 n/n)$. Thus these two kinds of approximation results are very close to each other. They differ only if we want to prove the optimal rate.

REMARK 2. It is worthwhile to compare our results with the analogous ones about the behaviour of the maximum likelihood estimate in the parametric case. Our results can be interpreted in such a way that for the PL estimator $F_n(x)$, $F_n(x) - F(x)$ can be approximated by the average of i.i.d. random processes with the rate $O(1/n)$ and with a Gaussian process with the rate $O(\log n/n)$. On the other hand, if η_n is the maximum likelihood estimator of a parameter θ with the help of a sample of n elements, then under very general conditions $\eta_n - \theta$ can be approximated by the average of i.i.d. random variables or by a Gaussian random variable with the rate $O(1/n)$. Thus our results mean that the PL estimator can be approximated by the average of i.i.d. processes with as good a rate as the maximum likelihood estimator by the average of i.i.d. random variables in the parametric case. If we want to get a Gaussian approximation, then the rate of the optimal approximation of the PL estimator is slightly weaker than that in the parametric case because a logarithmic factor appears.

It is an interesting question whether this property is a peculiarity of the PL estimator or if a similar result holds for a large class of nonparametric estimators.

REMARK 3. The case in which the distributions F and G may have jumps can be reduced to the case in which they are continuous. We suppose that the distribution functions are right continuous. First we must give the right definition of the PL estimator in the general case. Put

$$\begin{aligned} \gamma(u) &= \sum_{j=1}^{n} I(Z_j = u, \delta_j = 1), \qquad \gamma_i = \gamma(Z_i), \\ \mu(u) &= \sum_{j=1}^{n} I(Z_j = u, \delta_j = 0), \qquad \mu_i = \mu(Z_i). \end{aligned} \tag{1.15}$$

Then $1 - F_n(x)$ is defined by

$$1 - F_n(x) = \prod_{\substack{i=1 \\ \gamma_i \neq 0}}^{n} \left(\frac{N(Z_i) + \mu_i}{N(Z_i) + \mu_i + \gamma_i} \right)^{(I(Z_i \le x,\, \delta_i = 1))/\gamma_i}, \qquad \text{if } u \le \max(Z_1, \dots, Z_n). \tag{1.16}$$

It is easy to verify that if F and G are continuous, then (1.16) is equal to (1.2). Now we briefly explain how the general case can be reduced to the case when F and G are continuous.

Let $x_1, x_2, \dots$ be the set of all points where either F or G or both have a jump. Let us define the function

$$h(x) = x + \sum_{j:\, x_j < x} \frac{1}{j^2}, \qquad x \in \mathbb{R}^1,$$

the intervals

$$\Delta_{j,1} = \left[x_j + \sum_{k:\, x_k < x_j} \frac{1}{k^2};\ x_j + \frac{1}{2j^2} + \sum_{k:\, x_k < x_j} \frac{1}{k^2} \right],$$

$$\Delta_{j,2} = \left(x_j + \frac{1}{2j^2} + \sum_{k:\, x_k < x_j} \frac{1}{k^2};\ x_j + \frac{1}{j^2} + \sum_{k:\, x_k < x_j} \frac{1}{k^2} \right],$$

$$\Delta_j = \Delta_{j,1} \cup \Delta_{j,2}$$

and the sets $\Delta = \bigcup_{j=1}^{\infty} \Delta_j$, $\mathscr{C}(\Delta) = \mathbb{R}^1 - \Delta$. The function $h(x)$ maps $\mathbb{R}^1$ invertibly onto $\mathscr{C}(\Delta)$. Define the distribution funtions $\hat{F}$ and $\hat{G}$ as

$$\hat{F}(x)=\begin{cases} F(h^{-1}(x)), & \text{for } x\in\mathscr{C}(\Delta),\\ 2j^2\left(x-x_j-\sum\limits_{x_k<x_j}\dfrac{1}{k^2}\right) & \\ \quad\times\big(F(x_j)-F(x_j-0)\big)+F(x_j-0), & \text{for } x\in\Delta_{j,1},\ j=1,2,\dots,\\ F(x_j), & \text{for } x\in\Delta_{j,2},\ j=1,2,\dots, \end{cases}$$

$$\hat{G}(x)=\begin{cases} G(h^{-1}(x)), & \text{for } x\in\mathscr{C}(\Delta),\\ G(x_j-0), & \text{for } x\in\Delta_{j,1},\ j=1,2,\dots,\\ 2j^2\left(x-x_j-\dfrac{1}{2j^2}-\sum\limits_{k:\,x_k<x_j}\dfrac{1}{k^2}\right) & \\ \quad\times\big(G(x_j)-G(x_j-0)\big)+G(x_j-0), & \text{for } x\in\Delta_{j,2},\ j=1,2,\dots. \end{cases}$$

Obviously $\hat{F}(x)$ and $\hat{G}(x)$ are continuous.

Now we explain how a connection can be made between the original random censorship problem with the pairs F and G and its version where F and G are replaced by the continuous $\hat{F}$ and $\hat{G}$. Let us consider the random censorship problem described in the introduction. We observe a sample (Z_i,δ_i), $i=1,\dots,n$, where $Z_i=\min(X_i,Y_i)$ and $\delta_i=I(X_i\le Y_i)$. Let $\varepsilon_1,\varepsilon_2,\dots,\varepsilon_{2n}$ be a sequence of i.i.d. r.v.s with uniform distribution on $[0;1]$, which is independent of the sequences X_i and Y_i, $i=1,2,\dots,n$. Define the random variables $\hat{X}_i$ and $\hat{Y}_i$ by the formulas

$$\hat{X}_i=\begin{cases} h(X_i), & \text{if } X_i\ne x_j,\ j=1,2,\dots,\\ h(x_j)+\dfrac{1}{2j^2}\varepsilon_{2i-1}, & \text{if } X_i=x_j, \end{cases}$$

$$\hat{Y}_i=\begin{cases} h(Y_i), & \text{if } Y_i\ne x_j,\ j=1,2,\dots,\\ h(x_j)+\dfrac{1}{2j^2}+\dfrac{1}{2j^2}\varepsilon_{2i}, & \text{if } Y_i=x_j. \end{cases}$$

Then $\hat{X}_1,\dots,\hat{X}_n$ and $\hat{Y}_1,\dots,\hat{Y}_n$ are two independent sequences of i.i.d. r.v.s with distribution functions $\hat{F}$ and $\hat{G}$, respectively. Then since $\hat{F}$ and $\hat{G}$ are continuous distribution functions, our results can be applied for the sample $(\hat{Z}_i,\delta_i)$, where $\hat{Z}_i=\min(\hat{X}_i,\hat{Y}_i)$, $\hat{\delta}_i=I(\hat{X}_i\le\hat{Y}_i)=I(X_i\le Y_i)=\delta_i$. Denote the estimator by $\hat{F}_n(x)$. Observe that

$$F(x)=\hat{F}(h(x)),\qquad F_n(x)=\hat{F}_n(h(x))\quad\text{if } x\ne x_j,$$

and

$$F(x_j - 0) = \hat{F}(h(x_j)), \qquad F(x_j) = \hat{F}(h(x_j + 0)),$$
$$F_n(x_j - 0) = \hat{F}_n(h(x_j)), \qquad F_n(x_j) = \hat{F}_n(h(x_j + 0)) \quad \text{for } x = x_j.$$

As a consequence, the processes $F_n(t) - F(t)$ and $\hat{F}_n(h(t)) - \hat{F}(h(t))$, $-\infty < t < T$, coincide. In such a way we get a representation of $F_n(t) - F(t)$ via the representation of $\hat{F}_n(t) - \hat{F}(t)$. Then the representation for $F_n - F$ given in Theorem 1 can be obtained for general (possibly noncontinuous) distribution functions F and G by first applying it for $\hat{F}$ and $\hat{G}$. Since we want to express $F_n - F$ directly with the help of the original distribution functions and sample, we have to rewrite the expression $\hat{A}_n(u) + \hat{B}_n(u)$ corresponding to $\hat{F}_n(u) - \hat{F}(u)$. By formula (1.3) the expression $\hat{A}(n, u) + \hat{B}(n, u)$ can be written as

$$\begin{aligned}\hat{A}(n,u) + \hat{B}(n,u) = {} & \frac{\sqrt{n}\left(\hat{\tilde{H}}_n(u) - \hat{\tilde{H}}(u)\right)}{\sqrt{n}\,(1 - \hat{H}(u))} \\ & - \frac{1}{\sqrt{n}} \int_{-\infty}^{u} \frac{\sqrt{n}\left(\hat{\tilde{H}}_n(y) - \hat{\tilde{H}}(y)\right)}{(1 - \hat{H}(y))^2}\, d\hat{\tilde{\tilde{H}}}(y) \\ & + \frac{1}{\sqrt{n}} \int_{-\infty}^{u} \frac{\sqrt{n}\left(\hat{\tilde{\tilde{H}}}_n(y) - \hat{\tilde{\tilde{H}}}(y)\right)}{(1 - \hat{H}(y))^2}\, d\hat{\tilde{H}}(y),\end{aligned}$$

where the functions with carets ($\hat{\ }$) correspond to the variables $\hat{X}_i, \hat{Y}_i, \hat{Z}_i$. We define $A(n, u) + B(n, u)$ in the general case as $A(n, u) + B(n, u) = \hat{A}(n, h(u)) + \hat{B}(n, h(u))$. Thus for the explicit expression we have to compute the integrals over all the intervals Δ_j for which $x_j \le u$. After a simple computation, $A(n, u) + B(n, u)$ is given as

$$\begin{aligned}A(n,u) + B(n,u) = {} & \frac{\sqrt{n}\left(\tilde{H}_n(u) - \tilde{H}(u)\right)}{\sqrt{n}\,(1 - H(u))} \\ & - \frac{1}{\sqrt{n}} \int_{-\infty}^{u} \frac{\sqrt{n}\left(\tilde{H}_n(y) - \tilde{H}(y)\right) dG(y)}{(1 - H(y))(1 - G(y - 0))} \qquad (1.17) \\ & + \frac{1}{\sqrt{n}} \int_{-\infty}^{u} \frac{\sqrt{n}\left(\tilde{\tilde{H}}_n(y - 0) - \tilde{\tilde{H}}(y - 0)\right) dF(y)}{(1 - H(y - 0))(1 - F(y))},\end{aligned}$$

where the functions $\tilde{H}$ and $\tilde{\tilde{H}}$ can be written as

$$\begin{aligned}\tilde{H}(u) &= P(Z_i \le u, \delta_i = 1) = \int_{-\infty}^{u} (1 - G(y - 0))\, dF(y), \\ \tilde{\tilde{H}}(u) &= P(Z_i \le u, \delta_i = 0) = \int_{-\infty}^{u} (1 - F(y))\, dG(y).\end{aligned} \qquad (1.18)$$

In this way we give a representation of the process $F_n(u) - F(u)$ when both distributions F and G may have jumps. The statement is given as follows.

COROLLARY 1. *Suppose that the distribution functions F and G may have jumps and let T be such that $1 - H(T) > \delta$ with some $\delta > 0$. Then the process $F_n(u) - F(u)$ for $-\infty < u < +\infty$, $1 - H(u) > 0$, can be represented as*

$$F_n(u) - F(u) = \left(\frac{\sqrt{n}\,(\tilde{H}_n(u) - \tilde{H}(u))}{\sqrt{n}\,(1 - H(u))} - \frac{1}{\sqrt{n}} \int_{-\infty}^{u} \frac{\sqrt{n}\,(\tilde{H}_n(y) - \tilde{H}(y))\, dG(y)}{(1 - H(y))(1 - G(y-0))} \right.$$

$$\left. + \frac{1}{\sqrt{n}} \int_{-\infty}^{u} \frac{\sqrt{n}\left(\tilde{\tilde{H}}_n(y-0) - \tilde{\tilde{H}}(y-0)\right)}{(1 - H(y-0))(1 - F(y))}\, dF(y) \right) \tag{1.19}$$

$$\times (1 - F(u)) + R(n, u),$$

where $R(n, u)$ can be bounded for $u \le T$ by (1.9).

In a similar way, we obtain the embedding theorem, i.e., Proposition 1, for the noncontinuous case. The general form of the proposition is given as follows.

COROLLARY 2. *Suppose that the distribution functions F and G may have jumps. Then the stochastic process $A(n, u) + B(n, u)$ defined by* (1.17) *can be approximated by an appropriate Gaussian process $W(u)$, $-\infty < u < +\infty$, $EW(u) = 0$, with covariance function*

$$E(W(s)W(t)) = \int_{-\infty}^{s} \frac{dF(u)}{(1 - F(u))(1 - F(u-0))(1 - G(u-0))} \tag{1.20}$$

$$\text{for } -\infty < s \le t < +\infty,$$

in such a way that

$$P\left(\sup_{-\infty < u < T} \left| \sqrt{n}\,(A(n, u) + B(n, u)) - W(u) \right| > \frac{C}{\delta} \log n + x \right) \tag{1.21}$$

$$< Ke^{-\lambda \delta^2 x},$$

where $C > 0$, $K > 0$ and $\lambda > 0$ are some universal constants; T and δ are the same as in Corollary 1.

In Section 2 of our paper we prove the theorems with the help of some lemmas, whose proofs are given in Section 3.

2. Proof of the theorems. The function $F_n(u)$ is defined in (1.2), but its asymptotic behaviour cannot be seen directly from this formula. Hence we approximate it with another expression which has a simpler structure.

Let us introduce the functions

$$T_n(u) = -\log(1 - F_n(u)), \qquad T(u) = -\log(1 - F(u)).$$

A simple Taylor expansion yields

$$F_n(u) - F(u) = \exp(-T(u))[1 - \exp(T(u) - T_n(u))]$$
$$= \exp(-T(u))[T_n(u) - T(u)] + O\big((T_n(u) - T(u))^2\big),$$

or more precisely

$$F_n(u) - F(u) = (1 - F(u))(T_n(u) - T(u)) + R_1(u) \tag{2.1}$$

with

$$|R_1(u)| \le |T_n(u) - T(u)|^2 \quad \text{if } |T_n(u) - T(u)| \le 1. \tag{2.1$'$}$$

As we shall see, $T_n(u) - T(u)$ is rather small [typically $O(1/\sqrt{n})$]; therefore, $R_1(u)$ is negligibly small. We shall also see that the probability of the event $|T_n(u) - T(u)| \le 1$ is exponentially close to 1; therefore, this is not a serious restriction in the applicability of (2.1). Thus it is enough to investigate the term $T_n(u) - T(u)$. By (1.2),

$$T_n(u) = -\sum_{i=1}^{n} I(Z_i \le u, \delta_i = 1)\log\left(1 - \frac{1}{1 + N(Z_i)}\right).$$

Let us define the set

$$A_n = \left\{\sum_{i=1}^{n} I(Z_i > T) \le \frac{\delta}{2}n\right\}.$$

The relation $-\log(1 - x) \sim x$ for small x suggests approximating the process $T_n(u)$ by the process

$$\tilde{T}_n(u) = \sum_{i=1}^{n} \frac{I(Z_i \le u, \delta_i = 1)}{N(Z_i)} = \sum_{i=1}^{n} \frac{I(Z_i \le u, \delta_i = 1)}{\sum_{j=1}^{n} I(Z_j > Z_i)}.$$

In Section 3 we prove

LEMMA 1.

$$T_n(u) = \tilde{T}_n(u) + R_2(u),$$

where

$$P\left(\sup_{u \le T} |nR_2(u)| > \frac{2}{\delta}\right) \le \exp(-\lambda\delta n).$$

(The constants λ, δ, etc., denote some universal constants. The same letter may denote different constants in different formulas.)

The expression $\tilde{T}_n(u)$ is still not appropriate for our purposes. Since the denominators $N(Z_i) = \sum_{j=1}^{n} I(Z_j > Z_i)$ are dependent on different i – s, we cannot see directly the limit behaviour of $\tilde{T}_n(u)$. On the other hand, by exploiting the fact that the conditional distribution of $N(Z_i)$ given Z_i fixed is a binomial distribution with parameters $n - 1$ and $1 - H(Z_i)$, we can rewrite $\tilde{T}_n(u)$ in a more appropriate form. Indeed, by writing

$$\sum_{j=1}^{n} I(Z_j > Z_i) = n\overline{H}(Z_i)\left[1 + \frac{\sum_{j=1}^{n} I(Z_j > Z_i) - n\overline{H}(Z_i)}{n\overline{H}(Z_i)}\right]$$

and noting that $|1/(1+z) - 1 + z| < 2z^2$ for $|z| < \frac{1}{2}$, we get that

$$\tilde{T}_n(u) = \sum_{i=1}^{n} \frac{I(Z_i \le u, \delta_i = 1)}{n\overline{H}(Z_i)}\left[1 - \frac{\sum_{j=1}^{n} I(Z_j > Z_i) - n\overline{H}(Z_i)}{n\overline{H}(Z_i)}\right] + R_3(u)$$

with

$$|R_3(u)| \le 2\sum_{i=1}^{n} \frac{I(Z_i \le u, \delta_i = 1)}{n\overline{H}(Z_i)}\left[\frac{\sum_{j=1}^{n} I(Z_j > Z_i) - n\overline{H}(Z_i)}{n\overline{H}(Z_i)}\right]^2 \tag{2.2}$$

on the set

$$B_n = B_n(u) = I\left(\left|\sum_{j=1}^{n} I(Z_j > Z_i) - n\overline{H}(Z_i)\right| < \tfrac{1}{2}n\overline{H}(Z_i) \text{ or } Z_i > u \right.$$
$$\left. \text{for all } 1 \le i \le n\right). \tag{2.2'}$$

In Section 3 we prove

LEMMA 2.

$$P\left(\sup_{u \le T} |nR_3(u)| > x\right) \le K\exp(-\lambda\delta^2 x) \quad \textit{if } 0 \le x < \frac{2n}{\delta}.$$

Thus we get that

$$\tilde{T}_n(u) = 2\overline{A}(u) - \overline{B}(u) + R_3(u) \tag{2.3}$$

with

$$\overline{A}(u) = \overline{A}(n,u) = \sum_{i=1}^{n} \frac{I(Z_i \le u, \delta_i = 1)}{n\overline{H}(Z_i)}$$

and

$$\overline{B}(u) = \overline{B}(n,u) = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{I(Z_i \le u, \delta_i = 1)I(Z_j > Z_i)}{n^2\overline{H}^2(Z_i)}.$$

In another form,

$$\overline{A}(u) = \int_{-\infty}^{\infty} \frac{I(y \le u)}{1 - H(y)}\, d\tilde{H}_n(y) \tag{2.4}$$

and

$$\bar{B}(u) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{I(y \le u)I(x > y)}{(1 - H(y))^2}\, d\tilde{H}_n(y)\, dH_n(x).$$

Put

$$\begin{aligned}
\bar{B}(u) &= B_1(u) + B_2(u) + B_3(u) + B_4(u) \\
&= \frac{1}{\sqrt{n}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{I(y \le u)I(x > y)}{(1 - H(y))^2} d\left[\sqrt{n}\left(\tilde{H}_n(y) - \tilde{H}(y)\right)\right] dH(x) \\
&\quad + \frac{1}{\sqrt{n}}\int_{-\infty}^{\infty} \frac{I(y \le u)I(x > y)}{(1 - H(y))^2}\, d\tilde{H}(y) d\left[\sqrt{n}\left(H_n(x) - H(x)\right)\right] \\
&\quad + \frac{1}{n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{I(y \le u)I(x > y)}{(1 - H(y))^2} d\left[\sqrt{n}\left(\tilde{H}_n(y) - \tilde{H}(y)\right)\right] \\
&\qquad \times d\left[\sqrt{n}\left(H_n(x) - H(x)\right)\right] \\
&\quad + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{I(y \le u)I(x > y)}{(1 - H(y))^2}\, d\tilde{H}(y)\, dH(x).
\end{aligned} \tag{2.5}$$

Clearly

$$B_1(u) + B_4(u) = \bar{A}(u). \tag{2.6}$$

Put

$$R_4(u) = -B_3(u). \tag{2.7}$$

Since both processes $\sqrt{n}(\tilde{H}_n(y) - \tilde{H}(y))$ and $\sqrt{n}(H_n(x) - H(x))$ have a limit as $n \to \infty$, and there is a multiplier $1/n$ in the definition of $R_4(u)$, it is natural to expect that $R_4(u)$ is negligibly small. However, the hardest part of our proof is to bound the term $R_4(u)$. To solve this problem Major (1988) proved a more general inequality. We prove Lemma 3 with its help.

LEMMA 3.

$$P\Big(\sup_{u \le T} |nR_4(u)| > x\Big) < K \exp(-\lambda\delta^2 x) \quad \textit{for } x > 0.$$

The term $B_2(u)$ can be rewritten as

$$B_2(u) = \frac{1}{\sqrt{n}}\int_{-\infty}^{u} \frac{\sqrt{n}\left(H(y) - H_n(y)\right)}{(1 - H(y))^2}\, d\tilde{H}(y). \tag{2.8}$$

Formulas (2.3), (2.6) and (2.7) imply that

$$\tilde{T}_n(u) = \bar{A}(u) - B_2(u) + R_3(u) + R_4(u). \tag{2.9}$$

Put

$$\bar{A}(u) = A_0(u) + A_1(u)$$

(2.10)
$$= \int_{-\infty}^{u} \frac{d\tilde{H}(y)}{1 - H(y)} + \frac{1}{\sqrt{n}} \int_{-\infty}^{\infty} \frac{d\left[\sqrt{n}\,(\tilde{H}_n(y) - \tilde{H}(y))\right]}{1 - H(y)}.$$

By (1.5) and (1.6),

(2.11)
$$A_0(u) = \int_{-\infty}^{u} \frac{dF(u)}{1 - F(u)} = -\log(1 - F(u)) = T(u)$$

and integration by parts yields that

(2.12)
$$A_1(u) = \frac{\sqrt{n}\,(\tilde{H}_n(u) - \tilde{H}(u))}{\sqrt{n}\,(1 - H(u))} - \frac{1}{\sqrt{n}} \int_{-\infty}^{u} \frac{\sqrt{n}\,(\tilde{H}_n(y) - \tilde{H}(y))}{(1 - H(y))^2}\, dH(y).$$

Relations (2.9), (2.10), (2.11) and Lemma 1 imply that

(2.13)
$$T_n(u) - T(u) = A_1(u) - B_2(u) + R_2(u) + R_3(u) + R_4(u)$$

and this formula together with (2.1) yields that

(2.14)
$$F_n(u) - F(u) = (1 - F(u))(A_1(u) - B_2(u)) + R(n, u)$$

with

(2.14′)
$$R(n, u) = (1 - F(u))(R_2(u) + R_3(u) + R_4(u)) + R_1(u).$$

Define $A(n, u) = A_1(u)$ and $B(n, u) = -B_2(u)$. First we prove Theorem 1 in the case $0 < x < 2n$. We can assume without violating the generality that $x > 50/(n\delta^2)$. Indeed, we can choose the constant K so large in (1.9) that for $x < 50/(n\delta^2)$, $K \exp(-\lambda\delta^2 x) \geq 1$ and for such $x - s$ relation, (1.9) holds automatically.

By Lemmas 1–3,

(2.15)
$$P\left(\sup_{u \leq T} |nR_2(u)| > \frac{2}{\delta}\right) \leq \exp(-\lambda\delta n) < \exp(-\lambda\delta^2 x) \quad \text{for } 0 < x < 2n,$$

(2.16)
$$P\left(\sup_{u \leq T} |nR_3(u)| > \frac{x}{4}\right) \leq K \exp(-\lambda\delta^2 x) \quad \text{for } 0 < x < 2n$$

and

(2.17)
$$P\left(\sup_{u \leq T} |nR_4(u)| > \frac{x}{4}\right) \leq K \exp(-\lambda\delta^2 x) \quad \text{for } 0 < x < 2n.$$

We claim that

$$P\Big(\sup_{u\le T} n|T_n(u)-T(u)|^2 > x\Big) \le K\exp(-\lambda\delta^2 x) \quad \text{for } \frac{50}{n\delta^2} < x < 2n. \tag{2.18}$$

First we show that (2.14) and (2.14′) together with (2.15)–(2.18) imply Theorem 1 for $0 < x < 2n$. Indeed, relation (2.18) implies that

$$P\Big(\sup_{u\le T} |T_n(u)-T(u)| > 1\Big) \le K\exp(-\lambda\delta^2 n)$$

and by (2.18) and (2.1′),

$$P\Big(\sup_{u\le T} |nR_1(u)| > \frac{x}{2}\Big) \le K\exp(-\lambda\delta^2 x) \quad \text{for } \frac{50}{n\delta^2} < x < 2n.$$

The estimates obtained for $nR_j(u)$ imply that

$$P\Big(\sup_{u\le T} n|R(n,u)| > x + \frac{C}{\delta}\Big) \le K\exp(-\lambda\delta^2 x) \quad \text{for } \frac{50}{n\delta^2} < x < 2n,$$

which together with (2.14) implies Theorem 1 for $0 < x < 2n$. To prove (2.18) we write

$$\begin{aligned} &P\Big(\sup_{u\le T} n|T_n(u)-T(u)|^2 > x\Big) \\ &\quad\le P\Big(\sup_{u\le T} n|A_1(u)| > \tfrac{1}{5}\sqrt{nx}\Big) \\ &\qquad + P\Big(\sup_{u\le T} n|B_2(u)| > \tfrac{1}{5}\sqrt{nx}\Big) + P\Big(\sup_{u\le T} n|R_2(u)| > \tfrac{1}{5}\sqrt{nx}\Big) \\ &\qquad + P\Big(\sup_{u\le T} n|R_3(u)| > \tfrac{1}{5}\sqrt{nx}\Big) + P\Big(\sup_{u\le T} n|R_4(u)| > \tfrac{1}{5}\sqrt{nx}\Big). \end{aligned} \tag{2.19}$$

Because of the relations $1 - H(u) \ge \delta$ and

$$\int_{-\infty}^{u} \frac{dH(y)}{(1-H(y))^2} = \frac{H(u)}{1-H(u)} \le \frac{1}{\delta} \quad \text{for } u \le T, \tag{2.20}$$

(2.12) implies that

$$\sup_{u\le T} |A_1(u)| \le \frac{2}{\delta} \sup_{u\le T} |\tilde{H}_n(u) - \tilde{H}(u)|.$$

Therefore,

$$\int_{-\infty}^{u} \frac{d\tilde{H}(y)}{(1 - H(y))^2} \le \int_{-\infty}^{u} \frac{dH(y)}{(1 - H(y))^2} \le \frac{1}{\delta} \quad \text{for } u \le T, \tag{2.21}$$

$$P\left(\sup_{u \le T} n|A_1(u)| > \frac{1}{5}\sqrt{nx} \right) \le P\left(\sup n|\tilde{H}_n(u) - \tilde{H}(u)| > \frac{1}{10}\delta\sqrt{nx} \right).$$

Lemma 2 of Dvoretzky, Kiefer and Wolfowitz (1956) gives that

$$P\left(\sup_{u \le T} n|A_1(u)| > \tfrac{1}{5}\sqrt{nx} \right) \le K\exp(-\lambda\delta^2 nx).$$

Similarly we get from (2.8) the relation

$$P\left(\sup_{u \le T} n|B_2(u)| > \tfrac{1}{5}\sqrt{nx} \right) \le K\exp(-\lambda\delta^2 nx).$$

Applying Lemmas 1–3 we can estimate the next three terms in (2.19) as

$$P\left(\sup_{u \le T} n|R_2(u)| > \tfrac{1}{5}\sqrt{nx} \right) \le \exp(-\lambda\delta n) \quad \text{if } x > 50/(n\delta^2),$$

$$P\left(\sup_{u \le T} n|R_3(u)| > \tfrac{1}{5}\sqrt{nx} \right) \le K\exp(-\lambda\delta^2\sqrt{nx}),$$

$$P\left(\sup_{u \le T} n|R_4(u)| > \tfrac{1}{5}\sqrt{nx} \right) \le K\exp(-\lambda\delta^2\sqrt{nx}).$$

Since

$$K\exp(-\lambda\delta^2\sqrt{nx}) \le K\exp\left(-\frac{\lambda\delta^2}{2}x \right)$$

and

$$\exp(-\lambda\delta n) \le \exp\left(-\frac{\lambda\delta^2}{2}x \right) \quad \text{for } x < 2n,$$

the preceding estimates imply (2.18).

To prove Theorem 1 for $x > 2n$, observe that $|F_n(u) - F(u)| \le 1$; hence by (2.14) the relation $|nR(n,u)| > x$ implies that $n|A_1(u) - B_2(u)| > x/2$ for $x > 2n$. Thus in this case,

$$\begin{aligned} P\left(\sup_{u \le T} |nR(n,u)| > x \right) &\le P\left(\sup_{u \le T} n|A_1(u)| > \frac{x}{4} \right) + P\left(\sup_{u \le T} n|B_2(u)| > \frac{x}{4} \right) \\ &\le P\left(\sup_{u \le T} |\tilde{H}_n(u) - \tilde{H}(u)| > \frac{x\delta}{8\sqrt{n}} \right) \\ &\quad + P\left(\sup_{u \le T} |H_n(u) - H(u)| > \frac{x\delta}{8n} \right) \\ &\le K\exp\left(-\lambda\frac{\delta^2 x^2}{n} \right) \le K\exp(-\lambda\delta^2 x). \end{aligned}$$

Theorem 1 is proved.

PROOF OF PROPOSITION 1. In Section 3 we shall prove with the help of the Komlós–Major–Tusnády approximation

LEMMA 4. *A Brownian bridge $S(t)$, $0 < t < 1$, can be constructed in such a way that*

$$P\Big(\sup_{-\infty<u<\infty} \sqrt{n}\,\big|S(\tilde{H}(u)) - \sqrt{n}\,(\tilde{H}_n(u) - \tilde{H}(u))\big| > (C\log n + z)\Big) < Ke^{-\lambda z}$$

and

$$P\Big(\sup_{-\infty<u<\infty} \sqrt{n}\,\Big|S\big(1-\tilde{\tilde{H}}(u)\big) - \sqrt{n}\,\big(\tilde{\tilde{H}}(u) - \tilde{\tilde{H}}_n(u)\big)\Big| > (C\log n + z)\Big) < Ke^{-\lambda z}$$

for all $z > 0$.

A sequence of independent Brownian bridges $S_n(t)$, $n = 1,2,\dots$, $0 < t < 1$, can be constructed in such a way that

$$P\left(\sup_{0\le k\le n}\ \sup_{-\infty<u<\infty}\left|\sum_{j=1}^{k} S_j(\tilde{H}(u)) - k(\tilde{H}_k(u) - \tilde{H}(u))\right| > (C\log n + z)\log n\right) \le Ke^{-\lambda z}$$

and

$$P\left(\sup_{0\le k\le n}\ \sup_{-\infty<u<\infty}\left|\sum_{j=1}^{k} S_j\big(1-\tilde{\tilde{H}}(u)\big) - k\big(\tilde{\tilde{H}}(u) - \tilde{\tilde{H}}_k(u)\big)\right| > (C\log n + z)\log n\right) \le Ke^{-\lambda z}.$$

Define the Gaussian processes $W(u)$ and $W_n(u)$, $n = 1,2,\dots$, as

$$\begin{aligned}W(u) = {}& \frac{S(\tilde{H}(u))}{1-H(u)} - \int_{-\infty}^{u} \frac{S(\tilde{H}(y))}{(1-H(y))^2}\,dH(y)\\ &+ \int_{-\infty}^{u} \frac{S(\tilde{H}(y)) - S\big(1-\tilde{\tilde{H}}(y)\big)}{(1-H(y))^2}\,d\tilde{H}(y)\end{aligned} \tag{2.22}$$

and $W_n(u)$ is defined in the same way with S replaced by S_n. Here S and S_n are the same Brownian bridges which appear in Lemma 4. Then a comparison of (2.22) with (2.8) and (2.12) together with (2.20) and (2.21) and the application of Lemma 4 with $z = \delta x$ yields relations (1.11) and (1.12). Then Lemma 5 [proved in the Appendix of Breslow and Crowley (1974)] completes the proof of the proposition. □

LEMMA 5. *The Gaussian process defined in* (2.22) *has the covariance function given in* (1.10).

Theorem 2 is a straightforward consequence of Theorem 1 and Proposition 1. To prove (1.13), it is enough to observe that

$$P\left(\sup_{-\infty<u\le T}\left|\sqrt{n}\,(F_n(u)-F(u))-\overline{W}(u)\right|>\frac{2C}{\delta}\log n+x\right)$$

$$\le P\left(\sup_{-\infty\le u\le T}\left|\sqrt{n}\,(A(n,u)+B(n,u))-W(u)\right|>\frac{C}{\delta}\log n+\frac{x}{2}\right)$$

$$+P\left(\sup_{-\infty<u\le T}|R(n,u)|>\frac{x}{2}+\frac{C}{\delta}\right)$$

and to apply Theorem 1 and Proposition 1. The proof of (1.14) is the same.

3. Proof of the lemmas.

PROOF OF LEMMA 1. Since $|\log(1-x)-x|\le x^2$ if $|x|<\frac{1}{2}$, hence

$$\sup_{u\le T}|R_2(u)|\le\sum_{i=1}^{n}\frac{I(Z_i\le T)}{N^2(Z_i)}$$

if $N(Z_i)>1$ for all $Z_i\le T$. On the other hand, we claim that

$$P(A_n)=P\left(\sum_{i=1}^{n}I(Z_i>T)\le\frac{\delta}{2}n\right)\le\exp(-\lambda\delta n). \tag{3.1}$$

Since $P(Z_i>T)\ge\delta$, $P(A_n)$ is less than the probability of the event that a random variable with binomial distribution with parameters n and δ takes a value less than $\frac{1}{2}\delta n$. The probability of this event can be bounded by $\exp(-\lambda\delta n)$; hence, (3.1) holds. Since $N(Z_i)$ are different integers for different i, we see that on the complement of the set A_n,

$$\sup_{u\le T}|nR_2(u)|\le n\sum_{j=\frac{1}{2}\delta n}^{\infty}\frac{1}{j^2}\le\frac{2}{\delta}.$$

Thus

$$P\left(\sup_{u\le T}|nR_2(u)|>\frac{2}{\delta}\right)\le P(A_n)$$

and this implies Lemma 1. □

PROOF OF LEMMA 2. Because of (2.2),

$$
\begin{aligned}
P\Big(\sup_{u\le T}|nR_3(u)| > x\Big) &\le (1 - P(B_n(T))) \\
&\quad + P\Bigg(\sup_{u\le T} 2n\sum_{i=1}^{n}\frac{I(Z_i\le u,\,\delta_i=1)}{n\overline{H}(Z_i)} \\
&\qquad\times\Bigg[\sum_{j=1}^{n}\frac{I(Z_j>Z_i)-\overline{H}(Z_i)}{n\overline{H}(Z_i)}\Bigg]^2 > x\Bigg).
\end{aligned}
\tag{3.2}
$$

By (2.2),

$$
1 - P(B_n(T)) \le nP\left(\frac{\sum_{j=1}^{n} I(Z_j > Z_1) - n(1-H(Z_1))}{n(1-H(Z_1))} > 2,\ Z_1 < T\right). \tag{3.3}
$$

We estimate the right-hand side of (3.3) by first conditioning on Z_1. Applying a general form of the Bernstein inequality [see, e.g., Rényi (1970)], we get that

$$
\begin{aligned}
P\left(\frac{\sum_{j=1}^{n} I(Z_j > Z_1) - n(1-H(Z_1))}{n(1-H(Z_1))} > 2,\ Z_1 < T \mid Z_1 = t\right) \\
\le \exp(-Cn(1-H(t))), &\quad \text{if } t < T, \\
= 0, &\quad \text{if } t > T.
\end{aligned}
$$

Integrating this relation we get that

$$
\begin{aligned}
1 - P(B_n(T)) &\le n\int_{-\infty}^{T}\exp(-Cn(1-H(u)))\,dH(u) \\
&\le K\exp\left(-\frac{C}{2}n\delta\right).
\end{aligned}
\tag{3.4}
$$

On the other hand,

$$
\begin{aligned}
&P\left(\sup_{u\le T} 2\sum_{i=1}^{n}\frac{I(Z_i\le u,\,\delta_i=1)}{\overline{H}(Z_i)}\left(\sum_{j=1}^{n}\frac{I(Z_j>Z_i)-\overline{H}(Z_i)}{n\overline{H}(Z_i)}\right)^2 > x\right) \\
&\quad\le P\left(2\sum_{i=1}^{n}\frac{I(Z_i\le T,\,\delta_i=1)}{\overline{H}(Z_i)n}\left(\sum_{j=1}^{n}\frac{I(Z_j>Z_i)-\overline{H}(Z_i)}{\sqrt{n}\,\overline{H}(Z_i)}\right)^2 > x\right) \\
&\quad\le P\left(\frac{2}{\delta}\sup_{u\le T}\left(\frac{\sqrt{n}\,(H_n(u)-H(u))}{1-H(u)}\right)^2 > x\right) \\
&\quad\le K\exp(-\lambda\delta^2 x).
\end{aligned}
\tag{3.5}
$$

The last inequality follows, e.g., from Lemma 3 of Wellner (1978). Lemma 2 follows from (3.3)–(3.5). □

PROOF OF LEMMA 3. We shall deduce Lemma 3 from

LEMMA 6. *Let F_n be the empirical distribution function of a sample with the uniform distribution and $h(x, y)$ a measurable function on $\mathbb{R}^2$ such that $|h(x, y)| \le K$ for all $x \in \mathbb{R}$ and $y \in \mathbb{R}$. Define*

$$B(u) = \int_0^u \int_0^1 h(x, y) d\big(\sqrt{n}(F_n(x) - x)\big) d\big(\sqrt{n}(F_n(y) - y)\big).$$

Then

$$P\Big(\sup_{0\le u\le 1} |B(u)| > x\Big) \le C\exp\Big(-\frac{A}{K}x\Big) \quad \textit{for all } x > 0.$$

Lemma 6 is a special case of Theorem 1 in Major (1988).

Define the transformations T_1 and T_2 from $\mathbb{R}^1$ to $[0,1]$ by the formulas

$$T_1(x) = \tilde{H}(x), \qquad T_2(x) = 1 - \tilde{\tilde{H}}(x) \tag{3.6}$$

and the random variables $U_1, \dots, U_n$ with the help of the sample $(Z_1, \delta_1), \dots, (Z_n, \delta_n)$ as

$$U_i = \begin{cases} T_1(Z_i), & \text{if } \delta_i = 1, \\ T_2(Z_i), & \text{if } \delta_i = 0. \end{cases} \tag{3.7}$$

Then $U_1, \dots, U_n$ is a sequence of independent random variables with uniform distribution on $[0,1]$. Let us denote its empirical distribution function by $F_n(x)$. Define

$$T_1^{-1}(x) = \tilde{H}^{-1}(x) = \min\{y, \tilde{H}(y) = x\}$$

and

$$T_2^{-1}(x) = \left[1 - \tilde{\tilde{H}}(x)\right]^{-1} = \max\left\{y, 1 - \tilde{\tilde{H}}(y) = x\right\}.$$

Observe that T_1^{-1} is a measure preserving transformation from

$$\big([0, \tilde{H}(\infty)], d\big[\sqrt{n}(F_n(x) - x)\big]\big) \quad \text{to} \quad \big(\mathbb{R}^1, d\big[\sqrt{n}(\tilde{H}_n(x) - \tilde{H}(x))\big]\big)$$

and T_2^{-1} from

$$\big([\tilde{H}(\infty), 1], d\big[\sqrt{n}(F_n(x) - x)\big]\big) \quad \text{to} \quad \Big(\mathbb{R}^1, d\Big[\sqrt{n}\Big(\tilde{\tilde{H}}_n(x) - \tilde{\tilde{H}}(x)\Big)\Big]\Big).$$

Clearly

$$\sqrt{n}(H_n(x) - H(x)) = \sqrt{n}(\tilde{H}_n(x) - \tilde{H}(x)) + \sqrt{n}\Big(\tilde{\tilde{H}}_n(x) - \tilde{\tilde{H}}(x)\Big). \tag{3.8}$$

By exploiting the measure preserving property of T_1^{-1} and T_2^{-1} and by decomposing $d[\sqrt{n}(H_n(x) - H(x))]$ with the help of (3.8), we get from (2.5) and (2.7) that

$$-nR_4(u) = nB_3(u)$$
$$= \int_0^{\tilde H(\infty)} \int_0^1 \left[\frac{I(T_1^{-1}(y) \le u) I(T_1^{-1}(x) > T_1^{-1}(y), x \le \tilde H(\infty))}{(1 - H(T_1^{-1}(y)))^2} + \frac{I(T_1^{-1}(y) \le u) I(T_2^{-1}(x) > T_1^{-1}(y), x > \tilde H(\infty))}{(1 - H(T_1^{-1}(y)))^2} \right]$$
$$\times d\left[\sqrt{n}\,(F_n(x) - x)\right] d\left[\sqrt{n}\,(F_n(y) - y)\right]$$
$$= \int_0^1 \int_0^1 \frac{I(y \le \tilde H(u)) I\left(y < x < 1 - \tilde{\tilde H}(\tilde H^{-1}(y))\right)}{(1 - H(\tilde H^{-1}(y)))^2}$$
$$\times d\left[\sqrt{n}\,(F_n(x) - x)\right] d\left[\sqrt{n}\,(F_n(y) - y)\right].$$

This implies that

$$\sup_{u \le T} |nR_4(u)| = \sup_{u \le T} \left| \int_0^{H(u)} \int_0^1 f(x, y) d\left[\sqrt{n}\,(F_n(x) - x)\right] d\left[\sqrt{n}\,(F_n(y) - y)\right] \right| \tag{3.9}$$

with

$$f(x, y) = \frac{I\left(y < x < 1 - \tilde{\tilde H}(\tilde H^{-1}(y))\right) I(y \le \tilde H(T))}{(1 - H(\tilde H^{-1}(y)))^2}.$$

Observe that $|f(x, y)| \le 1/\delta^2$ since $y \le \tilde H(T)$ implies that $1 - H(\tilde H^{-1}(y)) \ge 1 - H(T) \ge \delta$. We get Lemma 3 by applying Lemma 6 for the expression (3.9). □

PROOF OF LEMMA 4. Let $U_1, \dots, U_n$ be defined by formulas (3.6) and (3.7) and let $F_n(t)$ be its empirical distribution function just as in Lemma 3. By Theorem 3 [Komlós, Major and Tusnády (1975)] a Brownian bridge $S(t)$ can be constructed in such a way that

$$P\left(\sup_{0 \le t \le 1} \sqrt{n}\,|S(t) - \sqrt{n}\,(F_n(t) - t)| > C \log n + z \right) < Ke^{-\lambda z}.$$

Since

$$\sqrt{n}\left(\tilde H_n(u) - \tilde H(u)\right) = \sqrt{n}\left(F_n(\tilde H(u)) - \tilde H(u)\right)$$

and

$$\sqrt{n}\left(\tilde{\tilde H}_n(u) - \tilde{\tilde H}(u)\right) = \sqrt{n}\left(\left(1 - \tilde{\tilde H}(u)\right) - F_n\left(1 - \tilde{\tilde H}(u)\right)\right),$$

this relation implies the first statement of Lemma 4. The proof of the second statement is the same, only Theorem 4 of Komlós, Major and Tusnády (1975) must be applied instead of Theorem 3. □

REFERENCES

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 approximations of some biometric estimates under random censorship. *Z. Wahrsch. verw. Gebiete* **56** 87–112.

CSÖRGŐ, S. and HORVÁTH, L. (1983). The rate of strong uniform consistency for the product-limit estimator. *Z. Wahrsch. verw. Gebiete* **62** 411–426.

DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and the classical multinomial estimator. *Ann. Math. Statist.* **27** 642–669.

FÖLDES, A. and REJTŐ, L. (1981). Strong uniform consistency for nonparametric survival estimates from randomly censored data. *Ann. Statist.* **9** 122–129.

GILL, R. D. (1980). *Censoring and Stochastic Integrals. Math. Centre Tracts* **124**. Math. Centrum, Amsterdam.

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

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent r.v.'s and the sample d.f. I. *Z. Wahrsch. verw. Gebiete* **32** 111–132.

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

MAJOR, P. (1988). On the tail behaviour of the distribution function of multiple stochastic integrals. *Probab. Theory Related Fields*. To appear.

RÉNYI, A. (1970). *Probability Theory*. Akademia, Budapest.

SUSARLA, V. and VAN RYZIN, J. (1978). Large sample theory for a Bayesian nonparametric survival curve estimator based on censored samples. *Ann. Statist.* **6** 755–768.

WELLNER, J. A. (1978). Limit theorems for the ratio of the empirical distribution function to the true distribution function. *Z. Wahrsch. verw. Gebiete* **45** 73–88.

MATHEMATICAL INSTITUTE OF THE
HUNGARIAN ACADEMY OF SCIENCES
P.O. BOX 127
H-1364 BUDAPEST
HUNGARY