The Annals of Statistics
1986, Vol. 14, No. 3, 1240–1245

LOCAL CONVERGENCE OF EMPIRICAL MEASURES IN THE RANDOM CENSORSHIP SITUATION WITH APPLICATION TO DENSITY AND RATE ESTIMATORS

BY HELMUT SCHÄFER

University of Heidelberg

In this paper, we study the local deviations of the empirical measure defined by the Kaplan–Meier (1958) estimator for the survival function. The results are applied to derive best rates of convergence for kernel estimators for the density and hazard rate function in the random censorship model.

1. Introduction. In the random censorship model, instead of the random variables T_i of interest, one observes variables $X_i = \min(T_i, C_i)$ and indicators $\delta_i = I(T_i < C_i)$, $i \in \mathbb{N}$. The T_i are i.i.d. and nonnegative, and so are the censoring variables C_i. T_i and C_i are assumed to be independent for all i.

In this situation, the product-limit estimator F_n introduced by Kaplan and Meier (1958) is widely used to estimate the distribution (survival) function $F(x) = P(T > x)$ from the observations. Földes and Rejtö (1981) proved strong uniform convergence of this estimator with rate of $O(\sqrt{\log(n)/n})$. In many applications, however, the convergence of the empirical measure dF_n is only needed locally, i.e., on intervals $I_n \subset \mathbb{R}$ with probability mass p_n tending to 0 as the sample size n increases. Exploiting the faster decrease of the variance of $\int_{I_n} dF_n$, it is possible to derive smaller rates of convergence on such sets of intervals. For the empirical probability measure defined by i.i.d. random variables, Stute (1982a) proves $\sup_{\int_I dF \leqq p_n} |\int_I dF_n - \int_I dF| = O(\sqrt{p_n \log(n)/n})$ a.s. under appropriate conditions on the sequence $p_n \to 0$, where the sup is taken over all intervals $I \subset \mathbb{R}$ with probability mass $\leqq p_n$.

In the present paper, we show that this result remains true for the Kaplan–Meier estimator F_n in the random censorship model. In other words, we study the local oscillation behaviour of a certain empirical process $F_n(t) - F(t)$, $t \in \mathbb{R}$, having dependent increments. The estimator H_n for the cumulative hazard function $H(x) = \int_0^x (dF(t))/(F(t))$, introduced by Nelson (1969), may be treated the same way. Indeed, the Kaplan–Meier estimator is even somewhat clumsier.

In Section 3, these results are applied to kernel estimators for the density function f and the hazard rate h of T. In a general form, these estimators may be written as

$$f_n(x) = \int R_n^{-1}(t) K((x-t)/R_n(t))\, dF_n(t)$$

(for density estimation) with a kernel function K integrating to 1 and a random

Received February 1985; revised November 1985.
AMS 1980 *subject classifications*. 62G05, 62P10.
Key words and phrases. Random censorship model, empirical measures, convergence rates, kernel density estimation, sample-point-dependent bandwidths.

process R_n. For deterministic fixed bandwidths $R_n(t) = d_n \to 0$, asymptotic properties of hazard rate estimators were studied by Ramlau-Hansen (1983), Tanner and Wong (1983), and Yandell (1983). Tanner (1983) showed pointwise consistency for random bandwidths R_n depending on the point of interest x. Schäfer (1985) proved strong uniform consistency for an estimator with bandwidths depending on the sample point

$$R_n(t) = \inf\{r > 0 | F_n(t - r/2) - F_n(t + r/2) \geqq p_n\}$$

($p_n \to 0$ again a sequence of positive real numbers), modelled upon the variable kernel estimator of Breiman, Meisel, and Purcell (1977) and Victor (1976).

It is well known that kernel estimators (with fixed bandwidths) converge uniformly with a rate of $O(\sqrt{\log(n)/(np_n)} + p_n)$ in the uncensored case (Silverman (1978); Stute (1982b)). We show that the same rate holds for randomly censored data, taking the example of the aforementioned data-adaptive estimator, which has found little attention in the literature. We remain in the context of density estimation. For hazard rates, simply replace the Kaplan–Meier estimator F_n by the Nelson estimator H_n and apply the corresponding convergence result.

2. Notation and assumptions. We make the usual general assumptions: The distribution (survival) functions $F(x) = P(T > x)$ of T and $G(x) = P(C > x)$ of C and the subsurvival function $\tilde{F}(x) = P(X > x \text{ and } \delta = 1)$ are continuous. The results are restricted to an interval $[0, B]$ with B such that $0 < F(B)G(B) < 1$. T is supposed to have a density function f which is positive and bounded on $[0, B]$ (i.e., $0 < m < f < M$) and satisfies a Lipschitz condition $|f(x) - f(y)| \leqq L_f|x - y|$ for $x, y \in [0, B]$.

For density estimation, the kernel function K is assumed to have compact support on $[-\frac{1}{2}, \frac{1}{2}]$, to integrate to 1, and to satisfy also a Lipschitz condition with constant denoted by L_K. (This might be replaced by monotonicity conditions, for example.) The sequence (p_n) must be chosen such that $p_n \to 0$ and $np_n/\log(n) \to \infty$.

In the absence of ties, the Kaplan–Meier estimator is well defined by

$$F_n(x) = \prod_{X_\iota \leqq x} \left((N_n(X_\iota) - 1)/N_n(X_\iota)\right)^{\delta_\iota},$$

with $N_n(x) = \sum_{\iota=1}^n I(X_i \geqq x)$ the number of individuals at risk at time $x - 0$. We denote by G_n the corresponding estimator for the censoring curve G, obtained by substituting $1 - \delta_\iota$ for δ_i. $\tilde{F}_n$ denotes the canonical estimator for $\tilde{F}$ defined by $\tilde{F}_n(x) = 1/n\sum_{\iota=1}^n \delta_\iota I(X_i \geqq x)$.

Finally, by abuse of notation, we write $F(A) := \int_A dF$ for the measure defined by any distribution function F.

3. Local convergence of F_n. Clearly, representing F_n as a sum of random variables suitable for approximation by i.i.d. ones is basic to the aim of investigating local properties of F_n. An easy transformation of $F_n(X_{(i)} - 0) - F_n(X_{(\iota)})$ for an uncensored observation $X_{(i)}$ using the above definition of F_n yields the representation

$$dF_n = G_n^{-1}\, d\tilde{F}_n,$$

which is fundamental to our procedure. We will approximate dF_n by

$$dF_n^* := G^{-1}\, d\tilde{F}_n$$

using the result by Földes and Rejtö (1981) cited in the introduction.

THEOREM 3.1. *For* $0 \leqq p \leqq 1$ *and* $0 < \varepsilon < 1$,

$$P\Big(\sup_{F(I)\leqq p} |F_n^*(I) - F(I)| > \varepsilon\Big) < C\varepsilon^{-2}\exp\big(-Cn\varepsilon^2(p+\varepsilon)^{-1}\big),$$

with constants C only depending on F and G. The sup *is taken over all intervals* $I \subset [0, B]$ *with mass less than or equal to p.*

PROOF. For a fixed single $I \subset [0, B]$, $F_n^*(I) - F(I)$ is the mean $\overline{Y}_n$ of n i.i.d. random variables distributed as $Y = G^{-1}(T)I(T < C)I(T \in I) - F(I)$. The calculation of the expectation and the variance is straightforward (use the independence of T and C and $\int_{c>t} G^{-1}(t)\, dG(c) = 1$):

$$E(Y) = 0,$$

$$\sigma^2 := \operatorname{var}(Y) = \int_I G^{-1}(t)\, dF(t) - F(I)^2 \leqq bF(I),$$

$$|Y| \leqq b,$$

where $b := G^{-1}(B)$. In this situation, it is standard to use the exponential bound

$$P(|F_n^*(I) - F(I)| > \varepsilon) \leqq 2\exp\big(-n\varepsilon^2/2(\sigma^2 + b\varepsilon)\big)$$

derived from the inequality of Bernstein (1924) as cited by Bennett (1962).

Now consider the finite system of compact intervals

$$S = \big\{[F^{-1}(i\varepsilon), F^{-1}(j\varepsilon)]\,|\,i = 0,\dots,\varepsilon^{-1}+1;\ j = i,\dots,i+p\varepsilon^{-1}+3\big\}.$$

(Take the integer part of ε^{-1}, $p\varepsilon^{-1}$ and always $F^{-1}(i\varepsilon) \leqq F^{-1}(j\varepsilon) \leqq B$).

Since $F(I) \leqq p + 3\varepsilon$ for every $I \in S$, and so $\sigma^2 \leqq bp + 3b\varepsilon$, and since card $(S) \leqq \text{const}/\varepsilon^2$, we obtain the desired exponential bound for $P(\sup_{I\in S}|F_n^*(I) - F(I)| > \varepsilon)$. It remains to remark that the inverse of this last inequality implies $|F_n^*(I) - F(I)| \leqq 3\varepsilon$ for all intervals I with $F(I) \leqq p$: Indeed, for such I there exist $I_1, I_2 \in S$ with $I_1 \subset I \subset I_2$ and $F(I_2) - 2\varepsilon \leqq F(I) \leqq F(I_1) + 2\varepsilon$. □

COROLLARY 3.2. *Let* $p_n \to 0$ *with* $np_n/\log(n) \to \infty$. *Then*

$$\sup_{F(I)\leqq p_n} |F_n^*(I) - F(I)| = O\big(\log(n)^{1/2} p_n^{1/2} n^{-1/2}\big) \quad a.s., \tag{1}$$

$$\sup_{F(I)\leqq p_n} |F_n(I) - F_n^*(I)| = O\big(\log(n)^{1/2} p_n n^{-1/2}\big) \quad a.s. \tag{2}$$

Again, the sup *are taken over all intervals with mass* $\leqq p_n$.

PROOF. The condition on p_n implies $p_n + \log(n)^{1/2}p_n^{1/2}n^{-1/2} = O(p_n)$. (1) then follows from the theorem by Borel–Cantelli. For (2),

$$|F_n(I) - F_n^*(I)| = \int_I |G_n^{-1} - G^{-1}|\, d\tilde{F}_n$$

$$= \int_I |G_n^{-1} - G^{-1}| G\, dF_n^* \leqq \sup_I |G_n - G| G_n^{-1}(B) F_n^*(I).$$

By Theorem 3.2 of Földes and Rejtö (1981), $\sup_I |G_n - G| = O(\log(n)^{1/2}n^{-1/2})$ and $G_n^{-1}(B) = b + O(\log(n)^{1/2}n^{-1/2})$. By (1),

$$F_n^*(I) = p_n + O\big(\log(n)^{1/2}p_n^{1/2}n^{-1/2}\big) = O(p_n). \square$$

4. Convergence results for the variable kernel estimator. Let $r_n(t) = \inf\{r > 0 | F(t - r/2) - F(t + r/2) \geqq p_n\}$ be the deterministic analogue of the bandwidths $R_n(t)$ defined in Section 1, and define

$$f_n^1(x) = \int r_n^{-1}(t) K((x - t)/r_n(t))\, dF_n(t)$$

and

$$f_n^2(x) = \int r_n^{-1}(t) K((x - t)/r_n(t))\, dF(t).$$

THEOREM 4.1. *Let $0 < a < b < B$. Under the conditions assumed in Section 2*

$$\sup_{x \in [a, b]} |f_n(x) - f_n^1(x)| = O\big(\log(n)^{1/2}n^{-1/2}p_n^{-1/2}\big) \quad a.s., \tag{3}$$

$$\sup_{x \in [a, b]} |f_n^1(x) - f_n^2(x)| = O\big(\log(n)^{1/2}n^{-1/2}p_n^{-1/2}\big) \quad a.s., \tag{4}$$

$$\sup_{x \in [a, b]} |f_n^2(x) - f(x)| = O(p_n). \tag{5}$$

PROOF. The following statements are valid uniformly in $x \in [a, b]$ (constants depend on f and K only) for large n, and with probability 1 as far as random variables are involved. The result of Corollary 3.2. is used for $2Mp_n/m$ (see Section 2 for notations) in the place of p_n:

$$\sup_{|x-y| \leqq 2Mp_n/m} |F_n[x, y] - F[x, y]| < C\log(n)^{1/2}p_n^{1/2}n^{-1/2}. \tag{*}$$

The sequence on the right-hand side will be denoted by ε_n in the following. Note $\varepsilon_n/p_n \to 0$ by hypothesis on p_n. For $x \in [a, b]$, put $a_n(x) = x - 2p_n/m$, $b_n(x) = x + 2p_n/m$, and $I_n(x) = [a_n(x), b_n(x)]$. Since, for example, $F[x, b_n(x)] \geqq 2p_n$ and, by (*), $F_n[x, b_n(x)] > p_n$, we get $|x - t| > r_n(t)/2$ and

$|x - t| > R_n(t)/2$ for $t \notin I_n(x)$. Due to $\mathrm{Supp}(K) \subset [-\frac{1}{2}, \frac{1}{2}]$, the integrals defining $f_n(x)$, $f_n^1(x)$, and $f_n^2(x)$ may thus be restricted to integration over $I_n(x)$.

(3) (*) implies

$$\inf_{t \in [a,b]} F_n\left[t - \tfrac{1}{2}\left(r_n(t) + \varepsilon_n/m\right), t + \tfrac{1}{2}\left(r_n(t) + \varepsilon_n/m\right)\right] > p_n$$

and the corresponding upper bound for the interval defined by subtracting ε_n/m. Hence, by definition of $R_n(t)$, $\sup_{t \in [a,b]} |R_n(t) - r_n(t)| \leqq \varepsilon_n/m$. In combination with boundedness of $r_n(t)/p_n$ and of K, and the Lipschitz condition for K, this shows

$$\sup_{t \in I_n(x)} |R_n^{-1}(t)K((x-t)/R_n(t)) - r_n^{-1}(t)K((x-t)/r_n(t))| = O(\varepsilon_n p_n^{-2})$$

(insert the mixed term $r_n^{-1}(t)K((x-t)/R_n(t))$). This difference has to be integrated over $I_n(x)$, so (3) follows by $F_n(I_n(x)) = O(p_n)$ (use (*) again).

(4) We abbreviate $k_n^x(t) := r_n^{-1}(t)K((x-t)/r_n(t))$ and $\alpha_n^x(t) := F_n[a_n(x), t] - F[a_n(x), t]$ for $t \in I_n(x)$. Obviously, $|r_n(t) - r_n(s)| \leqq |t - s|$ for all $t, s \in [a, b]$. Combined with $r_n \geqq p_n/M$, this implies that the total variation $V_{I_n(x)}(r_n^{-1})$ of $1/r_n$ over $I_n(x)$ is bounded by $4M^2/\mathrm{mp}_n$. Repeated application of the formulae $V(uv) \leqq \sup|u|V(v) + \sup|v|V(u)$ and $V(K \circ u) \leqq L_K \cdot V(u)$ leads to $V_{I_n(x)}(k_n^x) = O(p_n^{-1})$. Now, by the standard argument using signed measures and integration by parts,

$$\begin{aligned} |f_n^1(x) - f_n^2(x)| &= \left|\int_{I_n(x)} k_n^x(t)\, d\alpha_n^x(t)\right| \\ &= \left|-\int_{I_n(x)} \alpha_n^x(t)\, dk_n^x(t)\right| \\ &\leqq \sup_{t \in I_n(x)} |\alpha_n^x(t)|\, V_{I_n(x)}(k_n^x). \end{aligned}$$

The above bound for $V(k_n^x)$ and (*) for the first factor complete the proof.

(5) The Lipschitz condition for f yields the further approximation

$$\sup_{t \in I_n(x)} |p_n/f(x) - r_n(t)| = O(p_n^2)$$

for the bandwidths. Now write

$$f(x) = \int_{I_n(x)} (f(x)/p_n)K((f(x)/p_n)(x-t))f(x)\, dt$$

and

$$f_n^2(x) = \int_{I_n(x)} r_n^{-1}(t)K(r_n^{-1}(t)(x-t))f(t)\, dt$$

and use this approximation together with $|f(t) - f(x)| = O(p_n)$ for $t \in I_n(x)$. □

REFERENCES

AALEN, O. (1978). Nonparametric estimation of partial transition probabilities in multiple decrement models. *Ann. Statist.* **6** 534–545.

BENNETT, G. (1962). Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* **57** 33–45.

BERNSTEIN, S. (1924). Sur une modification de l'inéqualité de Tchebichef. *Ann. Sci. Inst. Sav. Ukraine, Sect. Math.* **1**.

BREIMAN, L., MEISEL, W. and PURCELL, E. (1977). Variable kernel estimates of multivariate densities. *Technometrics* **19** 135–144.

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

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

NELSON, W. (1969). Hazard plotting for incomplete failure data. *J. Qual. Tech.* **1** 27–52.

RAMLAU-HANSEN, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11** 453–466.

SCHÄFER, H. (1985). A note on data-adaptive kernel estimation of the hazard and density function in the random censorship situation. *Ann. Statist.* **13** 818–820.

SILVERMAN, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.* **6** 177–184.

STUTE, W. (1982a). The oscillation behaviour of empirical processes. *Ann. Probab.* **10** 86–107.

STUTE, W. (1982b). A law of the logarithm for kernel density estimators. *Ann. Probab.* **10** 414–422.

TANNER, M. A. (1983). A note on the variable kernel estimator of the hazard function from randomly censored data. *Ann. Statist.* **11** 994–998.

TANNER. M. A. and WONG, W. H. (1983). The estimation of the hazard function from randomly censored data by the kernel method. *Ann. Statist.* **11** 989–993.

VICTOR, N. (1976). Non-parametric allocation rules. In *Decision Making and Medical Care* (De Dombal and Grémy, eds.) 515–527. North Holland, Amsterdam.

YANDELL, B. S. (1983). Non-parametric inference for rates and densities with censored serial data. *Ann. Statist.* **11** 1119–1135.

INSTITUT FÜR MEDIZINISCHE
STATISTIK UND DOKUMENTATION
UNIVERSITÄT HEIDELBERG
IM NEUENHEIMER FELD 325
D-6900 HEIDELBERG
WEST GERMANY