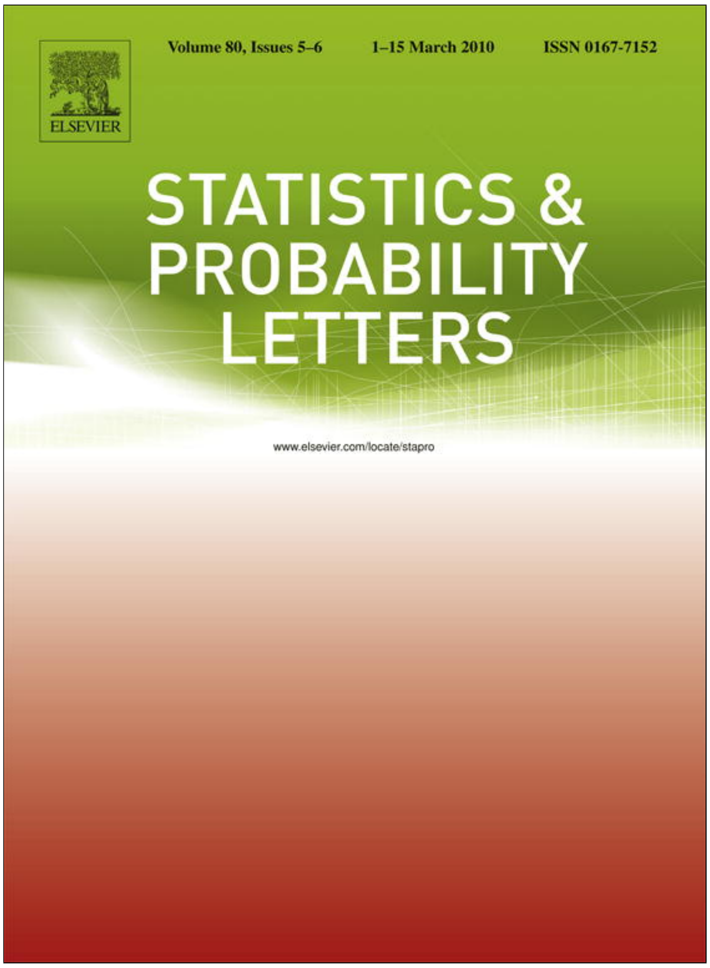
Volume 80, Issues 5–6
1–15 March 2010
ISSN 0167-7152
ELSEVIER
STATISTICS &
PROBABILITY
LETTERS
www.elsevier.com/locate/stapro

Statistics and Probability Letters 80 (2010) 318–323

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Strong uniform consistency of kernel density estimators under a censored dependent model

V. Fakoor

Department of Statistics, Ordered and Spatial Data Center of Ferdowsi University of Mashhad, P.O. Box: 1159-91775, Iran

ARTICLE INFO

Article history:
Received 5 July 2009
Received in revised form 7 November 2009
Accepted 10 November 2009
Available online 18 November 2009

ABSTRACT

Problems with censored data arise frequently in survival analyses and reliability applications. The estimation of the density function of the lifetimes is often of interest. In this paper, the estimation of the density function by the kernel method is considered, when censored data show some kind of dependence. We apply the strong Gaussian approximation technique for studying the strong uniform consistency for kernel estimators of the density function under a censored dependent model.

1. Introduction and main result

In medical follow-up or in engineering life testing studies, one may not be able to observe the variable of interest, referred to hereafter as the lifetime. Let $X_1, \ldots, X_n$ be a sequence of lifetimes, having a common unknown continuous marginal distribution function (d.f.) F, with a density function $f = F'$. The random variables are not assumed to be mutually independent (see Assumption (1) for the kind of dependence stipulated). Let the random variable X_i be censored on the right by the random variable Y_i, so that one observes only

$$Z_i = X_i \wedge Y_i \quad \text{and} \quad \delta_i = I(X_i \le Y_i),$$

where $\wedge$ denotes the minimum and $I(.)$ is the indicator of the event specified in parentheses. In this random censorship model, we assume that the censoring random variables $Y_1, \ldots, Y_n$ are not mutually independent (see Assumption (2) for the kind of dependence stipulated), having a common unknown continuous d.f. G, and that they are independent of the X_i's. Since censored data traditionally occur in lifetime analysis, we assume that X_i and Y_i are nonnegative. The actually observed Z_i's have a distribution function H satisfying

$$\overline{H}(t) = 1 - H(t) = (1 - F(t))(1 - G(t)).$$

Denote by

$$F_*(t) = P(Z \le t, \delta = 1),$$

the sub-distribution function for the uncensored observations, and by f_* the corresponding sub-density. Define by

$$N_n(t) = \sum_{i=1}^{n} I(Z_i \le t, \delta = 1) = \sum_{i=1}^{n} I(X_i \le t \wedge Y_i),$$

the number of uncensored observations less than or equal to t, and by

$$Y_n(t) = \sum_{i=1}^{n} I(Z_i \ge t),$$

the number of censored or uncensored observations greater than or equal to t and also the empirical distribution functions of $\overline{H}(t)$ and $F_*(t)$ are respectively defined as

$$\overline{Y}_n(t) = n^{-1} Y_n(t), \qquad \overline{N}_n(t) = n^{-1} N_n(t).$$

E-mail addresses: fakoor@math.um.ac.ir, vfbafandeh@yahoo.com.

0167-7152/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2009.11.005

Then the Kaplan–Meier estimator for $1-F(t)$, based on n pairs $\{(Z_i,\delta_i),\ 1\le i\le n\}$ is given by

$$1-\widehat{F}_n(t)=\prod_{s\le t}\left(1-\frac{\mathrm{d}N_n(s)}{Y_n(s)}\right), \tag{1.1}$$

where $\mathrm{d}N_n(t)=N_n(t)-N_n(t^-)$ and $N_n(t^-)=\lim_{\epsilon\to 0^+}N_n(t-\epsilon)$.

In the independence framework with no censoring, the kernel estimate f_n of a real univariate density f introduced by Rosenblatt (1956) and defined by

$$f_n(t)=\sum_{i=1}^{n}\frac{1}{nh_n}K\left(\frac{t-X_i}{h_n}\right),$$

where $X_1,\ldots,X_n$ are independent observations from the density, K is a kernel function, and h_n is a sequence of (positive) "bandwidths" tending to zero as $n\to\infty$. Parzen (1962) showed that under some mild smoothness conditions on K (and f), $f_n(t)$ is in any respect a consistent estimator of $f(t)$ for each $t\in\mathbb{R}$. The weak and strong uniform consistency properties of f_n have been considered by several authors, including Nadaraya (1965), Schuster (1969) and Van Ryzin (1969). In these papers the condition placed on the bandwidth for strong uniform consistency include $\sum\exp(-cnh_n{}^2)<\infty$ for all positive c. Silverman (1978) established the strong uniform consistency for f_n-f using the strong approximation technique developed by Komlós et al. (1975) for the ordinary empirical process. In censored case, based on the Kaplan–Meier estimator $\widehat{F}_n$, Blum and Susarla (1980) proposed to estimate the density function f by a sequence of kernel estimators f_n defined by

$$f_n(t)=\frac{1}{h_n}\int_0^\infty K\left(\frac{t-s}{h_n}\right)\mathrm{d}\widehat{F}_n(s), \tag{1.2}$$

where K is a kernel function having finite support on $(-1,1)$ and h_n is a sequence of positive bandwidths tending to 0 as $n\to\infty$. The properties of the kernel estimator f_n have been examined by Blum and Susarla (1980), Földes et al. (1981) and Mielniczuk (1986), among others. Zhang (1998) established the strong uniform consistency for f_n-f using the strong approximation technique developed by Burke et al. (1981, 1988) for the product-limit process $Z_n(t):=\sqrt{n}[\widehat{F}_n(t)-F(t)]$.

In the case where $\{X_i,i\ge 1\}$ and $\{Y_i,i\ge 1\}$ are two independent α-mixing sequences (see Definition 1), by the strong representation for the density estimators, Cai (1998) established uniform consistency (with rate) of the kernel estimators for density.

We consider a sequence of kernel estimators f_n defined by (1.2) for estimate of, $f(x)$, density function. The main aim of this paper is to derive strong uniform consistency of kernel density, for the case in which the underling lifetimes are assumed to be α-mixing whose definition is given below.

Definition 1. Let $\{X_i,i\ge 1\}$ denote a sequence of random variables. Given a positive integer m, set

$$\alpha(m)=\sup_{k\ge 1}\{|P(A\cap B)-P(A)P(B)|;A\in\mathcal{F}_1^k,B\in\mathcal{F}_{k+m}^\infty\}, \tag{1.3}$$

where $\mathcal{F}_i^k$ denote the σ-field of events generated by $\{X_j;i\le j\le k\}$. The sequence is said to be α-mixing (strongly mixing) if the mixing coefficient $\alpha(m)\to 0$ as $m\to\infty$.

Among various mixing conditions used in the literature, α-mixing, is reasonably weak and has many practical applications. There exists many processes and time series fulfilling the strong mixing condition. As a simple example we can consider the Gaussian AR(1) process for which

$$Z_t=\rho Z_{t-1}+\varepsilon_t,$$

where $|\rho|<1$ and ε_t's are independently identically distributed random variables with standard normal distribution. It can be shown (see Ibragimov and Linnik, 1971, pp. 312–313) that $\{Z_t\}$ satisfies strong mixing condition. The stationary autoregressive-moving average (ARMA) processes, which are widely applied in time series analysis, are α-mixing with exponential mixing coefficient, i.e., $\alpha(n)=\mathrm{e}^{-\nu n}$ for some $\nu>0$. The threshold models, the EXPAR models (see Ozaki, 1979), the simple ARCH models (see Engle, 1984; Masry and Tjostheim, 1995, 1997) and their extensions (see Diebolt and Guégan, 1993) and the bilinear Markovian models are geometrically strongly mixing under some general ergodicity conditions. Auestad and Tjostheim (1990) provided excellent discussions on the role of α-mixing for model identification in nonlinear time series analysis.

It is the purpose of this paper to study the strong uniform consistency for f_n-f, using the strong Gaussian approximation technique obtained by Fakoor and Nakhaei Rad (2009) for the product-limit process. Our approach is first to apply the strong approximation technique to establish the strong uniform consistency of $f_n-\tilde{f}_n$, where

$$\tilde{f}_n(t)=\frac{1}{h_n}\int_0^\infty K\left(\frac{t-s}{h_n}\right)\mathrm{d}F(s). \tag{1.4}$$

Now, for the sake of simplicity, the assumptions used in this paper are as follows.

Assumptions

(1) Suppose that $\{X_i,i\ge 1\}$ is a sequence of stationary α-mixing random variables with continuous distribution function F, survival function $S(.)$ and mixing coefficient $\alpha_1(n)$.

(2) Suppose that $\{Y_i, i \geq 1\}$ is a sequence of stationary α-mixing random variables with continuous distribution function G and mixing coefficient $\alpha_2(n)$. Moreover, we assume the censoring times are independent of $\{X_i, i \geq 1\}$.
(3) $\alpha(n) = O(\mathrm{e}^{-(\log n)^{1+\nu}})$ for some $\nu > 0$, with $\alpha(n) = \max(\alpha_1(n), \alpha_2(n))$ (see Remark 2.1 in Ould-Saïd and Sadki, 2005, for details).
(4) Suppose that f is continuous on $[0, \tau]$, where $\tau = \sup\{t : \overline{H}(t) > 0\}$.
(5) Suppose that the symmetric kernel function K satisfies $\int_{-1}^{1} K(t)\mathrm{d}t = 1$, $\int_{-1}^{1} tK(t)\mathrm{d}t = 0$, $K(t) = 0$ if $t \notin (-1, 1)$ and is of bounded variation on $(-1, 1)$ with total variation denoted by V_K.
(6) Suppose that f has a bounded second derivative on $[0, \tau]$.

Our main result is the following theorem.

Theorem 1. *Let h_n be a sequence of positive bandwidths tending to zero as $n \to \infty$. Suppose that Assumptions* (1)–(5) *hold and that*

$$\lim_{n\to\infty} \frac{(\log n)^{-\lambda}}{\sqrt{n}h_n} = 0, \tag{1.5}$$

for some $\lambda > 0$. Then, for any $\epsilon > 0$,

$$\lim_{n\to\infty} \sup_{0\leq t\leq \tau-\epsilon} |f_n(t) - f(t)| = 0 \quad a.s. \tag{1.6}$$

Proof. See the Appendix. □

An inspection of the proof of Theorem 1, gives the rate of strong uniform consistency for $f_n - \widetilde{f}_n$.

Lemma 1. *Under the same conditions as in Theorem 1, we have*

$$\sup_{0\leq t\leq \tau-\epsilon} |f_n(t) - \widetilde{f}(t)| = O\left(\sqrt{\frac{\log\log n}{n}}\right) + O\left(\frac{(\log n)^{-\lambda}}{\sqrt{n}h_n}\right) \quad a.s.$$

Remark 1. If the bandwidth h_n is chosen to be $h_n \sim \alpha n^{-\beta}$ with $\alpha > 0$ and $0 < \beta \leq \frac{1}{2}$, then condition (1.5) is satisfied.

Remark 2. In the independence framework with no censoring (with censoring), for suitable kernels, Silverman (1978) (Zhang, 1998) showed that the condition $h_n^{-1} = o(n/\log n)$ as $n \to \infty$ is sufficient for strong uniform consistency of kernel density estimates. In the α-mixing case with censoring, we cannot achieve the same rate as in the iid case.

Using strong Gaussian approximation in Lemma Fakoor and Nakhaei Rad (2009) for the product-limit process, we can find a two parameter Gaussian process which strongly uniform approximate the empirical density process as shown in this subsection. Let

$$\psi_n(t, s) = \frac{1}{h_n} K\left(\frac{t-s}{h_n}\right). \tag{1.7}$$

Theorem 2. *Let $\psi_n(t, s)$ be a sequence of functions defined on* (1.7)*. Suppose that Assumptions* (1)–(6) *hold, Then*

$$\sup_t |\sqrt{n}(f_n(t) - f(t)) - \Gamma(t, n)| = O\left(\frac{(\log n)^{-\lambda}}{h_n} + \sqrt{n}h_n^2\right) \quad a.s.,$$

where

$$\Gamma(t, n) = -\int_0^\infty S(x)B(x, n)\mathrm{d}\psi_n(t, x).$$

Proof. Applying Lemma, we have

$$\begin{aligned} f_n(t) - f(t) &= (f_n(t) - \widetilde{f}_n(t)) + (\widetilde{f}_n(t) - f(t)) \\ &= \int_0^\infty \psi_n(t, x)\mathrm{d}[\widehat{F}_n(x) - F(x)] + (\widetilde{f}_n(t) - f(t)) \\ &= -\frac{1}{\sqrt{n}} \int_0^\infty Z_n(x)\mathrm{d}\psi_n(t, x) + (\widetilde{f}_n(t) - f(t)) \\ &\overset{a.s.}{=} -\frac{1}{\sqrt{n}} \int_0^\infty S(x)B(x, n)\mathrm{d}\psi_n(t, x) + O\left(\frac{(\log n)^{-\lambda}}{\sqrt{n}h_n}\right) + (\widetilde{f}_n(t) - f(t)). \end{aligned}$$

The result follows from Lemma 6.1.2 in Csörgő and Révész (1981). □

Remark 3. Theorem 2 suggests the optimal rate $h_n \sim \left(n^{-1/2}(\log n)^{-\lambda}\right)^{1/3}$ for such approximation.

Acknowledgement

The author would like to sincerely thank the anonymous referees for their valuable comments on a previous draft, which resulted in the present version of this paper. This research was supported by a grant from Ferdowsi University of Mashhad; (No. MS88073FAK).

Appendix

In order to prove main theorem, we need the following lemmas.

Lemma 2 (*Theorem 3 in Dhompongsa, 1984*). *Under Assumptions* (1) *and* (3), *there exists a Kiefer process* $\{k(s,t), s \in \mathbb{R}, t \geq 0\}$ *with covariance function*

$$E[k(s,t)k(s',t')] = \Gamma(s,s')\min(t,t')$$

and $\Gamma(s,s')$ *is defined by*

$$\Gamma(s,s') = \mathrm{Cov}(g_1(s), g_1(s')) + \sum_{k=2}^{\infty}[\mathrm{Cov}(g_1(s), g_k(s')) + \mathrm{Cov}(g_1(s'), g_k(s))],$$

where $g_k(s) = I(Z_k \leq s) - H(s)$, *such that, for some* $\lambda > 0$ *depending only on* ν, *given in Assumption* (3),

$$\sup_{t\in\mathbb{R}} |\overline{Y}_n(t) - \overline{H}(t) - k(t,n)/n| = O(b_n), \quad a.s.$$

where

$$b_n = n^{-1/2}(\log n)^{-\lambda}. \quad \square$$

Consider the following Gaussian process

$$B(t,n) = \int_0^t \frac{k(x,n)/\sqrt{n}}{(\overline{H}(x))^2}\mathrm{d}F_*(x), \tag{A.1}$$

where $k(x,n)$ is the Kiefer process in Lemma 2.

Lemma 3 (*Theorem 3 in Fakoor and Nakhaei Rad, 2009*). *Suppose that Assumptions* (1)–(3) *are satisfied. On a rich probability space, there exists a two parameter mean zero Gaussian process* $\{B(u,v)u, v \geq 0\}$ *such that,*

$$\sup_{t\geq 0} |Z_n(t) - S(t)B(t,n)| = O((logn)^{-\lambda}) \quad a.s., \tag{A.2}$$

for some $\lambda > 0$.

To study strong uniform consistency of kernel density estimators, we also need to study the modulus of continuity of approximating process $B(u,v)$. In the next lemma, we prove the global modulus of continuity of the Gaussian process $B(u,v)$.

Lemma 4. *Let* h_n *be a sequence of positive numbers for which*

$$\lim_{n\to\infty} h_n\sqrt{\log\log n} = 0. \tag{A.3}$$

Then, for any $\epsilon > 0$

$$\sup_{0\leq t\leq \tau-\epsilon}\ \sup_{-1\leq u\leq 1} |B(t-h_n u, n) - B(t,n)| = O(h_n\sqrt{\log\log n}) \quad a.s. \tag{A.4}$$

Proof. It is easy to see that,

$$|B(t-h_n u, n) - B(t,n)| \leq \left(\sup_{0\leq x\leq \tau-\epsilon}\left|\frac{k(x,n)}{\sqrt{n}}\right|\right)\left|\frac{F_*(t-h_n u) - F_*(t)}{\overline{H}^2(\tau)}\right|.$$

Let $M_{f_*} = \sup_{0\leq t\leq\tau} f_*(t)$, then it follows from the Mean Value Theorem that $|F_*(t-h_n u) - F_*(t)| \leq M_{f_*}h_n$ for $u \in [-1,1]$ and $t \in [0, \tau-\epsilon]$. Now, by the law of iterated logarithm for the Kiefer process (see, Theorem A of Berkes and Philipp, 1977) the proof is completed. $\square$

Lemma 5. *Assuming the same conditions as in Theorem 1, we have*

$$\lim_{n\to\infty}\sup_{0\le t\le \tau-\epsilon}|f_n(t)-\tilde{f}_n(t)|=0 \quad a.s.$$

Proof. According to Lemma 3, there exists Gaussian process $B(t,n)$ such that, for large n and $t \in [0,\tau-\epsilon]$, we have

$$\begin{aligned}
f_n(t)-\tilde{f}_n(t) &= -\frac{1}{\sqrt{n}h_n}\int_0^\infty Z_n(x)\mathrm{d}K\left(\frac{t-x}{h_n}\right)\\
&= \frac{1}{\sqrt{n}h_n}\int_{-1}^1 S(t-uh_n)B(t-uh_n,n)\mathrm{d}K(u)+O\left(\frac{(\log n)^{-\lambda}}{\sqrt{n}h_n}\right) \quad a.s.\\
&= \frac{1}{\sqrt{n}h_n}S(t)\int_{-1}^1 [B(t-uh_n,n)-B(t,n)]\mathrm{d}K(u)\\
&\quad+\frac{1}{\sqrt{n}h_n}\int_{-1}^1 [S(t-uh_n)-S(t)][B(t-uh_n,n)-B(t,n)]\mathrm{d}K(u)\\
&\quad+\frac{1}{\sqrt{n}h_n}B(t,n)\int_{-1}^1 [S(t-uh_n)-S(t)]\mathrm{d}K(u)+O\left(\frac{(\log n)^{-\lambda}}{\sqrt{n}h_n}\right)\\
&= I_{1n}(t)+I_{2n}(t)+I_{3n}(t)+O\left(\frac{(\log n)^{-\lambda}}{\sqrt{n}h_n}\right) \quad a.s.
\end{aligned} \tag{A.5}$$

To deal with I_{1n}, we apply Lemma 4, so we have

$$\sup_{0\le t\le\tau-\epsilon}|I_{1n}(t)|=O\left(\sqrt{\frac{\log\log n}{n}}\right) \quad a.s. \tag{A.6}$$

Let $M_f=\sup_{0\le t\le\tau}f(t)$, then it follows from the Mean Value Theorem that

$$|S(t-h_nu)-S(t)|\le M_f h_n \tag{A.7}$$

for $u\in[-1,1]$ and $t\in[0,\tau-\epsilon]$. Now applying Lemma 4 yields

$$\sup_{0\le t\le\tau-\epsilon}|I_{2n}(t)|=O\left(h_n\sqrt{\frac{\log\log n}{n}}\right) \quad a.s. \tag{A.8}$$

According to the law of iterated logarithm for the Kiefer process (see Theorem A of Berkes and Philipp, 1977), we have

$$\sup_{0\le t\le\tau-\epsilon}|B(t,n)|=O\left(\sqrt{\log\log n}\right) \quad a.s. \tag{A.9}$$

It follows from (A.7) and (A.9)

$$\sup_{0\le t\le\tau-\epsilon}|I_{3n}(t)|=O\left(\sqrt{\frac{\log\log n}{n}}\right) \quad a.s. \tag{A.10}$$

Combining (A.5), (A.6), (A.8) and (A.10) completes the proof of the lemma. □

Proof of Theorem 1. Since f is continuous on $[0,\tau]$, f is uniformly continuous on $[0,\tau]$, and hence it is easy to show by the dominated convergence theorem that

$$\lim_{n\to\infty}\sup_{0\le t\le\tau-\epsilon}|\tilde{f}_n(s)-f(s)|=0.$$

Therefore, Theorem 1 is a straightforward consequence of Lemma 5 and the equality

$$f_n-f=f_n-\tilde{f}_n+\tilde{f}_n-f. \quad \square$$

References

Auestad, B., Tjostheim, D., 1990. Identification of nonlinear time series: First order characterisation and order determination. Biometrica 77, 669–687.
Berkes, I., Philipp, W., 1977. An almost sure invariance principle for the empirical distribution function of mixing random variables. Z. Wahrscheinlichkeitstheorie. Verwandte Geb. 41, 115–137.
Blum, J.R., Susarla, V., 1980. Maximal deviation theory of density and failure rate function estimates based on censored data. In: Krishnaiah, P.K. (Ed.), Multivariate Analysis. North-Holland, New York, pp. 213–222.
Burke, M.D., Csörgő, S., Horváth, L., 1981. Strong approximations of some biometric estimates under random censorship. Z. Wahrscheinlichkeitstheorie. Verwandte Geb. 56, 87–112.
Burke, M.D., Csörgő, S., Horváth, L., 1988. A correction to and improvement of strong approximations of some biometric estimates under random censorship. Probab. Theory Related Fields 79, 51–57.
Cai, Z., 1998. Kernel density and hazard rate estimation for censored dependent data. J. Multivariate Anal. 67, 23–34.
Csörgő, M., Révész, P., 1981. Strong Approximation in Probability and Statistics. Academic Press, New York.
Diebolt, J., Guégan, D., 1993. Tail behaviour of the stationary density of general nonlinear autoregressive processes of order 1. J. Appl. Probab. 30, 315–329.
Dhompongsa, S., 1984. A note on the almost sure approximation of the empirical process of weakly dependent random variables. Yokohama Math. J. 32, 113–121.
Engle, R.F., 1984. Autoregressive conditional heteroscedasticity with estimates of the variance of united Kingdom inflation. Econometrics 50 (4), 987–1007.
Fakoor, V., Nakhaei Rad, N., 2009. Strong Gaussian approximations of product-limit and quantil processes for strong mixing and censored data. Comm. Statist. Theory Methods (in press).
Földes, A., Rejtö, L., Winter, B.B., 1981. Strong consistency properties of nonparametric estimators for randomly censored data, II: Estimation of density and failure rate. Period. Math. Hungar. 12, 15–29.
Ibragimov, I.A., Linnik, Yu.V., 1971. Independent and Stationary Sequences of Random Variables. Walters-Noordhoff, Groningen, The Netherlands.
Komlós, J., Major, P., Tusnády, G., 1975. An approximation of partial sums of independent random variables and the sample d.f.. Z. Wahrscheinlichkeitstheorie. Verwandte Geb. 32, 111–132.
Ould-Saïd, E., Sadki, O., 2005. Strong approximation of quantile function for strong mixing and censored processes. Comm. Statist. Theory Methods 34, 1449–1459.
Masry, E., Tjostheim, D., 1995. Nonparametric estimation and identification of nonlinear ARCH time series: Strong convergence and asymptotic normality. Econom. Theory 11, 258–289.
Masry, E., Tjostheim, D., 1997. Additive nonlinear ARX time series and projection estimates. Econom. Theory 13, 214–256.
Mielniczuk, J., 1986. Some asymptotic properties of kernel estimators of a density function in case of censored data. Ann. Statist. 14, 766–773.
Nadaraya, E.A., 1965. On non-parametric estimates of density functions and regression curves. Theory Probab. Appl. 10, 186–190.
Ozaki, T., 1979. Nonlinear time series models for nonlinear random vibrations. Technical report. University of Manchester.
Parzen, E., 1962. On estimation of a probability density function and mode. Ann. Math. Statist. 33, 1065–1076.
Rosenblatt, M., 1956. Remarks on some non-parametric estimates of a density function. Ann. Math. Statist. 27, 832–837.
Silverman, B.M., 1978. Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. Ann. Statist. 6, 177–184.
Schuster, E.F., 1969. Estimation of a probability density and its derivatives. Ann. Math. Statist. 40, 1187–1196.
Van Ryzin, J., 1969. On strong consistency of density estiamtes. Ann. Math. Statist. 40, 1765–1772.
Zhang, B., 1998. A note on the strong uniform consistency of kernel density estimators under random censorship. Sankhyā Ser. A 60, 265–273.