Available online at www.sciencedirect.com
ScienceDirect
C. R. Acad. Sci. Paris, Ser. I 344 (2007) 651–656
COMPTES RENDUS MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Statistics/Probability Theory

Asymptotic properties of the kernel estimator of the conditional mode for the left truncated model

Elias Ould-Saïd [a], Abdelkader Tatachak [b]

[a] L.M.P.A. J. Liouville, Université du littoral côte d'opale, BP 699, 62228 Calais, France
[b] Laboratoire M.S.T.D, U.S.T.H.B, faculté de mathématiques, BP 32, El-Alia, 16111, Algeria

Received 27 June 2006; accepted after revision 23 March 2007

Available online 7 May 2007

Presented by Paul Deheuvels

Abstract

In this Note we propose a non-parametric kernel estimator of the conditional mode function, when the variable of interest is subject to random left-truncation. We establish the rate of the strong uniform consistency of the estimate as well as its asymptotic normality. ***To cite this article: E. Ould-Saïd, A. Tatachak, C. R. Acad. Sci. Paris, Ser. I 344 (2007).***

Résumé

Propriétés asymptotiques de l'estimateur à noyau du mode conditionnel pour un modèle tronqué à gauche. Dans cette Note nous proposons un estimateur non paramétrique de la fonction mode conditionnel, lorsque la variable d'intérêt est assujettie à une troncature aléatoire à gauche. Nous établissons dans un premier temps la convergence uniforme presque sûre de l'estimateur ainsi que sa vitesse, puis nous en donnons la normalité asymptotique. ***Pour citer cet article : E. Ould-Saïd, A. Tatachak, C. R. Acad. Sci. Paris, Ser. I 344 (2007).***

Version française abrégée

Dans le modèle de données tronquées aléatoirement à gauche (TAG), on considère deux suites $(Y_i)_{i\geqslant 1}$ et $(T_i)_{i\geqslant 1}$ mutuellement indépendantes de fonctions de répartition (f.d.r) F et G respectivement. On suppose que l'on dispose d'un échantillon de taille N. La variable d'intérêt Y est dite tronquée aléatoirement à gauche par la variable T si les seules variables aléatoires observables sont (Y_i, T_i) vérifiant $Y_i \geqslant T_i$, sinon rien n'est observé. Nous noterons par n la taille de l'échantillon observé qui est connue mais aléatoire, avec $n \leqslant N$. Sous ce modèle, soit à présent, $\mathbf{X} \in \mathbb{R}^d$ un vecteur aléatoire de covariables de f.d.r L possédant une densité ℓ. Dans toute la suite, la variable T sera supposée indépendante du vecteur $(\mathbf{X}, Y)$.

Il est clair que si $\alpha := \mathbb{P}(Y \geqslant T) = 1$, le mode conditionnel $\Theta(\mathbf{x})$ de Y sachant $\mathbf{X} = \mathbf{x}$, supposé unique, est donné par :

E-mail addresses: ouldsaid@lmpa.univ-littoral.fr (E. Ould-Saïd), atatachak@usthb.dz (A. Tatachak).

1631-073X/$ – see front matter
doi:10.1016/j.crma.2007.03.023

$$\Theta(\mathbf{x}) = \underset{-\infty<y<\infty}{\arg\max}\ g(y|\mathbf{x})$$

où $g(y|\mathbf{x}) = f(\mathbf{x}, y)/\ell(\mathbf{x})$ est la densité conditionnelle de Y sachant $\mathbf{X} = \mathbf{x}$, et $f(\cdot,\cdot)$ est la densité jointe de $(\mathbf{X}, Y)$.

Sous le modèle de données TAG, nous construisons un estimateur non paramétrique du mode conditionnel $\Theta(\mathbf{x})$, basé sur l'observation de n triplets $(\mathbf{X}_i, Y_i, T_i)$.

Dans cette Note, en fixant $d = 1$, nous établissons la convergence uniforme presque sûre de $\hat{\Theta}_n(\mathbf{x})$, solution de l'équation

$$\hat{\Theta}_n(x) = \underset{-\infty<y<\infty}{\arg\max}\ \hat{g}_n(y|x)$$

où $\hat{g}_n(\cdot|x)$ est un estimateur convergent de $g(\cdot|x)$. Enfin, nous établissons la normalité asymptotique de $\hat{\Theta}_n(x)$ convenablement normalisé et affirmons que :

$$\left(\frac{nh_n^4 G(\Theta(x))(f^{(0,2)}(x,\Theta(x)))^2}{\alpha f(x,\Theta(x))\int_{\mathbb{R}^2}[K_1(r)\Lambda^{(1)}(s)]^2\,\mathrm{d}r\,\mathrm{d}s}\right)^{1/2}\big(\hat{\Theta}_n(x) - \Theta(x)\big) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

où $K_1(\cdot)$ et $\Lambda(\cdot)$ sont des noyaux, h_n est une suite telle que $h_n \searrow 0$ lorsque $n \nearrow \infty$ et $\xrightarrow{\mathcal{D}}$ signifie la convergence en loi.

Des simulations ont été effectuées pour illustrer la consistance de notre estimateur ainsi que la normalité asymptotique à distance finie, qui ne figurent pas dans la Note par manque de place ; cependant, elles peuvent être obtenues auprès des auteurs.

1. Introduction

Let Y and T be independent random variables (rv) with distribution function (df) F and G respectively, both assumed to be continuous. Let $(Y_i, T_i)_{\{i\geqslant 1\}}$ be a sequence of independent and identically distributed (iid) as (Y, T), where the population size N is deterministic but unknown. In the random left-truncation model, the rv of interest Y is interfered by the truncation rv T, in such a way that Y_i and T_i are observable only if $Y_i \geqslant T_i$, whereas nothing is observed if $Y_i < T_i$. Without possible confusion, we still denote $(Y_i, T_i)_{1\leqslant i\leqslant n}$ $(n \leqslant N)$, these observed iid pairs. As a consequence of truncation, the size of the actually observed sample, n is a $Bin(N, \alpha)$ rv, with $\alpha := \mathbb{P}(Y \geqslant T) \neq 0$.

Let $\mathbf{X}$ be a $\mathbb{R}^d$-valued random covariates vector, assumed to be absolutely continuous with df L having density ℓ, and let $g(\cdot|\mathbf{x}) = f(\mathbf{x},\cdot)/\ell(\mathbf{x})$ denote the conditional probability density function (pdf) of Y given $\mathbf{X} = \mathbf{x}$, where $f(\cdot,\cdot)$ is the joint pdf of $(\mathbf{X}, Y)$. Assuming that $g(\cdot|\mathbf{x})$ has a unique mode at the point $\Theta(\mathbf{x})$, the conditional mode function of Y given $\mathbf{X} = \mathbf{x}$, is defined by the following equation:

$$\Theta(\mathbf{x}) = \underset{-\infty<y<\infty}{\arg\max}\ g(y|\mathbf{x}). \tag{1}$$

Since N is unknown and n is known (although random), our results will not be stated with respect to the probability measure $\mathbb{P}$ (related to the N-sample) but will involve the conditional probability $\mathbf{P}$ (related to the n-sample). Note also that $\mathbb{E}$ and $\mathbf{E}$ denote the expectation operators related to $\mathbb{P}$ and $\mathbf{P}$, respectively. In the sequel, we denote by a superscript (*) any df that is associated with the truncated rv. Under the left-truncation sampling scheme, the conditional joint distribution (Stute [5]) of (Y, T) becomes:

$$\begin{aligned}\mathbf{V}^*(y,t) &= \mathbf{P}\{Y \leqslant y, T \leqslant t\} = \mathbb{P}\{Y \leqslant y, T \leqslant t \mid Y \geqslant T\}\\ &= \alpha^{-1}\int_{-\infty}^{y} G(t \wedge v)\,\mathrm{d}F(v).\end{aligned}$$

The marginal distributions and their empirical versions are defined by

$$F^*(y) := \mathbf{V}^*(y,\infty) = \alpha^{-1}\int_{-\infty}^{y} G(v)\,\mathrm{d}F(v) \quad \text{and} \quad F_n^*(y) = n^{-1}\sum_{i=1}^{n}\mathbf{1}_{(Y_i\leqslant y)},$$

$$G^*(t) := \mathbf{V}^*(\infty, t) = \alpha^{-1} \int_{-\infty}^{+\infty} G(t \wedge v)\, \mathrm{d}F(v) \quad \text{and} \quad G_n^*(t) = n^{-1} \sum_{i=1}^{n} \mathbf{1}_{(T_i \leqslant t)}.$$

Since N is unknown, in the sequel we will use a consistent estimator α_n (He and Yang [1]) of α, defined by

$$\alpha_n = G_n(y)\bigl(1 - F_n(y^-)\bigr)\big/ R_n(y), \tag{2}$$

for all y such that $R_n(y) > 0$, where $G_n(\cdot)$ and $F_n(\cdot)$ are the product-limit estimators (Lynden-Bell [2]) for $F(\cdot)$ and $G(\cdot)$ and $R_n(y) = n^{-1} \sum_{i=1}^{n} \mathbf{1}_{(T_i \leqslant y \leqslant Y_i)}$.

Suppose that one observes the n triplets $(\mathbf{X}_i, Y_i, T_i)$ among the N ones and for any df W, define $a_W = \inf\{x\colon\ W(x) > 0\}$ and $b_W = \sup\{x\colon\ W(x) < 1\}$ as the endpoints of the W support. We point out that Woodroofe [6] proved that F and G can be estimated completely only if

$$a_G \leqslant a_F, \quad b_G \leqslant b_F \quad \text{and} \quad \int_{a_F}^{\infty} \frac{\mathrm{d}F}{G} < \infty.$$

Suppose now that T is independent of $(\mathbf{X}, Y)$ and let $\mathbf{F}(\cdot\,,\cdot)$ be the joint df of $(\mathbf{X}, Y)$ related to the N-sample, we have the following trivariate conditional distribution H^*:

$$H^*(x, y, t) = \mathbf{P}\{\mathbf{X} \leqslant \mathbf{x}, Y \leqslant y, T \leqslant t\} = \alpha^{-1} \int_{\mathbf{u} \leqslant \mathbf{x}} \int_{a_G \leqslant v \leqslant y} G(v \wedge t) \mathbf{F}(\mathbf{du}, \mathrm{d}v).$$

Taking $t = +\infty$, the observed pair $(\mathbf{X}, Y)$ then has the following df $\mathbf{F}^*(\cdot, \cdot)$:

$$\mathbf{F}^*(\mathbf{x}, y) = H^*(\mathbf{x}, y, \infty) = \alpha^{-1} \int_{\mathbf{u} \leqslant \mathbf{x}} \int_{a_G \leqslant v \leqslant y} G(v) \mathbf{F}(\mathbf{du}, \mathrm{d}v). \tag{3}$$

By differentiating in the statement (3) and integrating the result over y, we obtain

$$\mathbf{F}(\mathbf{dx}, \mathrm{d}y) = \bigl[\alpha^{-1} G(y)\bigr]^{-1} \mathbf{F}^*(\mathbf{dx}, \mathrm{d}y) \quad \text{for } y \geqslant a_G \tag{4}$$

and

$$L(\mathbf{x}) = \alpha \int_{\mathbf{u} \leqslant \mathbf{x}} \int_{a_G \leqslant y} \frac{1}{G(y)} \mathbf{F}^*(\mathbf{du}, \mathrm{d}y).$$

A natural estimator of $L(\mathbf{x})$ is then given by

$$\tilde{L}_n(\mathbf{x}) = \frac{\alpha}{n} \sum_{i=1}^{n} \frac{1}{G(Y_i)} \mathbf{1}_{(\mathbf{X}_i \leqslant \mathbf{x})}. \tag{5}$$

Note that in (5) and in the sequel, the sum is taken only for the i's such that $G(Y_i) \neq 0$. Finally (5) yields the estimator of the density function of the vector $\mathbf{X}$ as

$$\tilde{\ell}_n(\mathbf{x}) := \frac{1}{a_n^d} \int_{\mathbb{R}^d} K_d\left(\frac{\mathbf{x} - \mathbf{u}}{a_n}\right) \tilde{L}_n(\mathbf{du}) = \frac{\alpha}{n a_n^d} \sum_{i=1}^{n} \frac{1}{G(Y_i)} K_d\left(\frac{\mathbf{x} - \mathbf{X}_i}{a_n}\right), \tag{6}$$

where K_d is a fixed kernel with $\int_{\mathbb{R}^d} K_d = 1$, and $(a_n)_{n \geqslant 1} \searrow 0$ as $n \nearrow \infty$. Here, we point out that the estimate (6) has been already defined by Ould-Saïd and Lemdani [3] in the context of the estimation of the regression function. While adopting the same methodology, we get an estimator of $\mathbf{F}(\cdot, \cdot)$ as follows

$$\tilde{\mathbf{F}}_n(\mathbf{x}, y) = \frac{\alpha}{n} \sum_{i=1}^{n} \frac{1}{G(Y_i)} \mathbf{1}_{(\mathbf{X}_i \leqslant \mathbf{x}, Y_i \leqslant y)}. \tag{7}$$

Now, using (7) we define the kernel estimate of the joint pdf $f(\cdot\,,\cdot)$ by

$$\tilde{f}_n(\mathbf{x}, y) = \frac{1}{a_n^d b_n} \int\limits_{\mathbb{R}^d} \int\limits_{\mathbb{R}} K_d\left(\frac{\mathbf{x}-\mathbf{u}}{a_n}\right) \Lambda\left(\frac{y-v}{b_n}\right) \tilde{F}_n(\mathbf{du}, \mathrm{d}v)$$
$$= \frac{\alpha}{n a_n^d b_n} \sum_{i=1}^n \frac{1}{G(Y_i)} K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{a_n}\right) \Lambda\left(\frac{y-Y_i}{b_n}\right), \tag{8}$$

where Λ is a fixed kernel with $\int_{\mathbb{R}} \Lambda = 1$, and $(b_n)_{n\geqslant 1} \searrow 0$ as $n \nearrow \infty$. Finally, for all $\mathbf{x} \in \mathbb{R}^d$, we get a new estimator of the conditional density $g(y|\mathbf{x})$. Indeed, we have

$$\tilde{g}_n(y|\mathbf{x}) := \frac{\tilde{f}_n(\mathbf{x}, y)}{\tilde{\ell}_n(\mathbf{x})} = \frac{\sum_{i=1}^n (G(Y_i))^{-1} K_d((\mathbf{x}-\mathbf{X}_i)/a_n) \Lambda((y-Y_i)/b_n)}{b_n \sum_{i=1}^n (G(Y_i))^{-1} K_d((\mathbf{x}-\mathbf{X}_i)/a_n)}, \tag{9}$$

with the convention $\frac{0}{0} = 0$.

The natural estimator of $\Theta(x)$ is defined as the random variable $\hat{\Theta}_n(x)$, satisfying

$$\tilde{g}_n\big(\hat{\Theta}_n(x)|x\big) = \max_{-\infty<y<\infty} \tilde{g}_n(y|x). \tag{10}$$

However, the computation of $\hat{\Theta}_n(x)$ from (10) is not possible since both α and $G(\cdot)$ are usually unknown. One way to overcome this difficulty, is to replace $G(\cdot)$ by $G_n(\cdot)$ and α by α_n. Therefore, the feasible estimator of $\Theta(x)$ is then given as the solution of the equation

$$\hat{g}_n\big(\hat{\Theta}_n(x)|x\big) = \max_{-\infty<y<\infty} \hat{g}_n(y|x) = \max_{-\infty<y<\infty} \frac{\hat{f}_n(x, y)}{\hat{\ell}_n(x)}, \tag{11}$$

where $\hat{f}_n(\cdot,\cdot)$, $\hat{\ell}_n(\cdot)$ and the corresponding df's $\hat{\mathbf{F}}_n(\cdot\,,\cdot)$ and $\hat{L}_n(\cdot)$ are defined by the same expressions as in (8), (6), (7) and (5) respectively, in which α and $G(\cdot)$ are replaced by α_n and $G_n(\cdot)$.

2. Main results

With the aim of simplifying the results and the proofs, let us fix $d=1$, $a_n = b_n = h_n$, and set $\Xi_0 = \{x \in \mathbb{R} \mid \ell(x) > 0\}$. Let also $\Xi \subset \Xi_0$ be a compact set of $\mathbb{R}$, and $\mathcal{C}(\ell)$ be the set of continuity points of ℓ. The assumptions are gathered below for easy reference.

(A) **(a)** $f(\cdot\,,\cdot)$ is differentiable up to order 4 and $\sup_{x,y} |f^{(i,j)}(x,y)| < \infty$, $i+j \leqslant 4$;

(b) $f^{(0,2)}(\cdot,\cdot)$ is continuous and does not vanish in the neighborhood of $\Theta(\cdot)$;

(c) $\ell(\cdot)$ satisfies the Lipschitz condition and $\ell(x) \geqslant \gamma_0 > 0$;

(d) $g(y|x)$ is twice differentiable, uniformly continuous in y and the second derivative $g^{(2)}(\cdot|x)$ is continuous;

(e) $\Theta(\cdot)$ is such that for any $\varepsilon > 0$ and any function $\eta(\cdot)$, there exists a $\beta > 0$ such that $\sup_{x\in\Xi} |\Theta(x) - \eta(x)| \geqslant \varepsilon$ implies $\sup_{x\in\Xi} |g(\Theta(x)|x) - g(\eta(x)|x)| \geqslant \beta$.

(B) **(a)** The kernel $\mathcal{K}$ ($\mathcal{K} = K_1$ or $\mathcal{K} = \Lambda$) is continuous and satisfies $\lim_{|u|\to\infty} \mathcal{K}(u) = 0$ and $\int_{\mathbb{R}} \mathcal{K}(u)\,\mathrm{d}u = 1$;

(b) $\int_{\mathbb{R}} u^j \mathcal{K}(u)\,\mathrm{d}u = 0$; $j = 1, 2$;

(c) $\int_{\mathbb{R}} |u|^3 \mathcal{K}(u)\,\mathrm{d}u < \infty$;

(d) K_1 is differentiable and of bounded variations;

(e) Λ is three times differentiable and Λ, $\Lambda^{(1)}$ and $\Lambda^{(2)}$ are functions of bounded variations.

(H) **(a)** The sequence h_n is such that $h_n \searrow 0$ and $\frac{nh_n^8}{\log n} \to \infty$ as $n \nearrow \infty$;

(b) $nh_n^9 \to 0$ as $n \nearrow \infty$.

Remark 1. Assumptions **(A, a,b,e)** are needed to establish Theorem 2.1. The conditions **(B, a–c)** are very usual in the functional estimation; they allow us to study the asymptotic variance term of $\tilde{f}_n^{(0,1)}(\cdot\,,\cdot)$ while **(B, d,e)** are used to state the uniform convergence of $\hat{f}_n(\cdot\,,\cdot)$. The Assumptions **(H, a)** are needed to obtain the uniform convergence of the second derivative $\hat{f}_n^{(0,2)}$. The hypothesis **(H, b)** allows to state the convergence of the bias term of $\tilde{f}_n^{(0,1)}(\cdot\,,\cdot)$. Note that these assumptions are usual and not very restrictive. We point out that we obtain the same bandwidth as the one used in the complete data case (see, Samanta and Thavaneswaran [4]).

Our first result deals with the uniform almost sure (a.s.) convergence of the estimator (11):

Theorem 2.1. *Suppose that Assumptions* (**A: a,c**)*,* (**B: c,d,e**) *and* (**H: a**) *hold. Then,*

$$\lim_{n\to\infty}\sup_y\bigl|\hat g_n(y|x)-g(y|x)\bigr|=0,\quad a.s.$$

Moreover, we have for any compact set $\Xi\subset\Xi_0$,

$$\sup_{x\in\Xi}\sup_y\bigl|\hat g_n(y|x)-g(y|x)\bigr|=\mathrm{O}\bigl[\max\bigl\{\bigl(\bigl(nh_n^4\bigr)^{-1}\log n\bigr)^{1/2},h_n\bigr\}\bigr],\quad \mathbf{P}\text{-}a.s.$$

From this Theorem we deduce the following corollary:

Corollary 2.2. *Under Assumptions* (**A: a,b,d,e**)*,* (**B: d,e**)*,* (**H:a**)*, we have*

$$\lim_{n\to\infty}\sup_{x\in\Xi}\bigl|\hat\Theta_n(x)-\Theta(x)\bigr|=0,\quad a.s.$$

In addition, for all n large enough, we have

$$\sup_{x\in\Xi}\bigl|\hat\Theta_n(x)-\Theta(x)\bigr|=\mathrm{O}\Bigl[\max\bigl\{\bigl(\bigl(nh_n^4\bigr)^{-1}\log n\bigr)^{1/4},h_n^{1/2}\bigr\}\Bigr],\quad \mathbf{P}\text{-}a.s.$$

Remark 2. The proof of Theorem 2.1 is based mainly on the following classical decomposition, which holds with probability one and the convergence of each of the terms

$$\sup_{x\in\Xi}\sup_y\bigl|\hat g_n(y|x)-g(y|x)\bigr|\leqslant\frac{1}{\inf_{x\in\Xi}\hat\ell_n(x)}\Bigl\{\sup_{x\in\Xi}\sup_y\bigl|\hat f_n(x,y)-f(x,y)\bigr|+\kappa.\sup_{x\in\Xi}\bigl|\hat\ell_n(x)-\ell(x)\bigr|\Bigr\},\tag{12}$$

where κ is the upper bound of $g(y|x)$.

The following result states the asymptotic normality of the kernel conditional mode estimator:

Theorem 2.3. *Suppose that Assumptions* (**A: b**)*,* (**B: a,b,d,e**) *and* (**H: a,b**) *hold. Then, we have for any $x\in\mathcal{A}$,*

$$\left(\frac{nh_n^4(f^{(0,2)}(x,\Theta(x)))^2}{\mathrm{Var}(x,\Theta(x))}\right)^{1/2}\bigl(\hat\Theta_n(x)-\Theta(x)\bigr)\xrightarrow{\mathcal{D}}\mathcal{N}(0,1),$$

where $\mathcal{A}=\{x\colon x\in\mathcal{C}(f),\mathrm{Var}(x,\Theta(x))\neq0\}$*,*

$$\mathrm{Var}\bigl(x,\Theta(x)\bigr)=\frac{\alpha f(x,\Theta(x))}{G(\Theta(x))}\int_{\mathbb{R}^2}\bigl[K_1(r)\Lambda^{(1)}(s)\bigr]^2\,\mathrm{d}r\,\mathrm{d}s$$

and $\xrightarrow{\mathcal{D}}$ *denotes the convergence in distribution.*

Remark 3. A plug-in estimate for the asymptotic standard deviation $\sigma(x,\Theta(x)):=(\mathrm{Var}(x,\Theta(x)))^{1/2}$, can be easily obtained using the estimators $G_n(\cdot)$, (2), (11) and $\hat f_n(\cdot,\cdot)$ of $G(\cdot)$, α, $\Theta(x)$ and $f(\cdot,\cdot)$ respectively, which is given by $\hat\sigma(x,\hat\Theta_n(x))=(\frac{\alpha_n\hat f_n(x,\hat\Theta_n(x))}{G_n(\hat\Theta_n(x))}\int_{\mathbb{R}^2}[K_1(r)\Lambda^{(1)}(s)]^2\,\mathrm{d}r\,\mathrm{d}s)^{1/2}$. Then $\hat\sigma(x,\hat\Theta_n(x))$ can be used to get the following approximate $(1-\zeta)\%$ confidence interval for $\Theta(x)$.

$$\hat\Theta_n(x)\pm t_{1-\zeta/2}\times\left(\frac{\hat\sigma^2(x,\hat\Theta_n(x))}{nh_n^4(\hat f_n^{(0,2)}(x,\hat\Theta_n(x)))^2}\right)^{1/2}.$$

A simulation study has been performed to illustrate the behavior of our estimator and how good is the asymptotic normality in the finite sample. This study has not been included in this Note for lack of place, however, it can be obtained from the authors.

Acknowledgements

The authors wish to thank the two anonymous referees for their remarks and comments that permitted them to improve this Note.

References

[1] S. He, G.L. Yang, Estimation of the truncation probability in the random truncation model, Ann. Statist. 26 (1998) 1011–1027.
[2] D. Lynden-Bell, A method of allowing for known observational selection in small samples applied to 3CR quasars, Monthly Notices Roy. Astronom. Soc. 155 (1971) 95–118.
[3] E. Ould-Saïd, M. Lemdani, Asymptotic properties of a nonparametric regression function estimator with randomly truncated data, Ann. Inst. Statist. Math. 58 (2006) 357–378.
[4] M. Samanta, A. Thavaneswaran, Nonparametric estimation of the conditional mode, Comm. Statist. Theory Methods 19 (1990) 4515–4524.
[5] W. Stute, Almost sure representations of the product-limit estimator for truncated data, Ann. Statist. 21 (1993) 146–156.
[6] M. Woodroofe, Estimating a distribution function with truncated data, Ann. Statist. 13 (1985) 163–177.