Uniform Limit Laws of the Logarithm for Nonparametric Estimators of the Regression Function in Presence of Censored Data

B. Maillot* and V. Viallon**

1LSTA, Université Paris 6, Paris, France
2Unité de biostatistique et d’épidémiologie, Université Paris Descartes, Paris, France

Received September 2, 2007; in final form, March 9, 2009

Abstract—In this paper, we establish uniform-in-bandwidth limit laws of the logarithm for non-parametric Inverse Probability of Censoring Weighted (I.P.C.W.) estimators of the multivariate regression function under random censorship. A similar result is deduced for estimators of the conditional distribution function. The uniform-in-bandwidth consistency for estimators of the conditional density and the conditional hazard rate functions are also derived from our main result. Moreover, the logarithm laws we establish are shown to yield almost sure simultaneous asymptotic confidence bands for the functions we consider. Examples of confidence bands obtained from simulated data are displayed.

Key words: censored regression, kernel estimates, laws of the logarithm, inverse probability of censoring weighted estimates.

2000 Mathematics Subject Classification: 62G08, 62N01.

DOI: 10.3103/S1066530709020045

1. INTRODUCTION—MOTIVATIONS

Nonparametric estimators of functionals of the conditional law (such as the regression function or the conditional distribution function) are known to provide a suitable and efficient means to catch the possibly complex relation between a given variable of interest and some explanatory covariates. Because of this obvious practical interest, many authors have studied the (asymptotic) properties of such estimators (see, e.g., [3], [19], [18]). Fewer works deal with the special case where the variable of interest is censored ([14], [24]). Yet, this situation arises in many statistical applications, as medical research, reliability, etc., and it is therefore of paramount importance to build and study estimators adapted to the censored setting. When the variable of interest is subject to right-censoring, transformations of the observed data are generally needed to derive inference on the underlying (conditional) distribution (see [5]). Estimates based on these transformations are usually referred to as synthetic data estimates in the literature. In the case of the regression function estimation, Fan and Gijbels [14] especially proposed a transformation leading to a local version of the Buckley–James estimator. In this paper, we make use of an alternative transformation, which leads to Inverse Probability of Censoring Weighted [I.P.C.W.] estimators. I.P.C.W.-type estimators have recently gained popularity in the censored-data literature. To our mind, they basically present two particularly appealing properties. First, and as it will be especially made clear in the proofs of our forthcoming results (see also [6], [22], [4], [21] and the references therein), their asymptotic behavior can be easily deduced from that of analogous estimators in the uncensored case. Second, their computation is straightforward. In that sense, they are appealing for both theoretic and applied statistics purposes. It is however noteworthy that the methodology we propose here for I.P.C.W.-type estimates shall apply with minor modifications to cope with other synthetic data estimates.

The present paper is organized as follows. First, we introduce the main notations and hypotheses needed for our task. Then, following the methodology developed in the uncensored case by Einmahl and Mason [12], we establish a uniform-in-bandwidth law of the logarithm for a nonparametric I.P.C.W.

*E-mail: bertrand.maillot@upmc.fr
**E-mail: vivian.viallon@univ-paris5.fr
estimator of the regression function (see Theorem 3.1 below). This result corresponds to the almost sure and uniform-in-bandwidth version of Theorem 3.1 in [27]. At this point we shall stress the reader attention on the fact that, as was first shown by Deheuvels and Mason [9] (see also [13], [10] and the relevant references therein), such uniform-in-bandwidth limit laws turn out to be of particular interest in practice because they allow for establishing uniform consistency of data-driven (and hence random) bandwidth estimators. In Section 4 we derive a similar law of the logarithm for an estimator of the conditional distribution function and we establish the uniform-in-bandwidth consistency for estimators of the conditional density and the conditional hazard rate functions. As was especially pointed out by Deheuvels and Mason [9] in the uncensored case, limit laws of the logarithm prove themselves to be useful in the construction of simultaneous confidence bands for the true considered function. Such confidence bands, obtained from simulated data, are displayed in Section 4.2. Finally, Section 5 is devoted to the proofs of our results.

2. NOTATION AND ASSUMPTIONS

Consider a triple $(Y, C, X)$ of random variables defined in $\mathbb{R} \times \mathbb{R} \times \mathbb{R}^d, d \geq 1$. Here $Y$ is the variable of interest, $C$ a censoring variable and $X = (X_1, \ldots, X_d)$ a vector of concomitant variables. Throughout, we work with a sample \{$(Y_i, C_i, X_i)_{1 \leq i \leq n}$\} of independent and identically distributed replica of $(Y, C, X), n \geq 1$. Actually, in the right censorship model, the pairs $(Y_i, C_i), 1 \leq i \leq n$, are not directly observed and the corresponding information is given by $Z_i := \min\{Y_i, C_i\}$ and $\delta_i := 1_{\{Y_i \leq C_i\}}, 1 \leq i \leq n$, with $1_E$ standing for the indicator function of $E$. Accordingly, the observed sample is $D_n = \{(Z_i, \delta_i, X_i), i = 1, \ldots, n\}$.

In the sequel, we impose the following assumptions upon the distribution of $(X, Y)$. Denote by $I$ a given compact set in $\mathbb{R}^d$ with nonempty interior and set, for any $\alpha > 0$,

$$I^\alpha = \{x: \inf_{u \in I} |x - u|_{\mathbb{R}^d} \leq \alpha\}$$

with $|\cdot|_{\mathbb{R}^d}$ standing for the usual Euclidean norm on $\mathbb{R}^d$. We will assume that, for a given $\alpha > 0$, $(X, Y)$ (resp. $X$) has a density function $f_{X,Y}$ (resp. $f_X$) with respect to the Lebesgue measure on $I^\alpha \times \mathbb{R}$ (resp. $I^\alpha$). We will also assume that the Assumptions (F.1 – F.2) below hold. For $-\infty < t < \infty$, set $F(t) = \mathbb{P}(Y \leq t), G(t) = \mathbb{P}(C \leq t)$, and $H(t) = \mathbb{P}(Z \leq t)$, the right-continuous distribution functions of $Y, C$ and $Z$ respectively. For any right-continuous distribution function $L$ defined on $\mathbb{R}$, denote by $T_L = \sup\{t \in \mathbb{R}: L(t) < 1\}$ the upper point of the corresponding distribution.

(F.1) For all $x \in I^\alpha, \lim_{x' \to x, x' \in I^\alpha} f_{X,Y}(x', y) = f_{X,Y}(x, y)$ for almost every $y \leq T_H$.

(F.2) $f_X$ is continuous and strictly positive on $I^\alpha$.

Now consider a pointwise measurable class $\mathcal{F}$ (see p. 110 in [25]) of real measurable functions defined on $\mathbb{R}$. Throughout, $\mathcal{F}$ will be assumed to form a VC subgraph class (see §2.6.2 in [25]).

In this paper, we will mostly focus on the regression function of $\psi(Y)$ evaluated at $X = x$, for $\psi \in \mathcal{F}$ and $x \in I^\alpha$, given by

$$m_\psi(x) = \mathbb{E}(\psi(Y) \mid X = x).$$

To estimate $m_\psi$ when $Y$ is right-censored, the key idea of I.P.C.W. estimators is as follows. Introduce the real-valued function $\Phi_\psi$ defined on $\mathbb{R}^2$ by

$$\Phi_\psi(y, c) = \frac{1_{\{y \leq c\}} \psi(y \wedge c)}{1 - G(y \wedge c)}.$$  

Assuming the function $G$ to be known, first note that $\Phi_\psi(Y_i, C_i) = \delta_i \psi(Z_i)/(1 - G(Z_i))$ is observed for every $1 \leq i \leq n$. Moreover, under the Assumption (I) below,

(I) $C$ and $(Y, X)$ are independent;