On the acceleration of some empirical means with application to nonparametric regression

Bernard Delyon and François Portier

Let $(X_1, \ldots, X_n)$ be an i.i.d. sequence of random variables in $\mathbb{R}^d$, $d \geq 1$, for some function $\varphi : \mathbb{R}^d \to \mathbb{R}$, under regularity conditions, we show that

$$n^{1/2} \left( n^{-1} \sum_{i=1}^n \frac{\varphi(X_i)}{f^{(i)}(X_i)} - \int \varphi(x) dx \right) \overset{P}{\longrightarrow} 0,$$

where $\hat{f}^{(i)}$ is the classical leave-one-out kernel estimator of $f$ the density of $X_1$. This result is surprising because it speeds up traditional rates, in root $n$, derived from the central limit theorem when $\hat{f}^{(i)} = f$. As a consequence, it improves the classical Monte Carlo procedure for integral approximation. The paper is largely concerned with theoretical issues related to the later result (rates of convergence, bandwidth choice, regularity of $\varphi$) but is also interested in some statistical applications dealing with random design regression. In particular, we provide the asymptotic normality of the estimation of the linear functionals of a regression function on which the only requirement is the Hölder regularity. This leads us to a corrected version of the average derivative estimator introduced by Härdle and Stoker (hardle1989) that estimates the index of a regression with less variance.