APPLICATIONES MATHEMATICAE
34,4 (2007), pp. 401–404

Ryszard Zieliński (Warszawa)

KERNEL ESTIMATORS AND THE DVORETZKY–KIEFER–WOLFOWITZ INEQUALITY

Abstract. It turns out that for standard kernel estimators no inequality like that of Dvoretzky–Kiefer–Wolfowitz can be constructed, and as a result it is impossible to answer the question of how many observations are needed to guarantee a prescribed level of accuracy of the estimator. A remedy is to adapt the bandwidth to the sample at hand.

1. Dvoretzky–Kiefer–Wolfowitz inequality. Let $X_1, \ldots, X_n$ be a sample from an (unknown) distribution $F \in \mathcal{F}$ where $\mathcal{F}$ is the class of all continuous distribution functions. Let

$$F_n(x) = \frac{1}{n} \sum_{j=1}^{n} 1_{(-\infty,x]}(X_j)$$

be the empirical distribution function. The Dvoretzky–Kiefer–Wolfowitz inequality in its strongest version (Massart 1990) states that

$$P\{\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \geq \varepsilon\} \leq 2e^{-2n\varepsilon^2}. \tag{1}$$

Making use of this inequality, for every $\varepsilon > 0$ and every $\eta > 0$ one can easily calculate $N(\varepsilon, \eta)$ such that if $n \geq N(\varepsilon, \eta)$ then

$$P\{\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \geq \varepsilon\} \leq \eta.$$

2. Kernel estimators. The standard kernel density estimator is of the form (e.g. *Encyclopedia of Statistical Sciences* (2006))

$$\widehat{f}_n(x) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{h_n}\, k\left(\frac{x - X_j}{h_n}\right)$$

2000 *Mathematics Subject Classification*: 62G20, 62G30, 62G07.
Key words and phrases: kernel estimators, Dvoretzky–Kiefer–Wolfowitz inequality, bandwidth, adaptive estimators.

with appropriate h_n, $n = 1, 2, \ldots$. We shall consider kernel distribution estimators of the classical form

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{j=1}^{n} K\left(\frac{x - X_j}{h_n}\right)$$

where $K(x) = \int_{-\infty}^{x} k(t)\,dt$, and we shall show that no inequality like (1) with $\widehat{F}_n(x)$ instead of $F_n(x)$ can be constructed.

Proposition. *Let $k(\cdot)$ be any kernel such that $0 < K(0) < 1$ and $K^{-1}(t) < 0$ for some $t \in (0, K(0))$. Let $(h_n, n = 1, 2, \ldots)$ be any sequence of positive reals. Then there exist $\varepsilon > 0$ and $\eta > 0$ such that for every n there exists $F \in \mathcal{F}$ for which*

$$P\{\sup_{x\in\mathbb{R}} |\widehat{F}_n(x) - F(x)| \geq \varepsilon\} \geq \eta.$$

Proof. Obviously it is enough to demonstrate that under the assumptions of the Proposition there exist $\varepsilon > 0$ and $\eta > 0$ such that for every n there exists $F \in \mathcal{F}$ satisfying $P\{\widehat{F}_n(0) > F(0) + \varepsilon\} \geq \eta$.

Take $\varepsilon \in (0, t)$ and $\eta \in (t - \varepsilon, 1)$. Fix n. Given ε, η, and n, choose F such that $F(0) = t - \varepsilon$ and $F(-h_n K^{-1}(t)) = P\{X_j < -h_n K^{-1}(t)\} > \eta^{1/n}$. Then

$$P\left\{K\left(-\frac{X_j}{h_n}\right) > F(0) + \varepsilon\right\} > \eta^{1/n}$$

and due to the fact that

$$\bigcap_{j=1}^{n}\left\{K\left(-\frac{X_j}{h_n}\right) > F(0) + \varepsilon\right\} \subset \left\{\frac{1}{n}\sum_{j=1}^{n} K\left(-\frac{X_j}{h_n}\right) > F(0) + \varepsilon\right\}$$

we have

$$P\left\{\frac{1}{n}\sum_{j=1}^{n} K\left(-\frac{X_j}{h_n}\right) > F(0) + \varepsilon\right\} \geq \prod_{j=1}^{n} P\left\{K\left(-\frac{X_j}{h_n}\right) > F(0) + \varepsilon\right\} > \eta,$$

which ends the proof.

Remark. By the Proposition, $\sup_{x\in\mathbb{R}} |\widehat{F}_n(x) - F(x)|$ does not converge to zero in probability, uniformly in $F \in \mathcal{F}$.

3. Random bandwidth. Let $X_{1:n} \leq \cdots \leq X_{n:n}$ be order statistics from the sample $X_1, \ldots, X_n$. Define

$$H_n = \min\{X_{j:n} - X_{j-1:n} : j = 2, \ldots, n\}.$$

Define the kernel estimator

$$\widetilde{F}_n(x) = \frac{1}{n}\sum_{j=1}^{n} K\left(\frac{x - X_j}{H_n}\right)$$

where for K we assume:

$$K(t) = \begin{cases} 0 & \text{for } t \le -1/2, \\ 1 & \text{for } t \ge 1/2, \end{cases}$$

$K(0) = 1/2$, $K(t)$ is continuous and nondecreasing in $(-1/2, 1/2)$.

Now, for $k = 1, \ldots, n$ we have $|\widetilde{F}_n(X_{k:n}) - F_n(X_{k:n})| = 1/2n$. The kernel estimator $\widetilde{F}_n(x)$ is continuous and nondecreasing, the empirical distribution function $F_n(x)$ is a step function, and consequently $|\widetilde{F}_n(x) - F_n(x)| \le 1/2n$ for all $x \in (-\infty, \infty)$. By the triangle inequality

$$|\widetilde{F}_n(x) - F(x)| \le |F_n(x) - F(x)| + \frac{1}{2n}$$

we obtain

$$P\{\sup_{x\in\mathbb{R}} |\widetilde{F}_n(x) - F(x)| \ge \varepsilon\} \le P\left\{\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| + \frac{1}{2n} \ge \varepsilon\right\},$$

and hence by (1) we have

$$P\{\sup_{x\in\mathbb{R}} |\widetilde{F}_n(x) - F(x)| \ge \varepsilon\} \le 2e^{-2n(\varepsilon - 1/2n)^2}, \qquad n > \frac{1}{2\varepsilon}, \tag{2}$$

which enables us to calculate $N = N(\varepsilon, \eta)$ that guarantees the prescribed accuracy of the kernel estimator $\widetilde{F}_n(x)$.

Comment 1. Observe that the smallest $N = N(\varepsilon, \eta)$ that guarantees the prescribed accuracy is somewhat greater for the kernel estimator $\widetilde{F}_n$ than that for the crude empirical step function F_n. For example, $N(0.1, 0.1)$ is 150 for F_n and 160 for $\widetilde{F}_n$; $N(0.01, 0.01)$ is 26 492 for F_n and 26 592 for $\widetilde{F}_n$.

Comment 2. Another disadvantage of kernel smoothing has been discovered by Hjort and Walker (2001): "kernel density estimator with optimal bandwidth lies outside any confidence interval, around the empirical distribution function, with probability tending to 1 as the sample size increases". Perhaps a reason is that smoothing adds to observations something which is rather arbitrarily chosen and which may spoil the inference.

A generalization. Inequality (2) holds for every distribution function $\widetilde{F}_n(x)$ that satisfies $|\widetilde{F}_n(X_{k:n}) - F_n(X_{k:n})| = 1/2n$, $k = 1, \ldots, n$.

References

Encyclopedia of Statistical Sciences (2006), 2nd ed., Vol. 6, Wiley-Interscience (E. J. Wegman, Kernel estimators).

N. L. Hjort and S. G. Walker (2001), *A note on kernel density estimators with optimal bandwidths*, Statist. Probab. Lett. 54, 153–159.

P. Massart (1990), *The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality*, Ann. Probab. 18, 1269–1283.

Institute of Mathematics
Polish Academy of Sciences
00-956 Warszawa, Poland
E-mail: R.Zielinski@impan.gov.pl

Received on 11.10.2007;
revised version on 10.12.2007 (1893)